ESOP 2019

Counters in Kappa: Semantics, Simulation, and Static Analysis

P. Boutillier[1], **I. Cristescu**[2], **and** **J. Feret**[3]

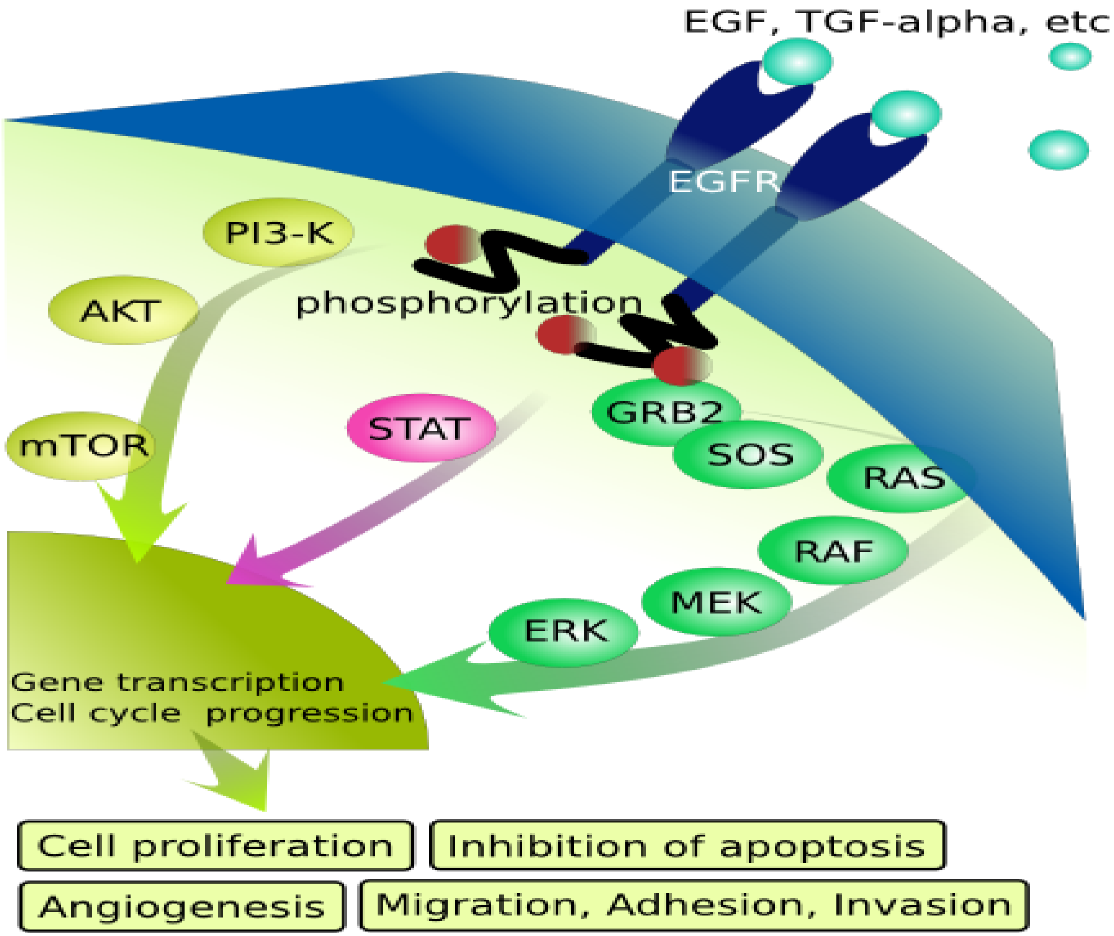
1 - Fontana Lab, Harvard Medical School, Boston, USA
2 - TAMIS, Inria - Bretagne Atlantique, Rennes, France
3 - Antique, DI-ÉNS (Inria/CNRS/ÉNS/PSL researh university), Paris, France

Prague, 2019 April 8

On the menu today

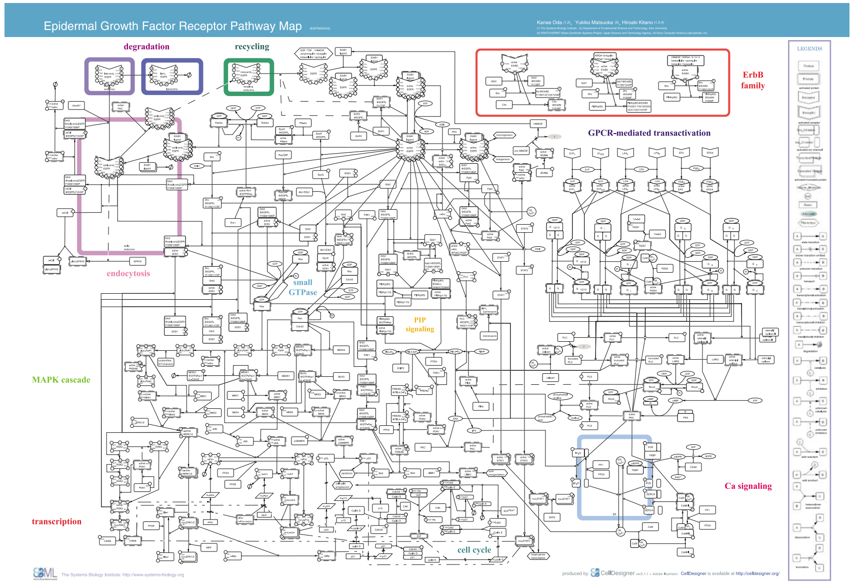
1. Mechanistic models
2. Kappa
3. Causality
4. Efficient simulation
5. Static analysis
6. Conclusion

Signalling pathways



Eikuch, 2007

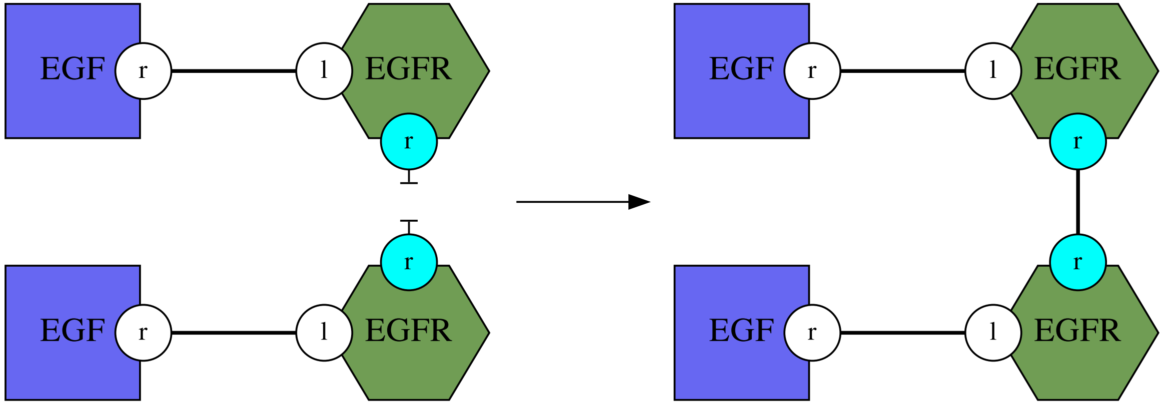
Bridge the gap between. . .



$$\begin{cases} \frac{dx_1}{dt} = -k_1 \cdot x_1 \cdot x_2 + k_{-1} \cdot x_3 \\ \frac{dx_2}{dt} = -k_1 \cdot x_1 \cdot x_2 + k_{-1} \cdot x_3 \\ \frac{dx_3}{dt} = k_1 \cdot x_1 \cdot x_2 - k_{-1} \cdot x_3 + 2 \cdot k_2 \cdot x_3 \cdot x_3 - k_{-2} \cdot x_4 \\ \frac{dx_4}{dt} = k_2 \cdot x_3^2 - k_2 \cdot x_4 + \frac{v_4 \cdot x_5}{p_4 + x_5} - k_3 \cdot x_4 - k_{-3} \cdot x_5 \\ \frac{dx_5}{dt} = \cdots \\ \quad \vdots \\ \frac{dx_n}{dt} = -k_1 \cdot x_1 \cdot c_2 + k_{-1} \cdot x_3 \end{cases}$$

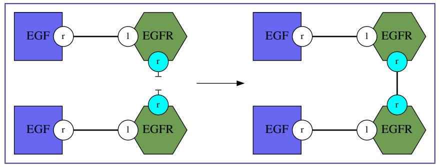
knowledge representation **and** **models of the behaviour of systems**

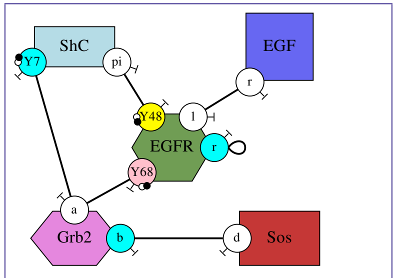
Site-graphs rewriting



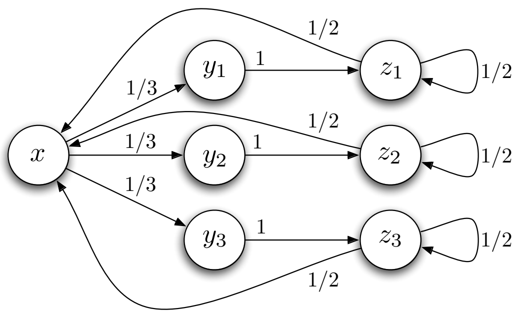
- a language close to knowledge representation;
- rules are easy to update;
- a compact description of models.

Choices of semantics





interaction map

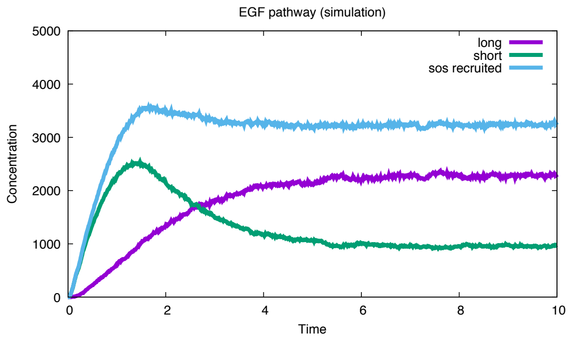


Markov chain

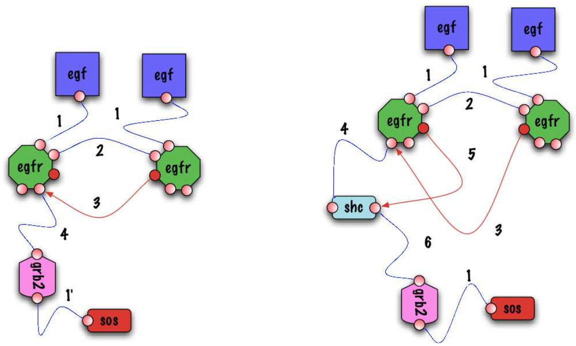
$$
\begin{cases}
\frac{dx_1}{dt} = -k_1 \cdot x_1 \cdot x_2 + k_{-1} \cdot x_3 \\
\frac{dx_2}{dt} = -k_1 \cdot x_1 \cdot x_2 + k_{-1} \cdot x_3 \\
\frac{dx_3}{dt} = k_1 \cdot x_1 \cdot x_2 - k_{-1} \cdot x_3 + 2 \cdot k_2 \cdot x_3 \cdot x_3 - k_{-2} \cdot x_4 \\
\frac{dx_4}{dt} = k_2 \cdot x_3^2 - k_2 \cdot x_4 + \frac{v_4 \cdot x_5}{p_4 + x_5} - k_3 \cdot x_4 - k_{-3} \cdot x_5 \\
\frac{dx_5}{dt} = \cdots \\
\vdots \\
\frac{dx_n}{dt} = -k_1 \cdot x_1 \cdot c_2 + k_{-1} \cdot x_3
\end{cases}
$$

ordinary differential equations

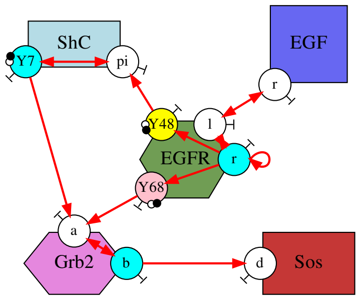
Abstractions offer different perspectives on models



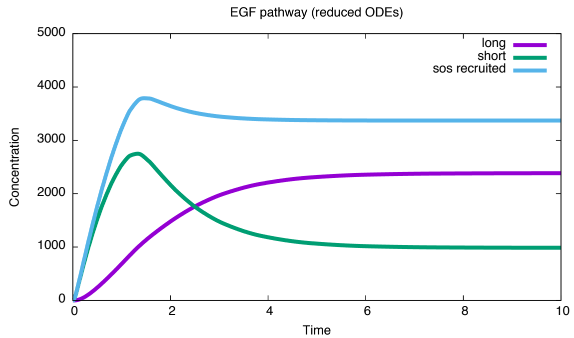
concrete semantics



causal traces



information flow



exact projection
of the ODE semantics

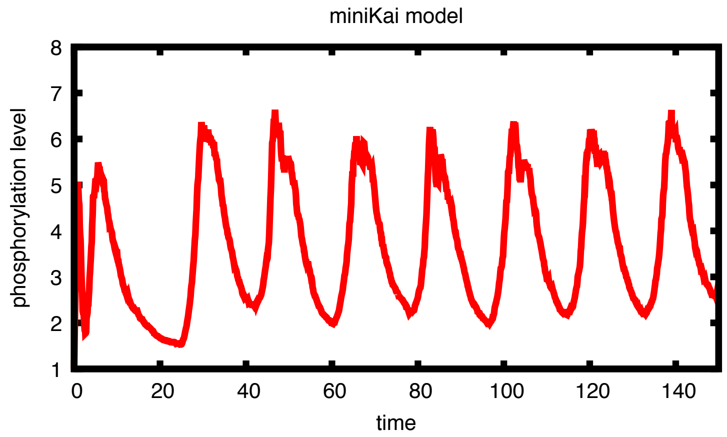
The need for counters

Modelers makes *a priori* simplification because:

- some knowledge may be missing;
- there is no way to describe compactly what is known;
- mechanistic details would make models intractable.

Example: (cirdadian clock)

- KaiC has six phosphorylation sites;
- phosphorylation rate depends on the number of sites already phosphorylated.

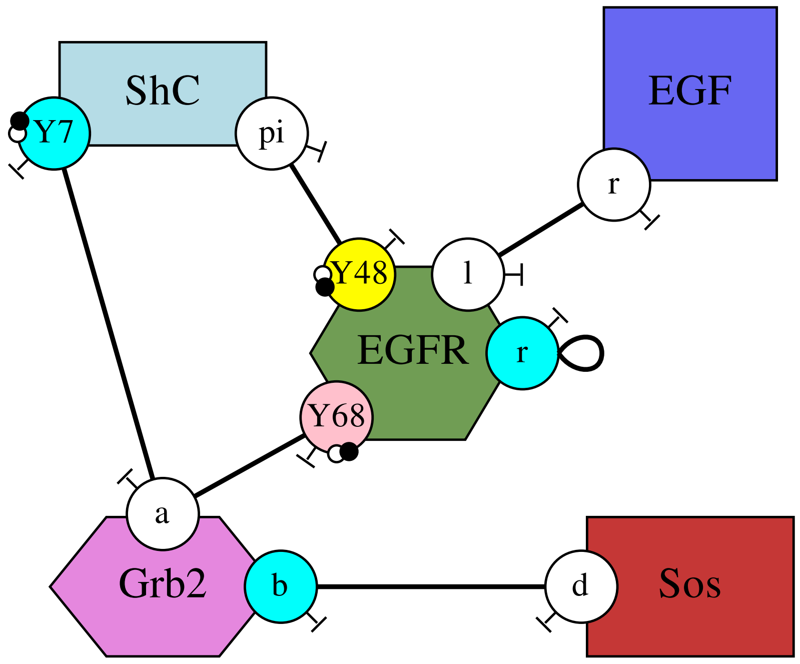


Thanks to counters, there is no need to enumerate all the potential configurations.

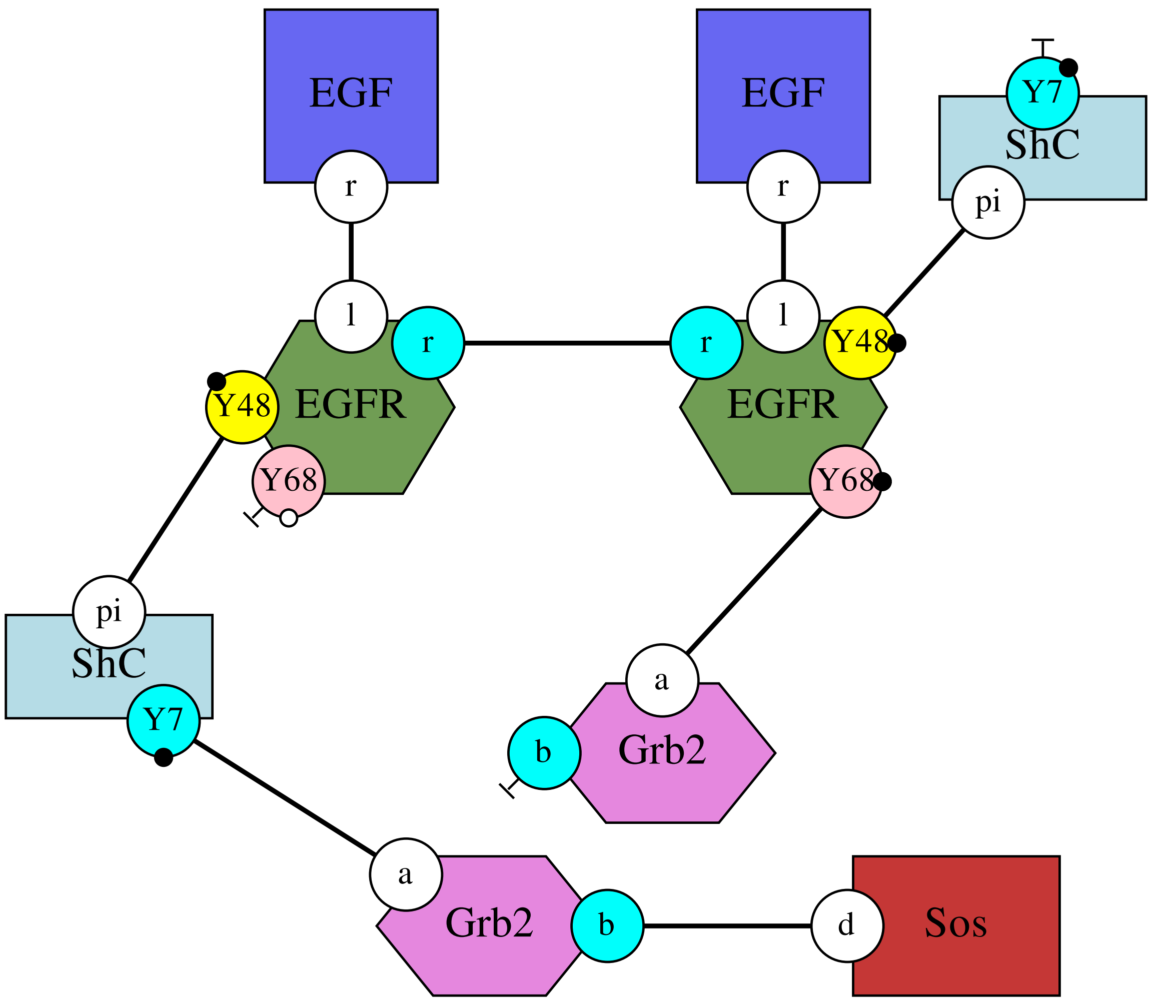
On the menu today

1. Mechanistic models
2. Kappa
3. Causality
4. Efficient simulation
5. Static analysis
6. Conclusion

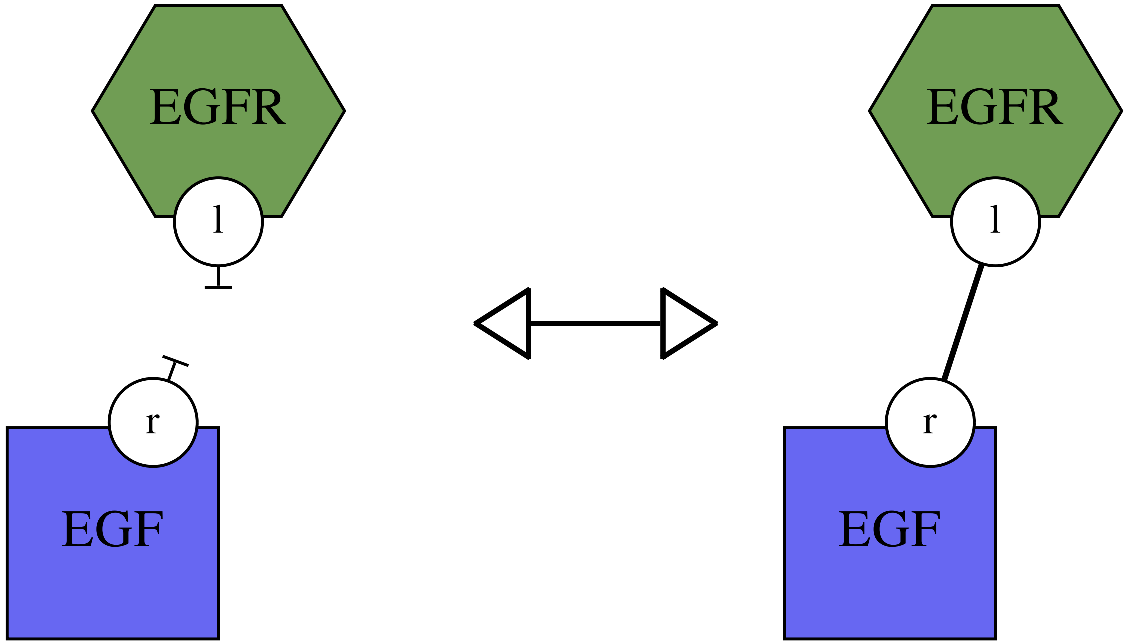
Signature



Bio-molecular complex

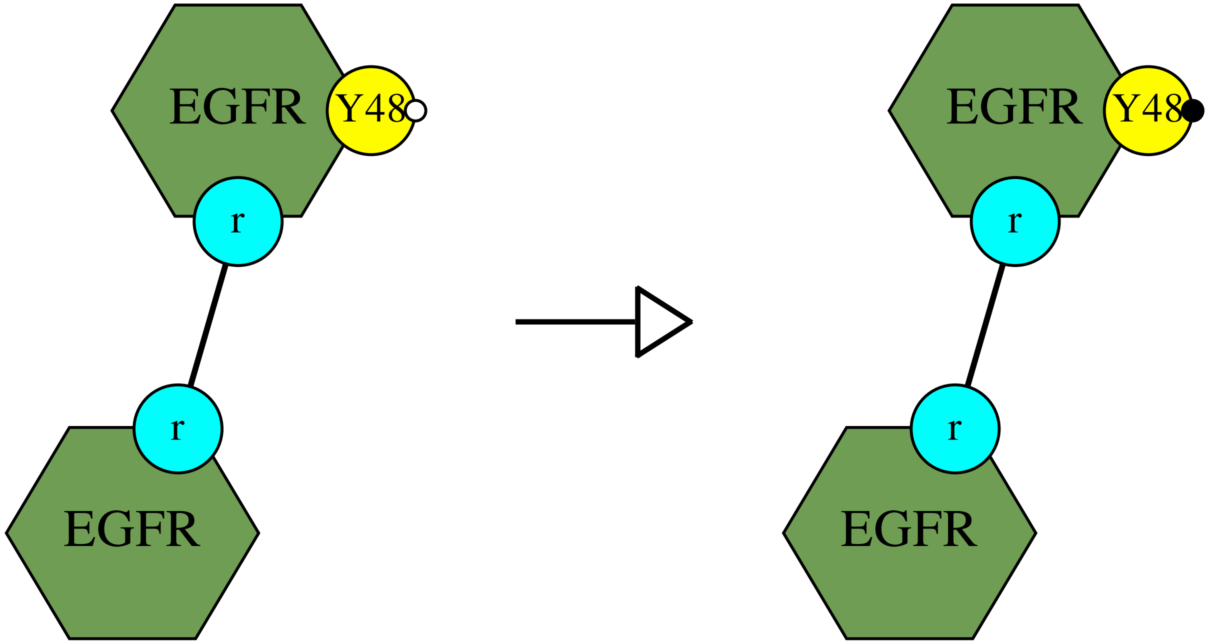


Binding unbinding



@ k, k_d

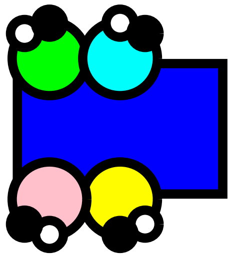
Phosphorylation rule



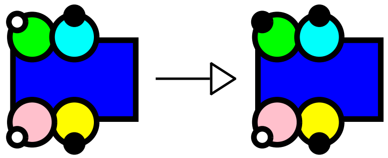
@ k_p

Case-study

- Signature:

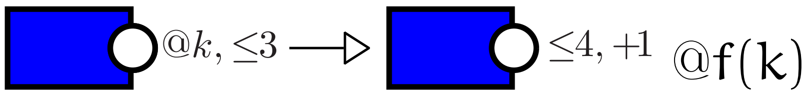


- Let $f(i)$ be the rate of phosphorylation of a given site assuming that exactly i sites are already phosphorylated.
- Phosphorylation of the site  when only the sites  and are phosphorylated.

 @f(2)

Overall we would need $4 \cdot 2^3$ rules to model phosphorylation.

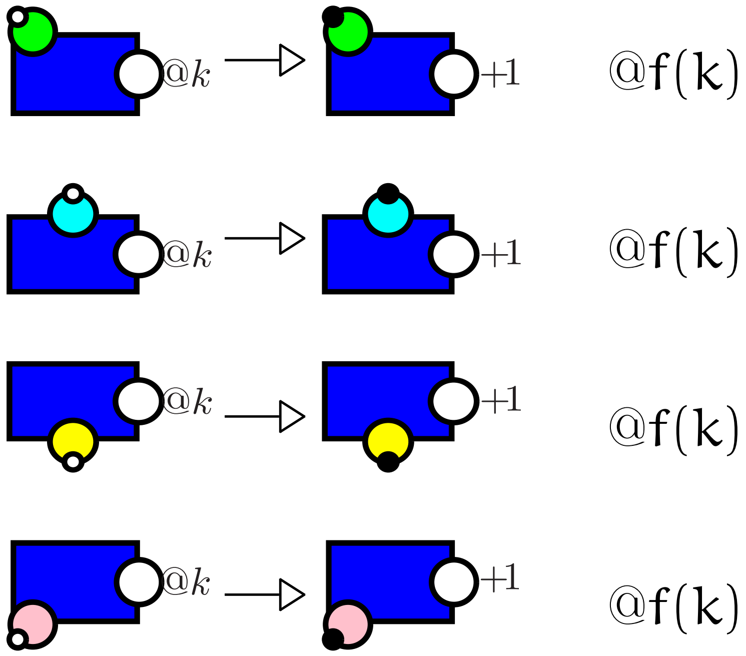
Counters



- $= \mathsf{2}$, $\leq \mathsf{3}$, $\geq \mathsf{1}$: preconditions and postconditions about the value of a counter;
- $@\mathsf{k}$ binds k to the value of the counter (to define the rule rate);
- $= \mathsf{2}$, $+\mathsf{1}$, $-\mathsf{3}$: action on the value of a counter.

Case study

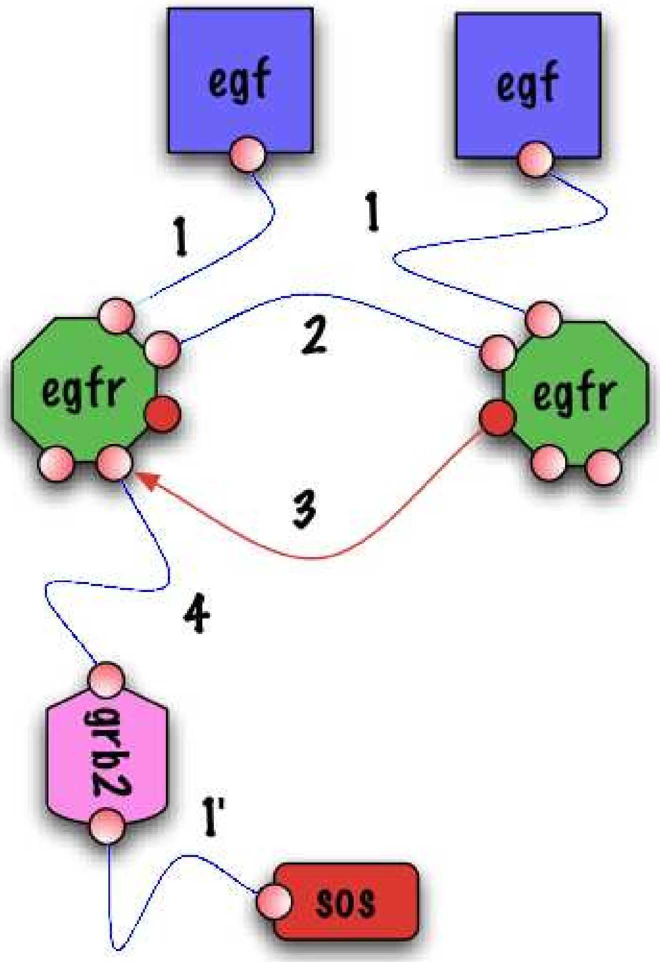
4 rules are enough to describe our case study:

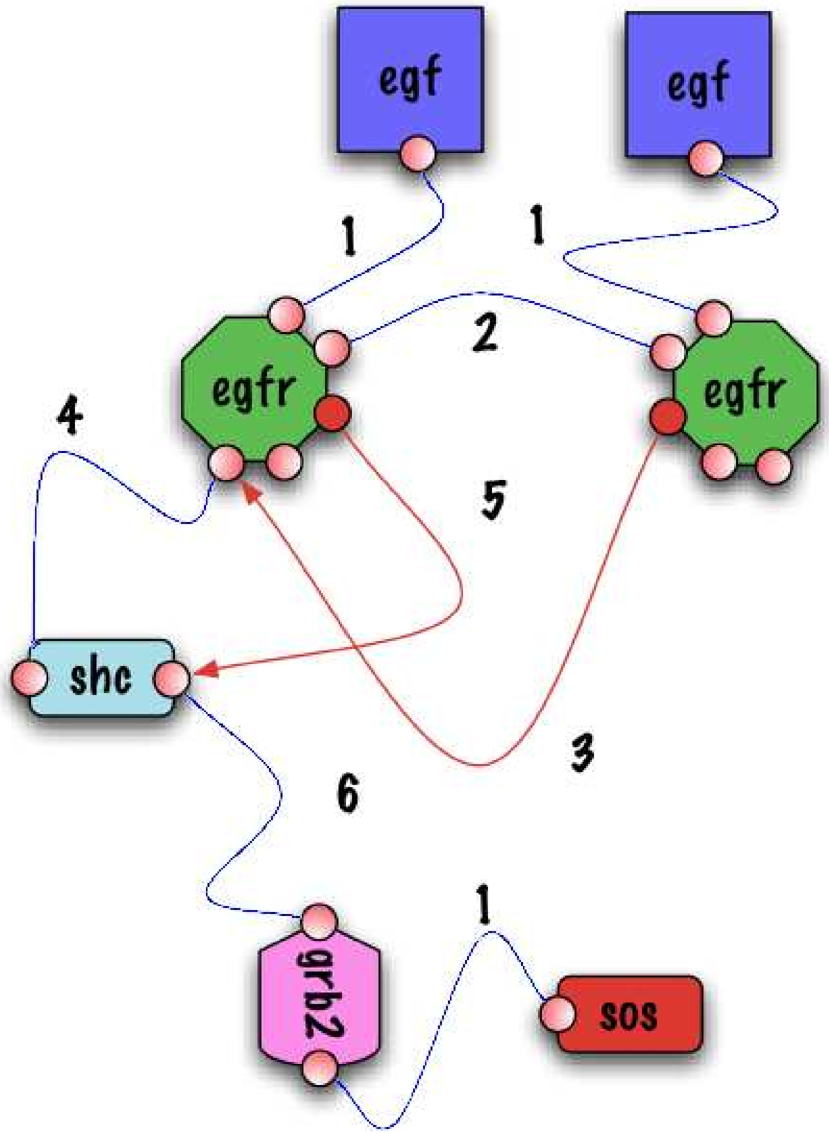


On the menu today

1. Mechanistic models
2. Kappa
3. Causality
4. Efficient simulation
5. Static analysis
6. Conclusion

Causal traces





Challenges

Compute minimal traces up to commutation of concurrent events.

Parametric with respect to:

- the notion of states
- the notion of event

which can be seen at different levels of abstraction.

The choices of the syntax and of the semantics for the modelling paradigm are crucial.

The biochemical structure is required

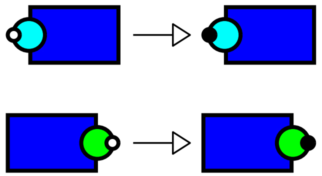
Reactions:

$$\begin{array}{rcl} A & \rightarrow & {}^{\bullet}A \\ A & \rightarrow & A^{\bullet} \\ {}^{\bullet}A & \rightarrow & {}^{\bullet}A^{\bullet} \\ A^{\bullet} & \rightarrow & {}^{\bullet}A^{\bullet} \end{array}$$

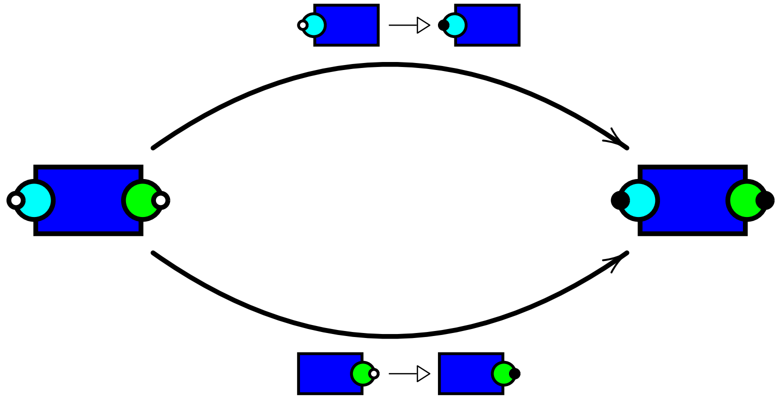
Causal traces:

$$\begin{array}{l} A \rightarrow {}^{\bullet}A \rightarrow {}^{\bullet}A^{\bullet} \\ A \rightarrow A^{\bullet} \rightarrow {}^{\bullet}A^{\bullet} \end{array}$$

Rules:

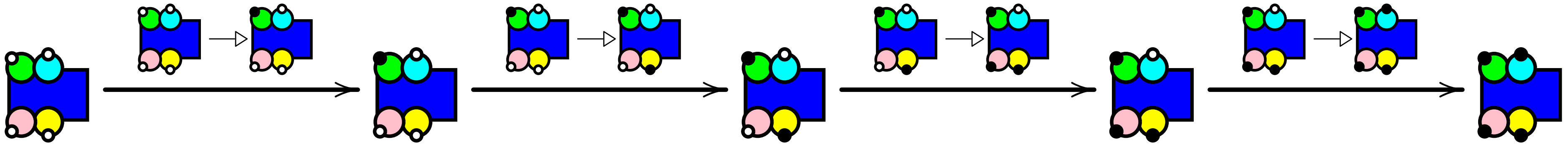


Causal traces:



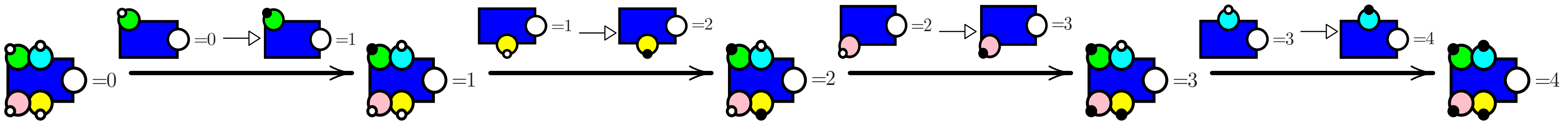
Causal traces for the circadian clock in classical Kappa

Example of causal trace:



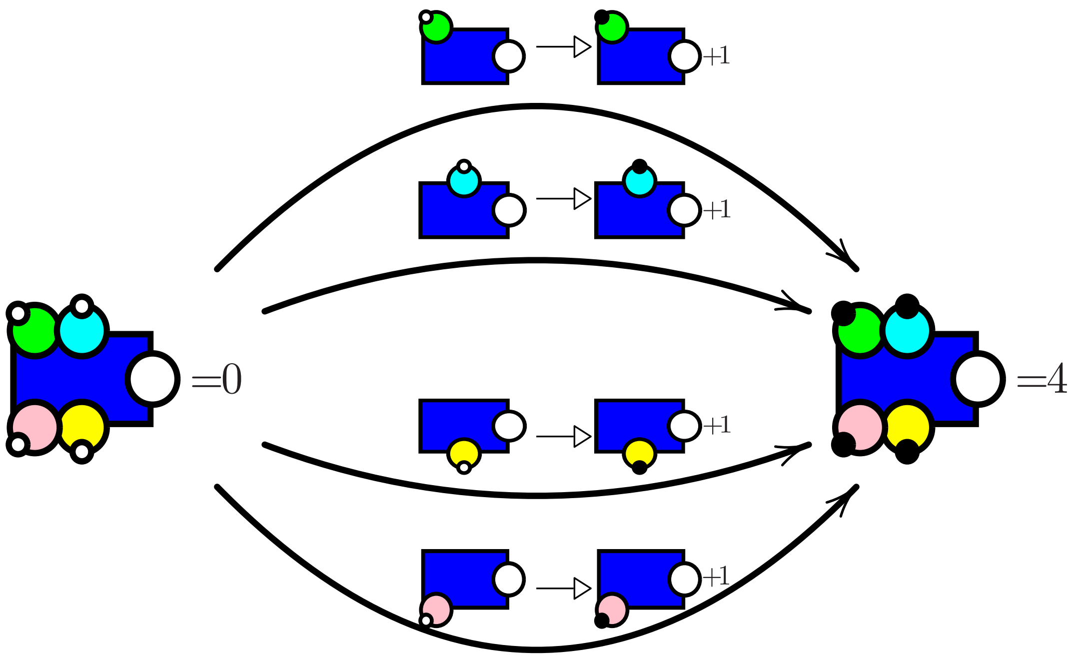
Causal traces for the circadian clock with flat counters

Example of causal trace:



Causal traces for the circadian clock with arithmetic counters

Only one causal trace:



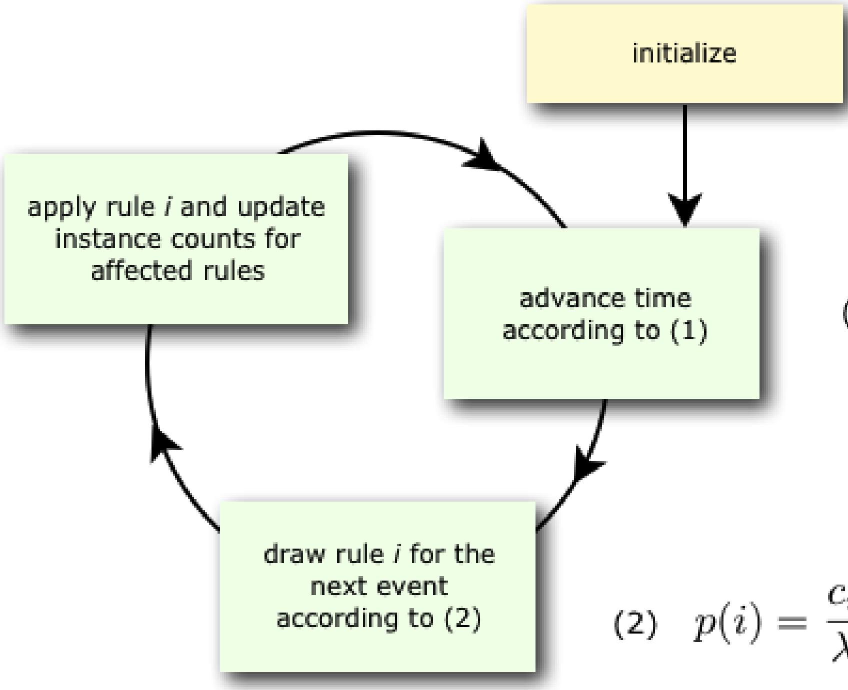
On the menu today

1. Mechanistic models
2. Kappa
3. Causality
4. Efficient simulation
5. Static analysis
6. Conclusion

Simulation

c_i ... "activity" of rule *i*

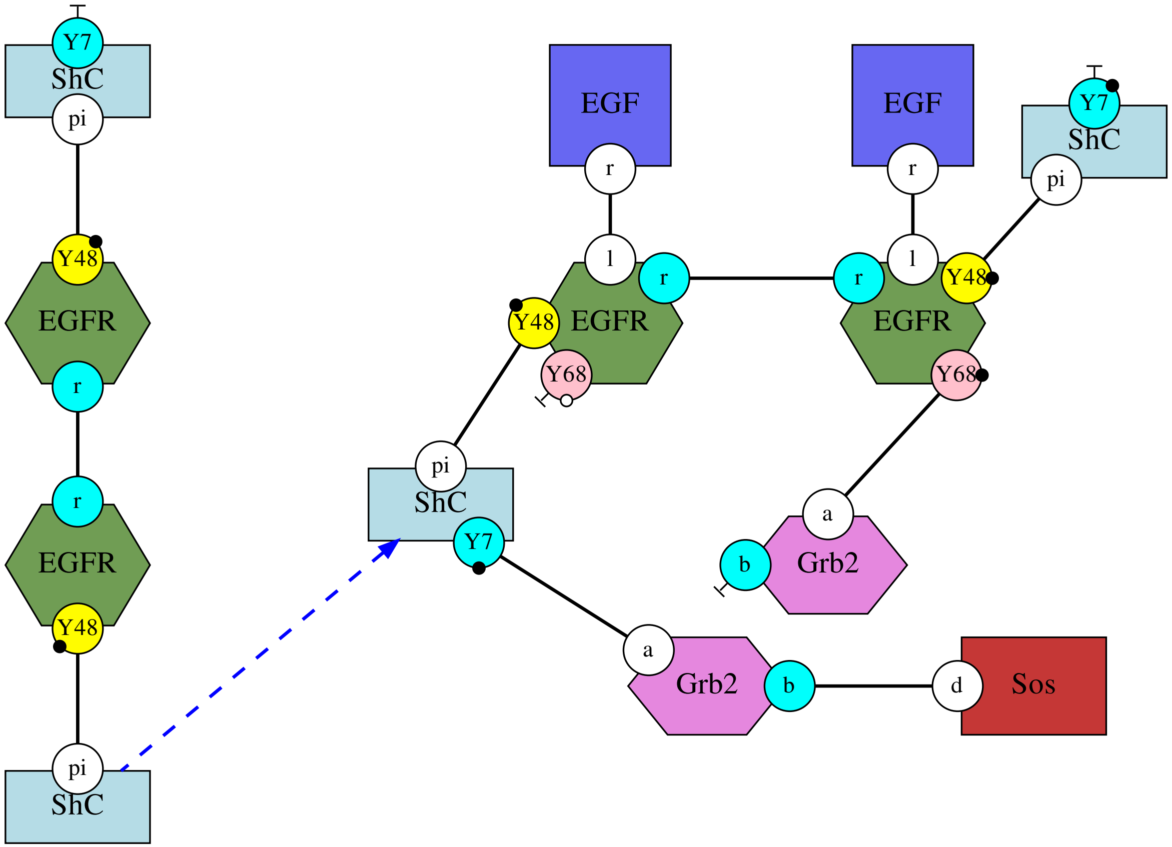
$\lambda = \sum_{i}^{n} c_i$ ... "system activity"



(1) $T \leftarrow T + t$ with t drawn from $p(t) = \lambda e^{-\lambda t}$

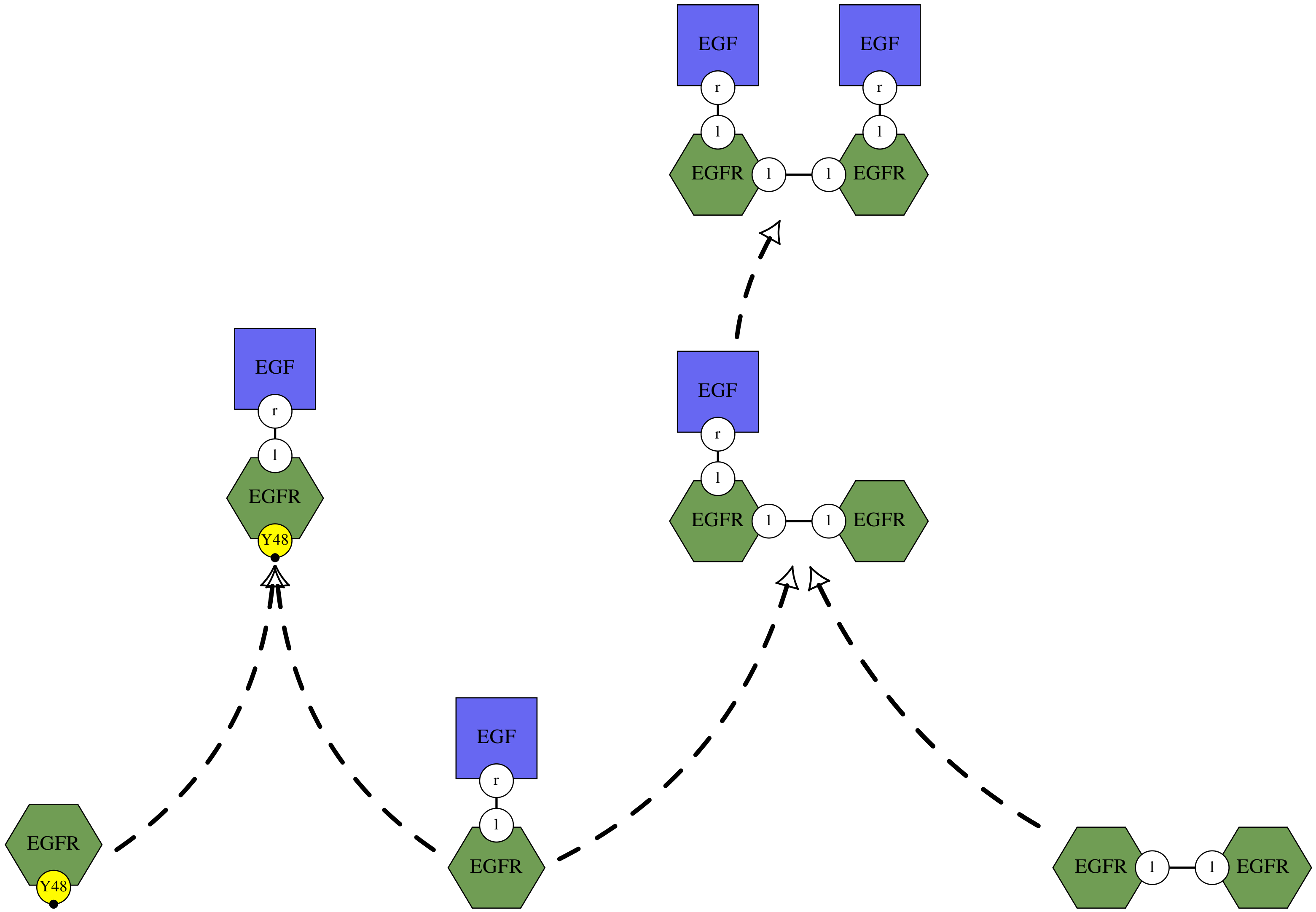
(2) $p(i) = \frac{c_i}{\lambda}$

Rigidity



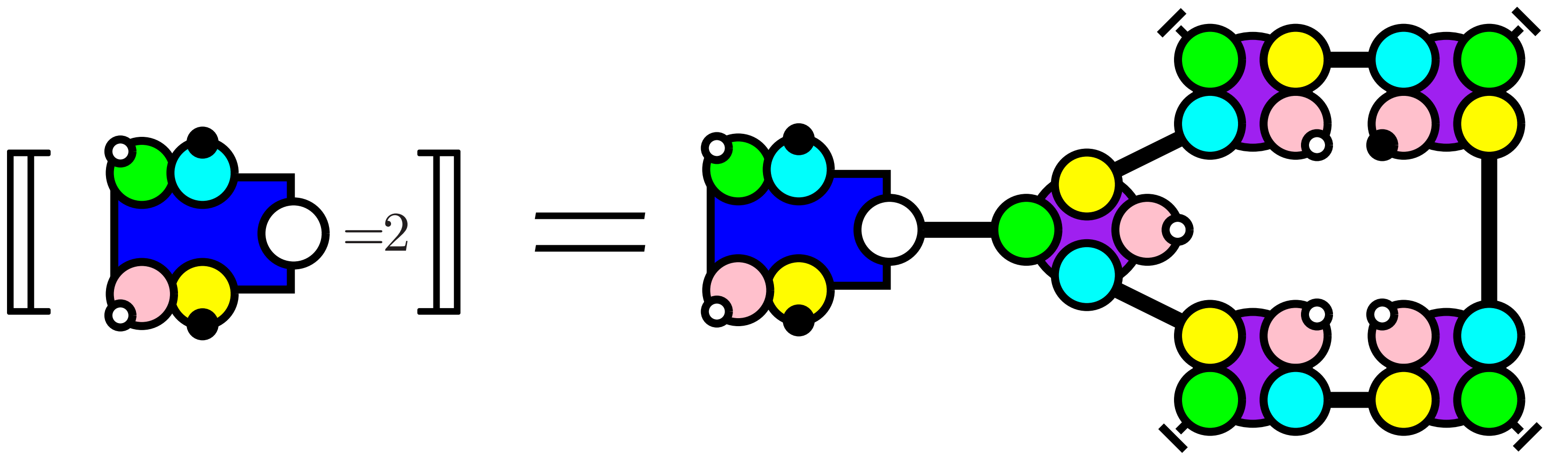
One matching from a connected pattern into a pattern is fully characterised by the image of an agent.

Extension bases

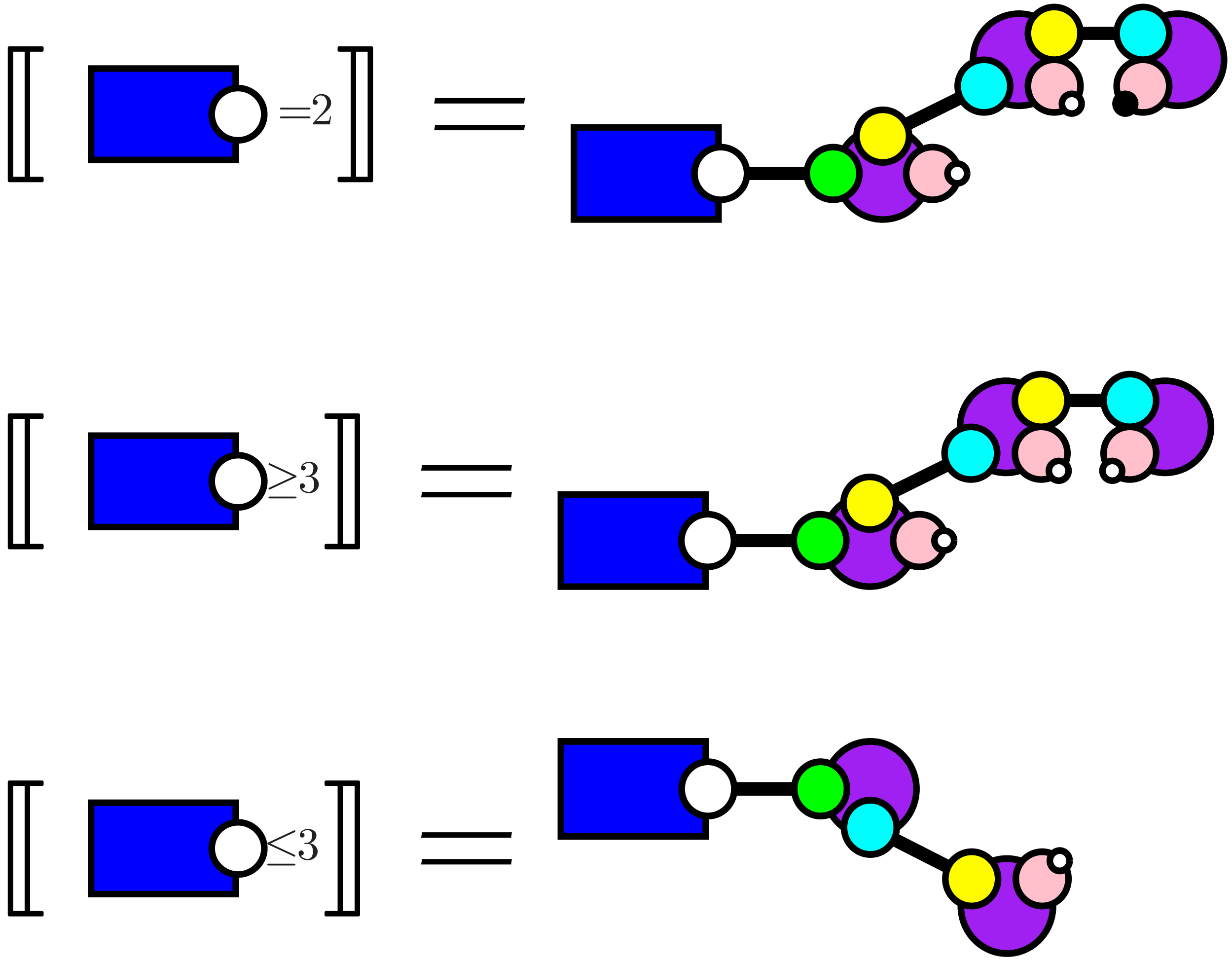


Counter encoding

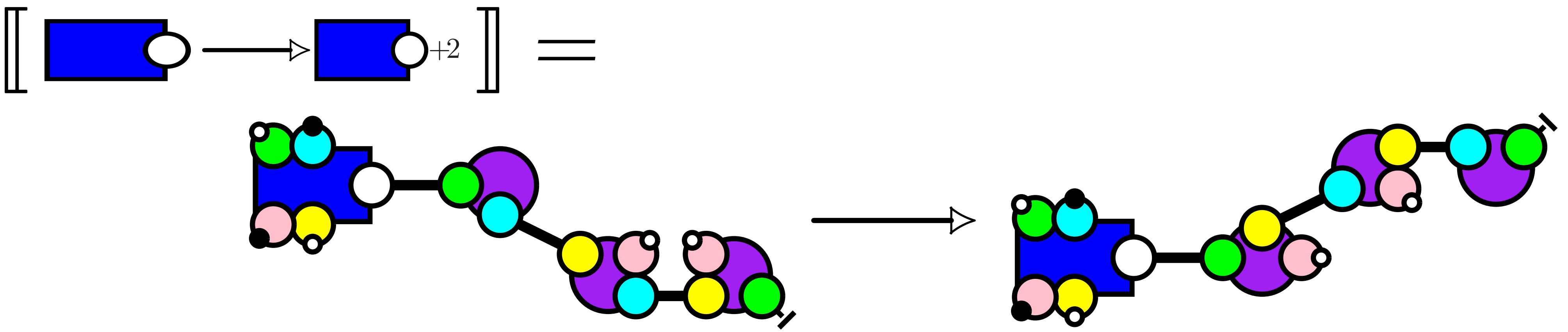
We assume that counters may range between 0 and n (here $n = 4$)

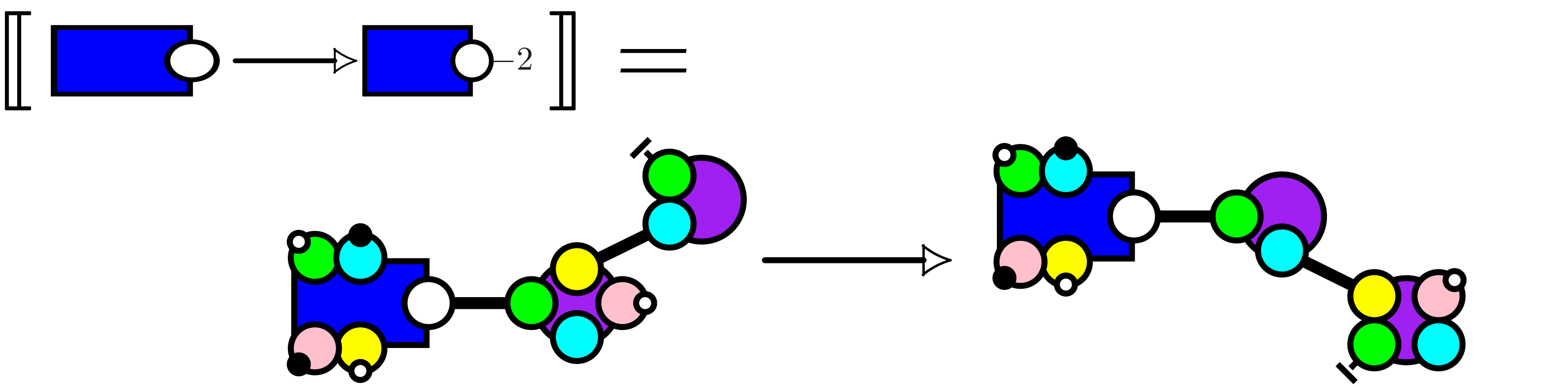


Tests



Updates





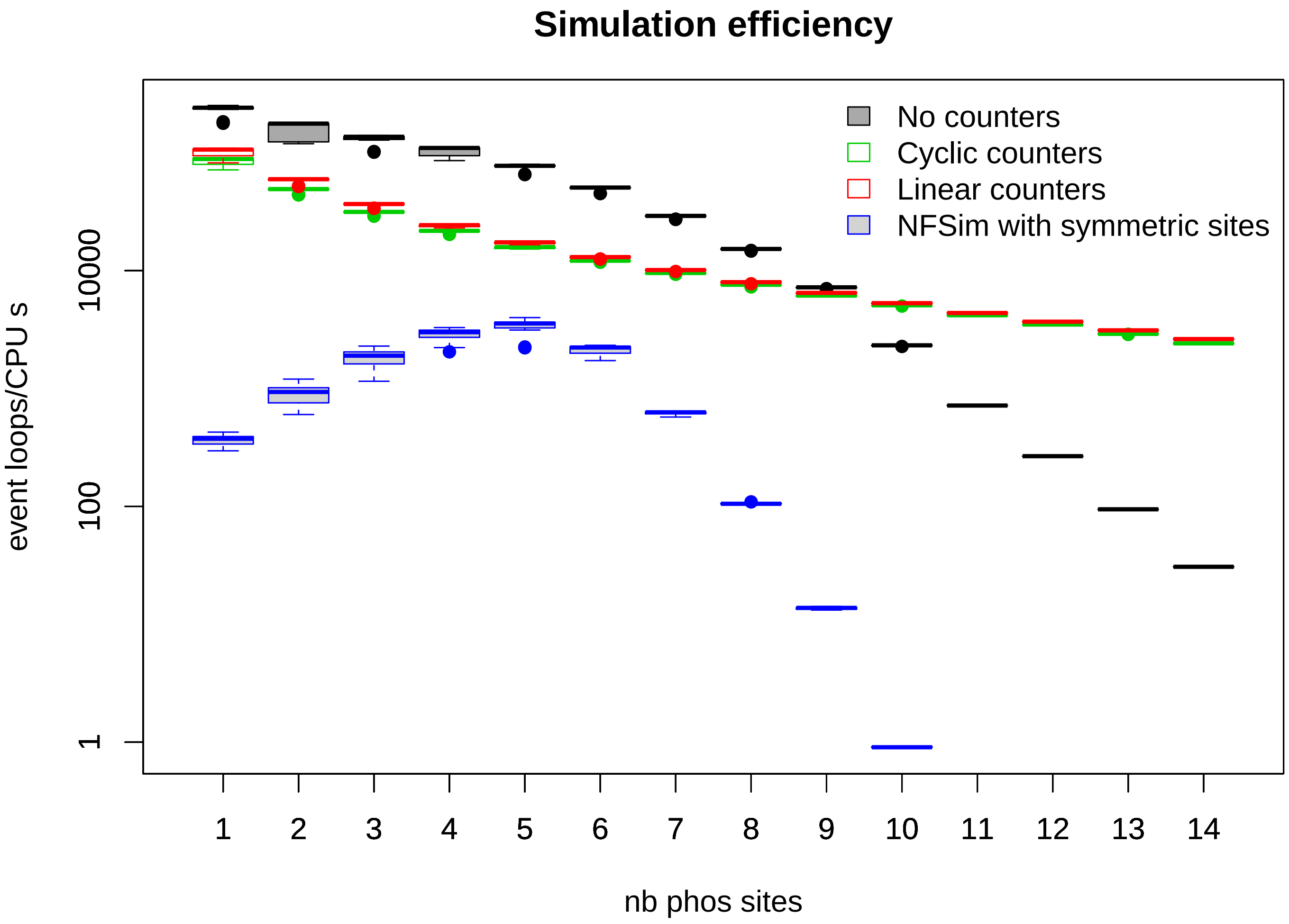
Coherence

Let $\mathbb{Q}_{\Sigma}$ be the set of states (initial signature).
Let Err_{Σ} be the set of erroneous states (whith some counters out of their bounds)
Let $\mathbb{Q}_{[\![\Sigma]\!]}$ be the set of states (signature of the encoding).

Theorem 1 (correspondence) Let $\mathsf{G} \in \mathbb{Q}_{\Sigma} \setminus Err_{\Sigma}$ and r be a rule.
Both following properties are satisfied:

1. if $\exists \mathsf{G}' \in \mathbb{Q}_{\Sigma} \setminus Err_{\Sigma}$ such that $\mathsf{G} \xrightarrow{\mathrm{r}} \mathsf{G}'$,
   then $[\![\mathsf{G}]\!] \xrightarrow{[\![\mathrm{r}]\!]} [\![\mathsf{G}']\!]$;
2. if $\exists \mathsf{E}' \in \mathbb{Q}_{[\![\Sigma]\!]}$ such that $[\![\mathsf{G}]\!] \xrightarrow{[\![\mathrm{r}]\!]} \mathsf{E}'$,
   then $\exists \mathsf{G}' \in \mathbb{Q}_{\Sigma} \setminus Err_{\Sigma}$ such that $\mathsf{G} \xrightarrow{\mathrm{r}} \mathsf{G}'$ and $[\![\mathsf{G}']\!] = \mathsf{E}'$.

Benchmarks



100 agents, 15 simulations of 10^5 events

On the menu today

1. Mechanistic models
2. Kappa
3. Causality
4. Efficient simulation
5. Static analysis
6. Conclusion

Goal of static analysis

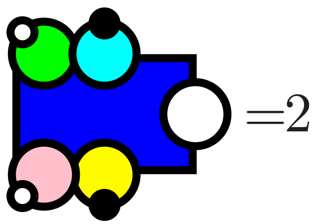
At edit time:

1. Prove that the values of counters are bounded;
2. Infer the ranges of counters;
3. Retrieve the meaning of counters.

Translation

Each agent instance is translated into a vector:

For instance, the following protein:



gets translated into the following vector:

$$\begin{cases} x^{\circ}_{\color{green}\bullet} = 1, x^{\bullet}_{\color{green}\bullet} = 0, \\ x^{\circ}_{\color{cyan}\bullet} = 0, x^{\bullet}_{\color{cyan}\bullet} = 1, \\ x^{\circ}_{\color{yellow}\bullet} = 0, x^{\bullet}_{\color{yellow}\bullet} = 1, \\ x^{\circ}_{\color{pink}\bullet} = 1, x^{\bullet}_{\color{pink}\bullet} = 0, \\ x_{\circ} = 2. \end{cases}$$

Abstract domain

We use a reduced product between:

- intervals (to express properties of interest);
- affine equalities (to make their proof).

In our example:

- inductive invariant:

$$x_{\circ} = x^{\bullet}_{\color{green}\bullet} + x^{\bullet}_{\color{cyan}\bullet} + x^{\bullet}_{\color{yellow}\bullet} + x^{\bullet}_{\color{pink}\bullet}$$

- invariant:

$$0 \leq x_{\circ} \leq 4.$$

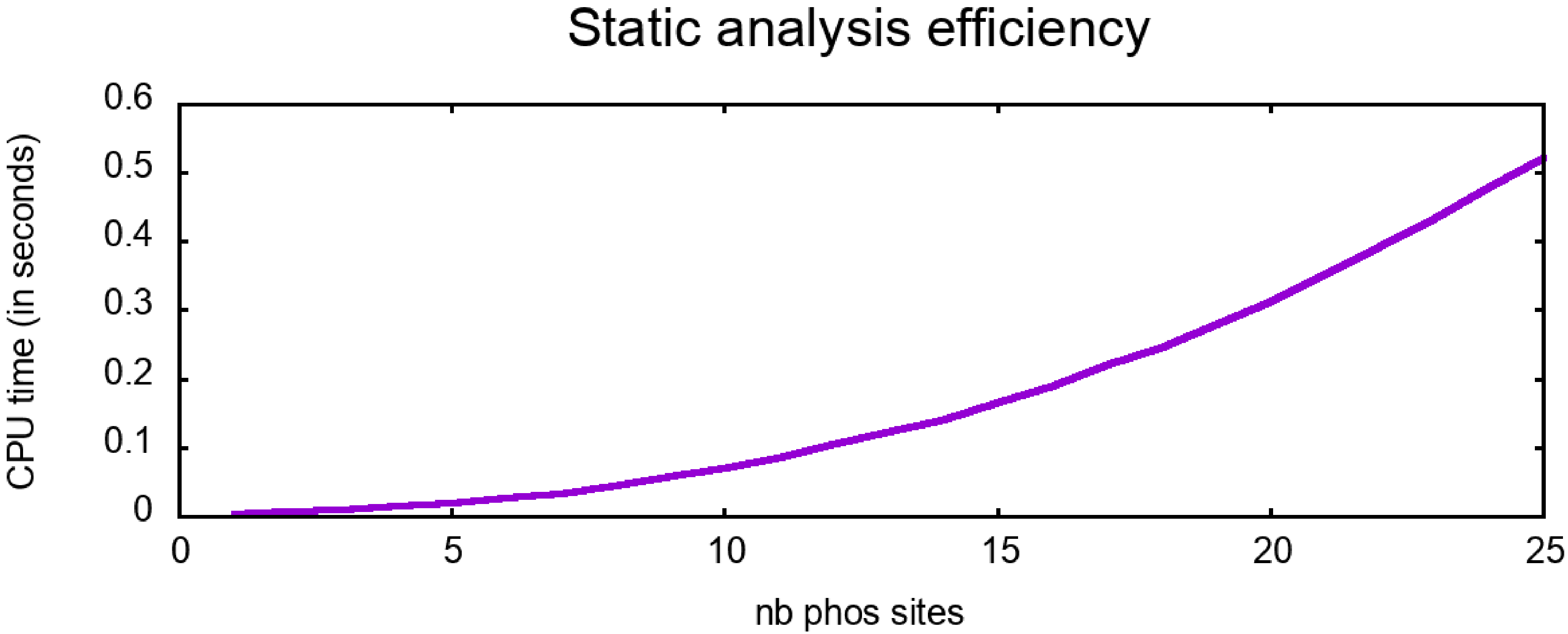
Approximate reduced product

An exact reduced product would be NP.
We use:

- Gaus reduction: $$\begin{cases} x+y+z=1 \\ x+y+t=2 \end{cases} \Longrightarrow \begin{cases} x+y+z=1 \\ t-z=1 \end{cases}$$
- Interval propagation: $$\begin{cases} x+y+z=3 \\ x\in[0;\infty) \\ y\in[0;\infty) \\ z\in[0;\infty) \end{cases} \Longrightarrow \begin{cases} x+y+z=3 \\ x\in[0;3] \\ y\in[0;\infty) \\ z\in[0;\infty) \end{cases}$$
- Redundancy introduction: $$\begin{cases} x+y-z=3 \\ x\in[1;2] \end{cases} \Longrightarrow \begin{cases} x+y-z=3 \\ y-z\in[1;2] \\ x\in[1;2] \end{cases}$$

to get a cubic approximated reduced product.

Benchmarks



Static analysis efficiency
CPU time (in seconds)
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20
25
nb phos sites

Conclusion

We are equipped Kappa with counters, including:

1. an extension of the SPO semantics;
2. a parsimonious notion of causality;
3. an efficient simulation;
4. static analysis to retrieve the meaning of counters and infer proofs obligations.

This provides a confortable model environment to describe and use models with a lot of symmetries.

Future works

Make counters implicit:

- no need to specify them,
- allow preconditions that counts the number of sites satisfying a given property.