Hardware Model Checking Competition 2011

HWMCC'11

Chairs

Armin Biere, Keijo Heljanko

Technical Advisors

Siert Wieringa, Niklas Sörensson

presented at

Formal Methods in Computer Aided Design 2011

FMCAD'11

Austin, TX, USA

November 2, 2011

aposterio updated version from 21st Nov. 2011

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

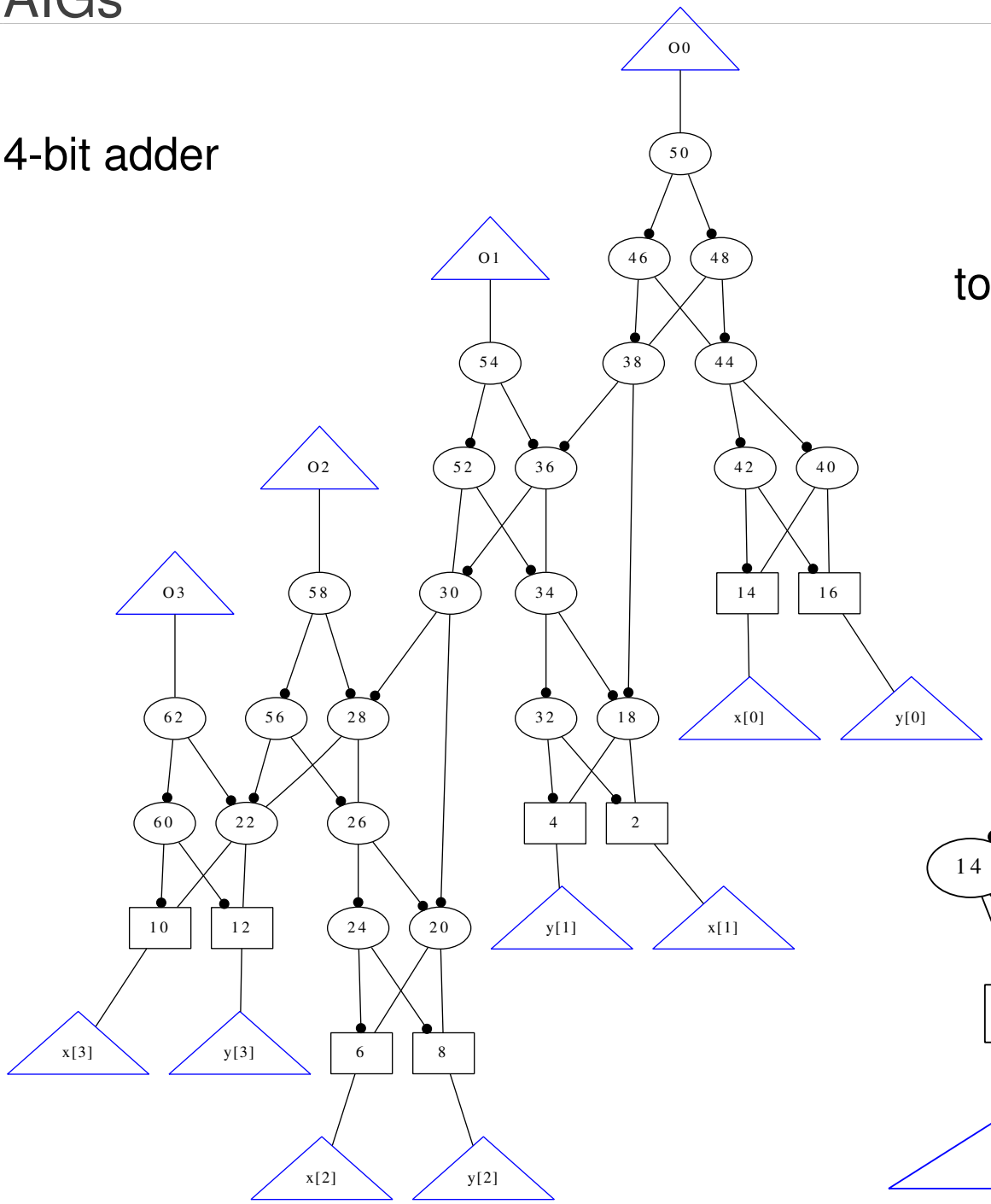
- relies on active support by submitters of benchmarks and model checkers

History

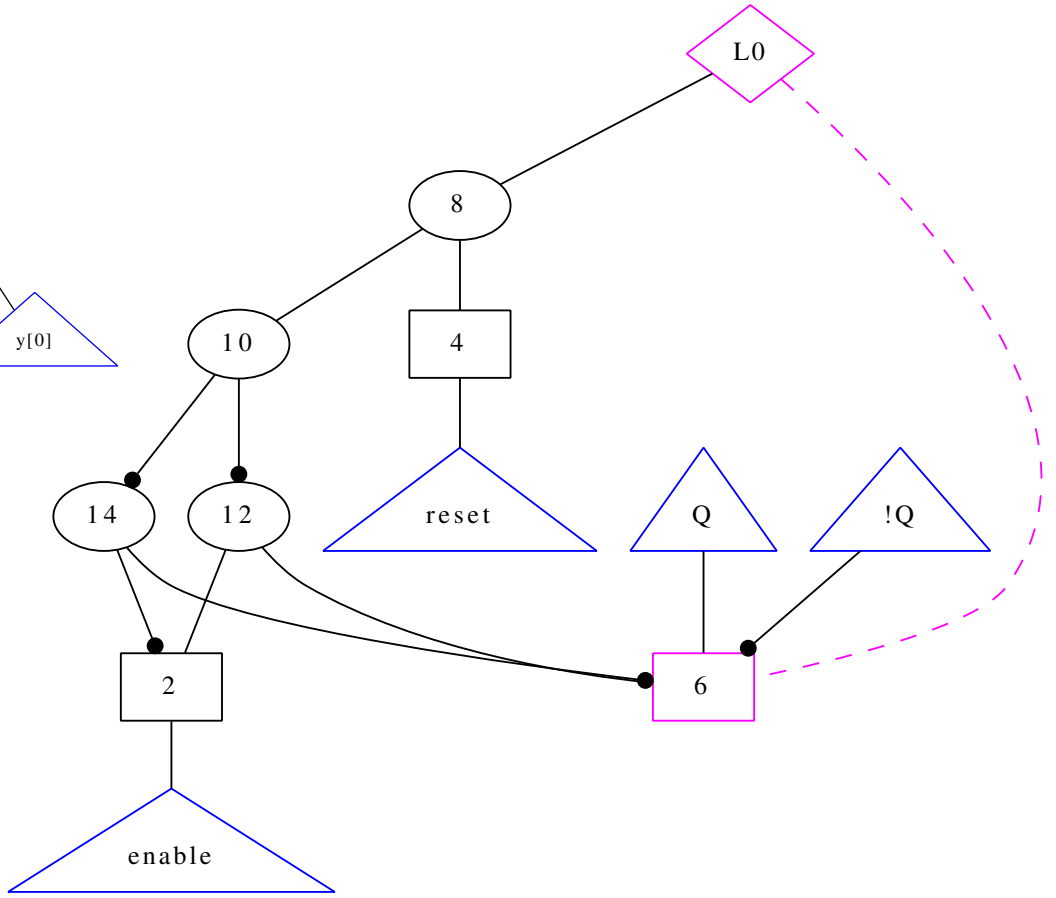
AIGER format AVM'06 Ascona	1st HWMCC CAV'07 Berlin	2nd HWMCC CAV'08 Princeton	3rd HWMCC CAV'10 FLOC'10 Edinburgh	**4th** **HWMCC** **FMCAD'11** **Austin**
Founding Lunch CAV'06 FLOC'06 Seattle		HWMCC Lunch FMCAD'08 Portland		
2006	2007	2008	2010	**2011**

- founding lunch CAV'06: Biere, Cimatti, Claessen, McMillan, Somenzi
- HWMCC lunch at FMCAD'08: should have benchmarks with multiple properties !!!
- HWMCC'10 with reduced committee: Biere, Claessen
  - still no *multiple properties* $\Rightarrow$ same competition mode as before
- HWMCC'11: old single property track, new live'ness and new multi property track

4-bit adder



toggle flip-flop with reset & enable



AIGER (ASCII Version)

add4 1st part

```
aag 31 8 0 4 23
2
4
6
8
10
12
14
16
50
54
58
62
18 4 2
20 8 6
22 12 10
24 9 7
26 25 21
28 26 22
30 29 21
```

add4 continued

```
32 5 3
34 33 19
36 34 31
38 37 19
40 16 14
42 17 15
44 43 41
46 44 39
48 45 38
50 49 47
52 35 30
54 53 37
56 27 23
58 57 29
60 13 11
62 61 23
i0 x[1]
i1 y[1]
i2 x[2]
i3 y[2]
i4 x[3]
i5 y[3]
i6 x[0]
i7 y[0]
```

togglere

```
aag 7 2 1 2 4
2              input 0
4              input 1
6 8            latch po/next
6              output 0
7              output 1
8 4 10          8 =   4 &  10
10 13 15       10 = !12 & !14
12 2 6         12 =   2 &   6
14 3 7         14 =  !2 &  !7
i0 enable      symbol table
i1 reset
o0 Q
o1 !Q
```

```
aig M I L O A
M = max vars
I = #inputs
L = #latches
O = #outputs
A = #ands
```

- 0 / 1 initialized latches or uninitialized latches — before only 0-initialized
- multiple properties and liveness properties — before only bad states properties
  - ‘b’ section of bad state properties — negation of safety
  - ‘j’ section of justice properties — negation of liveness
- environment / fairness constraints
  - ‘c’ section of invariant environment constraints — $c\,\mathbf{U}\,(c \wedge b)$
  - ‘f’ section of fairness constraints — $(\mathbf{G}c) \wedge (\bigwedge \mathbf{GF} f_i) \wedge \bigwedge \mathbf{GF} j_k$
- new witness / trace definition

```
aig M I L O A B C J F
```

Tools from README

```
aigand     conjunction of all outputs
aigbmc     new bounded model checker for format 1.9.x including liveness
aigdd      delta debugger for AIGs in AIGER format
aigflip    flip/negate all outputs
aigfuzz    fuzzer for AIGS in AIGER format
aiginfo    show comments of AIG
aigjoin    join AIGs over common inputs
aigmiter   generate miter of AIGER models
aigmove    treat non-primary outputs as primary outputs
aignm      show symbol table of AIG
aigor      disjunction of all outputs
aigsim     simulate AIG from stimulus or randomly
aigsplit   split outputs into separate files
aigstrip   strip simbols from AIG
aigtoaig   converts AIG formats (ascii, binary, stripped, compressed)
aigtocnf   translate combinational AIG into a CNF
aigtoblif  translate AIG into BLIF
aigtodot   visualizer for AIGs using 'dot' format
aigtosmv   translate sequential AIG to SMV format
andtoaig   translate file of AND gates into AIG
aigunroll  time frame expansion for bmc (previously called 'aigbmc')
bliftoaig  translate flat BLIF model into AIG
mc.sh      SAT based model checker for AIGER using these tools
smvtoaig   translate flat boolean encoded SMV model into AIG
soltostim  extract input vector from DIMACS solution
wrapstim   sequential stimulus from expanded combinational stimulus
```

Single Safety (Bad State) Property Benchmarks

- selected 297 = 73 sat + 198 unsat + 26 unsolved from HWMCC'10 out of 818
  sorted by: #solved, #solved in 100 sec, #solved in 10 sec, #solved in 1 sec
  removed instances which half of the model checkers solved within 10 seconds

- added 43 "negated" properties from HWMCC'11 detected by Håkan Hjort

- 58 single property **6s** benchmarks submitted by Jason Baumgartner
  includes 5 from 2 new multi property benchmarks

- 67 single property benchmarks from Torino submitted by Cabodi,Nocco,Quer
  except one all from 2 new multi property benchmarks

- 168 new benchmarks + 297 HWMCC'10 benchmarks = **465 benchmarks**

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

- 61 benchmarks used in LMCS'06 paper — submitted by Siert Wieringa
  actually from 14 benchmarks with multiple liveness properties
  but we do not have a multiple liveness properties track – at least this year

- 41 benchmarks used in FMCAD'11 paper — sub. by Hassan,Bradley,Somenzi
  all single liveness property benchmarks

- 16 arbiter benchmarks — submitted by Koen Claessen
  scalable benchmark set with (assumed) symmetric properties
  only picked some sizes and one property

- altogether 61 + 41 + 16 = **118 benchmarks**

Multiple Bad-State Properties

- was most requested new feature
  - lot of multiple properties in old benchmark set (e.g. HWMCC'10)
  - but already separated and **hard** to join
- we still "found" some:
  - 2 new from **6s** suite and 2 new from Torino
  - 4 from Bob Brayton's benchmarks suite submitted to HWMCC'10
    actually 8 including the flipped ones
  - 6 from NuSMV distribution
  - 1 submitted from Mentor Graphics to HWMCC'10
  - 5 from Bwolen Yang's benchmark set (from 1998!)

altogether 24

Details on Multiple Property Benchmarks

		M	I	L	O	A	B	C
6s40.aig	aig	36883	249	5608	0	31026	3	
6s48.aig	aig	934	72	66	0	796	2	
bob9234specmulti.aig	aig	815	36	111	0	668	8	
bob9234specnegmulti.aig	aig	815	36	111	0	668	8	
bobmiterbm1multi.aig	aig	3074	122	381	0	2571	1150	
bobmiterbm1negmulti.aig	aig	3074	122	381	0	2571	1150	
bobsynthmulti.aig	aig	18623	224	3015	0	15384	14	
bobsynthnegmulti.aig	aig	18623	224	3015	0	15384	14	
bobtuintmulti.aig	aig	2476	213	212	0	2051	32	
bobtuintnegmulti.aig	aig	2476	213	212	0	2051	32	
mentorbm1.aig	aig	36213	224	4376	0	31613	13	70
nusmvdme1d16multi.aig	aig	2225	288	321	0	1616	120	
nusmvdme1d3multi.aig	aig	379	54	61	0	264	3	
nusmvdme2d16multi.aig	aig	3144	293	326	0	2525	120	1
nusmvdme2d3multi.aig	aig	548	56	63	0	429	3	
nusmvsyncarb10multi.aig	aig	148	10	20	0	118	46	
nusmvsyncarb5multi.aig	aig	63	5	10	0	48	11	
pdtvsar8multip.aig	aig	7174	23	195	0	6956	33	
pdtvsarmultip.aig	aig	2890	17	130	0	2743	33	
sm98a7multi.aig	aig	10178	81	89	0	10008	5	1
sm98tcas16multi.aig	aig	5677	279	310	0	5088	6	1
sm98tcas16tmulti.aig	aig	5757	279	310	0	5168	6	1
sm98tcasmulti.aig	aig	2958	142	170	0	2646	6	1
sm98tcastmulti.aig	aig	3038	142	170	0	2726	6	1

							2824	

Model Checkers

alphabetically

- **aigbmc**, **blimc** by Biere (Linz) **new**
- **iimc** by Bradley, Somenzi, Hassan, Zhang, Cox (Boulder) new version
- **superprove**, **simpleprove**, **simplebip** from Brayton's group (Berkeley) new versions
- **tarmo** by Wieringa (Helsinki): 2 versions **new**
- **tip** by Sörensson, Claessen (Göteborg): 3 variants new version
- **pdtrav** by Cabodi, Nocco, Quer (Torino): 3 variants new versions

---

- **ic3*** + last winners **abcdsuperprove***,**abcbmc2***,**pdtrav*** **7-18 model checkers** from 6 groups

aigbmc + blimc

aigbmc

- 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
  no other sequential optimization

Safety: relatively naïve integration

- Some combinational and sequential reduction
- Followed by timed applications of ”reverse” IC3, BMC, forward BDD-based reachability, and finally IC3
- Goal is simply to be able to handle the benchmarks that are easy for BMC or BDDs but not IC3
- Reverse IC3 is sometimes better at finding counterexamples than either BMC or standard IC3

Liveness: only FAIR, as described in the FMCAD paper

- As suggested in the paper, FAIR and BDDs are complementary, so we expect that giving some time to a BDD-based solver would improve results

superprove

superprove is the most complete integrated model checker from Berkeley

- uses concurrency throughout
- initial simplification includes retiming, phase and temporal abstraction, signal correspondence, rewriting
- next, abstraction is attempted based on combined counter-example-based and proof-based abstraction as in FMCAD'10 paper last year by Niklas Eén et.al.
- next, speculation is attempted; the speculative miters are processed concurrently; counter-examples are used to refine the speculation
- if the above fails, the following engines PDR, BMC, Interpolation, and BDD reachability are run concurrently for the remaining time

simpleprove + simplebip

simpleprove simplified integrated model checker based on subset of ABC commands

- uses concurrency throughout
- uses initial simplification as in superprove
- tries to prove the property by running concurrently the following ABC engines (PDRm, INTRPm, BMC3, BMC3J, Reachy)

simplebip simplified integrated model checker based on provers developed by Niklas Eén

- uses concurrency throughout
- uses initial simplification as in superprove
- tries to prove the property by running concurrently the BIP engines (PDR, InterpMC, BMC, BMC3J, Reachy)

- Tip is a model checking tool consisting of a collection of inductive transformation and verification techniques:

  k-induction, signal correspondence, constraint extraction, temporal decomp., etc.

- This years submission to the competition is a work-in-progress rewrite that does not yet include as many transformations and engines:
  - Core engines are BMC and IC3
  - Circuit simplification by temporal decomposition

- Supports all new AIGER features:
  - Uninitialized latches
  - Constraints
  - Multiple properties — although not well yet
  - Liveness

- 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+.

General purpose parallel incremental SAT solver with clause sharing

- Runs multiple instances of MiniSAT-2.2.0 in parallel
- Different solver threads may be solving different jobs from the same incremental sequence
- HWMCC”11 version includes an AIGER front-end, which handles BMC encoding and counterexample printing

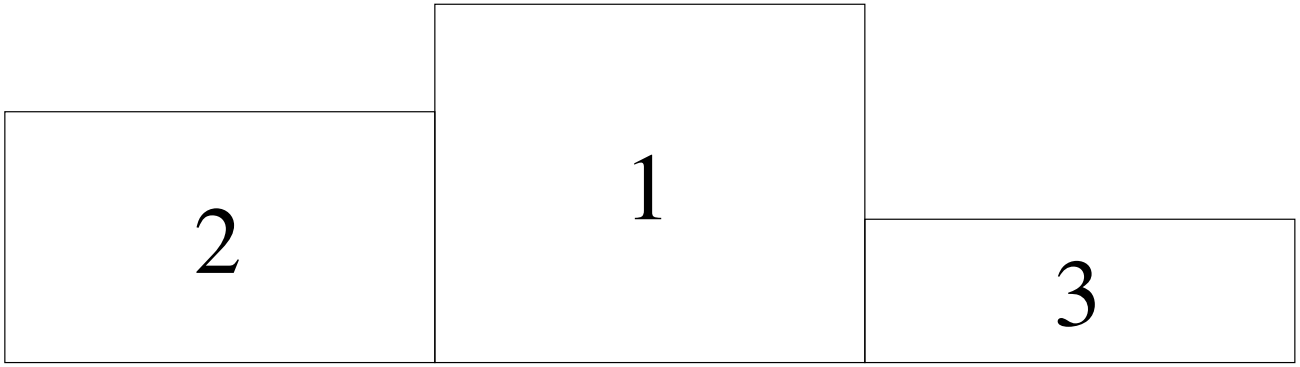
Setup

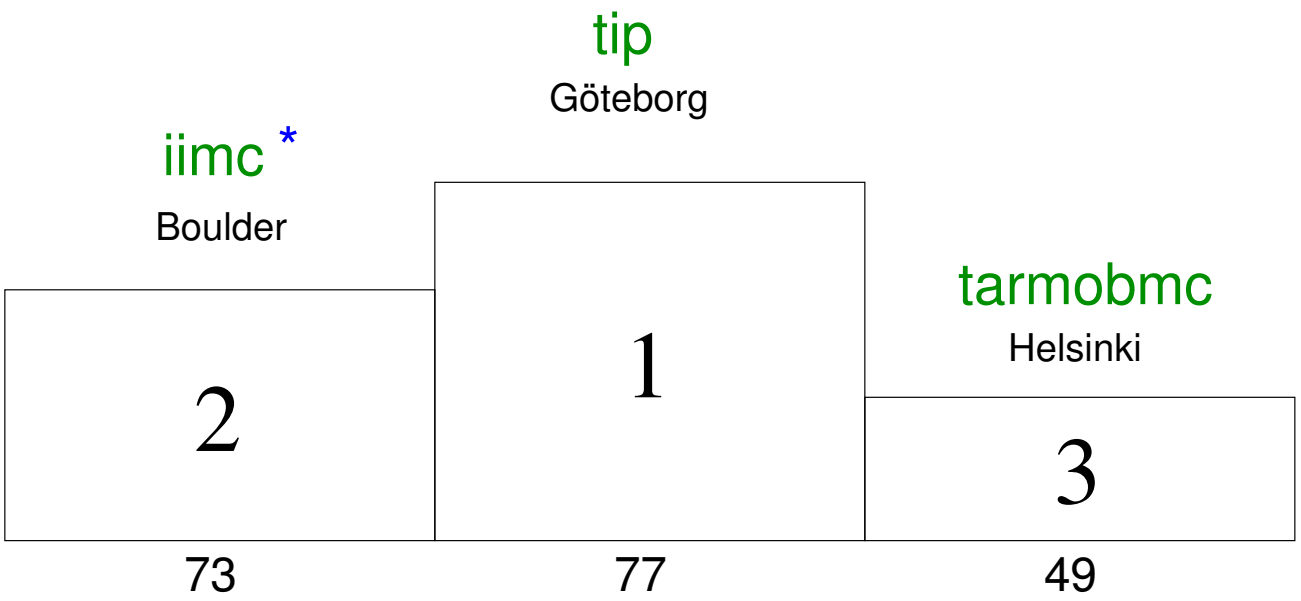
- single property benchmarks (single + live tracks) as in HWMCC'07 - HWMCC'10
  - bad state resp. fair SCC *reachable* $\Rightarrow$ instance *satisfiable* SAT
  - bad state resp. fair SCC *unreachable* $\Rightarrow$ instance *unsatisfiable* UNSAT
- multiple properties per benchmarks (multi track)
  - 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)

- 9 rankings
  - three 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/hwmcc11
  - you will also get spread sheets and all the log files there

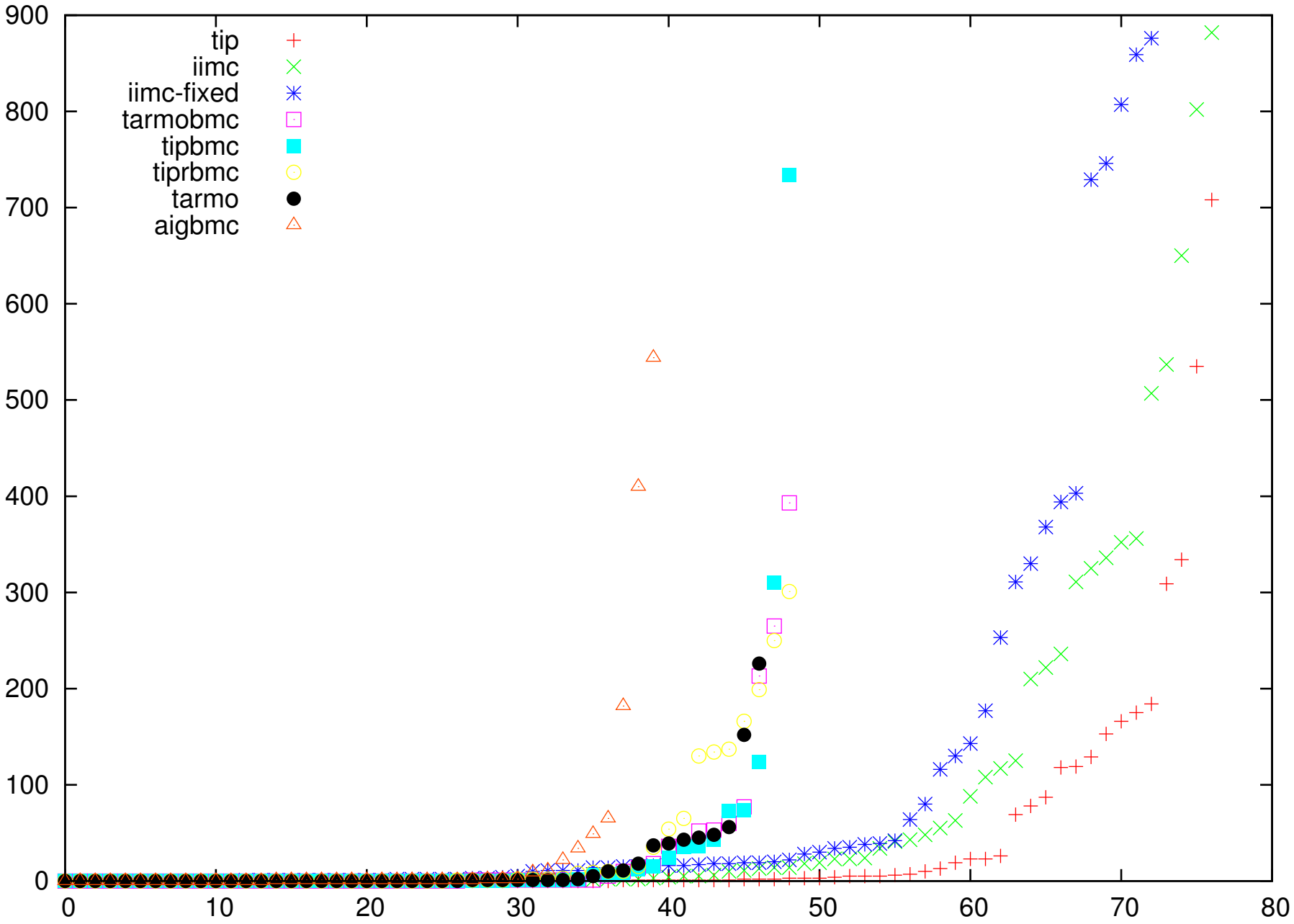

2
1
3

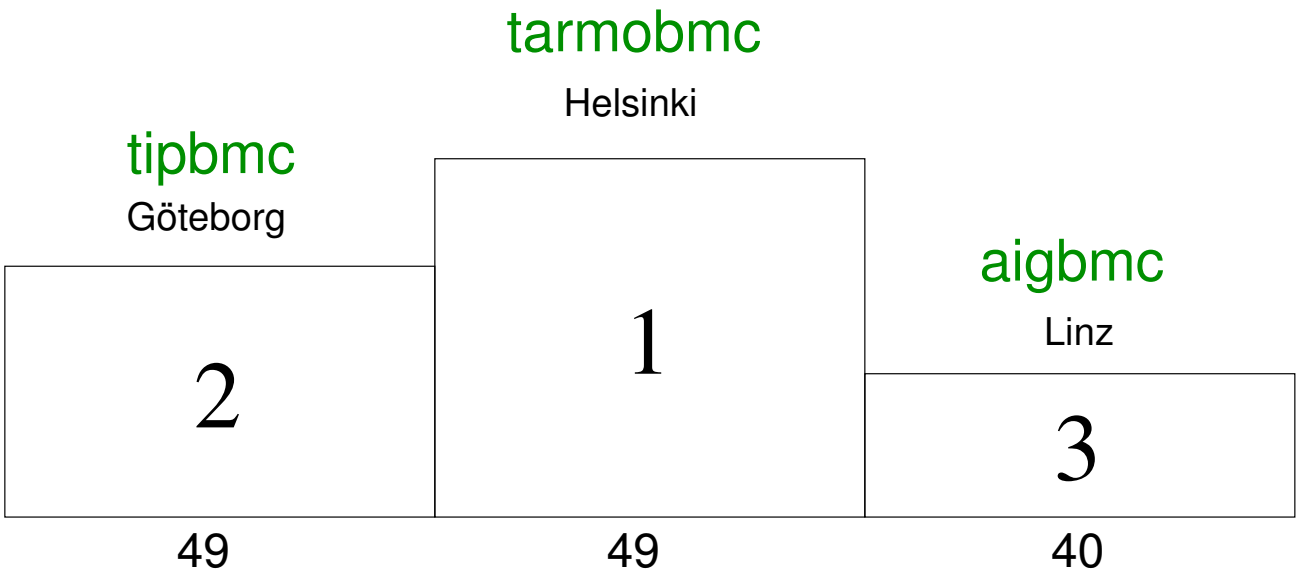


* aposterio results for the fixed **iimc**, see http://fmv.jku.at/hwmcc11/results.html#ltrack

SAT+UNSAT ranking (group ranking = solver ranking)

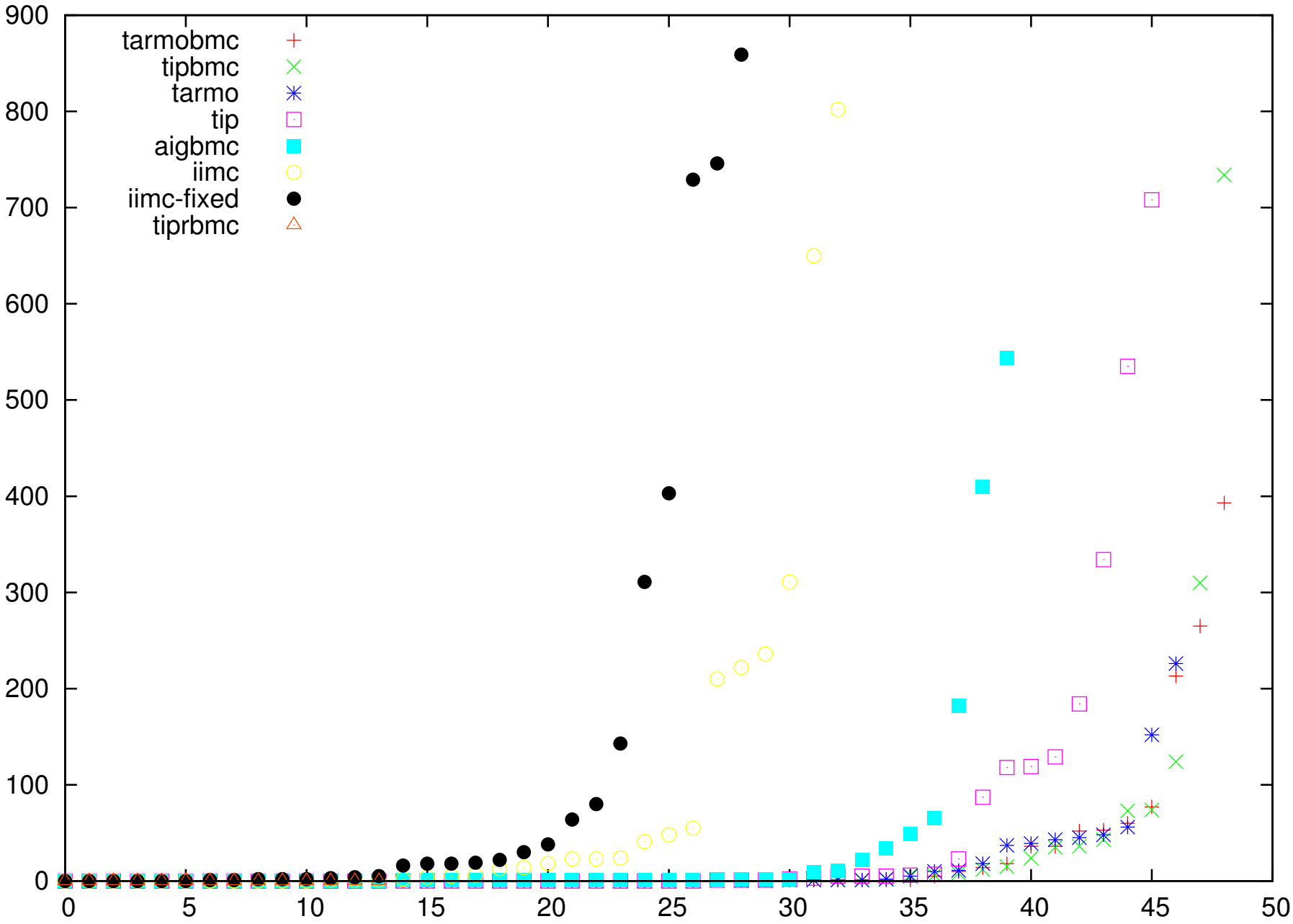
		solver	fnd	ok	sat	uns	fld	to	mo	real time	space	max
				SAT+UNS								
A	1	tip	118	77	46	31	41	41	0	3338 3306	1539	236
B	2	iimc-fixed	118	73	29	44	45	45	0	7354 7100	2605	240
C	3	tarmobmc	118	49	49	0	69	64	5	1243 4907	3870	477
	4	tipbmc	118	49	49	0	69	66	0	1510 1500	834	235
	5	tiprbmc	118	49	14	35	69	69	0	1539 1531	262	43
	6	tarmo	118	47	47	0	71	47	24	704 2740	3593	482
	7	aigbmc	118	40	40	0	78	78	0	1349 1326	313	52

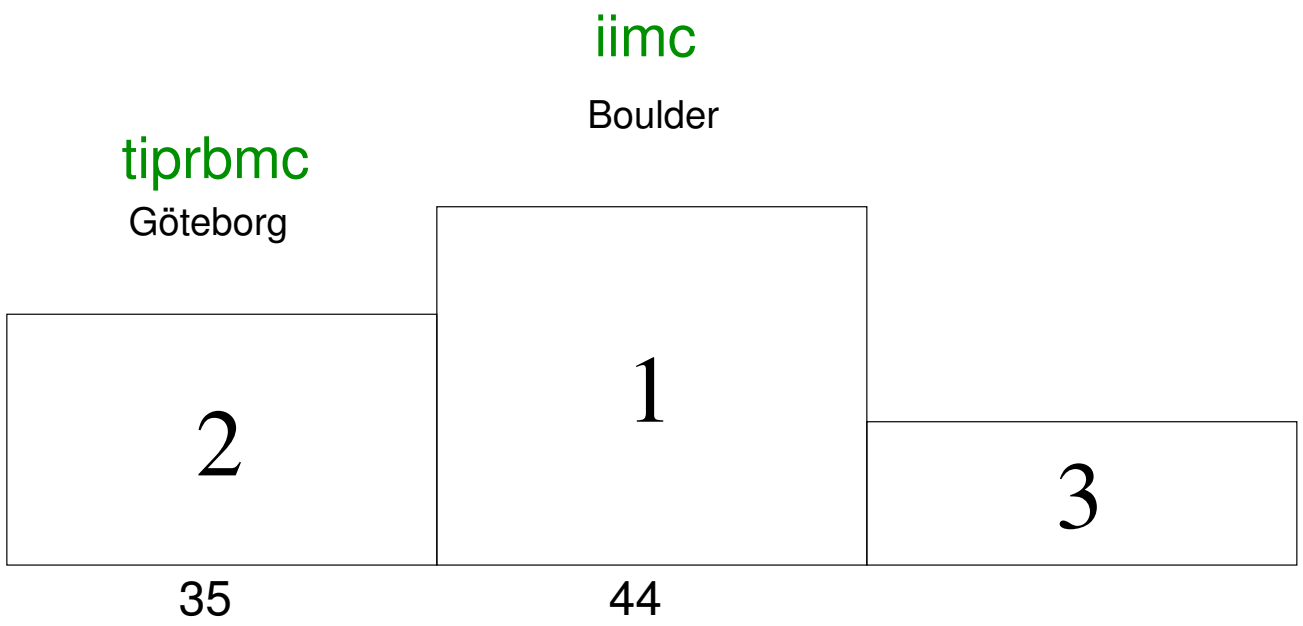

tip
iimc
iimc-fixed
tarmobmc
tipbmc
tiprbmc
tarmo
aigbmc
900
800
700
600
500
400
300
200
100
0
0
10
20
30
40
50
60
70
80


tarmobmc
Helsinki
tipbmc
Göteborg
aigbmc
Linz
2
1
3
49
49
40

SAT ranking (1st column group, 2nd solver)

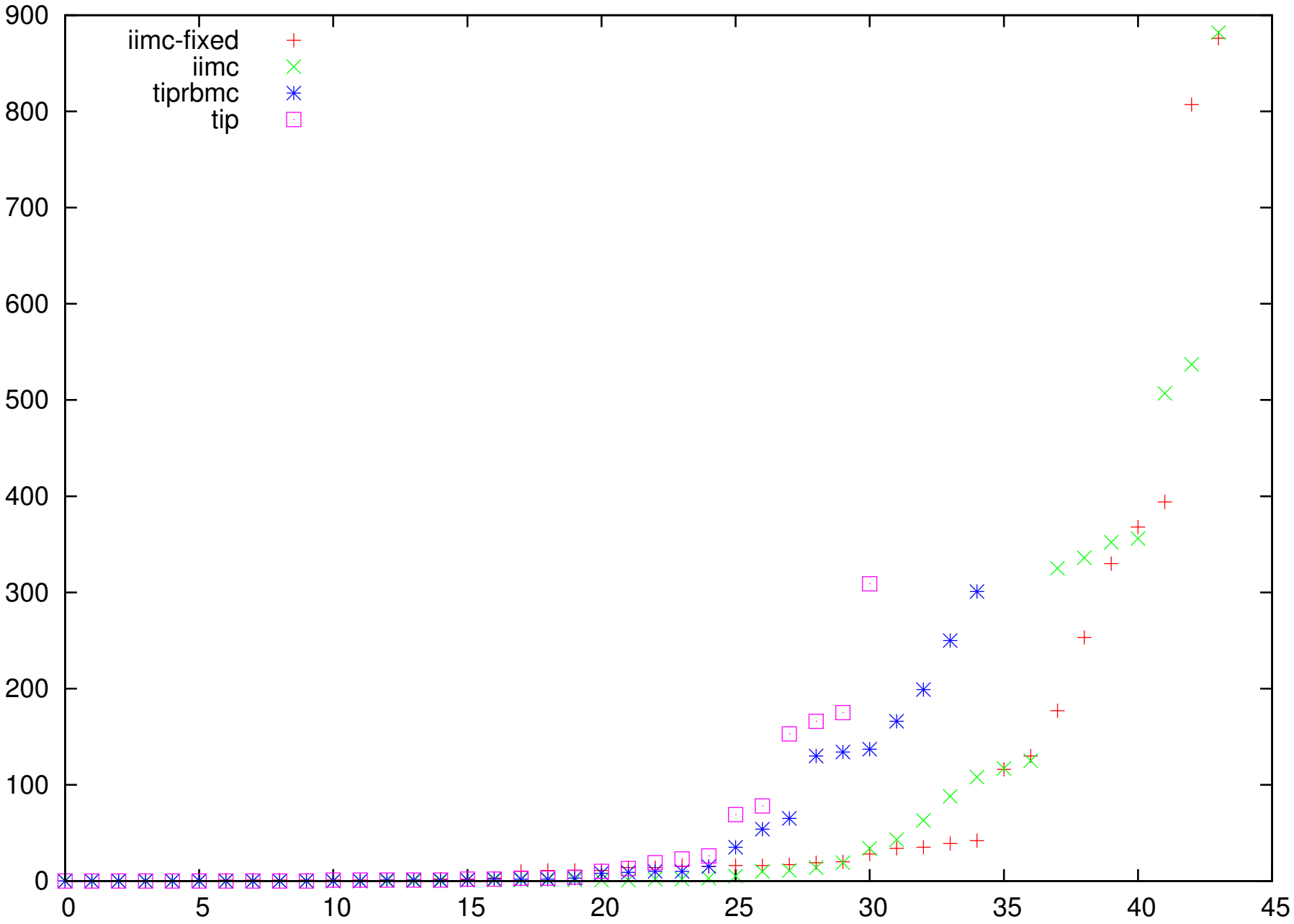
		solver	fnd	ok	SAT	uns	fld	to	mo	real	time	space	max
A	1	tarmobmc	118	49	49	0	69	64	5	1243	4907	3870	477
B	2	tipbmc	118	49	49	0	69	66	0	1510	1500	834	235
	3	tarmo	118	47	47	0	71	47	24	704	2740	3593	482
	4	tip	118	77	46	31	41	41	0	2278	2259	727	236
C	5	aigbmc	118	40	40	0	78	78	0	1349	1326	313	52
	6	iimc-fixed	118	73	29	44	45	45	0	7354	7100	2605	240
	7	tiprbmc	118	49	14	35	69	69	0	2	0	0	0


tarmobmc
tipbmc
tarmo
tip
aigbmc
iimc
iimc-fixed
tiprbmc
900
800
700
600
500
400
300
200
100
0
0
5
10
15
20
25
30
35
40
45
50


iimc
Boulder
tiprbmc
Göteborg
2
1
3
35
44

UNSAT ranking (no group on third place)

		solver	fnd	ok	sat	UNS	fld	to	mo	real	time	space	max
A	1	iimc-fixed	118	73	29	44	45	45	0	7354	7100	2605	240
B	2	tiprbmc	118	49	14	35	69	69	0	1538	1531	262	43
	3	tip	118	77	46	31	41	41	0	1060	1046	811	147
	4	tipbmc	118	49	49	0	69	66	0	0	0	0	0
	5	tarmo	118	47	47	0	71	47	24	0	0	0	0
	6	aigbmc	118	40	40	0	78	78	0	0	0	0	0
	7	tarmobmc	118	49	49	0	69	64	5	0	0	0	0


iimc-fixed
iimc
tiprbmc
tip
900
800
700
600
500
400
300
200
100
0
0
5
10
15
20
25
30
35
40
45

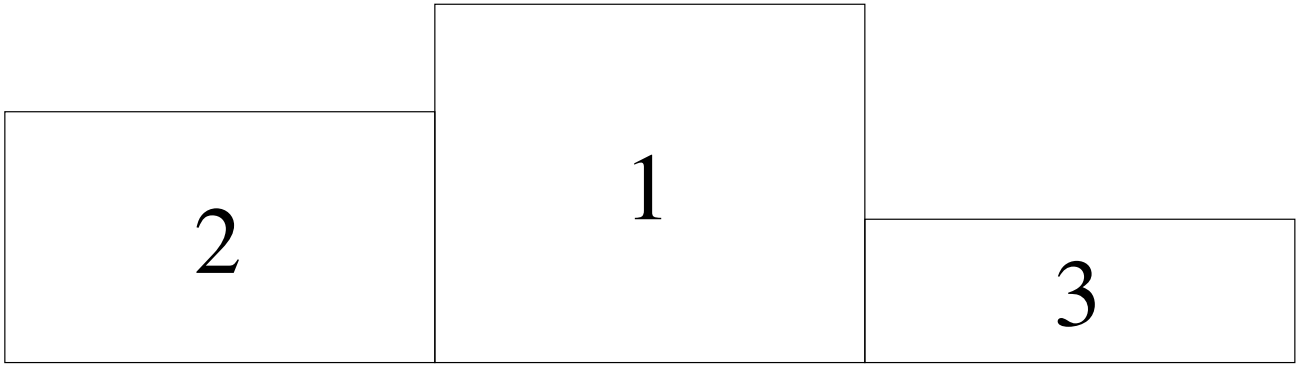
solver	all	sat	uns	SAT+UNSAT ranking	SAT ranking	UNSAT ranking
tarmo	2252	1253	999	A	A	A
tip	1368	1229	139	B	B	B
tarmobmc	1256	1256	0			
aigbmc	1216	1216	0	C	C	
tipbmc	1210	1210	0			
pdtmulti	412	275	137			*

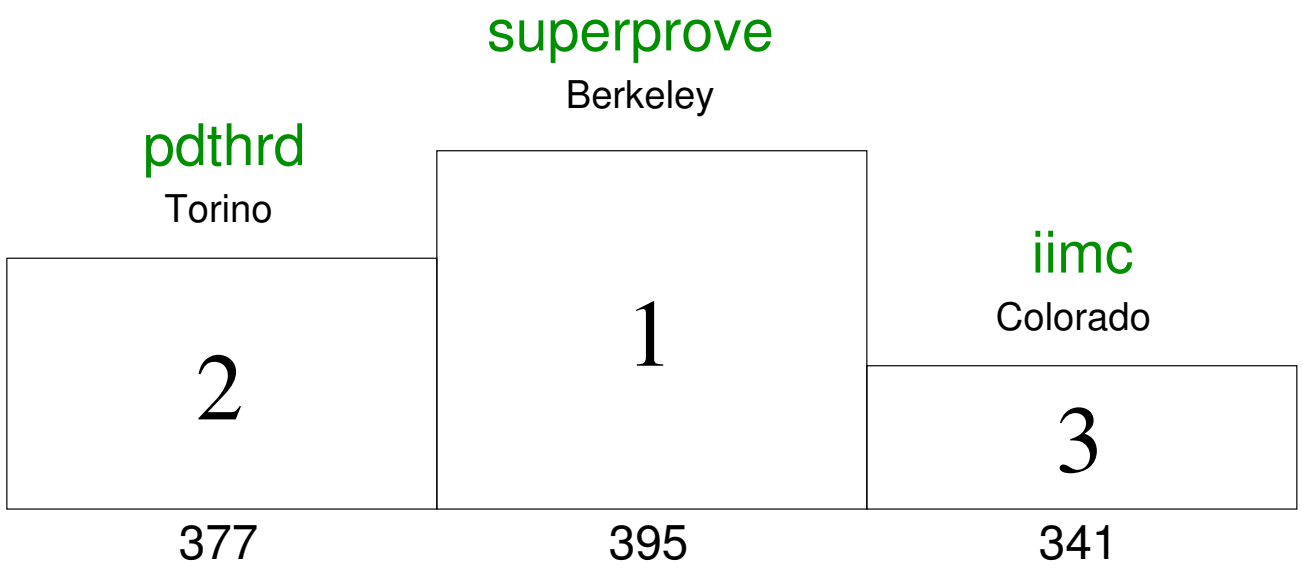
* pdtmulti gave UNSAT on two provable SAT instances

lot of SOTA properties: solved by exactly one "state-of-the-art" solver

terminology from automated theorem proving competition CASC

- 954 properties from bobmiterbm1multi only proved UNSAT by **tarmo**
- 245 properties from nusmvdme2d16multi only proved SAT by **pdtmulti**
- 9 properties from mentorbm1 only proved UNSAT by **pdtmulti**

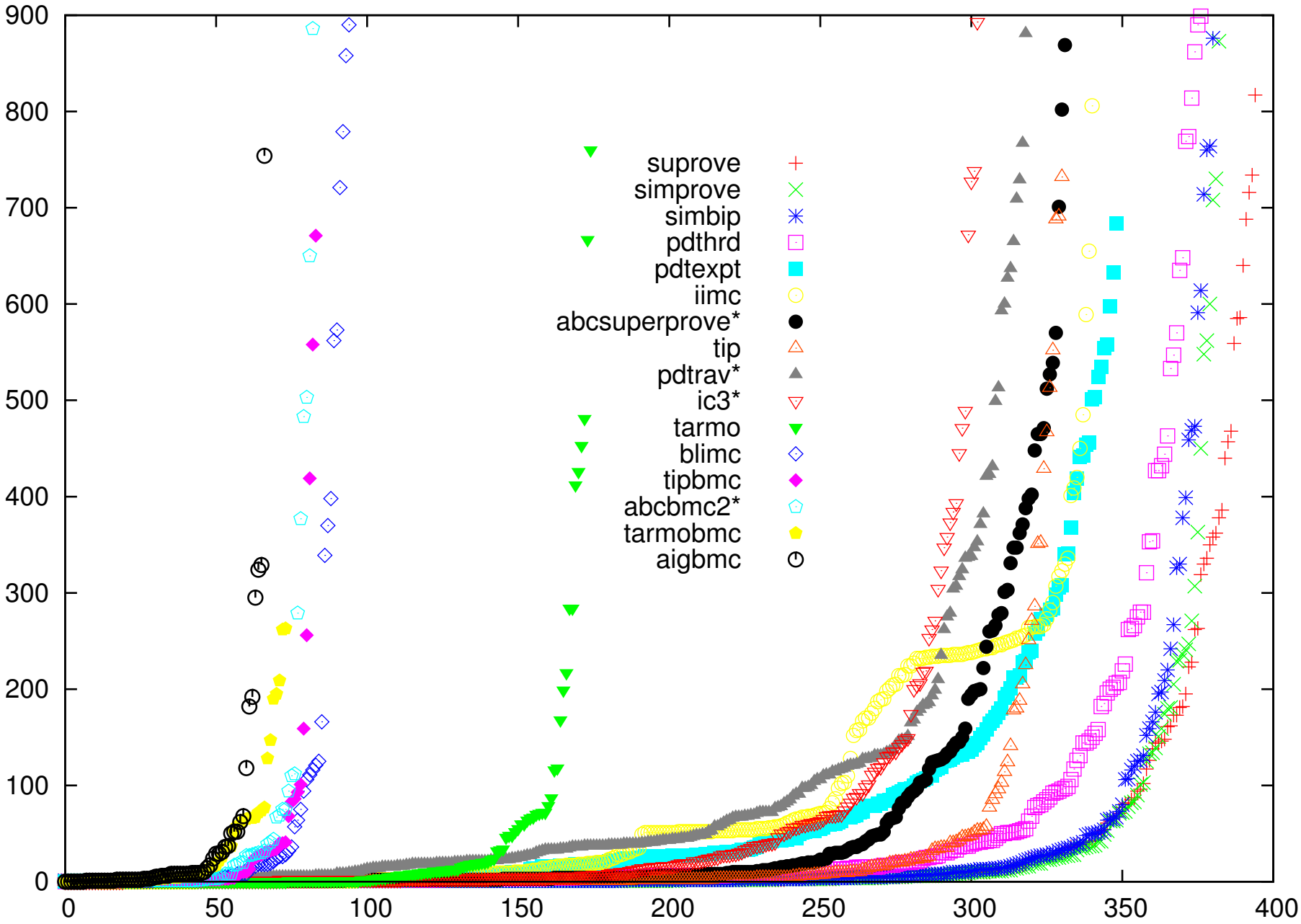

2
1
3

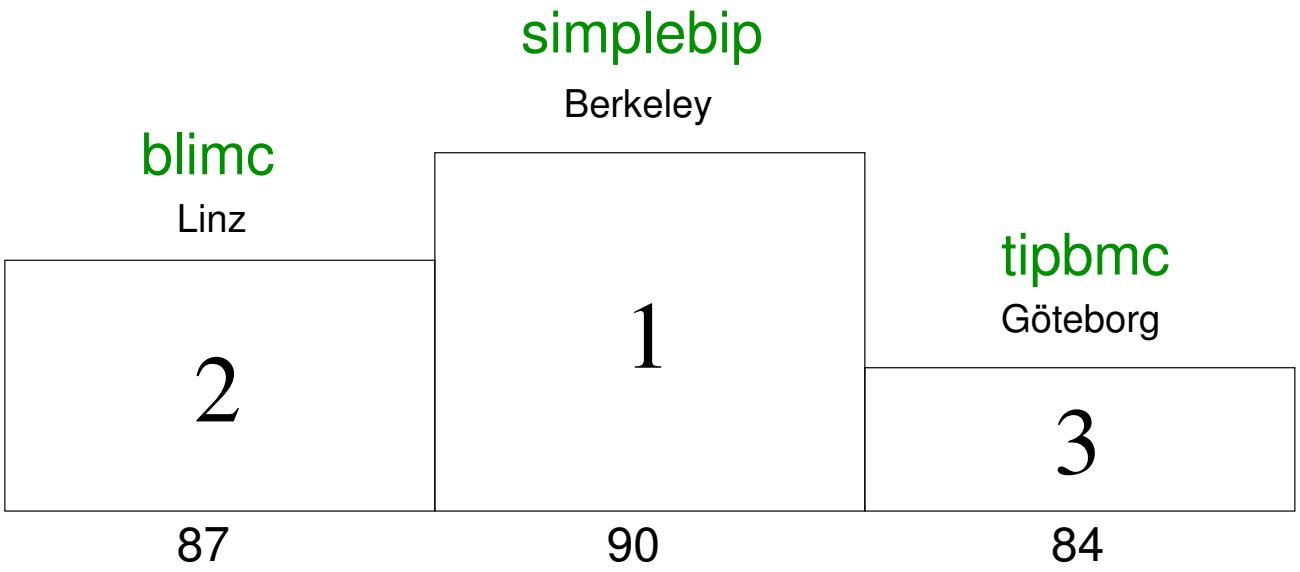

superprove
Berkeley
pdthrd
Torino
iimc
Colorado
2
1
3
377
395
341

SAT+UNSAT ranking (1st column per group, 2nd column per solver)

		solver	fnd	ok	sat	uns	fld	to	mo	s6	uk	real	time	space	max
				SAT+UNS											
A	1	suprove	465	395	83	312	70	70	0	0	0	15381	30904	95051	3406
	2	simprove	465	383	87	296	82	82	0	0	0	10697	35204	93851	3299
	3	simbip	465	381	90	291	84	84	0	0	0	12234	40909	83738	3074
B	4	pdthrd	465	377	81	296	88	74	0	9	0	19903	51735	102283	3219
	5	pdtexpt	465	349	68	281	116	106	0	10	0	23195	23051	34945	884
C	6	iimc	465	341	75	266	124	124	0	0	0	26540	26334	74975	4411
	7	abcsuperprove*	465	332	75	257	133	106	0	0	27	17412	14089	23438	1361
	8	tip	465	331	76	255	134	134	0	0	0	9072	8986	4619	164
	9	pdtrav*	465	319	61	258	146	142	0	3	1	25321	25141	36538	1078
	10	ic3*	465	303	58	245	162	162	0	0	0	14377	14224	10783	497
	11	tarmo	465	175	71	104	290	287	3	0	0	6216	24312	33269	2318
	12	blimc	465	95	87	8	370	369	0	1	0	7014	6945	9383	1231
	13	tipbmc	465	84	84	0	381	296	44	0	41	2981	2946	5579	581
	14	abcbmc2*	465	83	83	0	382	354	28	0	0	4265	4230	9760	1009
	15	tarmobmc	465	74	74	0	391	311	80	0	0	2442	9532	20944	1972
	16	aigbmc	465	67	67	0	398	344	54	0	0	2870	2839	8362	1255

abcsuperprove* winner HWMCC'10 SAT+UNSAT

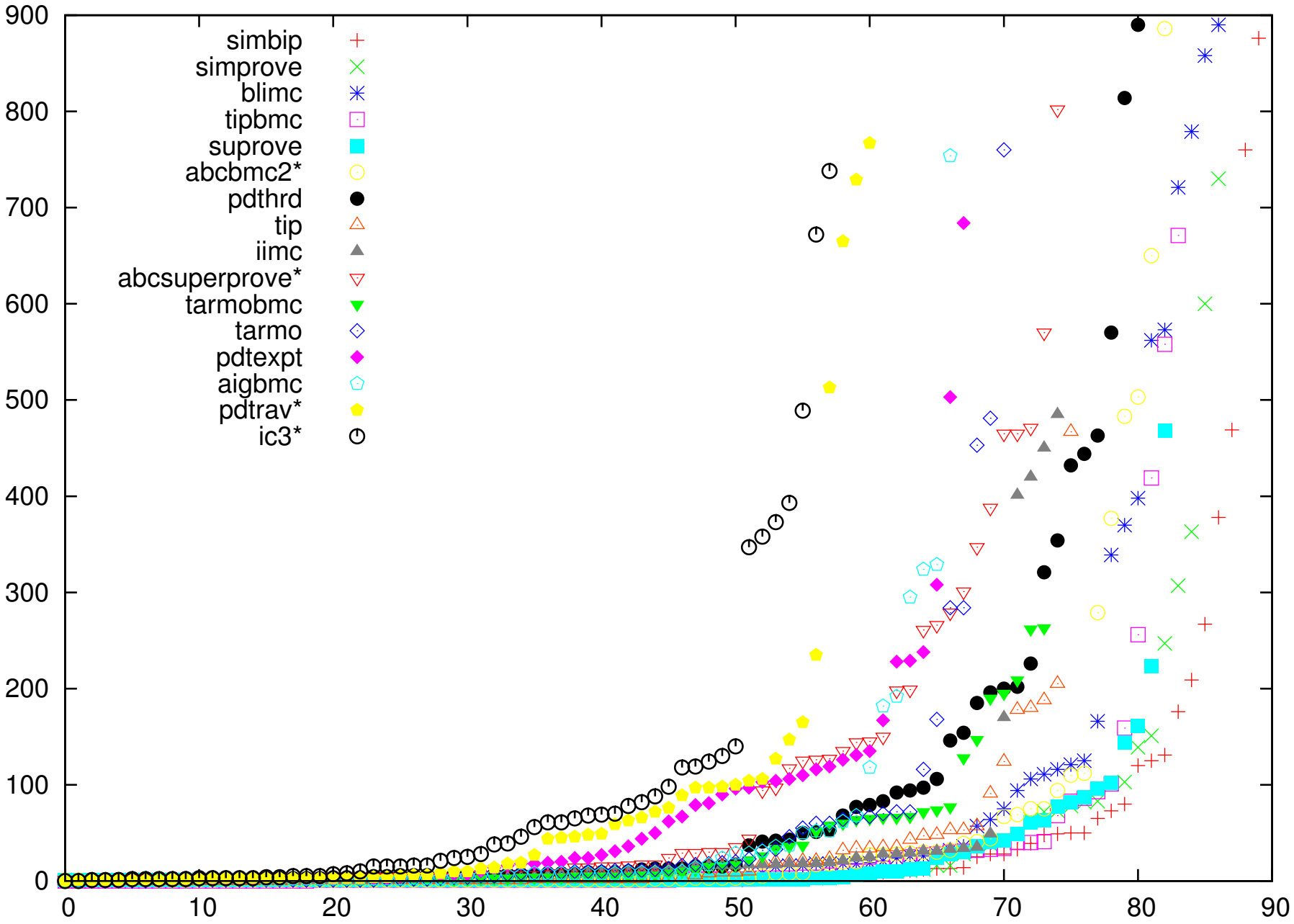

900
800
700
600
500
400
300
200
100
0
0
50
100
150
200
250
300
350
400
suprove
simprove
simbip
pdthrd
pdtexpt
iimc
abcsuperprove*
tip
pdtrav*
ic3*
tarmo
blimc
tipbmc
abcbmc2*
tarmobmc
aigbmc

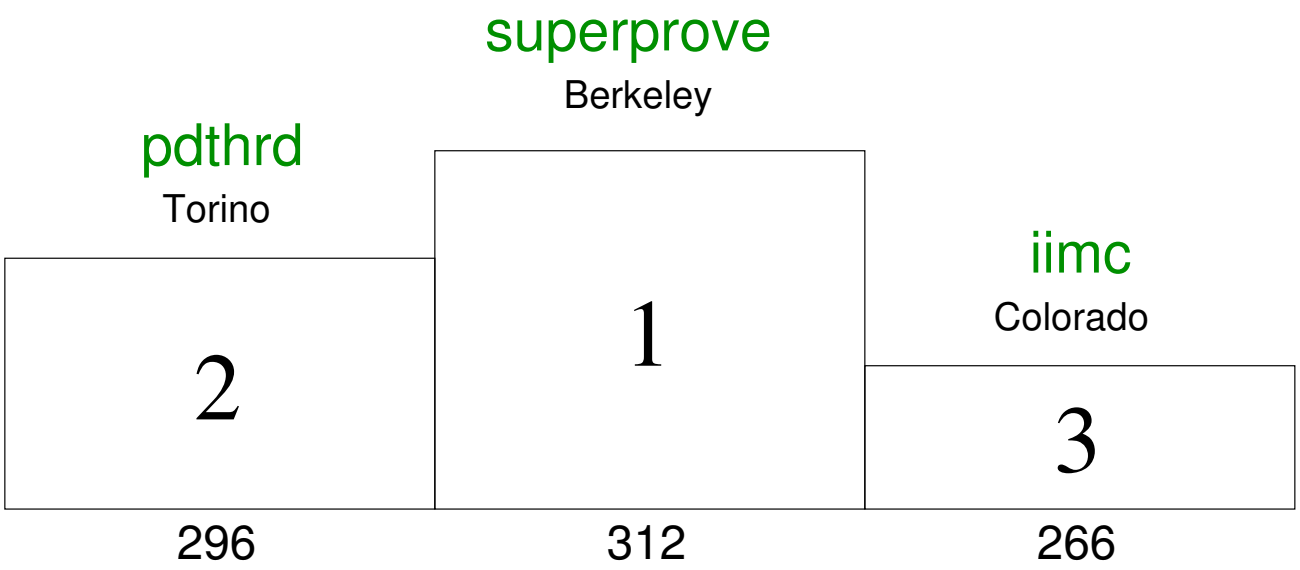

simplebip
Berkeley
blimc
Linz
tipbmc
Göteborg
2
1
3
87
90
84

SAT ranking (1st column per group, 2nd column per solver)

		solver	fnd	ok	SAT	uns	fld	to	mo	s6	uk	real	time	space	max
A	1	simbip	465	381	90	291	84	84	0	0	0	4267	13490	32419	2366
	2	simprove	465	383	87	296	82	82	0	0	0	3546	11224	35996	3299
B	3	blimc	465	95	87	8	370	369	0	1	0	7012	6944	9376	1231
C	4	tipbmc	465	84	84	0	381	296	44	0	41	2981	2946	5579	581
	5	suprove	465	395	83	312	70	70	0	0	0	1943	5101	28922	3298
	6	abcbmc2*	465	83	83	0	382	354	28	0	0	4265	4230	9760	1009
	7	pdthrd	465	377	81	296	88	74	0	9	0	6922	15208	28533	1233
	8	tip	465	331	76	255	134	134	0	0	0	2120	2097	936	68
	9	iimc	465	341	75	266	124	124	0	0	0	2621	2603	8976	2028
	10	abcsuperprove*	465	332	75	257	133	106	0	0	27	6728	5880	8727	1361
	11	tarmobmc	465	74	74	0	391	311	80	0	0	2442	9532	20944	1972
	12	tarmo	465	175	71	104	290	287	3	0	0	3527	13825	20851	2318
	13	pdtexpt	465	349	68	281	116	106	0	10	0	4413	4389	9402	884
	14	aigbmc	465	67	67	0	398	344	54	0	0	2870	2839	8362	1255
	15	pdtrav*	465	319	61	258	146	142	0	3	1	4796	4761	8491	1078
	16	ic3*	465	303	58	245	162	162	0	0	0	5169	5141	2019	159

abcbmc2* winner HWMCC'10 SAT

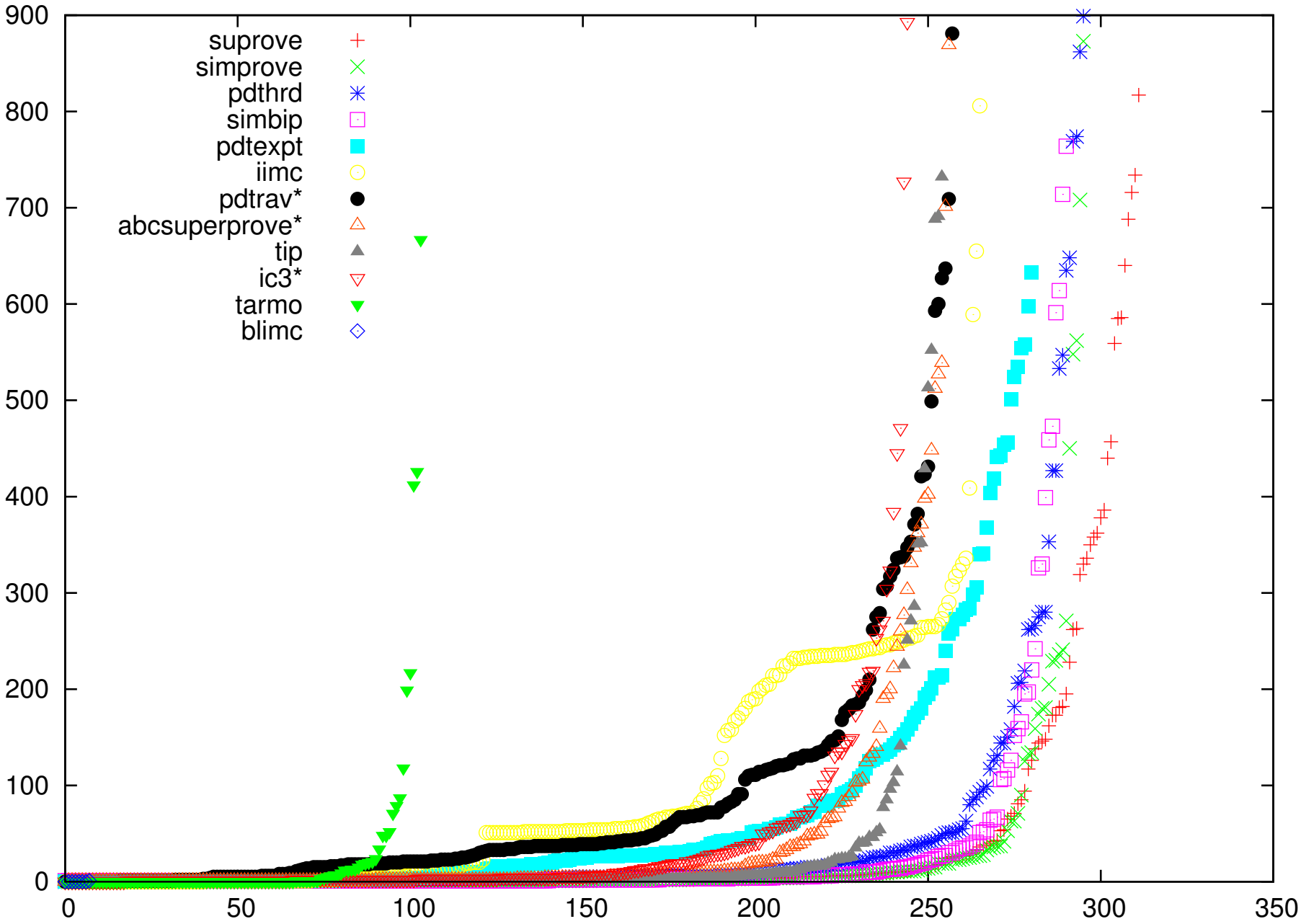

simbip
simprove
blimc
tipbmc
suprove
abcbmc2*
pdthrd
tip
iimc
abcsuperprove*
tarmobmc
tarmo
pdtexpt
aigbmc
pdtrav*
ic3*
0
100
200
300
400
500
600
700
800
900
0
10
20
30
40
50
60
70
80
90


superprove
Berkeley
pdthrd
Torino
iimc
Colorado
2
1
3
296
312
266

UNSAT ranking (1st column per group, 2nd column per solver)

		solver	fnd	ok	sat	UNS	fld	to	mo	s6	uk	real	time	space	max
A	1	suprove	465	395	83	312	70	70	0	0	0	13438	25803	66129	3406
	2	simprove	465	383	87	296	82	82	0	0	0	7151	23980	57856	3266
B	3	pdthrd	465	377	81	296	88	74	0	9	0	12981	36527	73750	3219
	4	simbip	465	381	90	291	84	84	0	0	0	7967	27419	51319	3074
	5	pdtexpt	465	349	68	281	116	106	0	10	0	18782	18662	25543	779
C	6	iimc	465	341	75	266	124	124	0	0	0	23919	23731	65999	4411
	7	pdtrav*	465	319	61	258	146	142	0	3	1	20525	20380	28047	1065
	8	abcsuperprove*	465	332	75	257	133	106	0	0	27	10684	8209	14711	652
	9	tip	465	331	76	255	134	134	0	0	0	6952	6889	3683	164
	10	ic3*	465	303	58	245	162	162	0	0	0	9209	9083	8763	497
	11	tarmo	465	175	71	104	290	287	3	0	0	2689	10486	12419	1186
	12	blimc	465	95	87	8	370	369	0	1	0	2	1	7	4
	13	tipbmc	465	84	84	0	381	296	44	0	41	0	0	0	0
	14	aigbmc	465	67	67	0	398	344	54	0	0	0	0	0	0
	15	tarmobmc	465	74	74	0	391	311	80	0	0	0	0	0	0
	16	abcbmc2*	465	83	83	0	382	354	28	0	0	0	0	0	0

pdtrav* winner HWMCC'10 UNSAT


suprove
simprove
pdthrd
simbip
pdtexpt
iimc
pdtrav*
abcsuperprove*
tip
ic3*
tarmo
blimc
900
800
700
600
500
400
300
200
100
0
0
50
100
150
200
250
300
350

Conclusion

- achievements this year
  - new AIGER 1.9 and new tracks
  - new benchmarks, new versions, new model checkers
  - state-of-the-art improved in all previous categories
  - IC3/PDR has been integrated in leading model checkers
  - clear trend towards parallel implementations (portfolio as this point)
- next time
  - (please) provide witnesses
  - AIGER 2.0 with new binary encoding, same semantics
  - more benchmarks, particularly multi & live