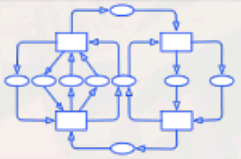
Model Checking Contest

Report for 2012

Fabrice Kordon - LIP6/MoVe, UPMC, France
Alban Linard - CUI/SMV, Univ. Genève, Switzerland
Franck Pommereau - IBISC, Univ. Evry Val d'Essonne

MCC 2012

Model Checking Contest @



Contents

Contents

Special thanks for those who helped to organize this MCC, in particular Nicolas Gibelin (Cluster), Lom Hillah (PNML), Emmanuel Paviot-Adet (models)

MCC 2012

Objectives

When it Comes to Deal with Large and Complex Systems...

- Lots of questions are raised...
  - To verify highly concurrent systems, should we use a symmetry-based or a partial order-based model checker?
  - For models with large variable domains, should we use decision diagram-based, or a symmetry-based model checker?
  - Can we combine structural reductions techniques with partial-order ones or symmetry-based ones?
  - ...
- A large variety of model checking techniques
  - and their potential combination
- A large variety of model categories
- A challenge with large scale specifications
- A need to evaluate in the fairest way current MC implementations

The Objectives...

- MCC is intended to:
  - Exchange experience between tool programmers,
  - Imagine some association of techniques, and thus better tools
  - Stimulate development of tools
  - Provide visibility to these tools
- MCC can also be of great help for the PN community (and users):
  - Define a common set of models for benchmarks
  - Identify experimentally classes of problems (in models)
    - identify the techniques able to cope with a given class of problems...
  - Improve communication between tools (and PNML ;-))
  - Provides raw data for comparison
- This is the second edition
  - We hope more editions for an enhanced analysis and evaluation of tools

MCC 2012

Evaluation Procedure

What to be measured?

- The «enemies» of model checking
  - Memory consumption
  - CPU consumption
- «Examinations» to be processed
  - State space generation
  - Formula evaluation
    - Structural Formulas
    - Reachability Formulas
    - CTL formulas
    - LTL formulas
- Another 2012 innovation
  - Models to be proposed by the community («call for model»)
    - 7 models in 2011
    - 19 models in 2012 (including the 7 from 2011)

What to be measured?

- The «enemies» of model checking
  - Memory consumption
  - CPU consumption
- «Examinations» to be processed
  - State space generation
  - Formula evaluation
    - Structural Formulas
    - Reachability Formulas
    - CTL formulas
    - LTL formulas

- Another 2012 innovation
  - Models to be proposed by the community («call for model»)
    - 7 models in 2011
    - 19 models in 2012 (including the 7 from 2011)

What to be measured?

- The «enemies» of model checking
  - Memory consumption
  - CPU consumption
- «Examinations» to be processed
  - State space generation
  - Formula evaluation
    - Structural Formulas
    - Reachability Formulas
    - CTL formulas
    - LTL formulas
- Another 2012 innovation
  - Models to be proposed by the community («call for model»)
    - 7 models in 2011
    - 19 models in 2012 (including the 7 from 2011)

Special thanks for the community who provided interesting models

12 new models coming from 5 institutions

Univ. Evry Val d'Essone, France
Univ. Geneva, Switzerland
Univ. P. & M. Curie France
Univ. Paris 13, France
Univ. Rostock, Germany

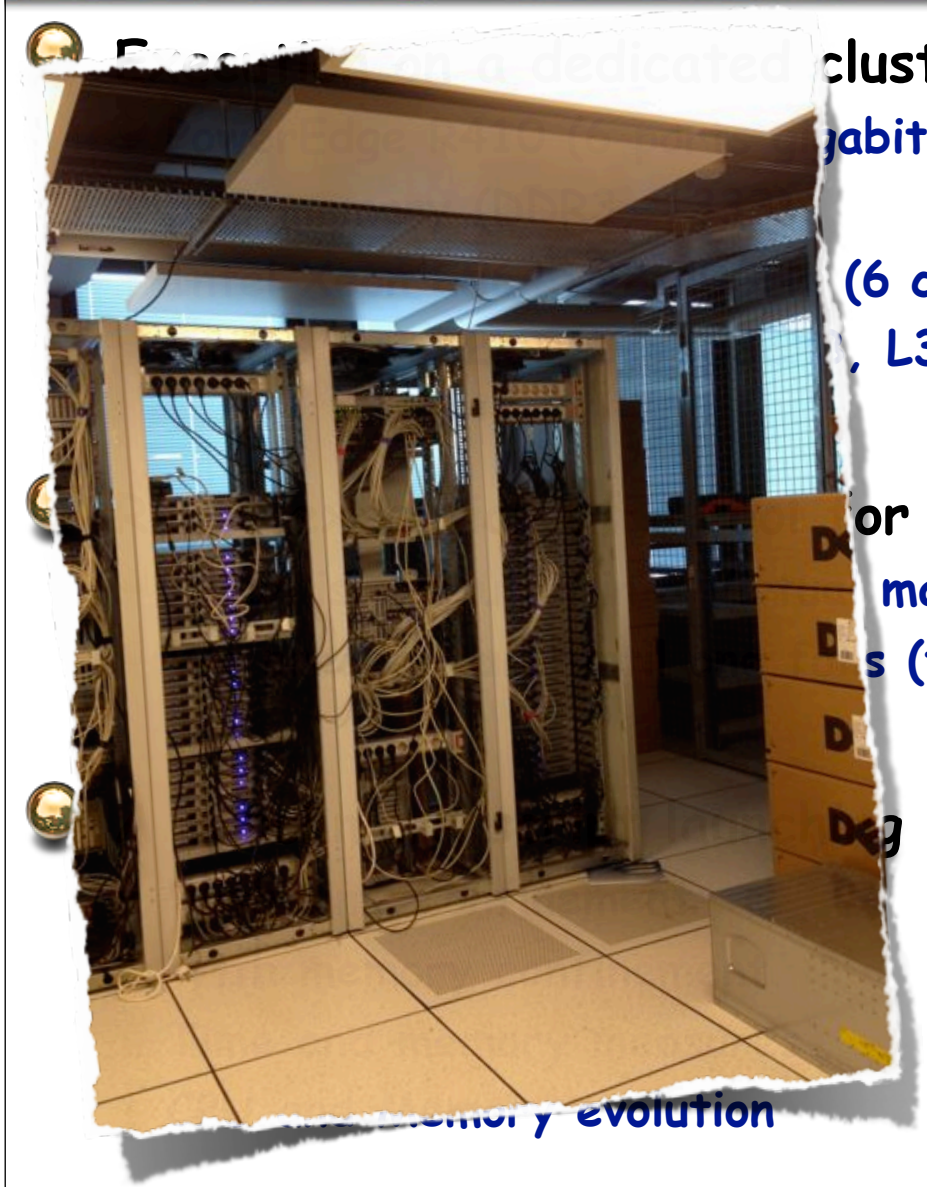
Evaluation procedure

- Execution on a dedicated cluster (23 nodes)
  - PowerEdge R410 (6 ports gigabits) and 1.5To local disks
  - 8GB memory (DDR3, 1333)
  - Intel Xeon E5645@2.40GHz (6 cores, 12 threads)
  - Cache L1=192kB, L2=1536kB, L3=12288kB

- Run = execution of a tool for one examination on one model/scale
  - A run is executed in a Virtual machine
  - We process runs until one fails (to check how far a tool goes)

- A benchmark script launching all runs
  - With time confinement 3600 sec per run
  - With memory confinement 4 GByte per run
  - Time and memory measures
  - CPU and Memory evolution

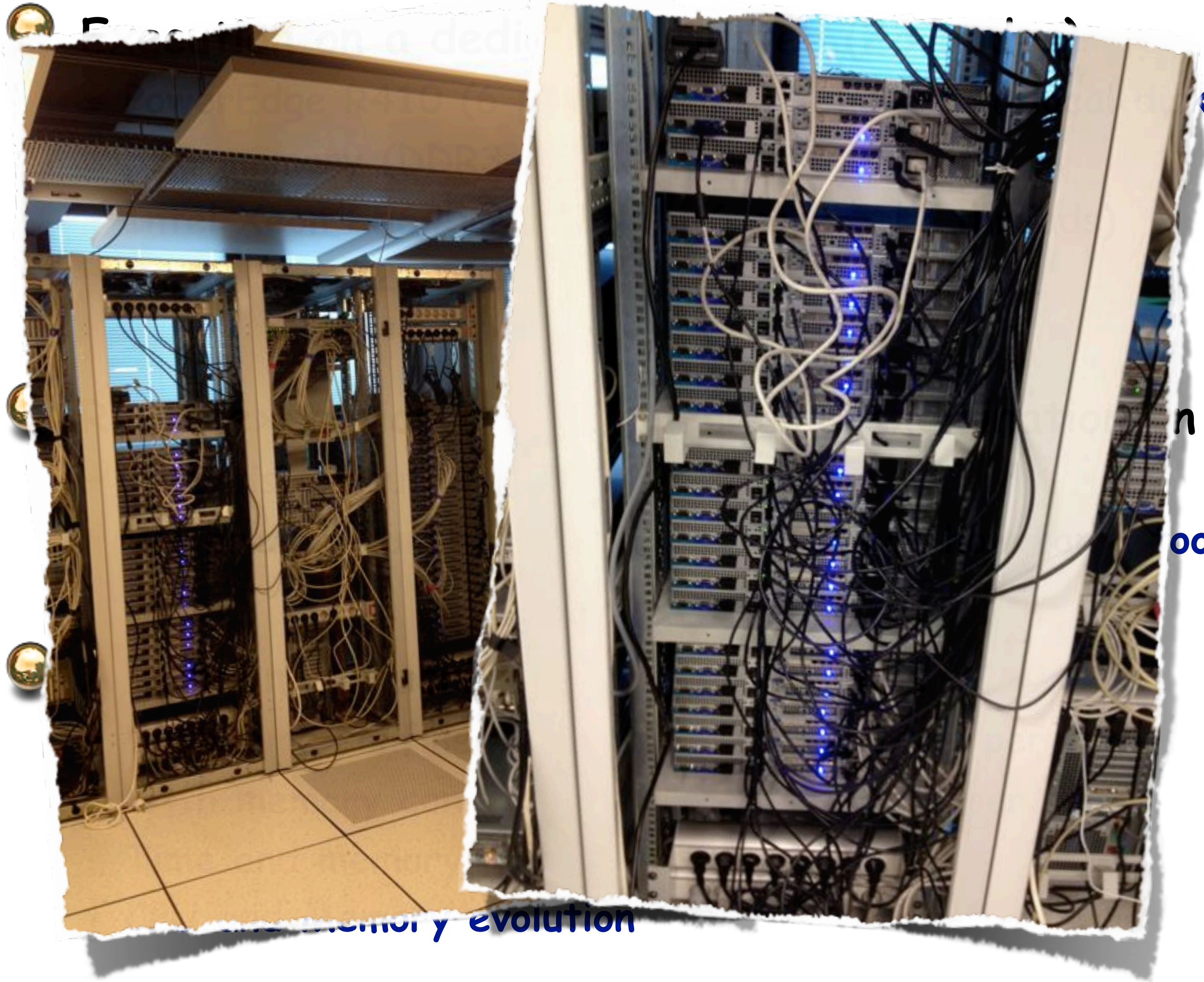
Evaluation procedure



- cluster (23 nodes)
  - gabits) and 1.5To local disks
  - (6 cores, 12 threads)
  - , L3=12288kB
- or one examination on one model/scale
  - machine
  - s (to check how far a tool goes)
- g all runs
  - 3600 sec per run
  - 4 GByte per run

memory evolution

Evaluation procedure



ks

n one model/scale

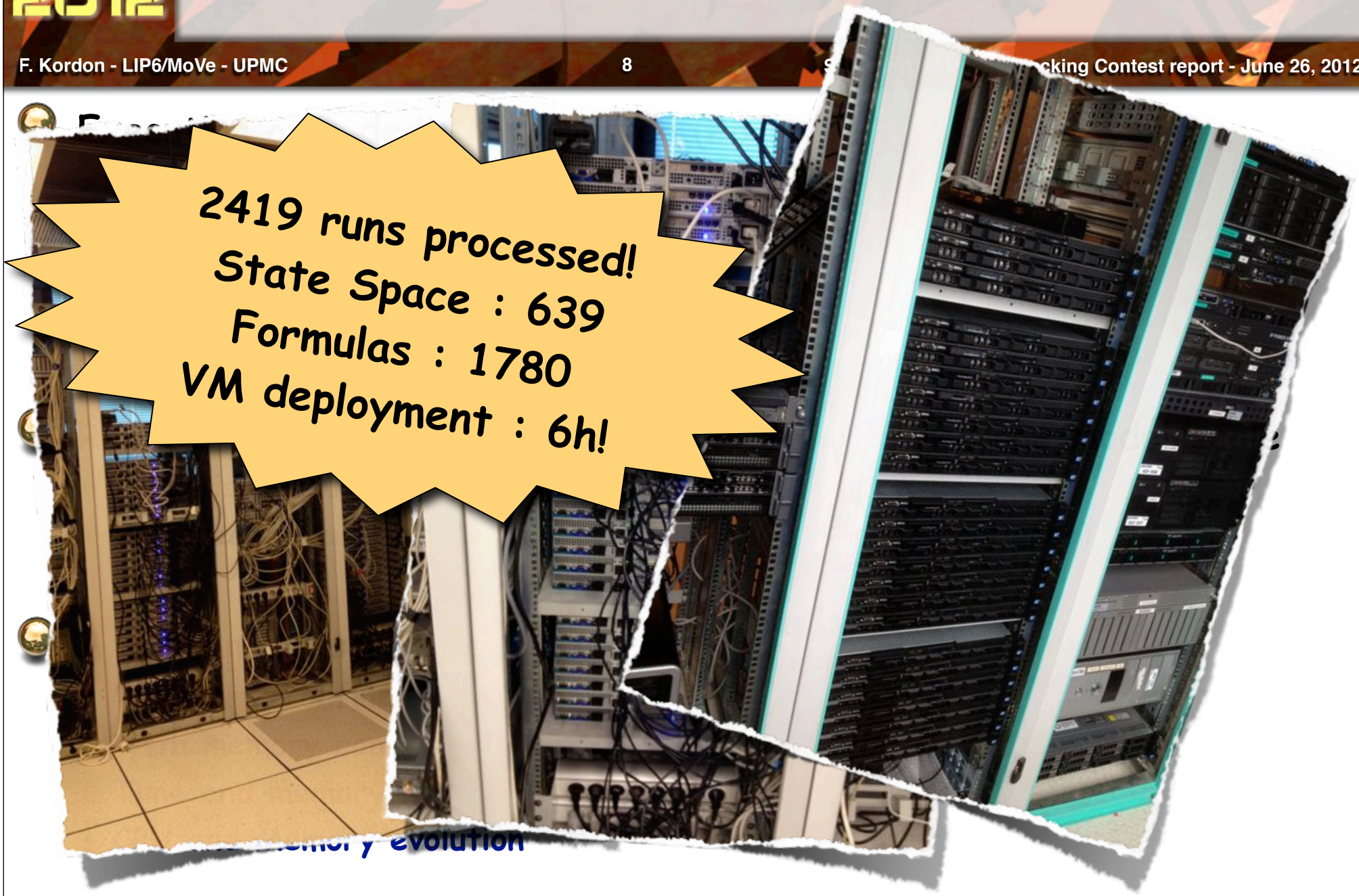
ool goes)

emory evolution

Evaluation procedure

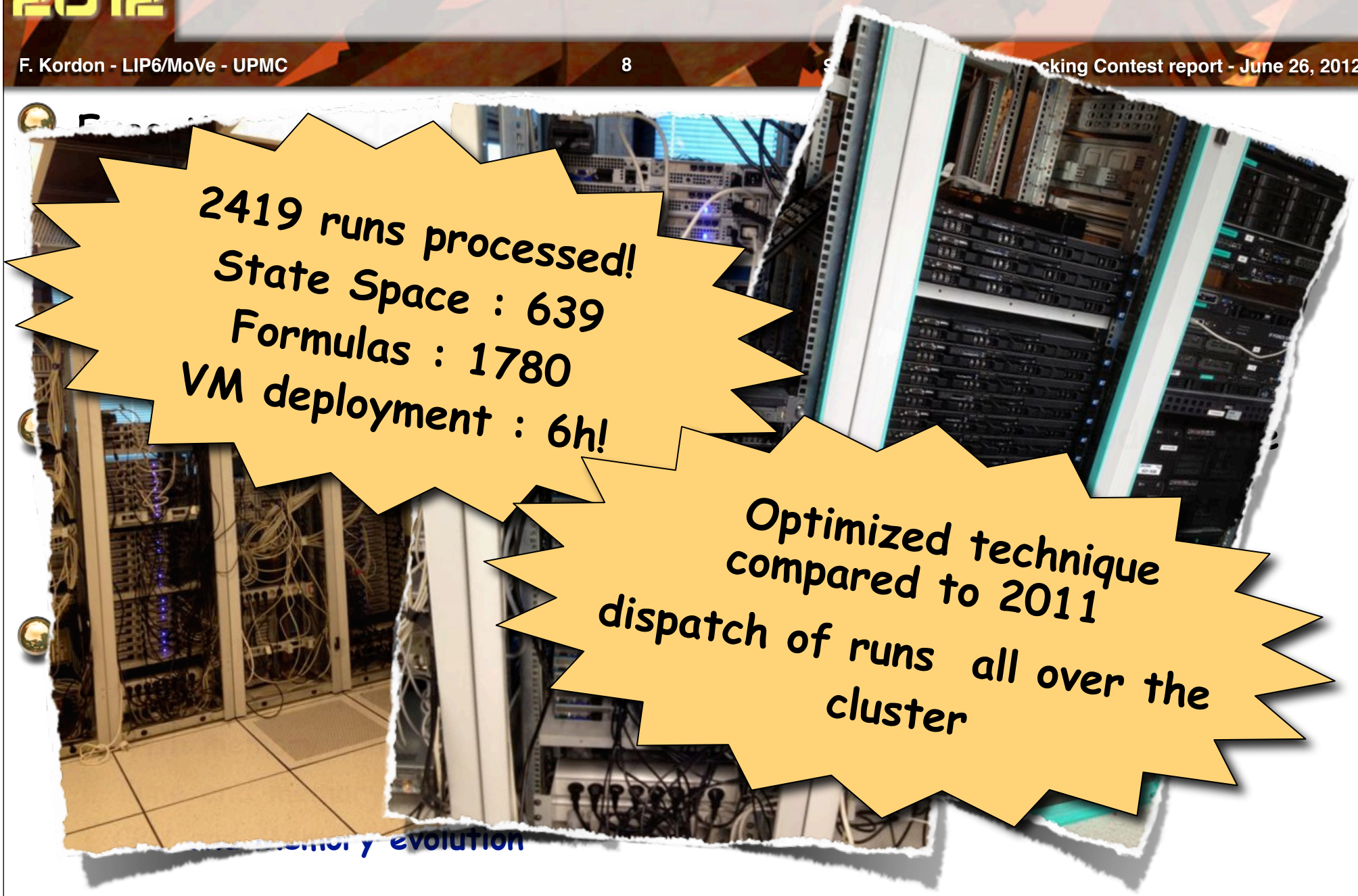
... evolution

Evaluation procedure



- Memory evolution

Evaluation procedure



Difficulties

- The Cluster
  - Was delivered later than expected
    - Old nodes could not operate virtualization
- The formulas
  - Last year solution was not satisfactory
    - Based on invariants
    - Too «easy» formulas
    - One set per model
  - This year solution
    - One set per run
    - Two formats, XML and textual (update of the grammar)
    - But...

... a nightmare 

- Other technical difficulties
  - Fighting with qemu
  - Change of structure for formulas
  - provide PNML form for submitted models

MCC 2012

The Models

Presentation of the Models

	Model Name	model type	safe	dead-lock	free choice	state machine	event graph	rever-sible
Models proposed for 2012	cs_repetitions	colored + P/T	✗	✗	?	?	?	?
	rwmutex	colored + P/T	✓	✗	✗	✗	✗	✓
	echo	colored + P/T	✓	✓	✗	✗	✗	✗
	eratosthenes	colored + P/T	✓	✓	?	?	?	?
	galloc_res	colored + P/T	✗	✗	?	?	?	?
	lamport_fmea	colored + P/T	✓	✗	✗	✗	✗	✗
	neoelection	colored + P/T	✓	✓	✗	✗	✗	✓
	philo_dyn	colored + P/T	✓	✓	?	?	?	?
	planning	colored + P/T	✗	?	✗	✗	?	?
	railroad	colored + P/T	✗	✗	?	?	?	?
	ring	colored + P/T	✓	?	✗	✗	?	?
	simple_lbs	colored + P/T	✓	✗	✗	✗	✗	✗
Models from 2011	FMS	P/T	✗	?	?	?	?	?
	Kanban	P/T	✗	?	?	?	?	?
	MAPK	P/T	✗	?	?	?	?	?
	Peterson	colorred	✓	?	?	?	?	?
	Philosophers	colorred	✓	?	?	?	?	?
	SharedMemory	colorred	✓	✗	?	?	?	?
	TokenRing	colorred	✓	?	?	?	?	?

Presentation of the Models

	Model Name	model type	safe	dead-lock	free choice	state machine	event graph	rever-sible
Models proposed for 2012	cs_repetitions	colored + P/T	✗	✗	?	?	?	?
	rwmutex	colored + P/T	✓	✗	✗	✗	✗	✓
	echo	colored + P/T	✓	✓	✗	✗	✗	✗
	eratosthenes	colored + P/T	✓	✓	?	?	?	?
	galloc_res	colored + P/T	✗	✗	?	?	?	?
	lamport_fmea	colored + P/T	✓	✗	✗	✗	✗	✗
	neoelection	colored + P/T	✓	✓	✗	✗	✗	✓
	philo_dyn	colored + P/T	✓	✓	?	?	?	?
	planning	colored + P/T	✗	?	✗	✗	?	?
	railroad	colored + P/T	✗	✗	?	?	?	?
	ring	colored + P/T	✓	?	✗	✗	?	?
	simple_lbs	colored + P/T	✓	✗	✗	✗	✗	✗
Models from 2011	FMS	P/T	✗	?	?	?		
	Kanban	P/T	✗	?	?	?		
	MAPK	P/T	✗	?	?	?		
	Peterson	colorred	✓	?	?	?		
	Philosophers	colorred	✓	?	?	?		
	SharedMemory	colorred	✓	✗	?	?	?	
	TokenRing	colorred	✓	?	?	?	?	?

Diversification of characteristics

MCC 2012

Participating Tools

The Submissions (order of arrival ;-)

#	Tool	from
1	lola-binstore	Univ. Rostock
2	lola-bloom	Univ. Rostock
3	sara	Univ. Rostock
4	neco	Univ. Evry Val D'essonne
5	pnxdd	Univ. P. & M. Curie
6	marcie	Univ. Cottbus
7	helena	Univ. Paris 13
8	AlPiNa	Univ Geneva
9	crocodile	Univ. P. & M. Curie
10	ITS-tools	Univ. P. & M. Curie

The Submissions (order of arrival ;-)

#	Tool	from
1	lola-binstore	Univ. Rostock
2	lola-bloom	Univ. Rostock
3	sara	Univ. Rostock
4	neco	Univ. Evry Val D'essonne
5	pnxdd	Univ. P. & M. Curie
6	marcie	Univ. Cottbus
7	helena	Univ. Paris 13
8	AlPiNa	Univ Geneva
9	crocodile	Univ. P. & M. Curie
10	ITS-tools	Univ. P. & M. Curie

The Submissions (order of arrival ;-)

#	Tool	from
1	lola-binstore	Univ. Rostock
2	lola-bloom	Univ. Rostock
3	sara	Univ. Rostock
4	neco	Univ. Evry Val D'essonne
5	pnxdd	Univ. P. & M. Curie
6	marcie	Univ. Cottbus
7	helena	Univ. Paris 13
8	AlPiNa	Univ Geneva
9	crocodile	Univ. P. & M. Curie
10	ITS-tools	Univ. P. & M. Curie

Provided in their VM

Nice hotline too ;-)

Participating tools: Supported techniques

- State space generation
  - Decision Diagrams (any kind)
  - Explicit representation of the state space
  - Exploitation of the system symmetries
  - Use of «topological» information (syphon, traps,invariant, etc)
- Formula evaluation
  - Abstractions (on the fly state elimination)
  - Decision Diagrams (any kind)
  - Explicit representation of the state space
  - Use of a constraint solver (SAT, SMT)
  - Use of structural reduction (Berthelot's, Haddad's, etc.)
  - Use of Partial order techniques

Participating tools: Supported techniques

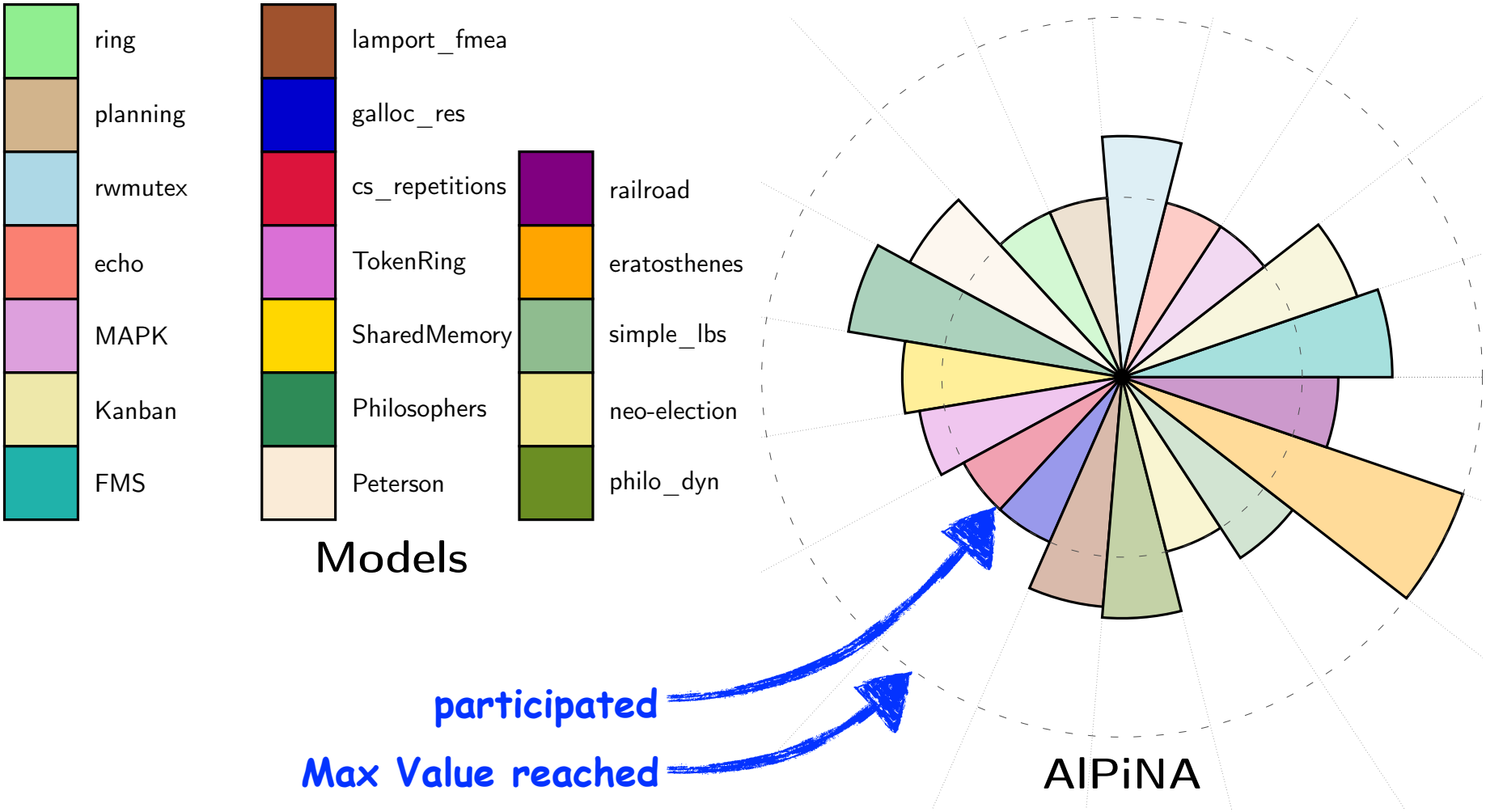
- State space generation
  - Decision Diagrams (any kind)
  - Exp
  - Exp
  - Us
- Forn
  - Ab
  - D
  - E
  - U
  - U
  - U

Also a combination of such techniques

- State space
  - ITS-Tool: Decision Diagrams + Symmetries
  - PNXDD, ITS-Tool: Decision Diagrams + Topological
- Formula evaluation
  - Lola*: Explicit + Partial Orders + Topological
  - Sara:Abstraction + SAT/SMT + Decision Diagrams

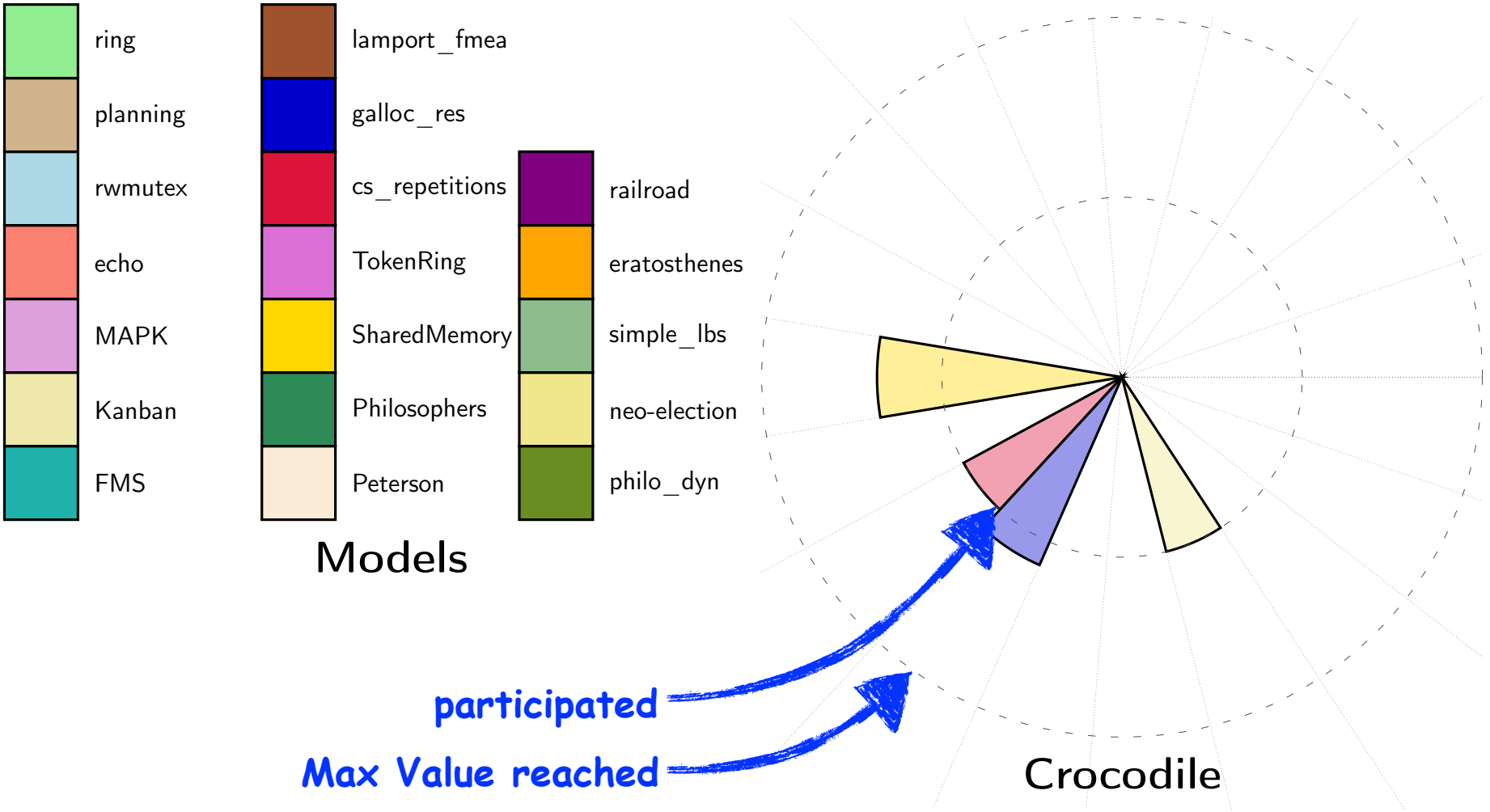
Models & examinations Processed by Tools (State Space)

LoLa* and Sara did not participated in the State Space generation



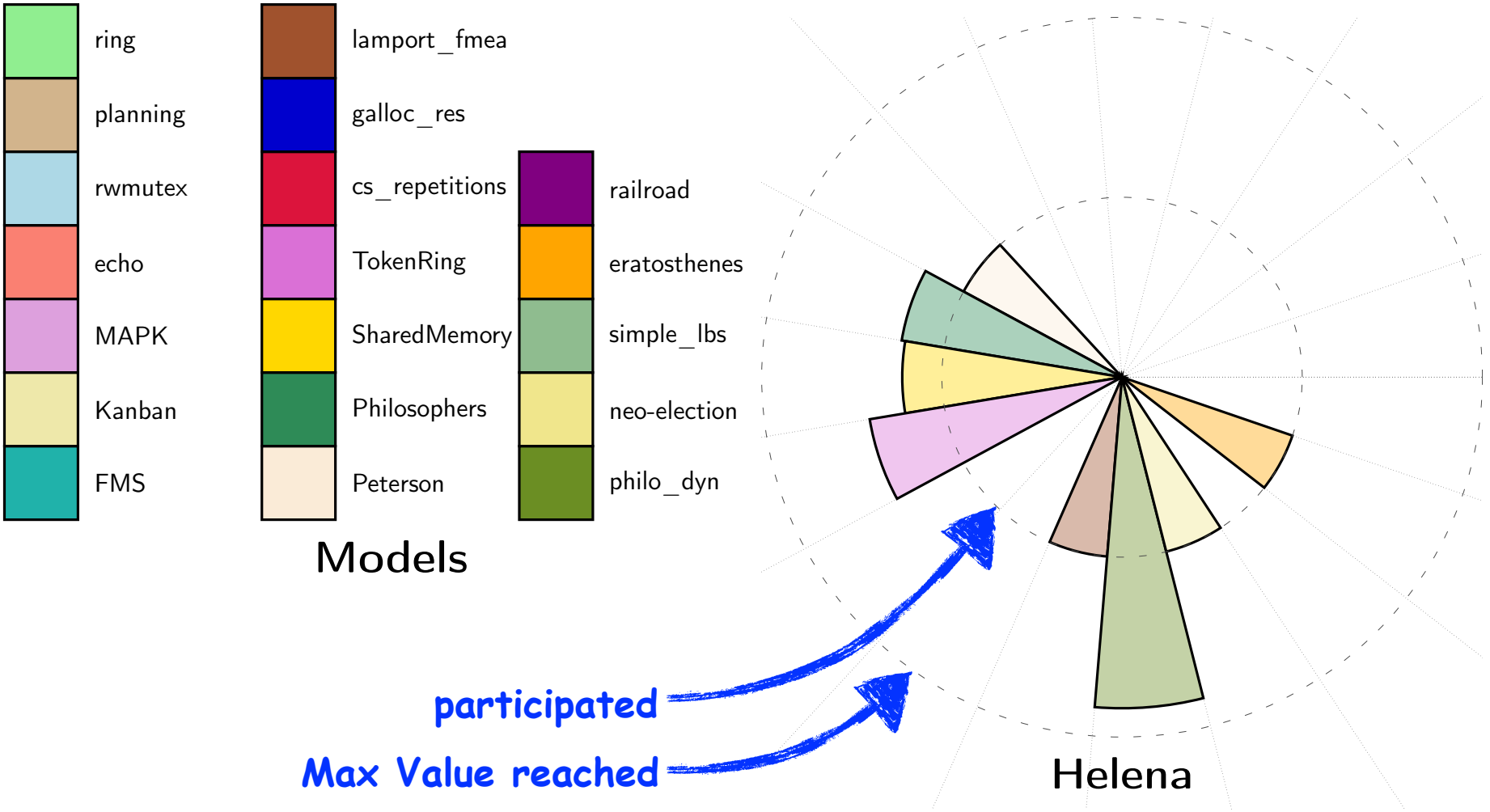
Models & examinations Processed by Tools (State Space)

LoLa* and Sara did not participated in the State Space generation



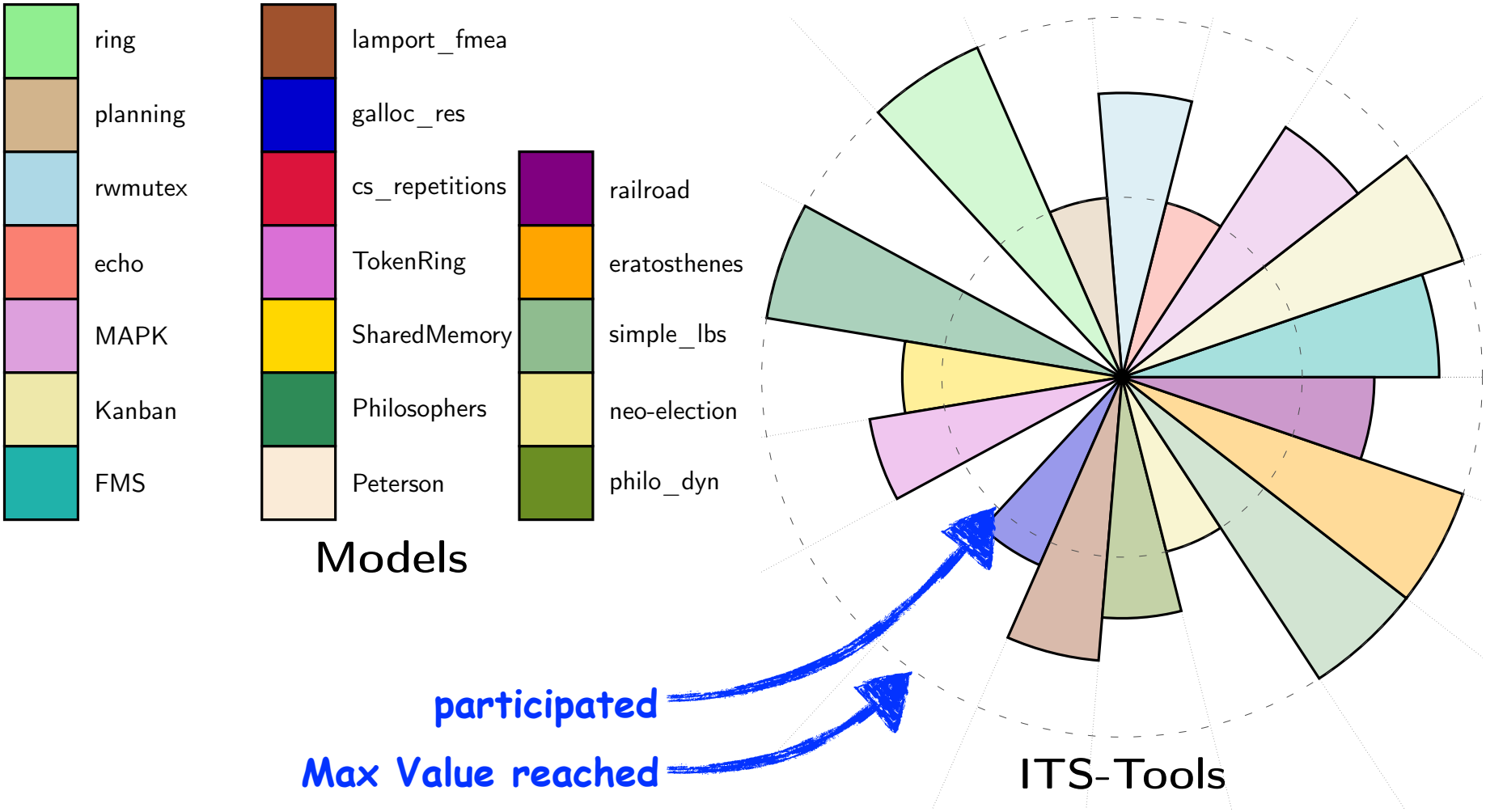
Models & examinations Processed by Tools (State Space)

LoLa* and Sara did not participated in the State Space generation



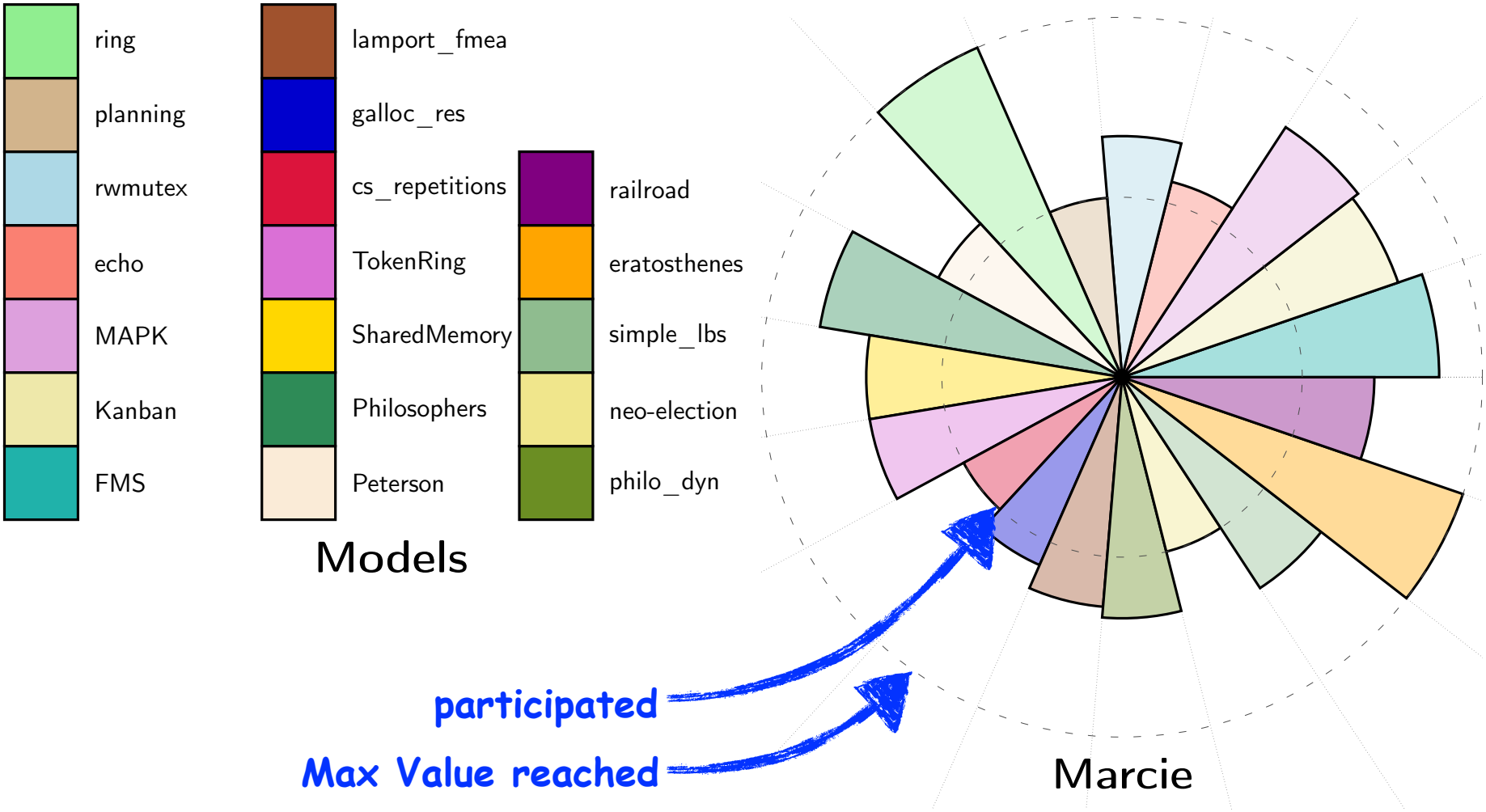
Models & examinations Processed by Tools (State Space)

LoLa* and Sara did not participated in the State Space generation



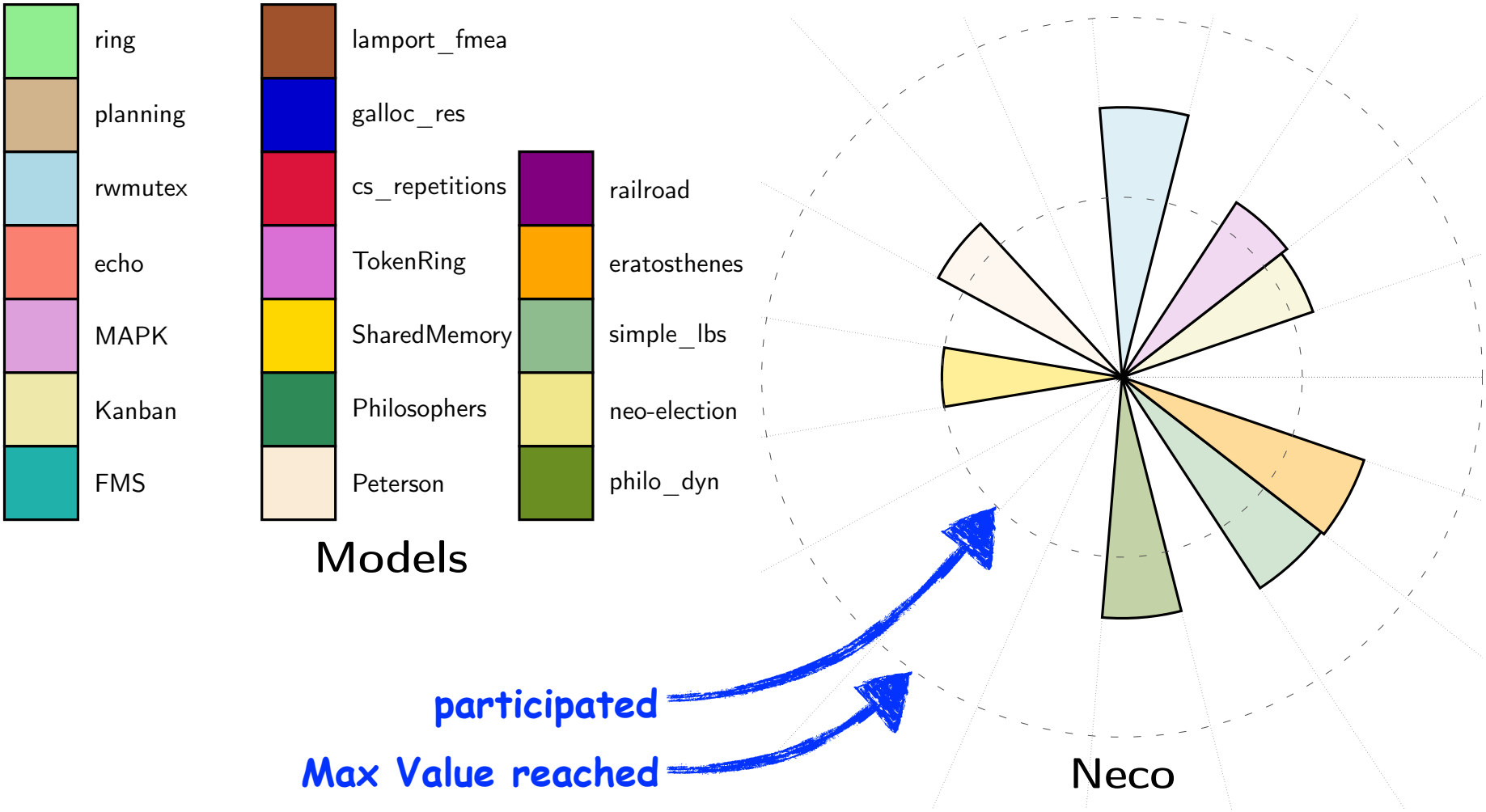
Models & examinations Processed by Tools (State Space)

LoLa* and Sara did not participated in the State Space generation



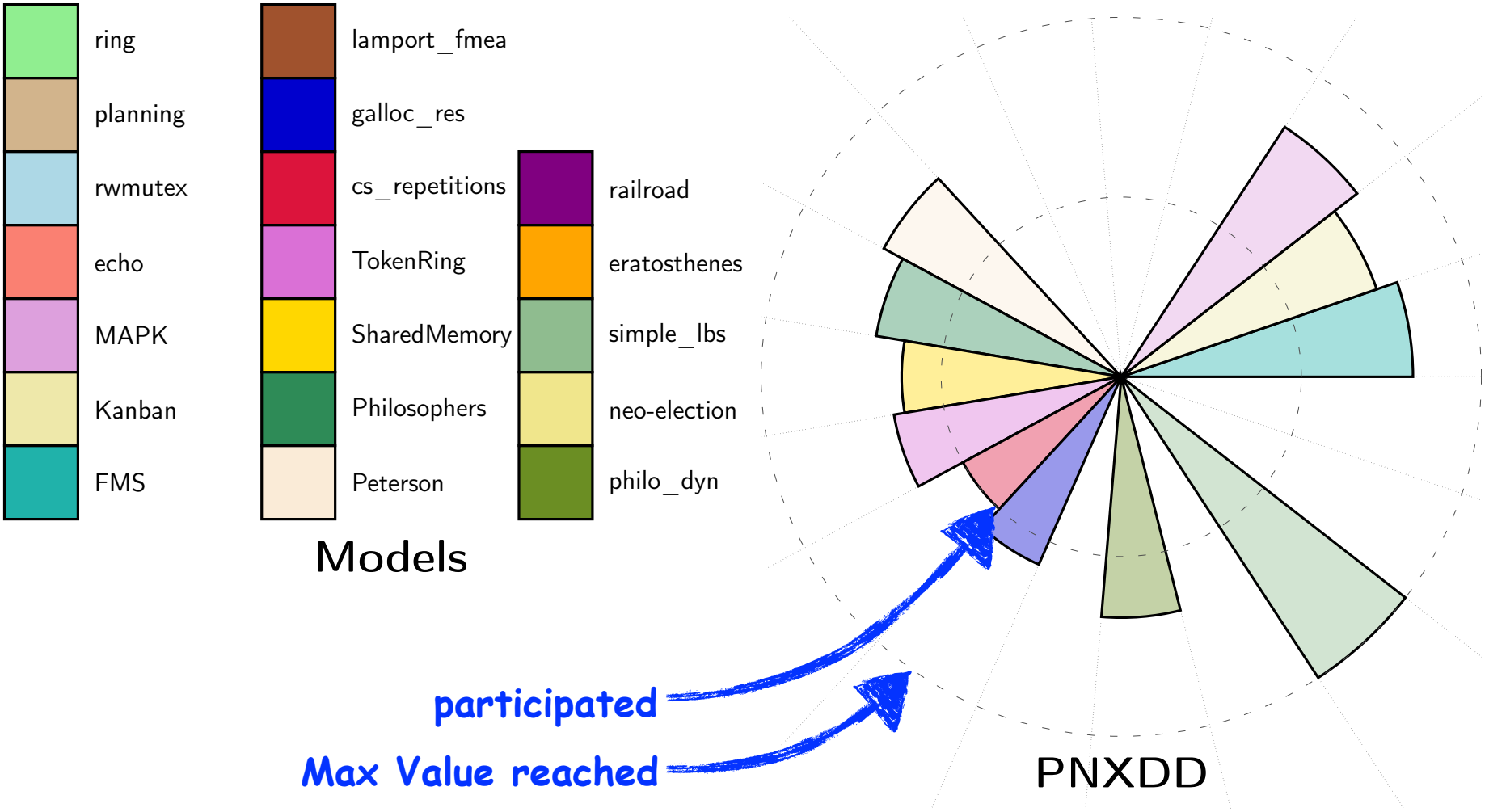
Models & examinations Processed by Tools (State Space)

LoLa* and Sara did not participated in the State Space generation



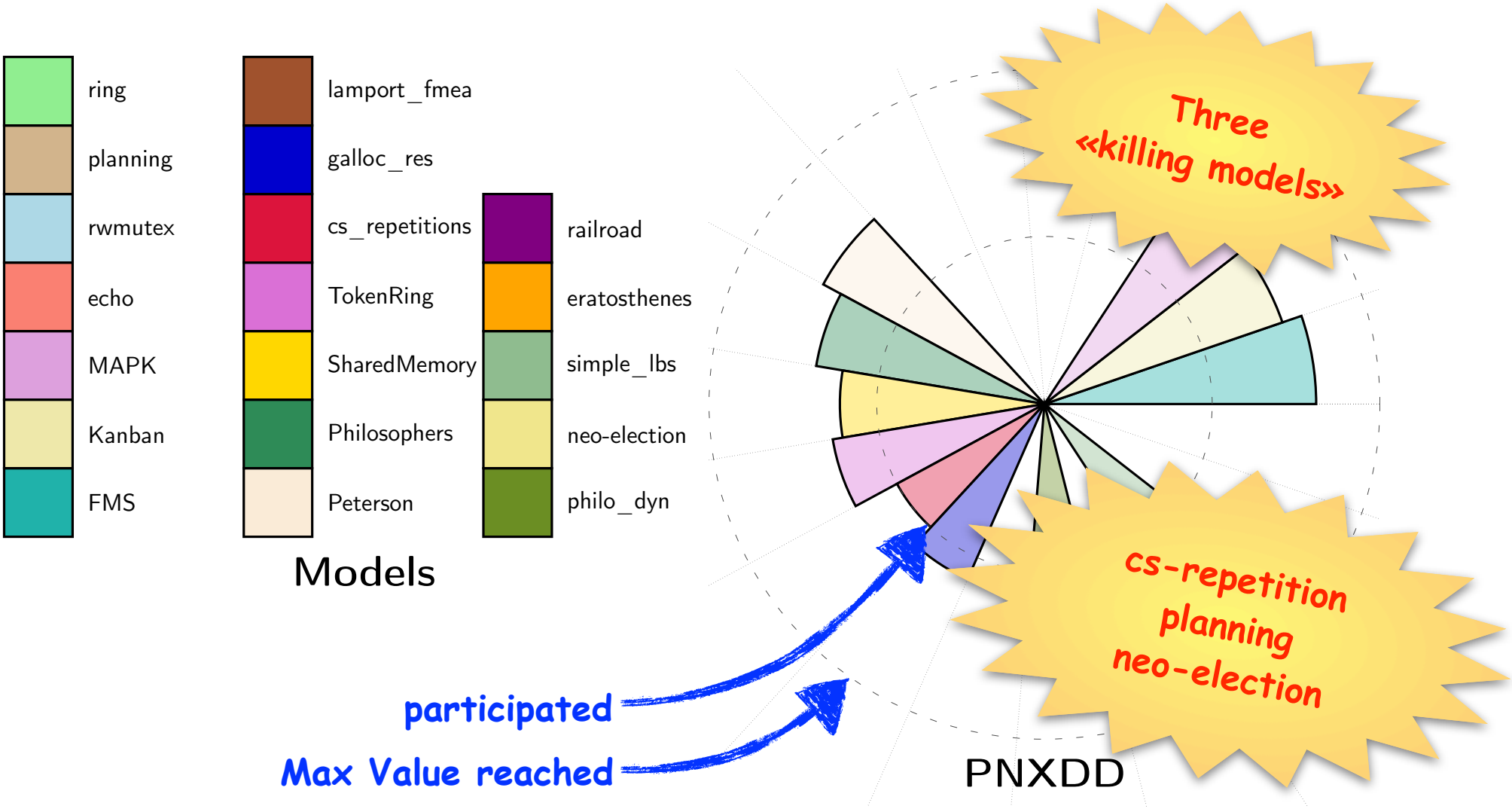
Models & examinations Processed by Tools (State Space)

LoLa* and Sara did not participated in the State Space generation



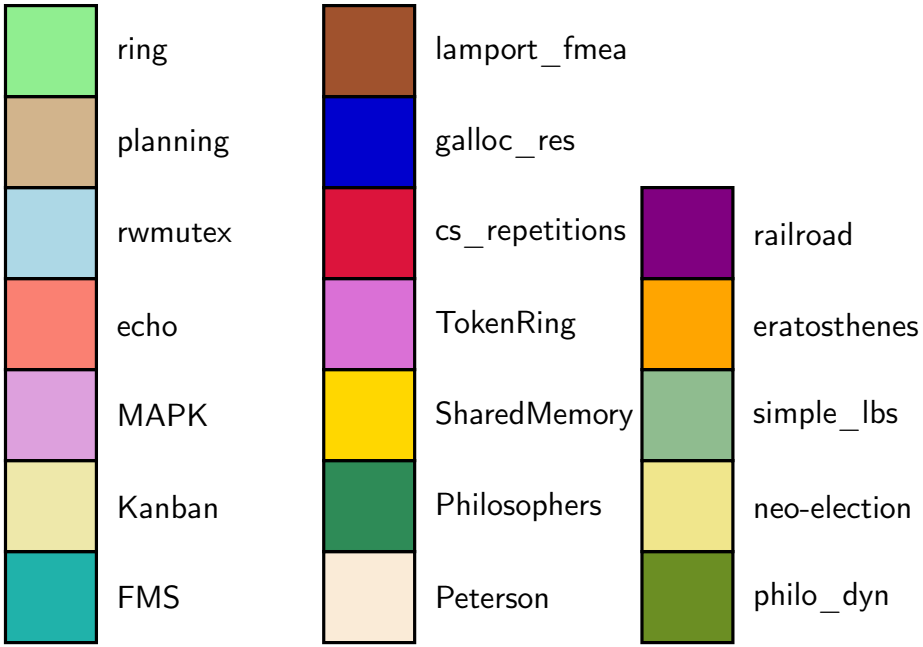
Models & examinations Processed by Tools (State Space)

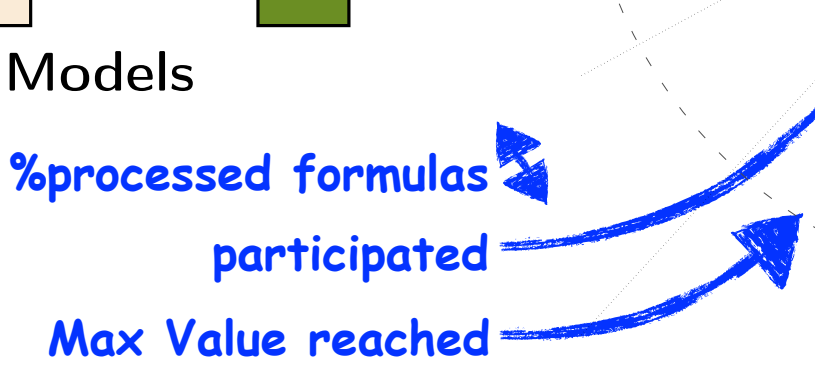
LoLa* and Sara did not participated in the State Space generation

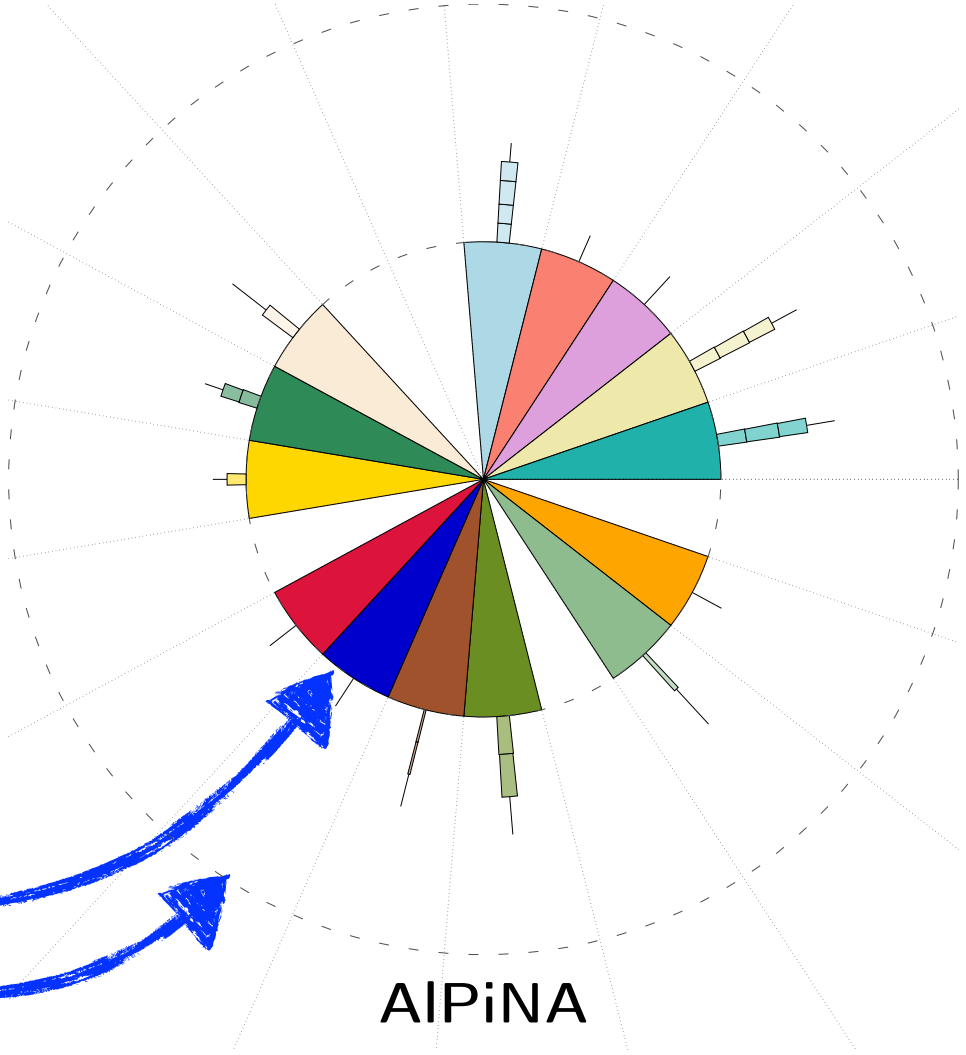


Models & examinations Processed by Tools (Reachability)

Only AlPiNa, Helena, LoLa* and Sara participated

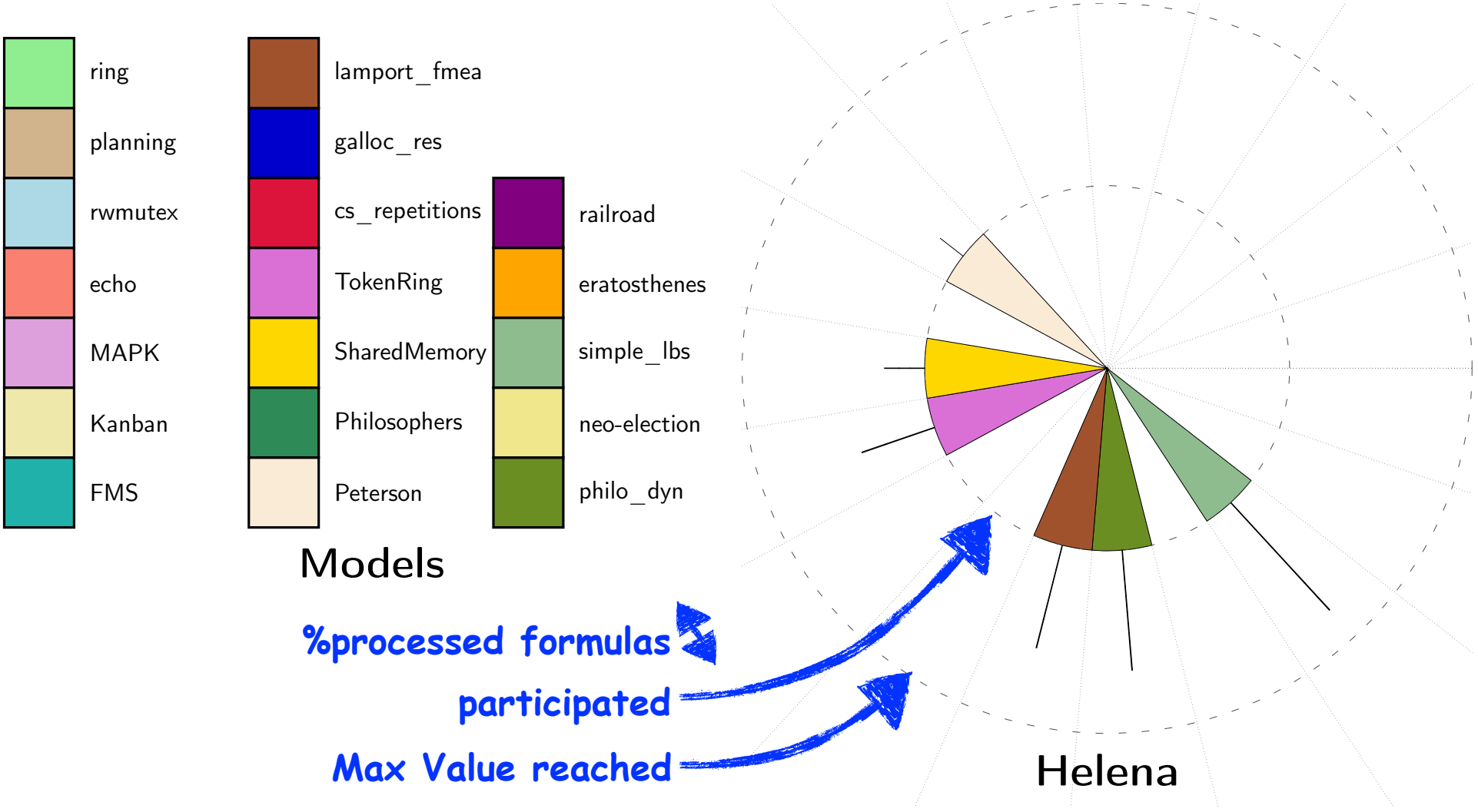






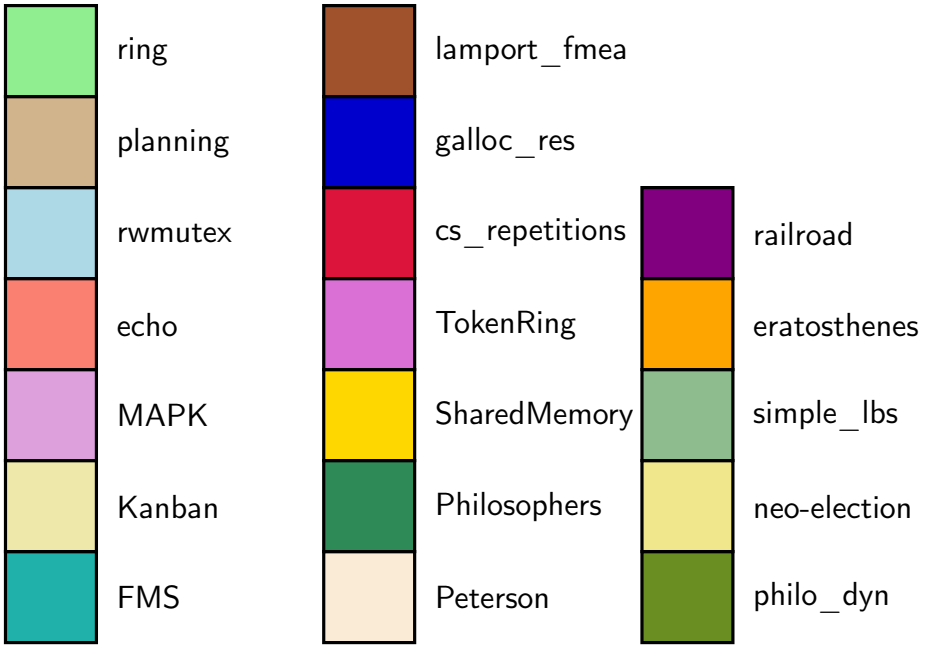
Models & examinations Processed by Tools (Reachability)

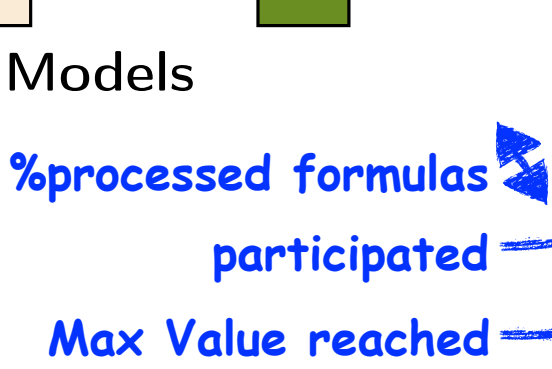
Only AlPiNa, Helena, LoLa* and Sara participated

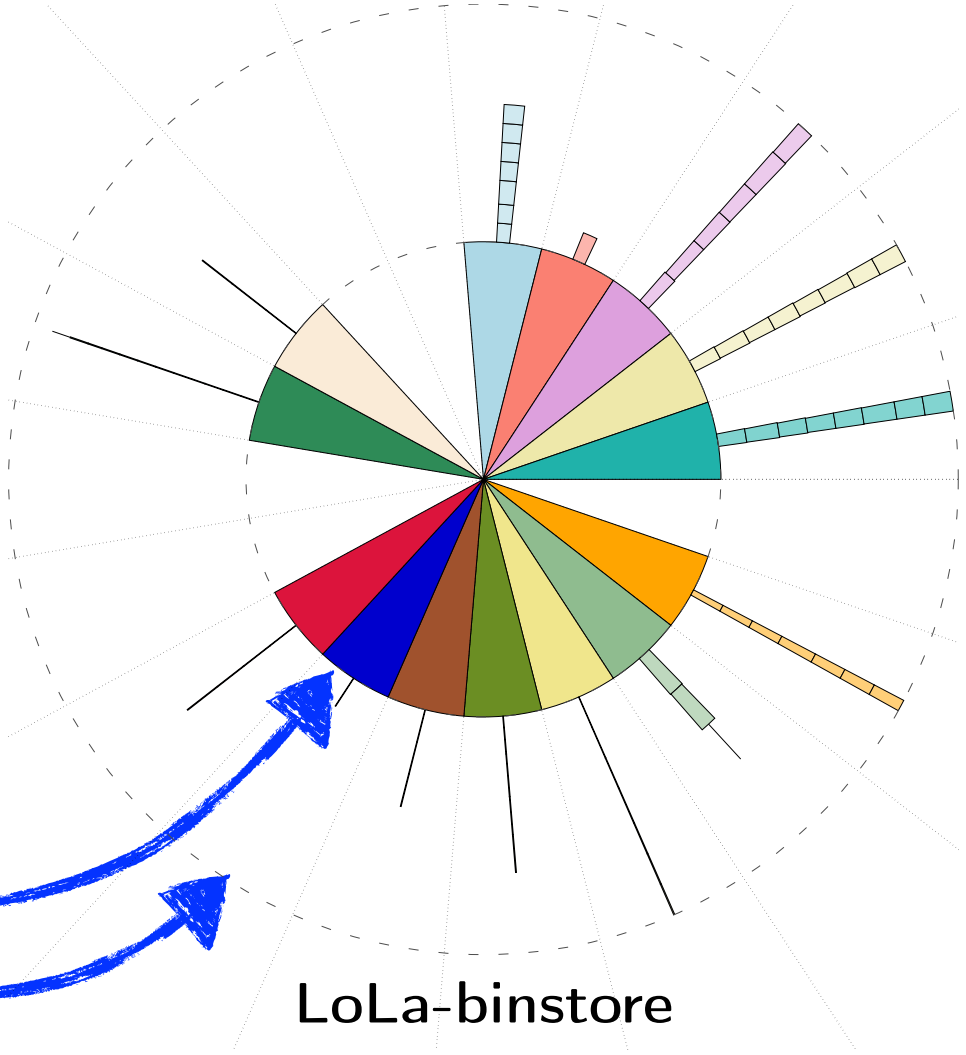


Models & examinations Processed by Tools (Reachability)

Only AlPiNa, Helena, LoLa* and Sara participated

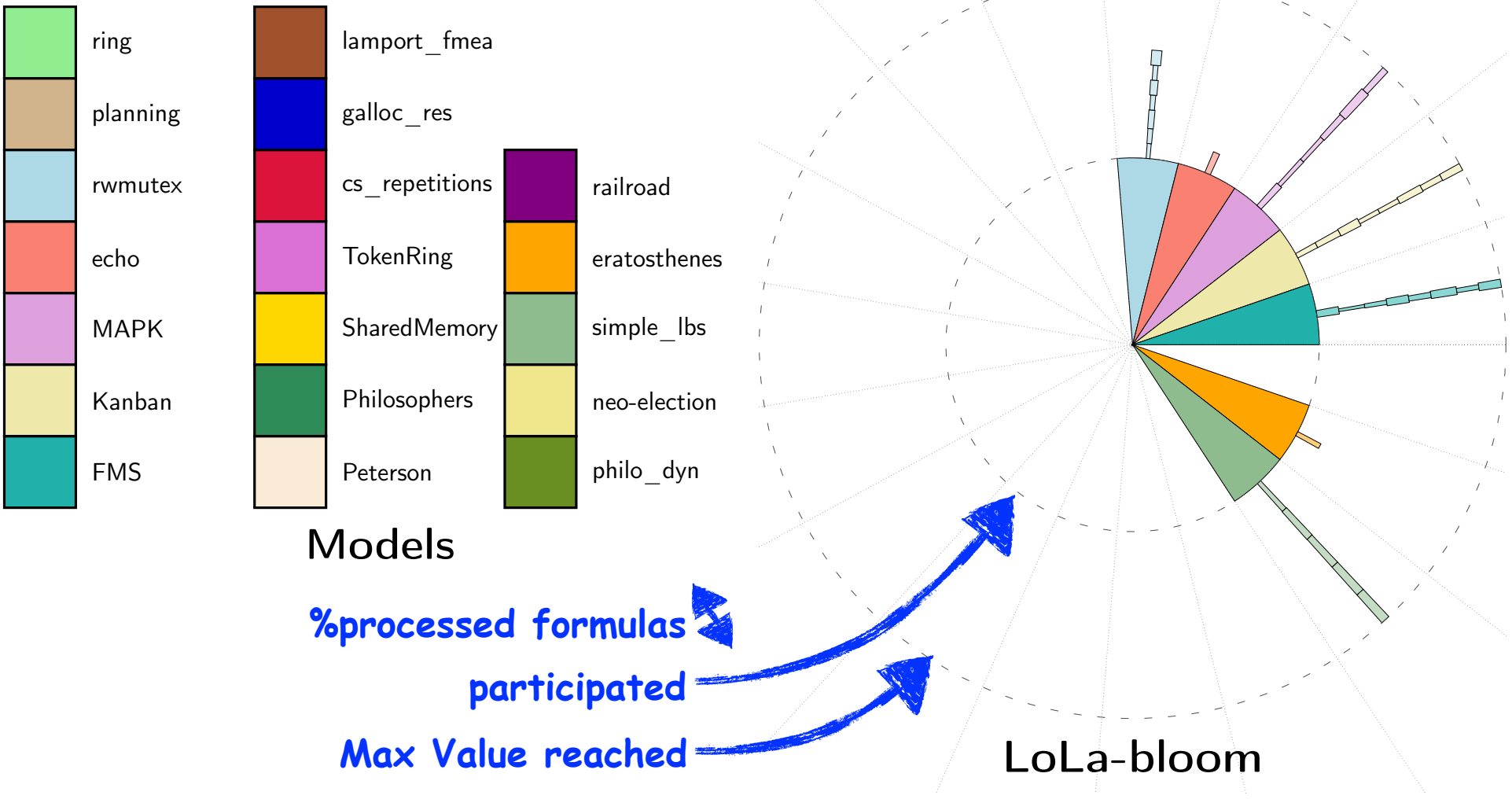






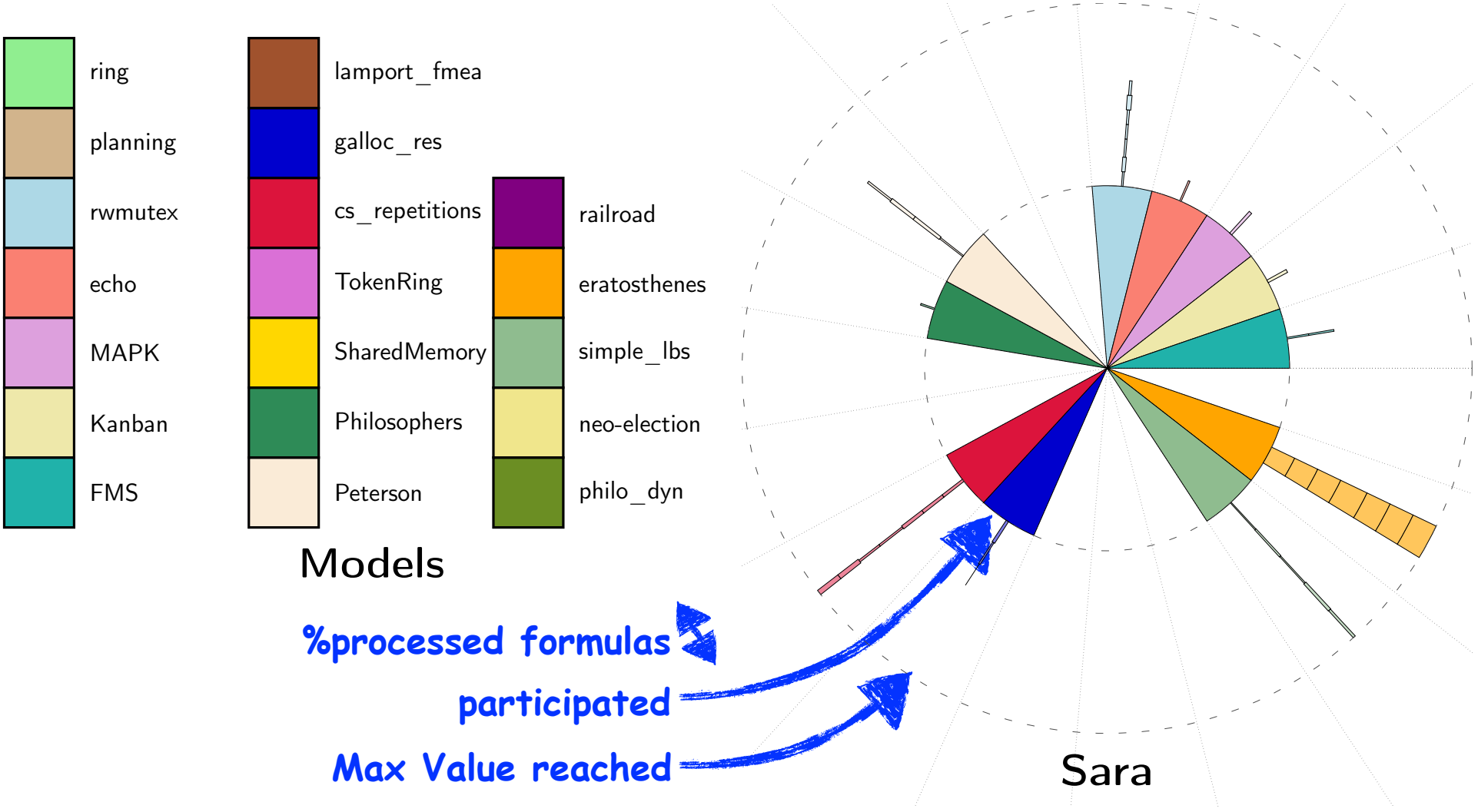
Models & examinations Processed by Tools (Reachability)

Only AlPiNa, Helena, LoLa* and Sara participated



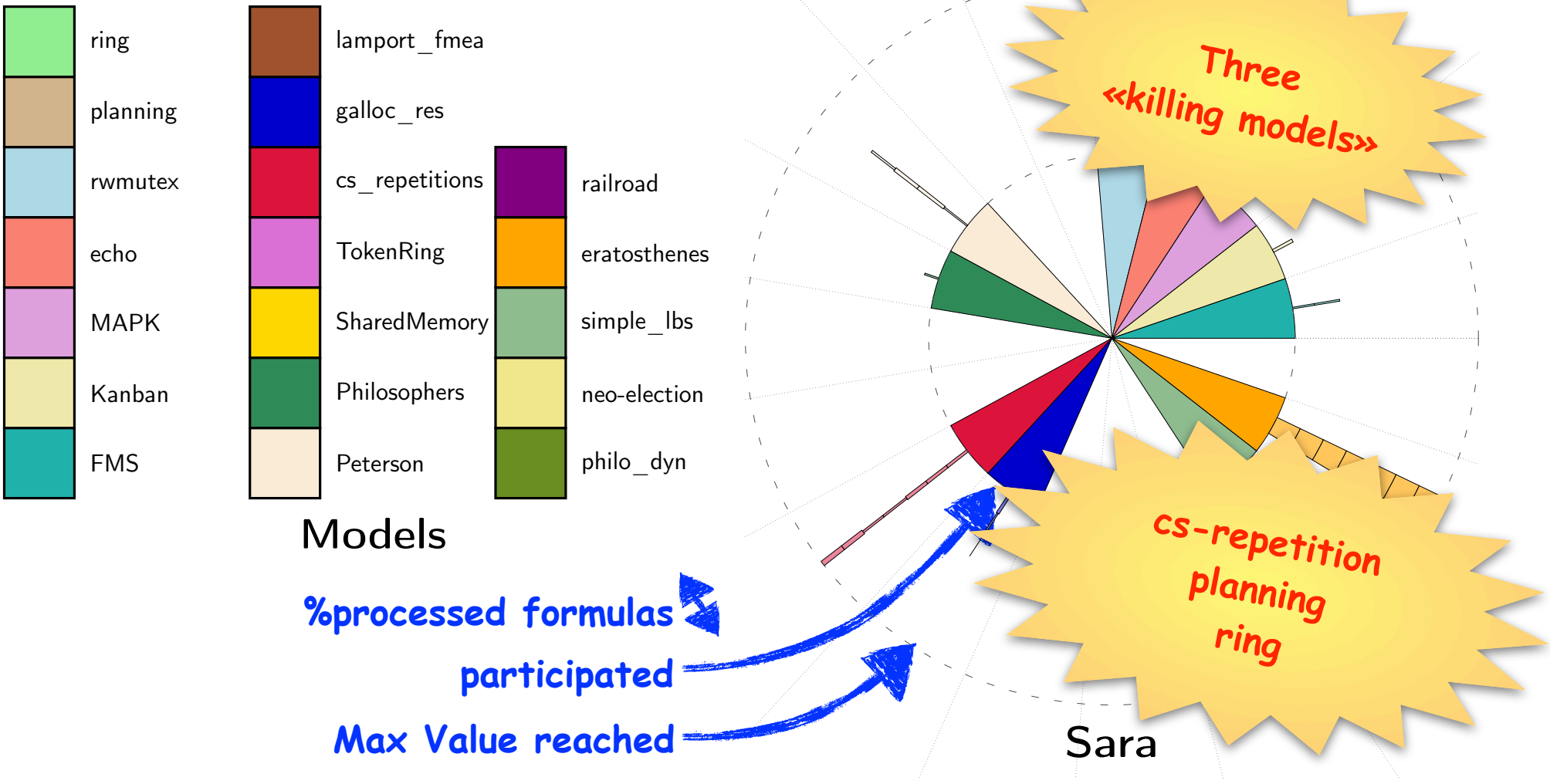
Models & examinations Processed by Tools (Reachability)

Only AlPiNa, Helena, LoLa* and Sara participated



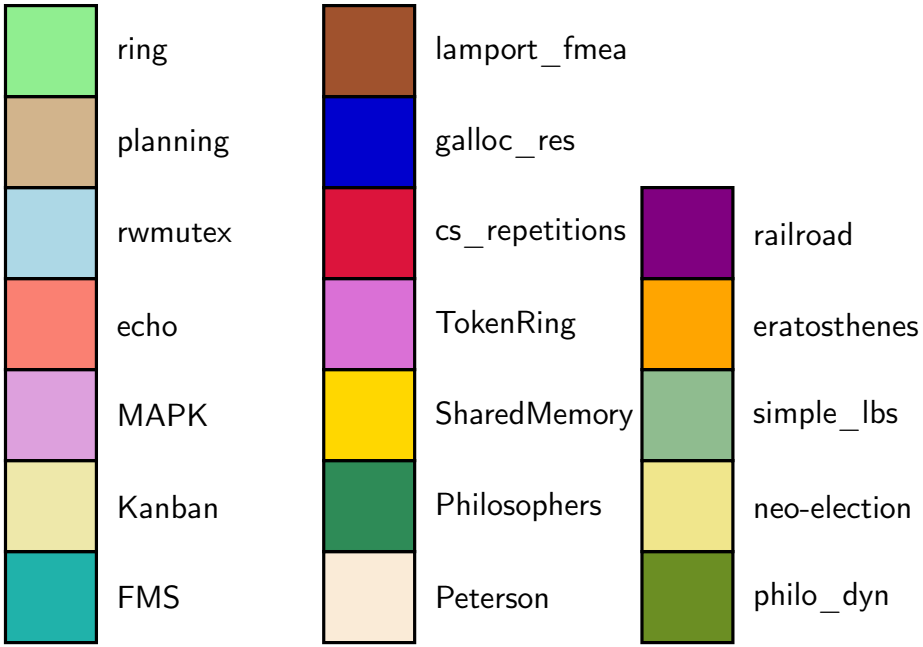
Models & examinations Processed by Tools (Reachability)

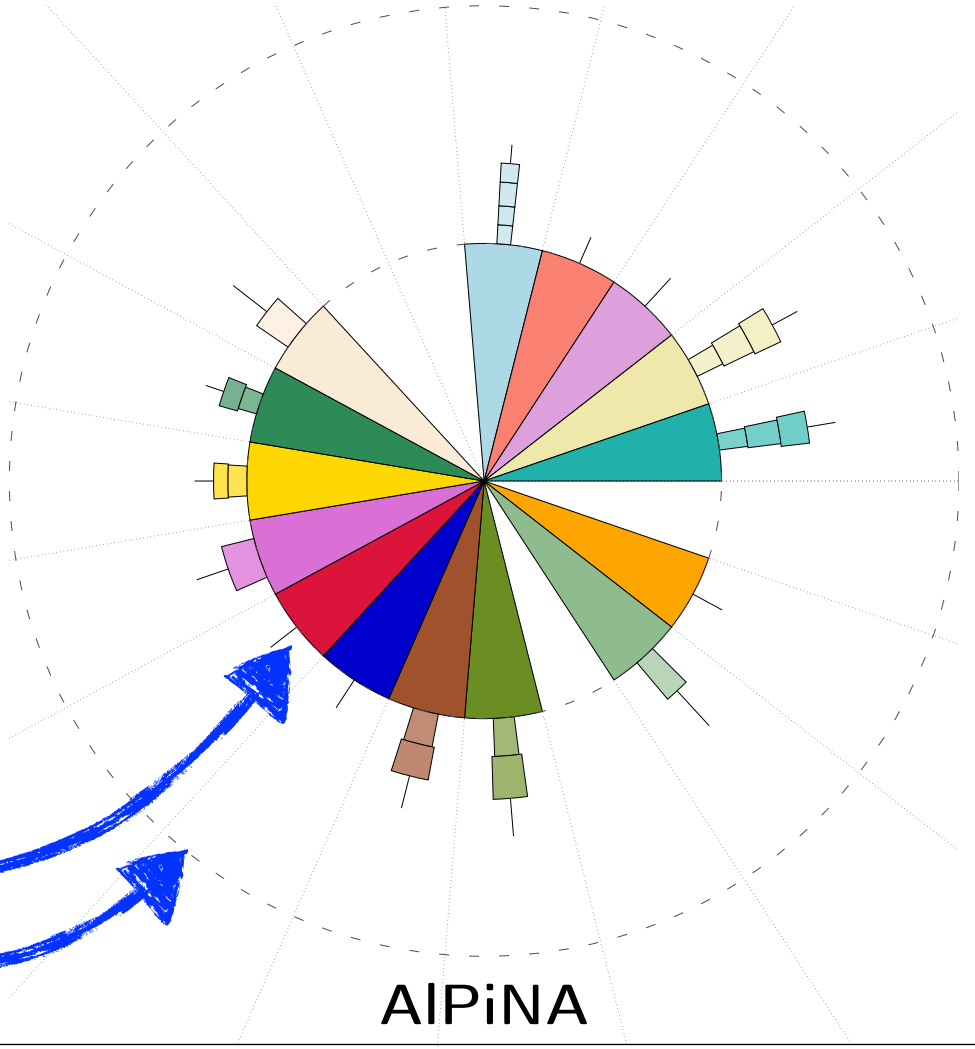
Only AlPiNa, Helena, LoLa* and Sara participated



Models & examinations Processed by Tools (Structural)

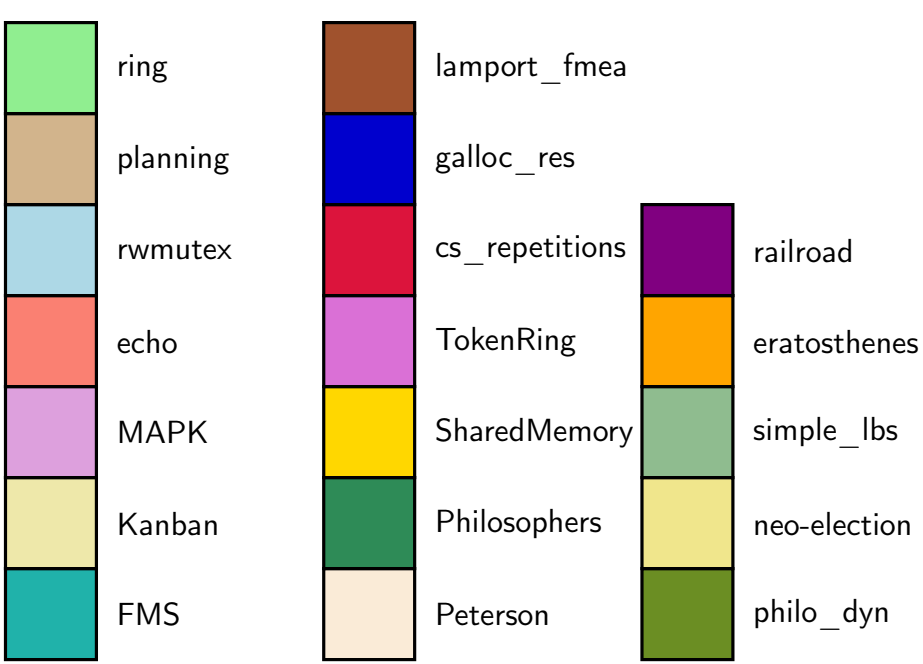
Only AlPiNa and Helena participated

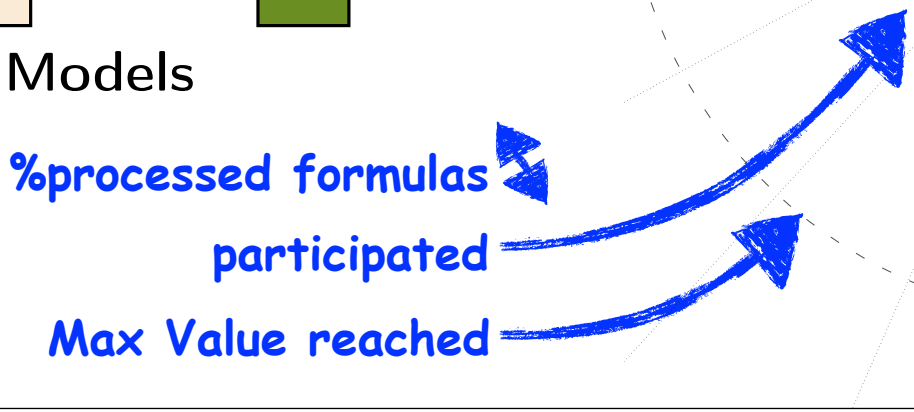


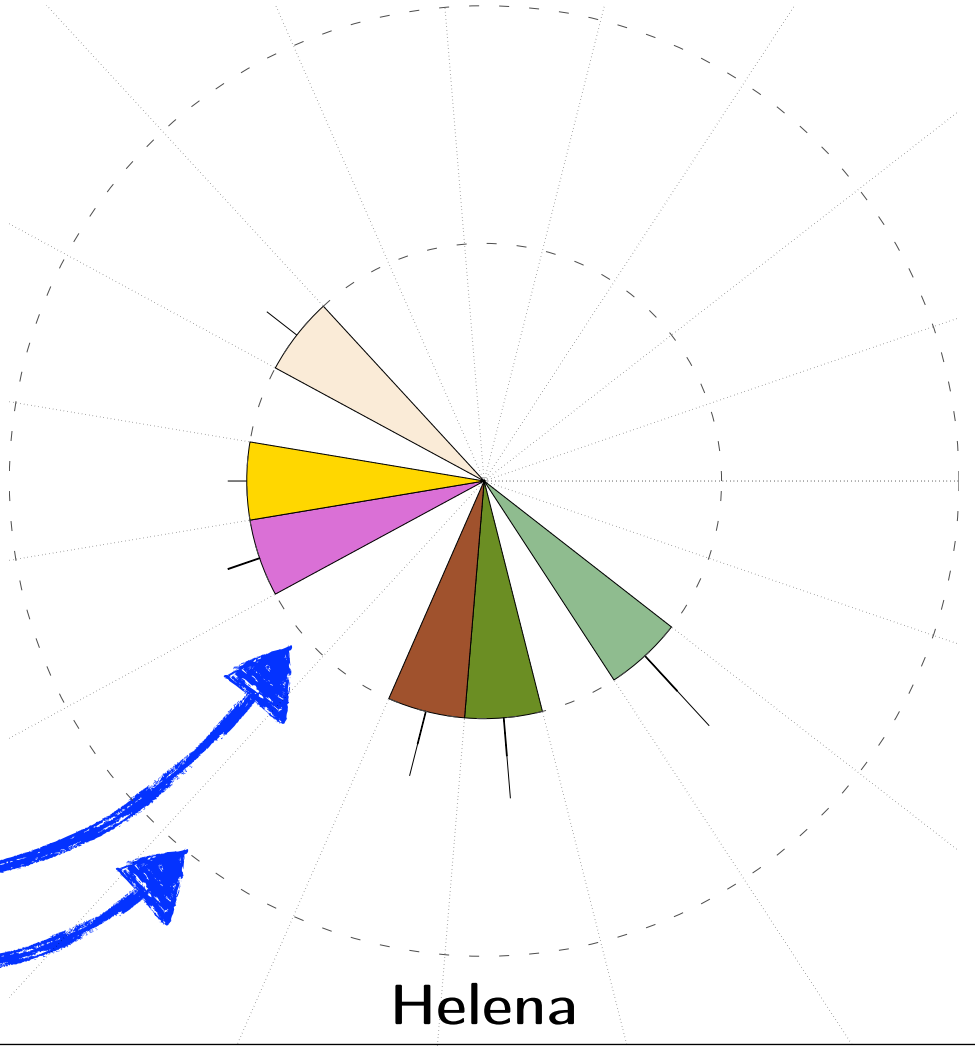


Models & examinations Processed by Tools (Structural)

Only AlPiNa and Helena participated

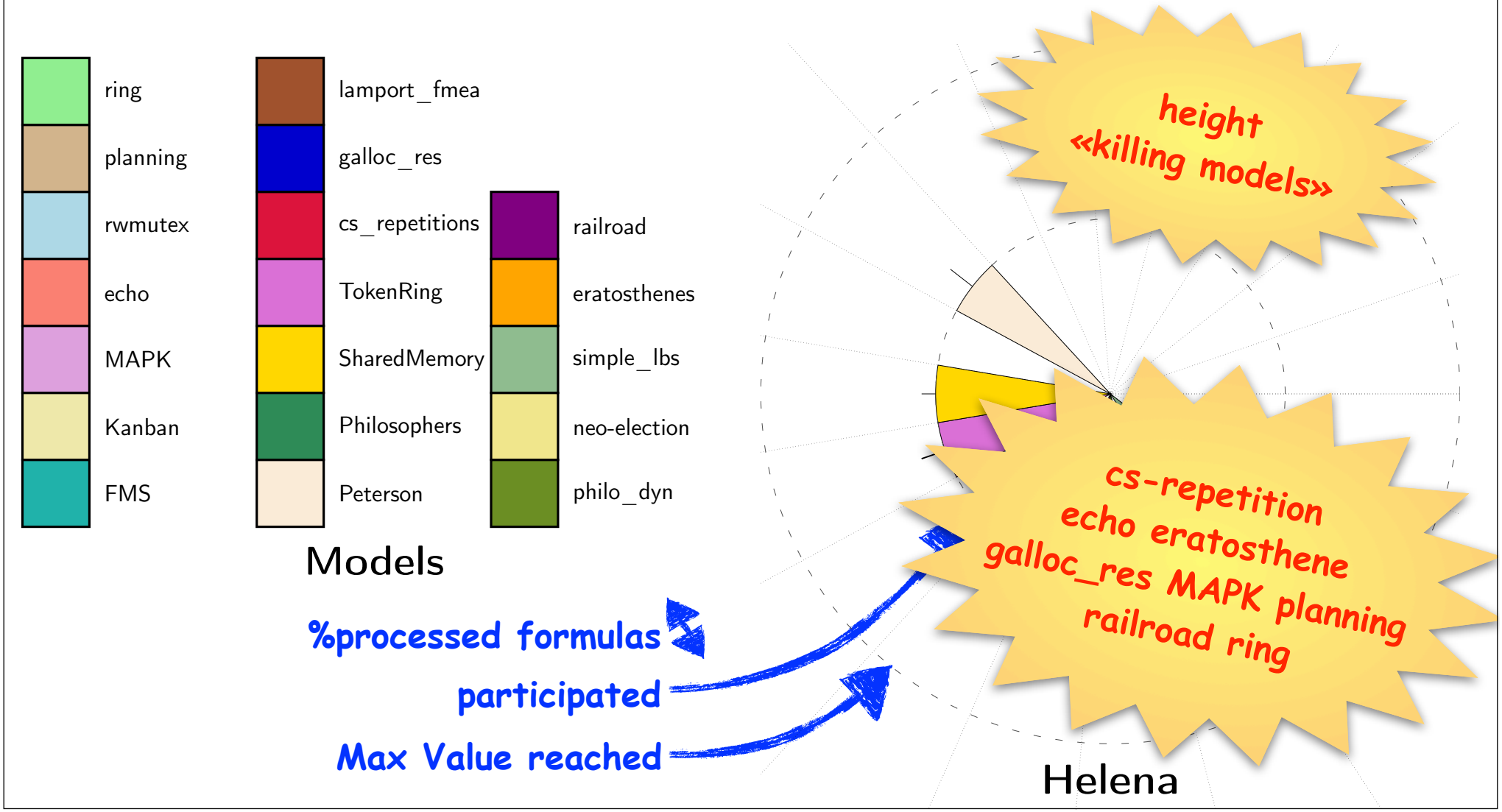






Models & examinations Processed by Tools (Structural)

Only AlPiNa and Helena participated



The «surprise model»

- We had a bench of «big ones»
  - From Paris and from Rostock
- Unfortunately...
  - Formula analysis ended last Sunday evening (about 23h00)
- Due to the number of «small problems» to be solved...
  - No possibility to operate this examination this year

MCC 2012

Analysis of the Results

Results were difficult to analyze

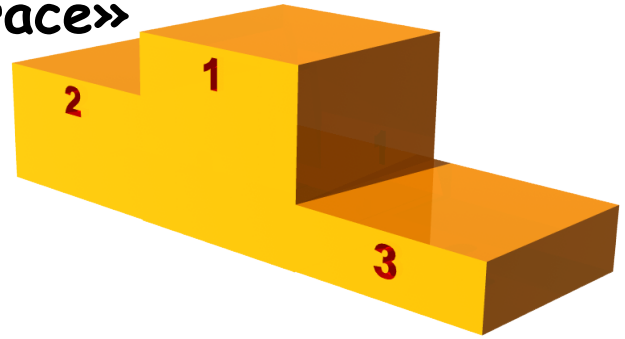
- The execution itself was shorter than expected
  - Around 6+8 hours (both examinations)
- But outcomes were much bigger (see next slide)

- State space analysis
  - This is possible
  - «comparison» is also possible
- Formula evaluation
  - This is more difficult
    - Some problems come from the original requirements
  - Comparison impossible
    - All tools do not process the same subset of formulas
    - Most formulas where false

- More work is needed on formulas for the next edition

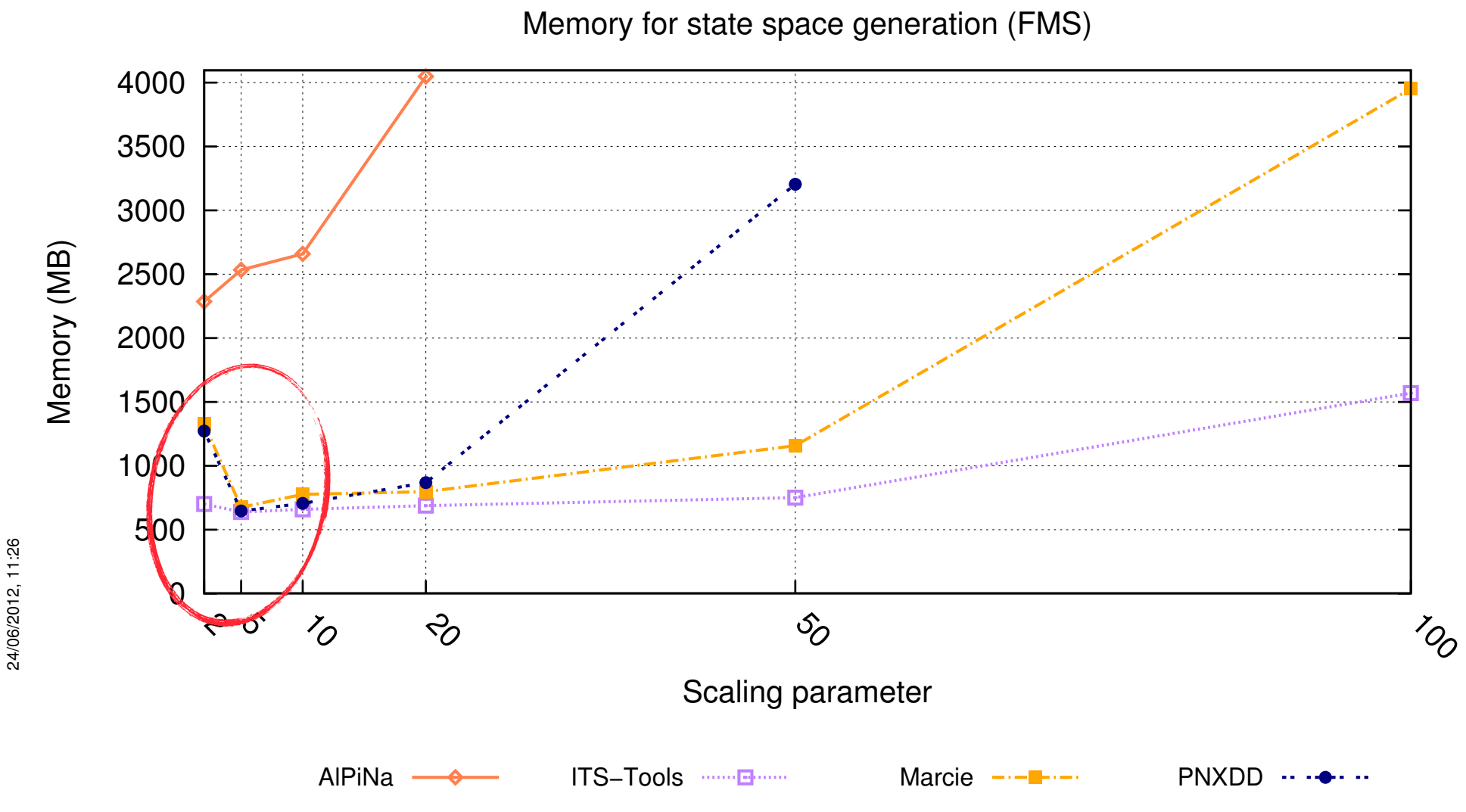
How to perform the analysis

- No (more than last year) interest in a «race»



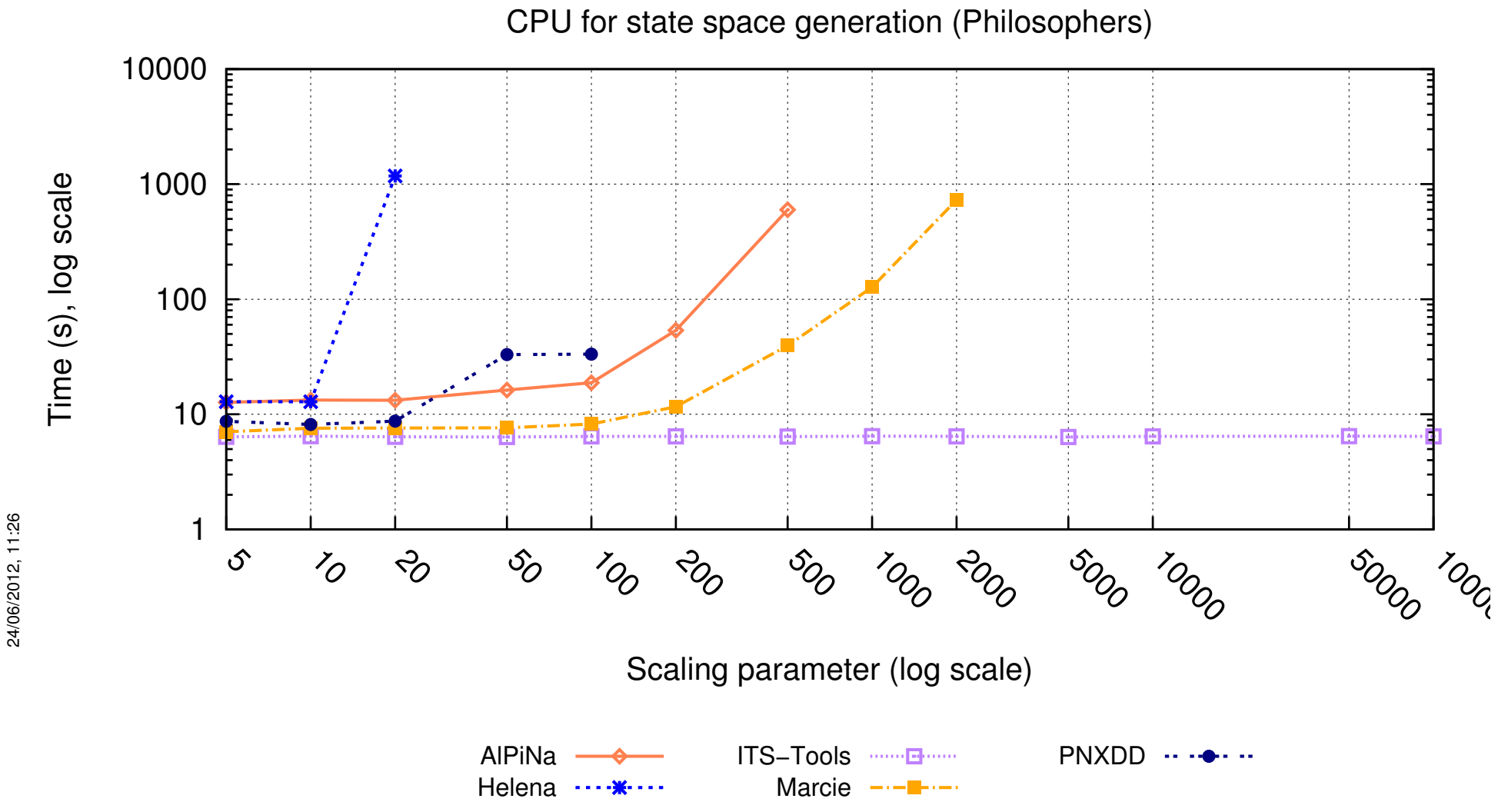
- 654 charts generated
  - 358 for the state space examination
    - Comparison of CPU, elapsed time, Memory,
    - Evolution of memory and CPU
    - Radars
  - 296 for the formulas examination (reachability and structural)
    - Comparison of CPU, elapsed time, Memory (no signification)
    - Radars
- Identification (partial) of some «surprises» discovered when test were processed
  - How tools scale up
    - P/T and colored
  - Some observations on time and memory consumption
  - Feed back with tools' characteristics

Interesting Facts, initial cost of some techniques

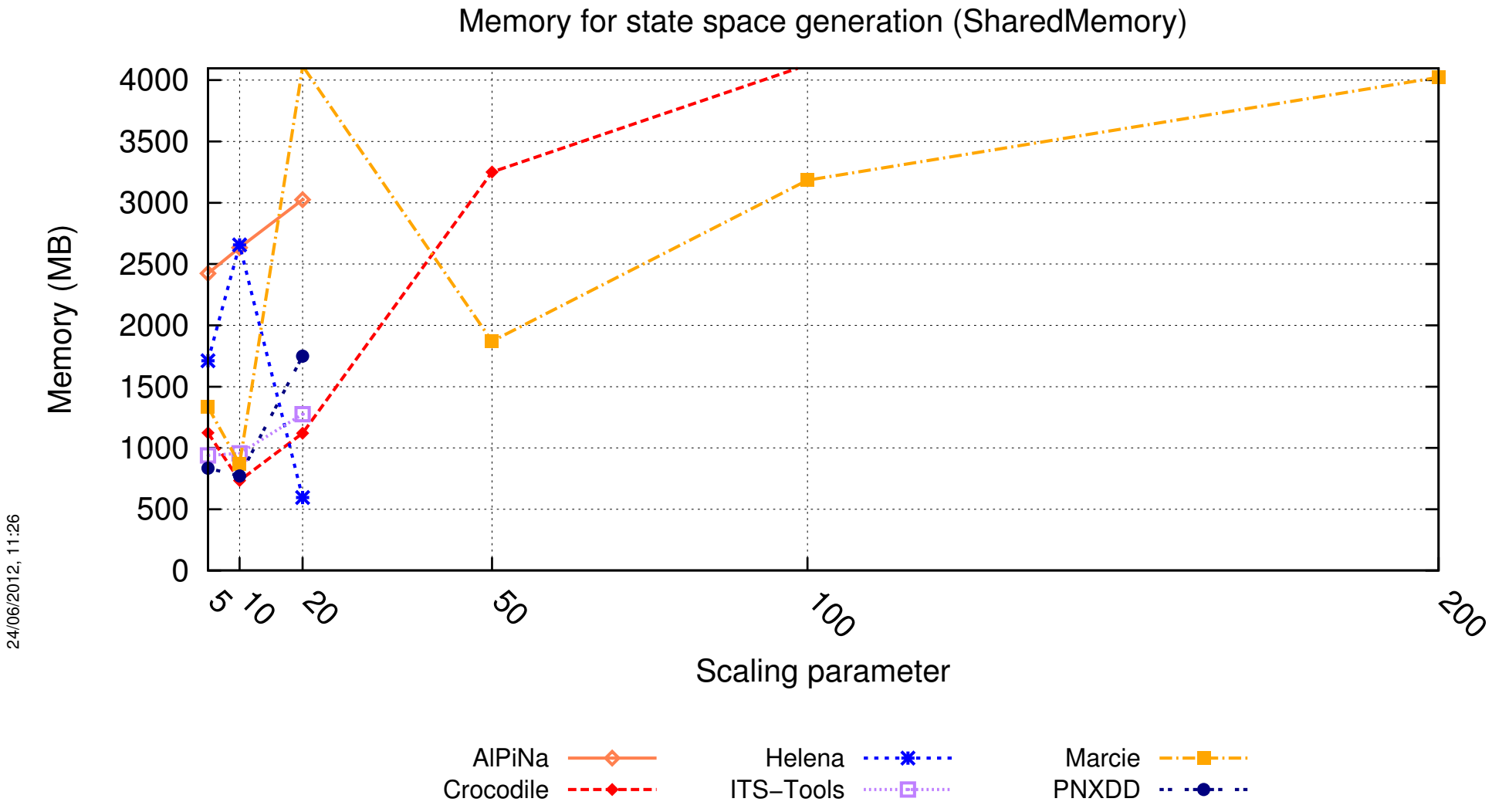


Philosophers, comparison with last year

Marcie (new tool) is doing quite well

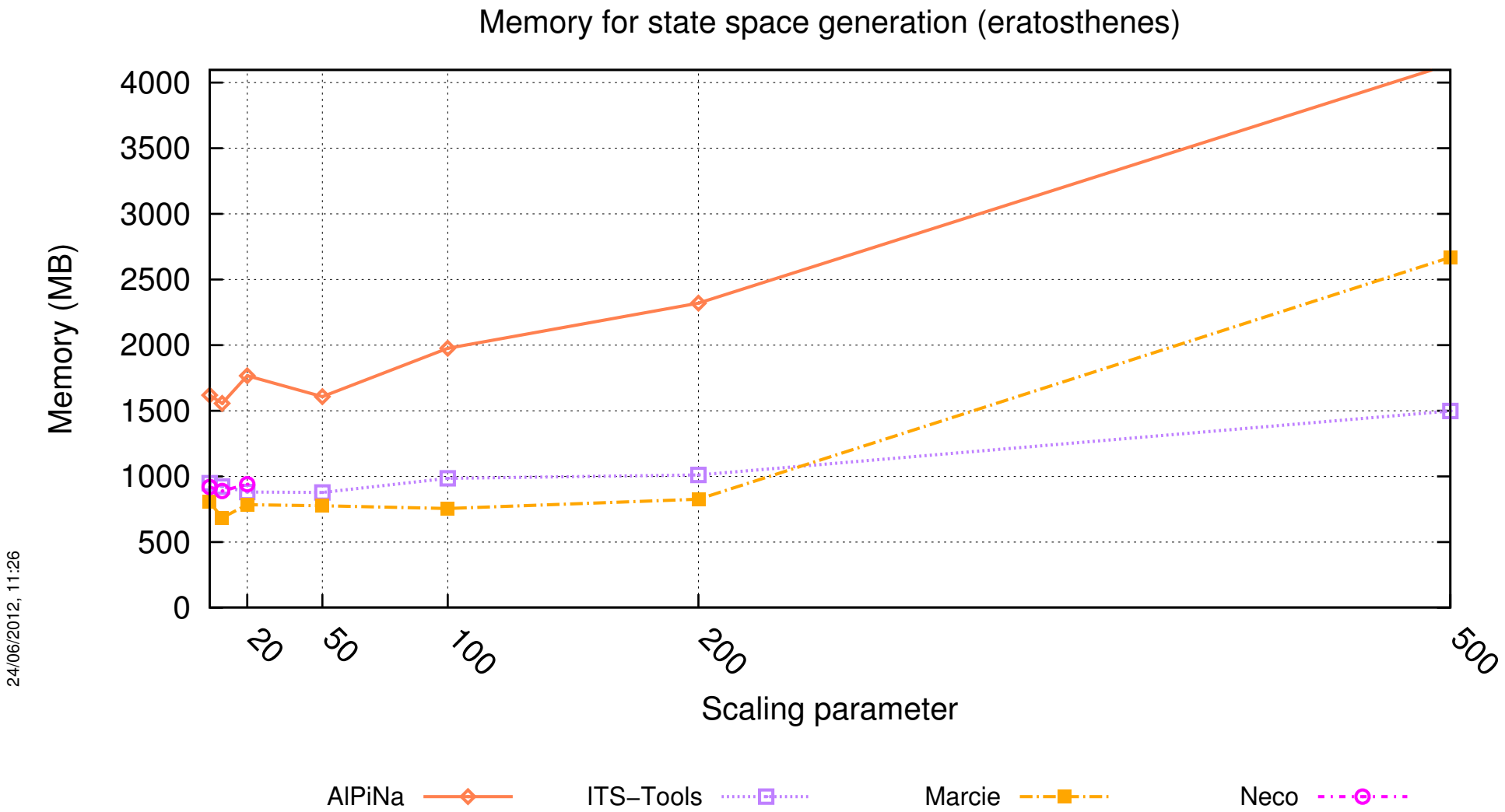


Decision diagrams for Marcie



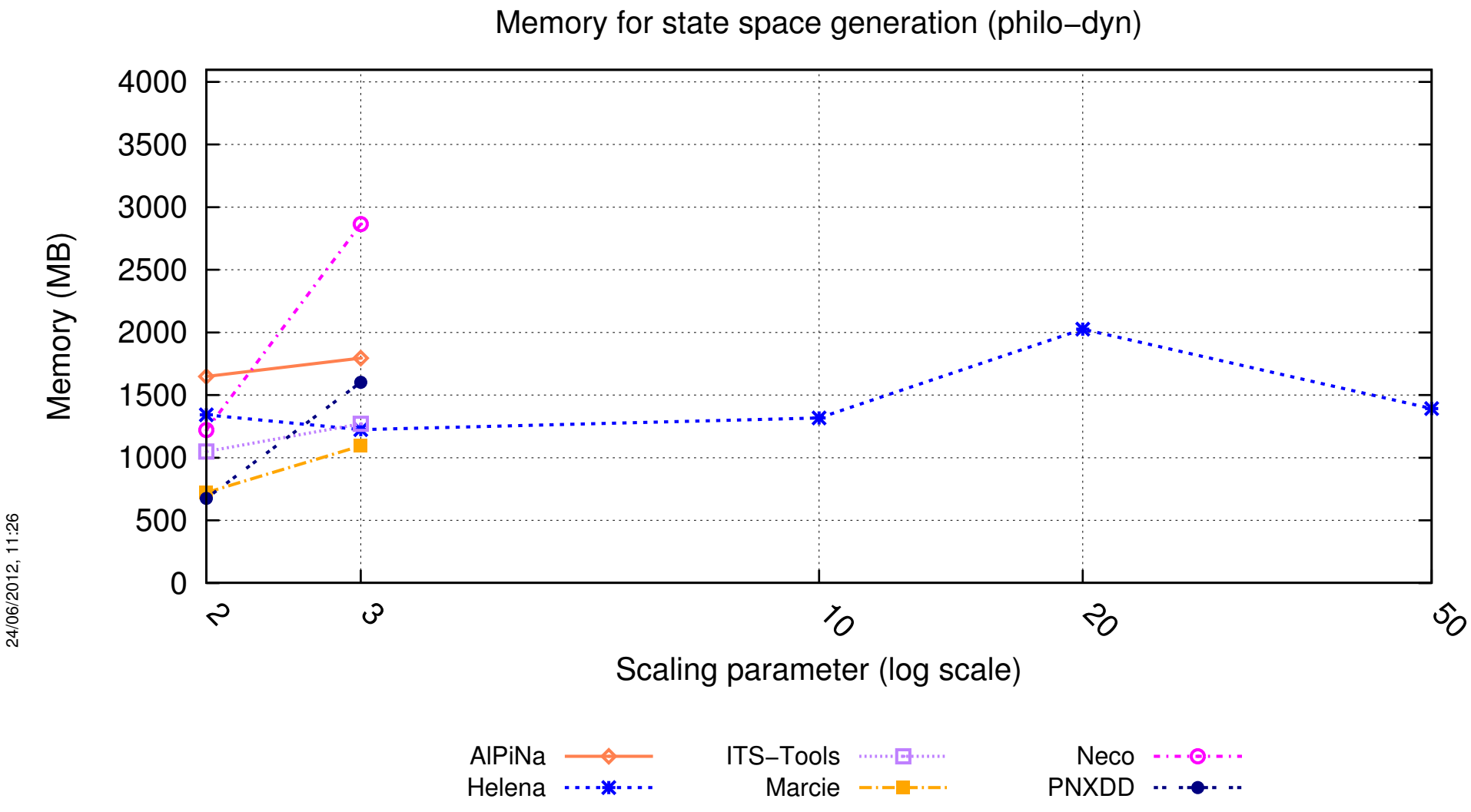
Eratosthene, Scalability of Decision diagrams?

Eratosthene 500 = $4,13 \times 10^{121}$ states

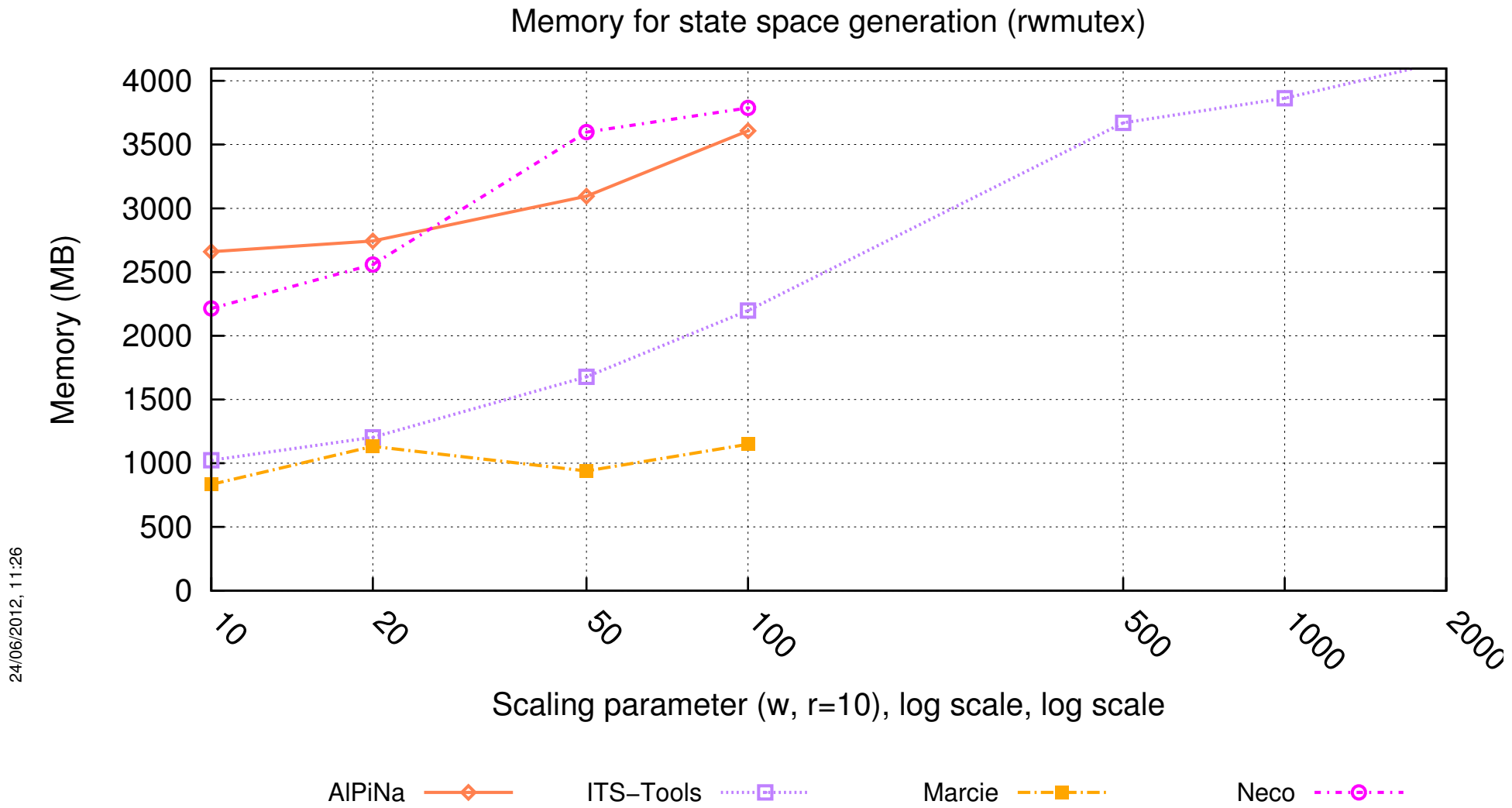


Sometimes, «simple» is better

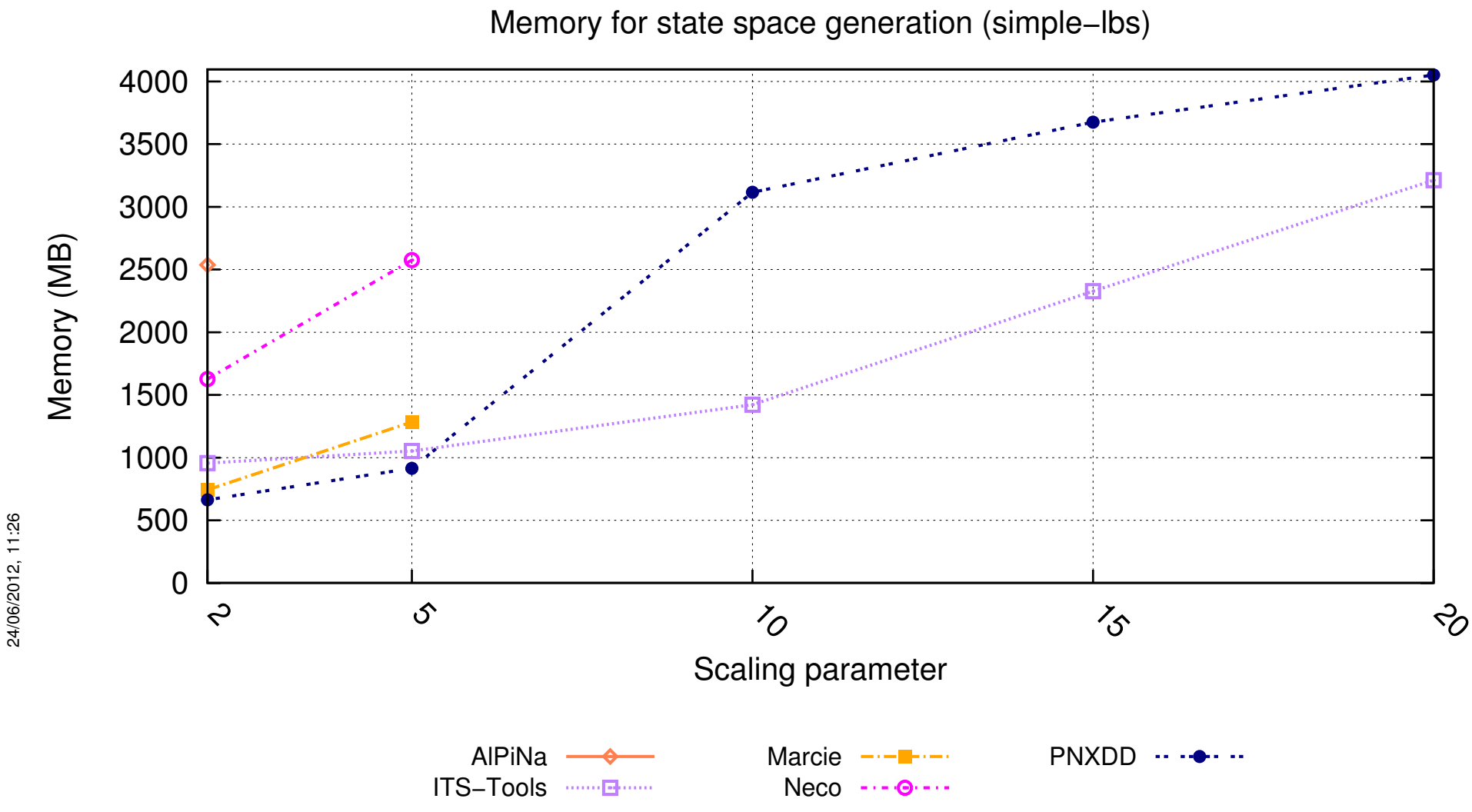
Helena beats them all (philo_dyn 50 = $2,26 \times 10^6$ states)



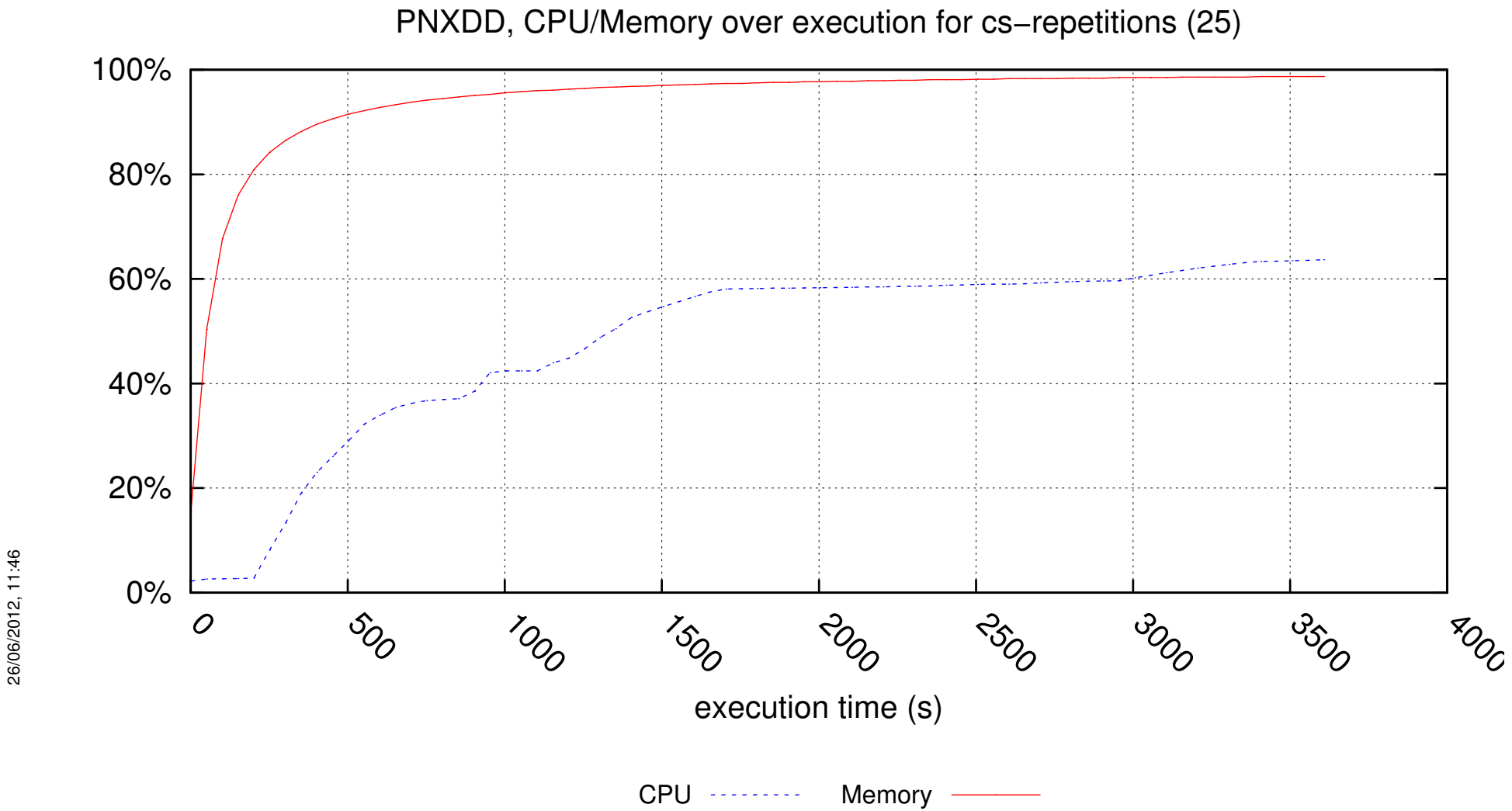
The tool going farer used a combination of techniques



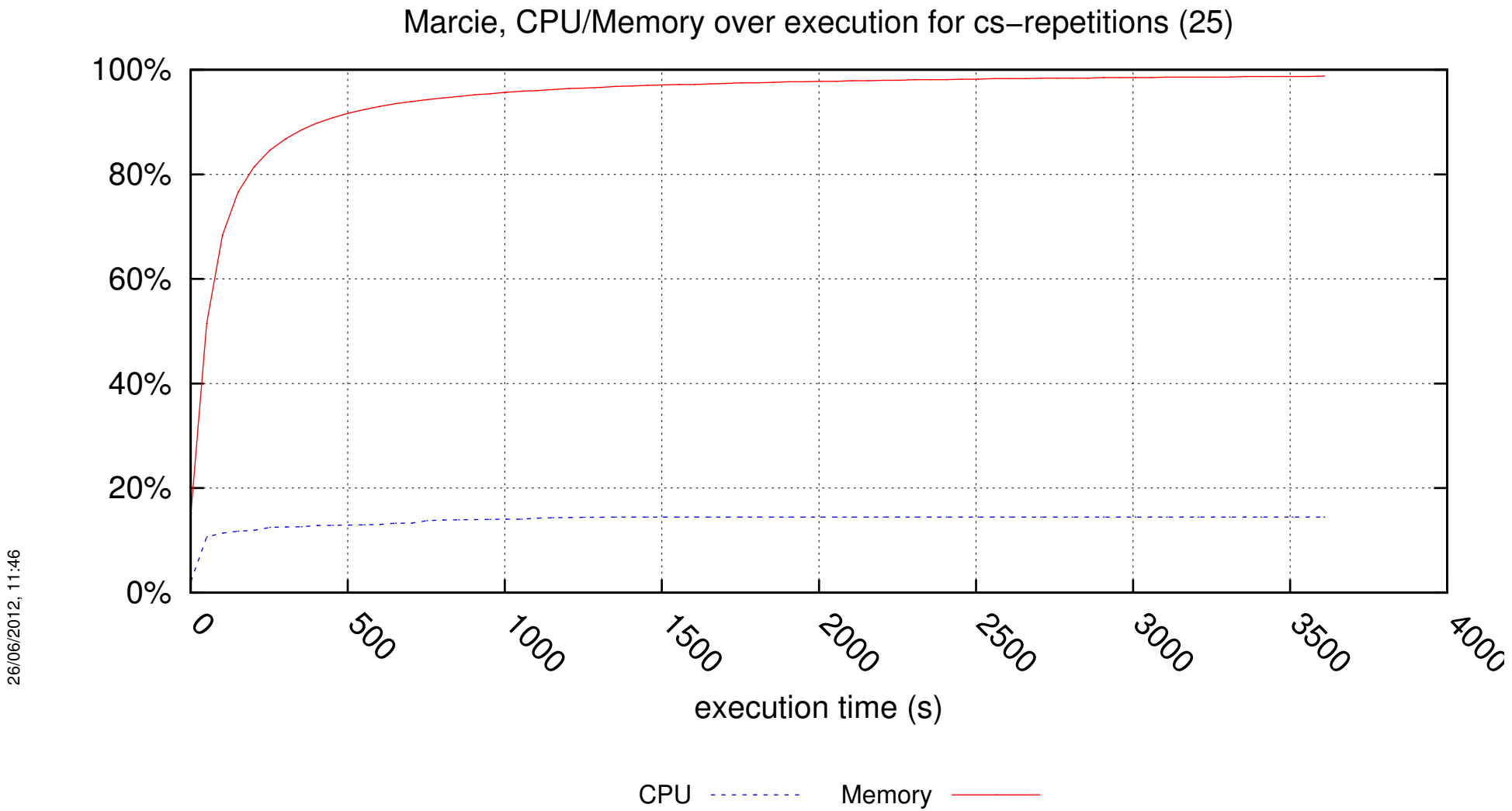
The tools that go farer also use a combination of techniques



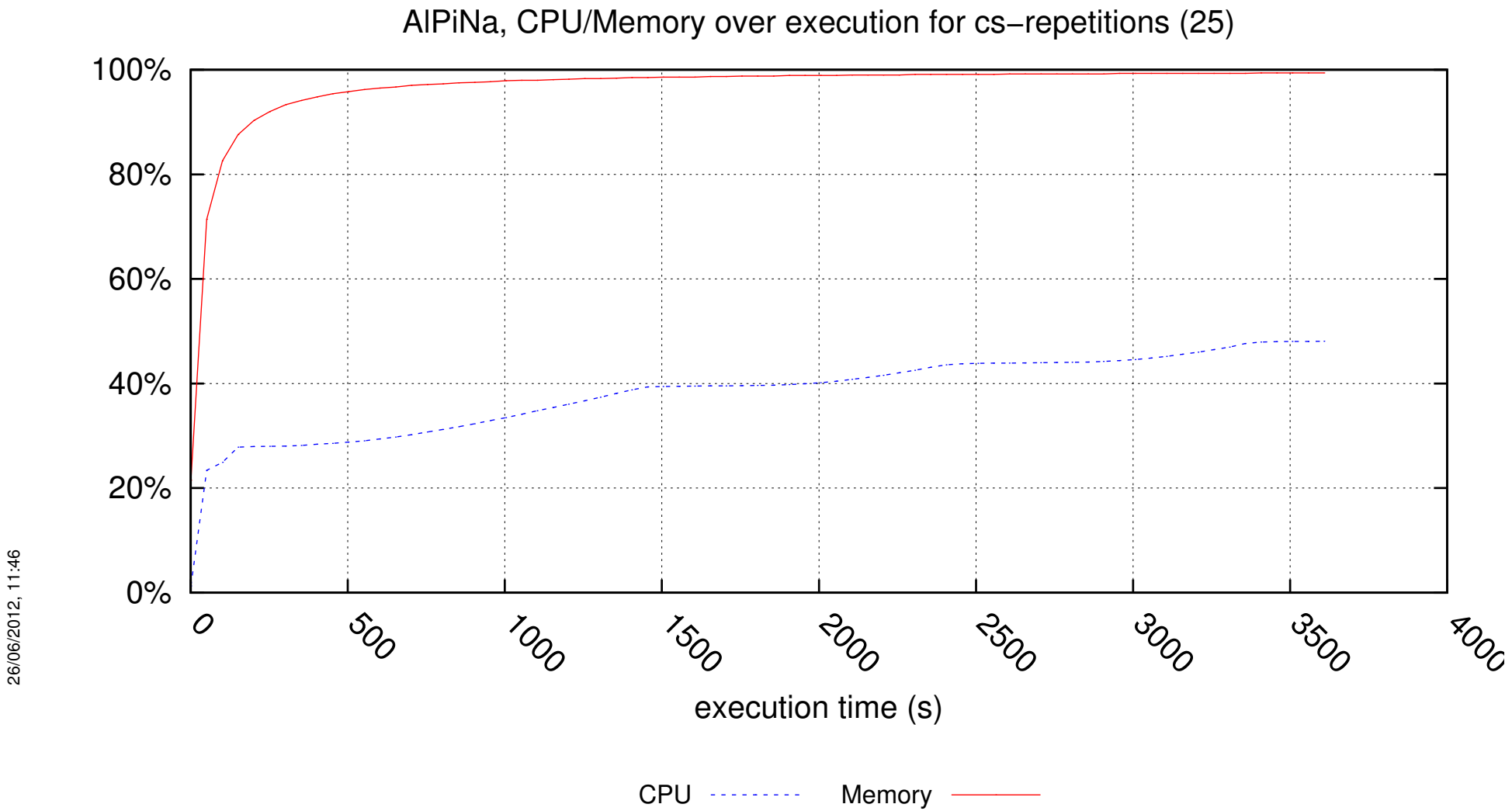
CS_REPETITION?



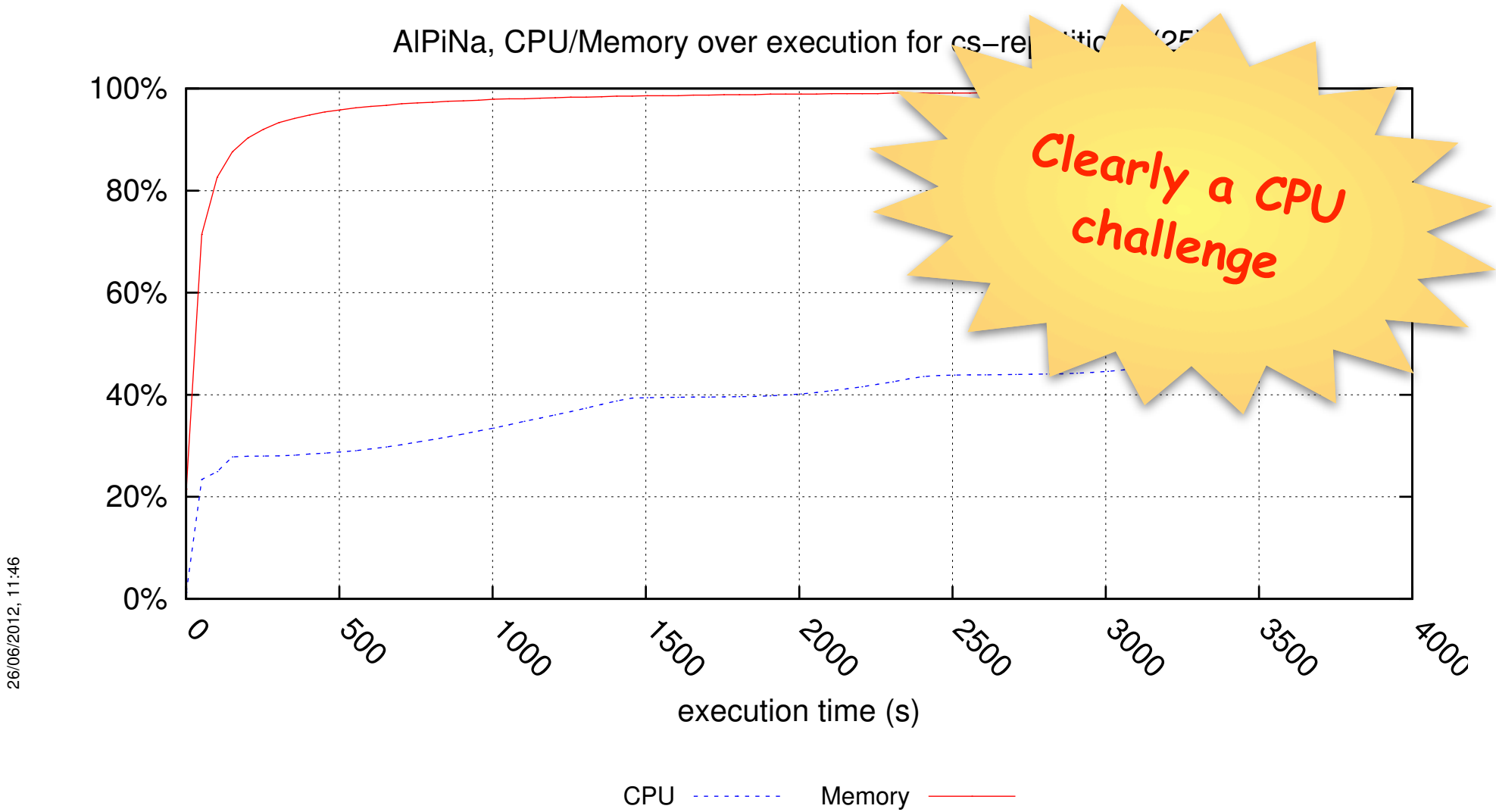
CS_REPETITION?

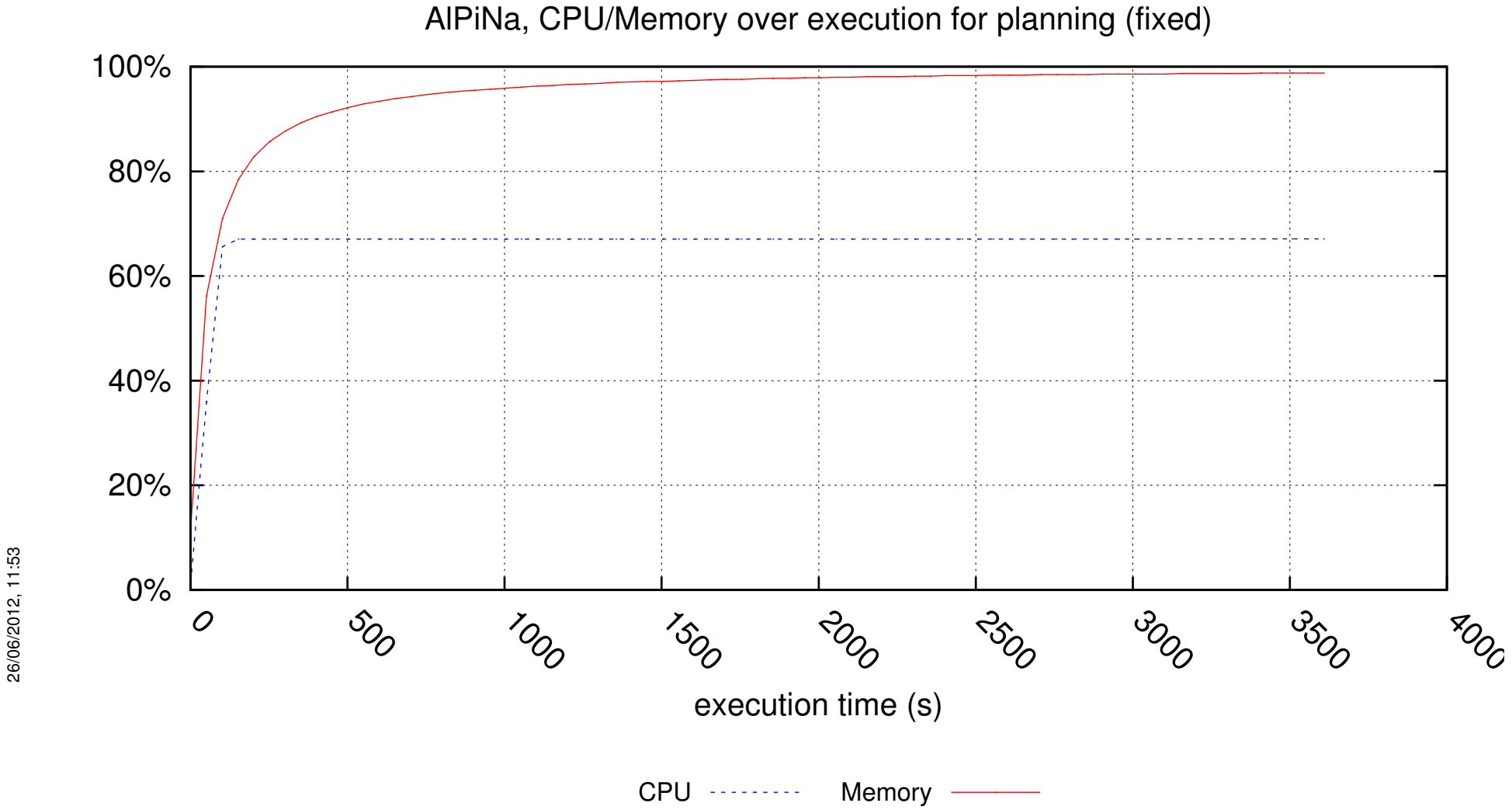


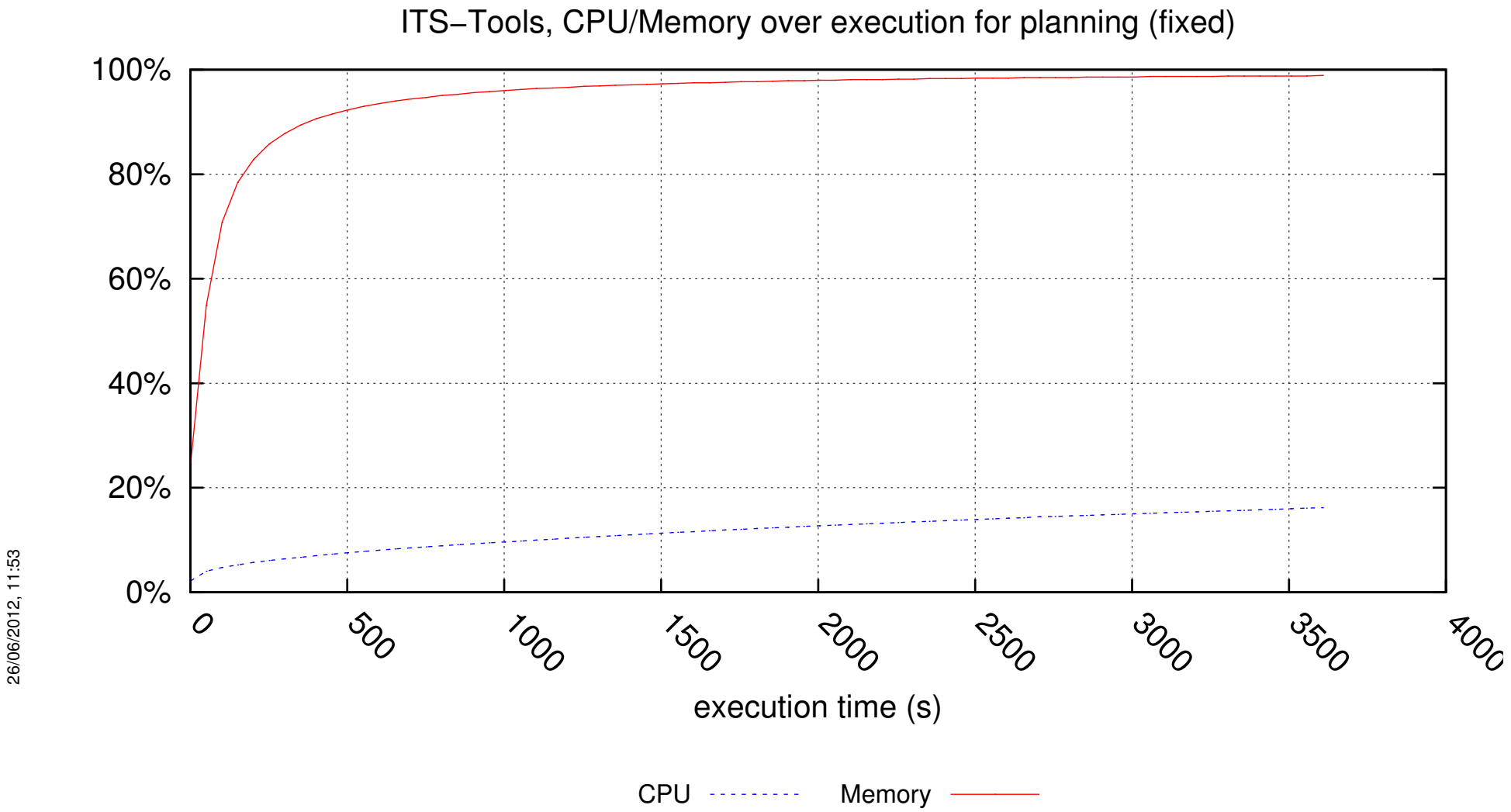
CS_REPETITION?

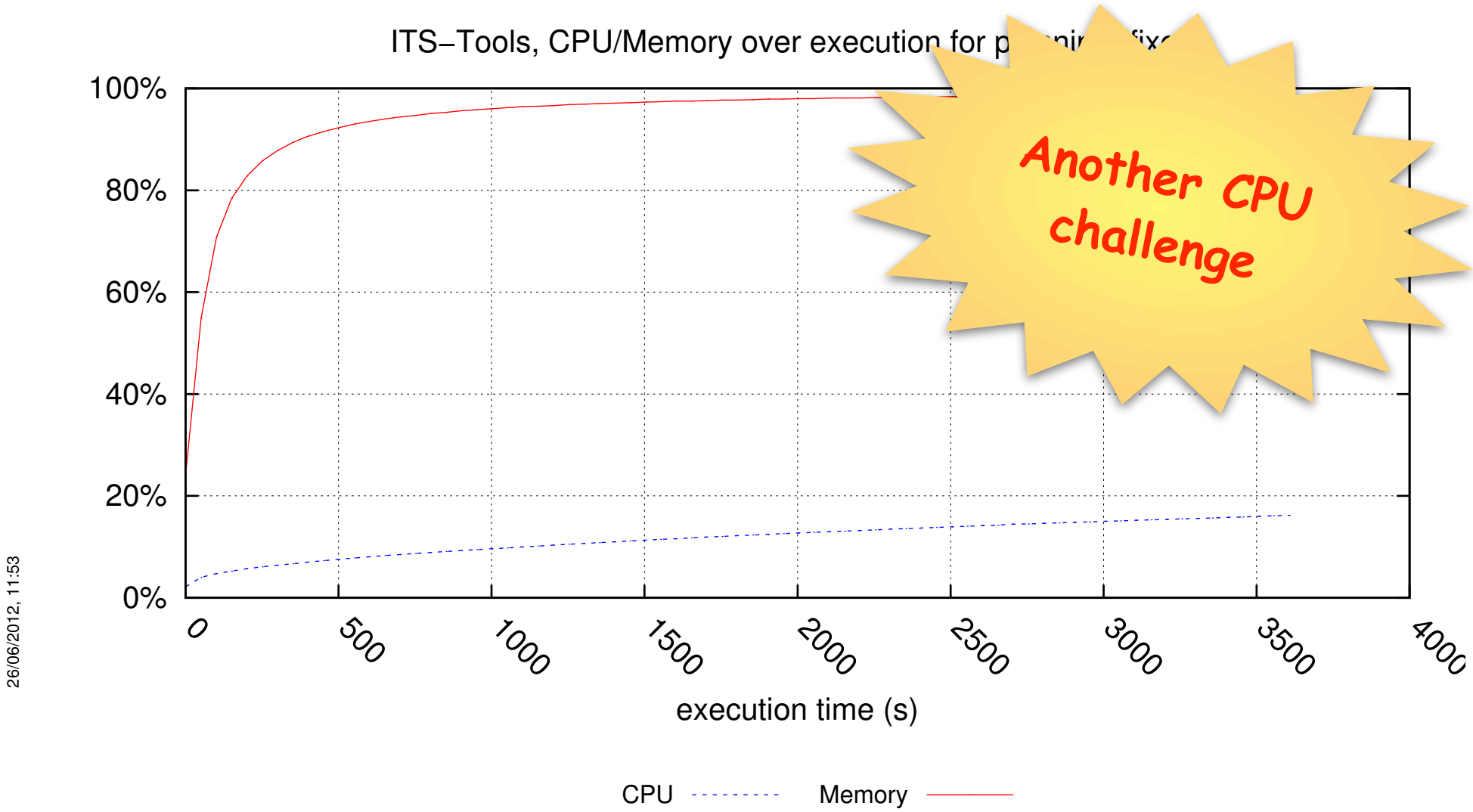


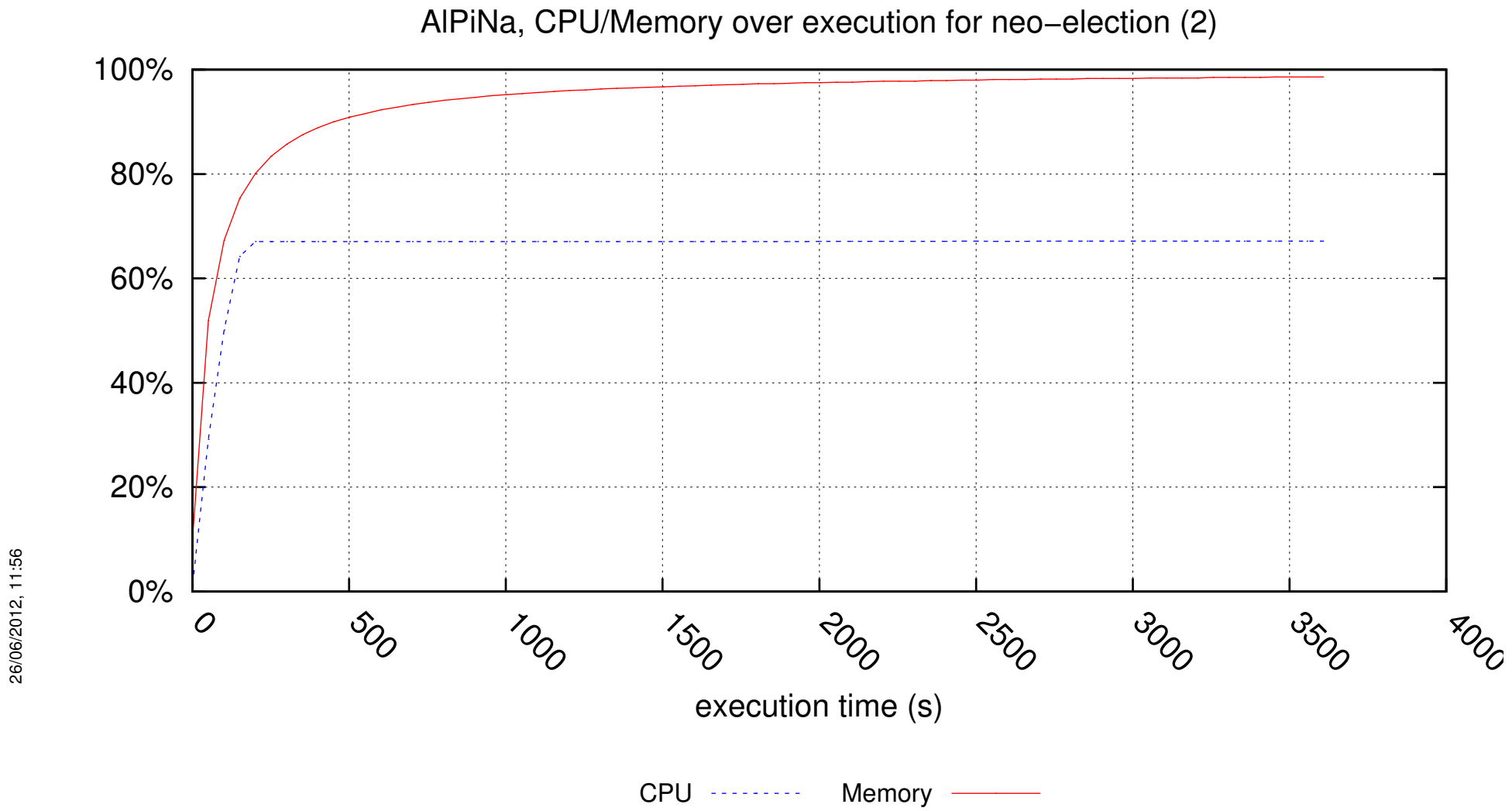
CS_REPETITION?

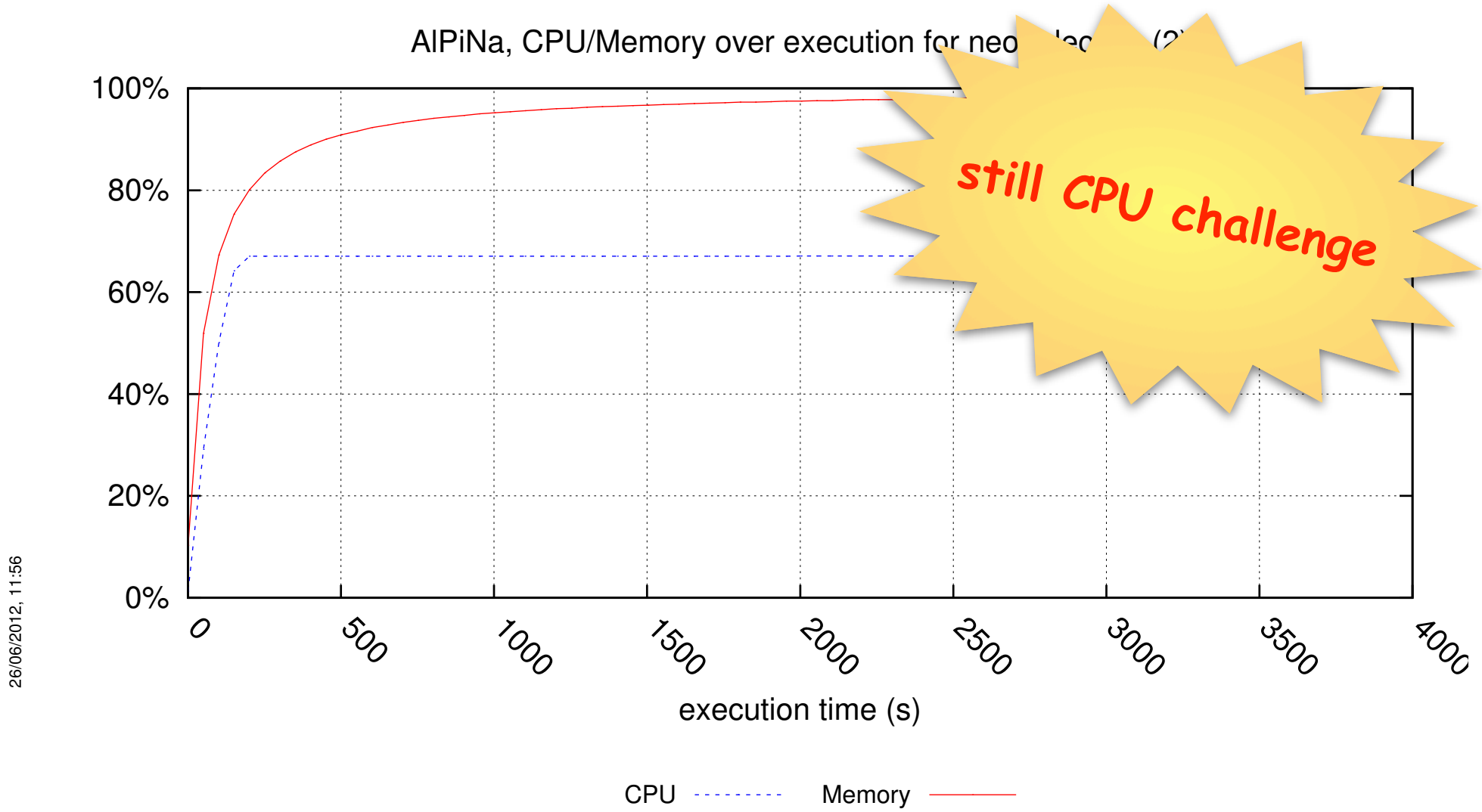



AlPiNa, CPU/Memory over execution for planning (fixed)
100%
80%
60%
40%
20%
0%
0
500
1000
1500
2000
2500
3000
3500
4000
execution time (s)
CPU
Memory
26/06/2012, 11:53


ITS–Tools, CPU/Memory over execution for planning (fixed)
100%
80%
60%
40%
20%
0%
0
500
1000
1500
2000
2500
3000
3500
4000
execution time (s)
26/06/2012, 11:53
CPU
Memory


ITS–Tools, CPU/Memory over execution for p
Another CPU challenge
100%
80%
60%
40%
20%
0%
0
500
1000
1500
2000
2500
3000
3500
4000
execution time (s)
26/06/2012, 11:53
CPU
Memory


AlPiNa, CPU/Memory over execution for neo−election (2)
100%
80%
60%
40%
20%
0%
0
500
1000
1500
2000
2500
3000
3500
4000
execution time (s)
26/06/2012, 11:56
CPU
Memory


AlPiNa, CPU/Memory over execution for neo
100%
80%
60%
40%
20%
0%
0
500
1000
1500
2000
2500
3000
3500
4000
execution time (s)
CPU
Memory
26/06/2012, 11:56
still CPU challenge

On formulas

- It is impossible to really evaluate
  - Charts have been generated but with no real meaning
  - No execution chart (tracking bug)
- On the number of evaluated reachability formulas
  - LoLa-binstore is clearly better (LoLa-bloom has more fails)
  - AlPiNa does not scale up well
- On the number of evaluated structural formulas
  - AlPiNa beats Helena (more formulas and less fails)
- Let us note that sometimes tool diverge
  - The vector of evaluated formula is not the same ;-)
- On the state space too
  - Buts the notion of state (symbolic, explicit) may not be the same.

MCC 2012

Concluding Remarks

Outcomes

- Mitigated results for formulas
  - Lots of difficulties... and lessons learned
  - No much time left for analysis (and big data to go through)
- MCC 2013 @ Petri Nets?
  - The team is ready to go
  - Need to provide more help to tools submitters
    - A way to execute a tool on the first instance of each model?
- What to be proposed in MCC 2013 @ Petri Nets?
  - No more complex things (we must stabilize the procedure)
  - More memory (but 64bits VM then)
  - More models: models from 2012 and more?
  - A finer classification of properties
    - Bounds + Deadlocks + mixed
    - Satisfiable + Unsatisfiable
  - The «surprise model»

Downloading Participating Tools

- AlPiNA - CUI, University of Geneva - Switzerland
  - http://cui.unige.ch/~buchs/Site/About_Me.html
- Crocodile - LIP6, Université P. & M. Curie - France
  - http://www.cosyverif.org (as a part of the environment)
- Helena - LIPN, Université Paris 13 - France
  - http://www-lipn.univ-paris13.fr/~evangelista/
- ITS_Tools - LIP6, Université P. & M. Curie - France
  - http://ddd.lip6.fr and http://www.cosyverif.org (as a part of the environment)
- LoLA Binstore - University of Rostock - Germany
  - http://www.informatik.uni-rostock.de/tpp/lola/
- LoLA Bloom - University of Rostock - Germany
  - http://www.informatik.uni-rostock.de/tpp/lola/
- Marcie - BTU-Cottbus - Germany
  - http://www-dssz.informatik.tu-cottbus.de/
- Neco - IBISC, Université Evry val d'Essonne - France
  - www.ibisc.fr/~lfronc/
- PNXDD - LIP6, Université P. & M. Curie - France
  - http://move.lip6.fr and http://www.cosyverif.org (as a part of the environment)
- Sara- University of Rostock - Germany
  - http://www.informatik.uni-rostock.de/tpp/lola

MCC 2012

Discussion