HDR defense

Static analysis and model reduction for site-graph rewriting

Jérôme Feret
DI - ÉNS (INRIA/ENS/CNRS/PSL)

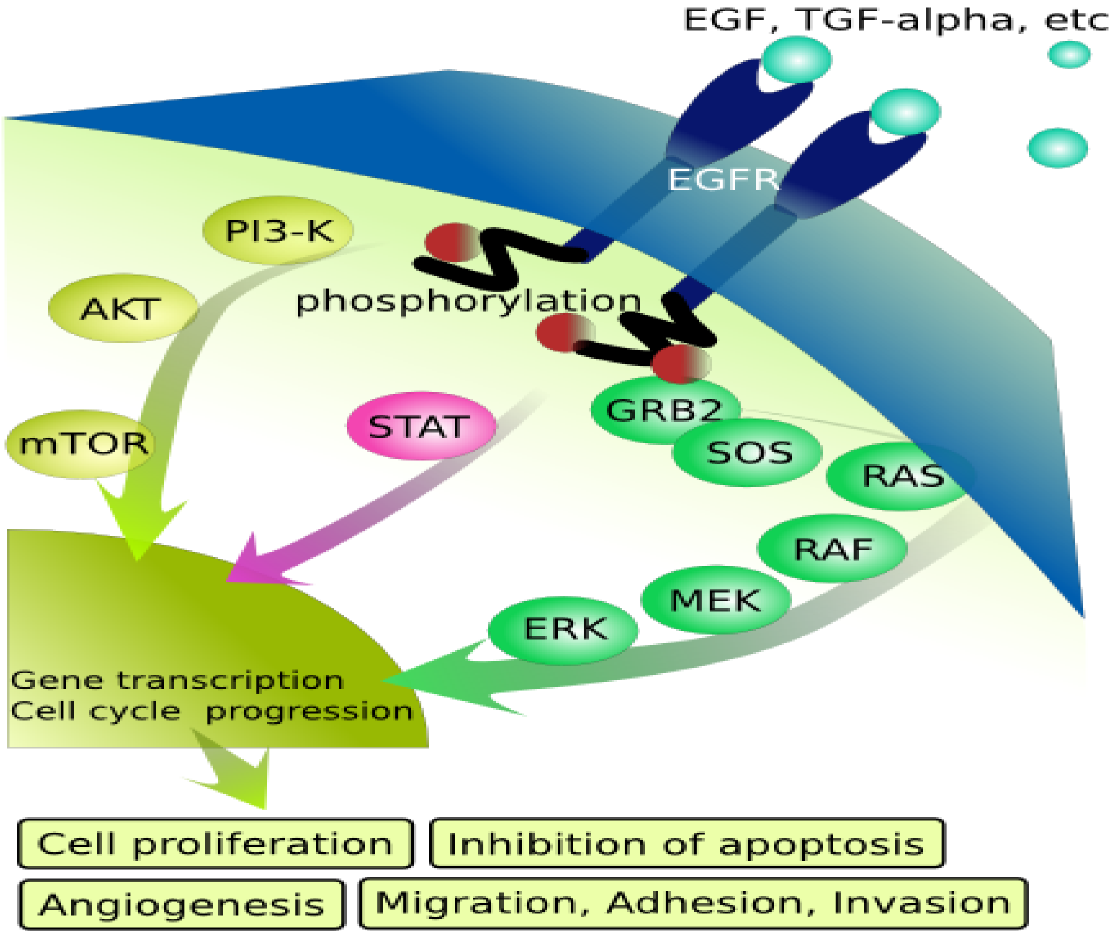
http://www.di.ens.fr/~feret

Tuesday, the 12th of December, 2023

Outline

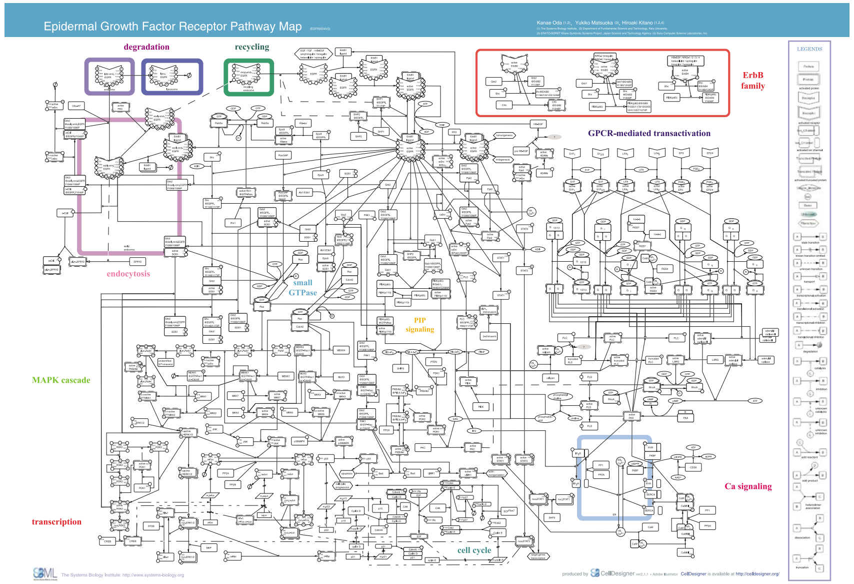
1. Context and motivations
2. Static analysis
3. Flow-based model reduction (differential)
4. Flow-based model reduction (stochastic)
5. Symmetries
6. Conclusion and perspectives

Intra-cellular signalling pathways



Eikuch, 2007

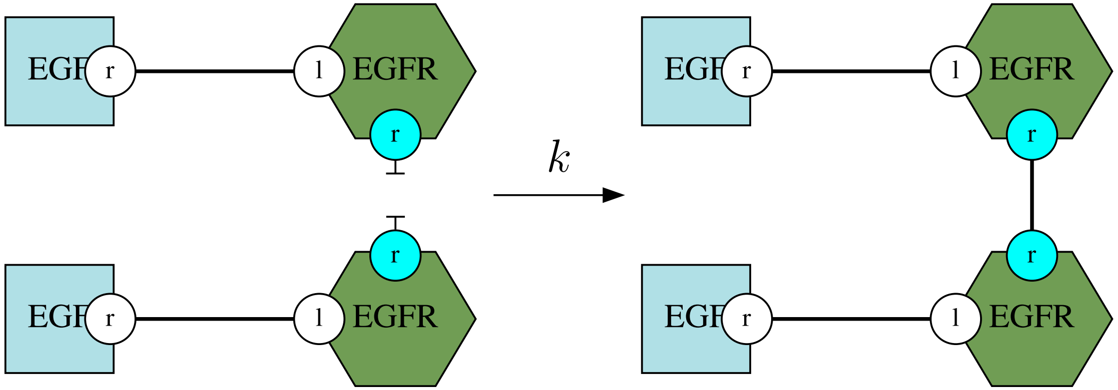
Bridge the gap between. . .



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

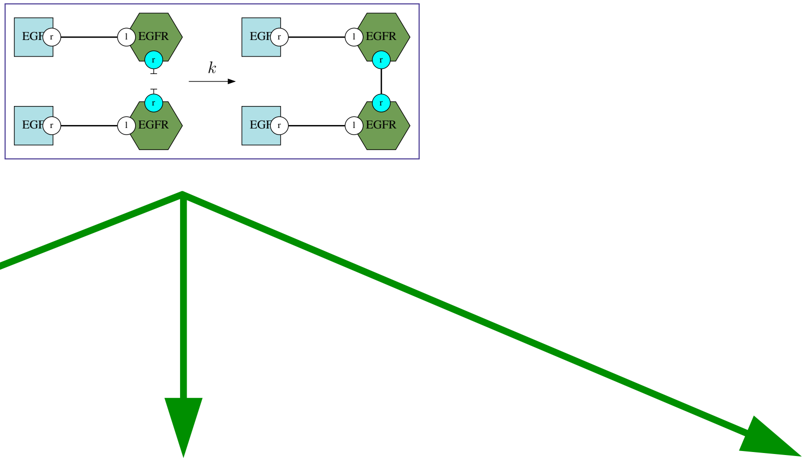
knowledge representation **and** **models of the behaviour of systems**

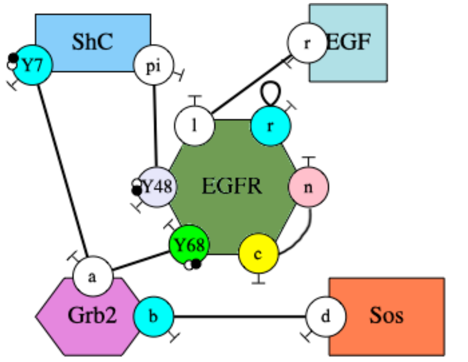
Site-graph rewriting



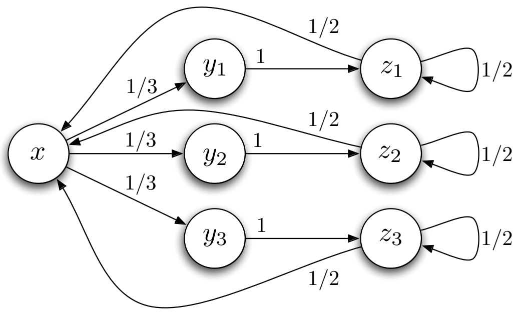
- a language close to knowledge representation;
- rules are easy to update;
- a compact description of models.

Choices of semantics





interaction map

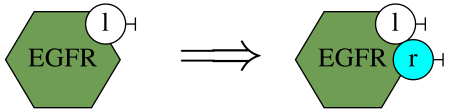


Markov chain

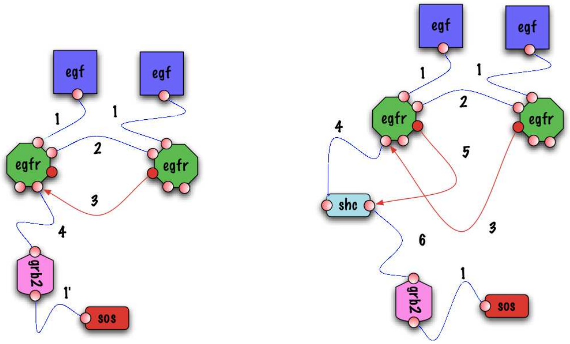
$$\begin{cases}
\frac{dx_1}{dt} = -k_1 \cdot x_1 \cdot x_2 + k_{-1} \cdot x_3 \\
\frac{dx_2}{dt} = -k_1 \cdot x_1 \cdot x_2 + k_{-1} \cdot x_3 \\
\frac{dx_3}{dt} = k_1 \cdot x_1 \cdot x_2 - k_{-1} \cdot x_3 + 2 \cdot k_2 \cdot x_3 \cdot x_3 - k_{-2} \cdot x_4 \\
\frac{dx_4}{dt} = k_2 \cdot x_3^2 - k_2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

ordinary differential equations

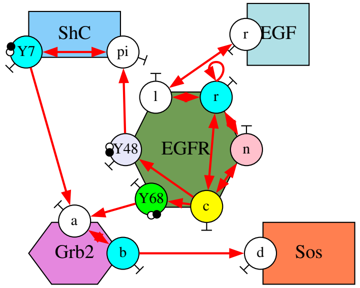
Abstractions offer different perspectives



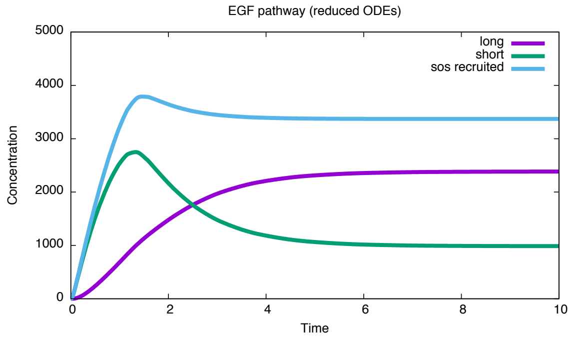
static analysis



causal traces



information flow



exact projection
of the ODE semantics

Bottom up approach in modeling

1. Reaction-based modeling: Reactions are enumerated explicitly.
   - Reaction-networks [Feinberf, 1979]
   - Petri-nets [Heiner, 2003]
   - BIOCHAM [Fages *et al.* 2004]
2. Agent-based modeling: Each agent describes its potential behaviors.
   - π-calculus [Cardelli *et al.*, 2009]
   - communicating automata [Cardelli, 2007]
   - BlenX [Priami *et al.*, 2008]
3. Rule-base modeling: The behaviors of agents emerge from the rules
   - π-calculus [Regev *et al.*, 2001, 2004]
   - Kappa [Danos *et al.*, 2003-], BNGL [Faeder *et al.*, 2005-], Mød [Andersen etal, 2014-]
   - ML-Space: with spatial diffusion [Uhrmacher *et al.*, 2008-]

Outline

1. Context and motivations
2. Static analysis
    - Motivations
    - Case studies
    - Reachable patterns analysis
    - Specialization to orthogonal sets
    - Conclusion and perspectives
3. Flow-based model reduction (differential)
4. Flow-based model reduction (stochastic)
5. Symmetries
6. Conclusion and perspectives

Motivations

Main goal : Increase confidence in models

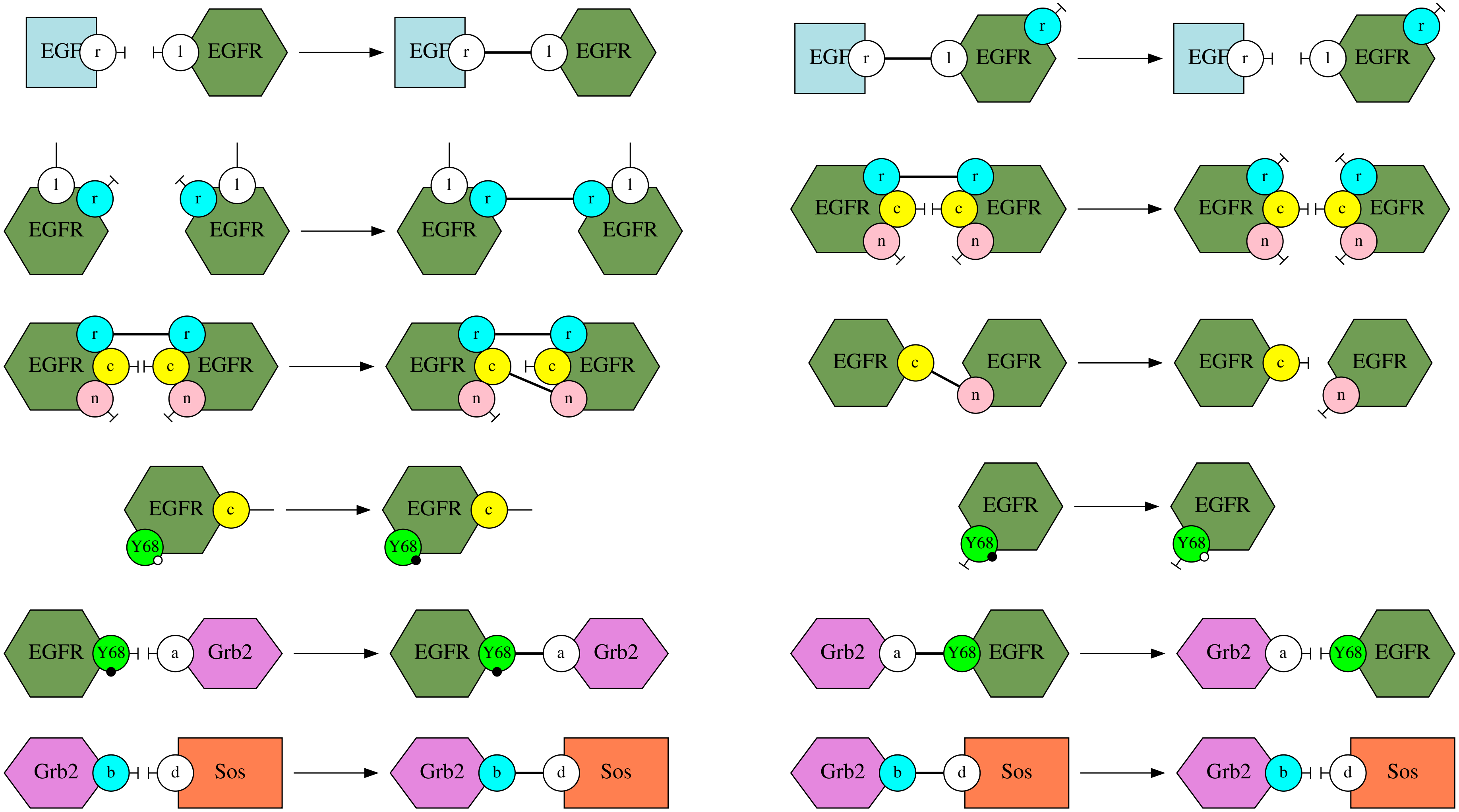
- Summarize structural properties :
  - Relationships between the states of sites in biochemical complexes;
  - Prove the absence of polymers.

- Detect dead rules :
  - due to typos;
  - due to complex causal properties ;
  - due to parts that may be missing.

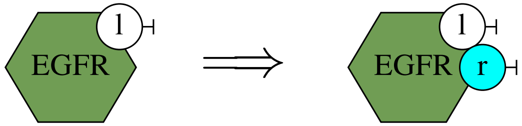
Outline

1. Context and motivations
2. Static analysis
    - Motivations
    - Case studies
    - Reachable patterns analysis
    - Specialization to orthogonal sets
    - Conclusion and perspectives
3. Flow-based model reduction (differential)
4. Flow-based model reduction (stochastic)
5. Symmetries
6. Conclusion and perspectives

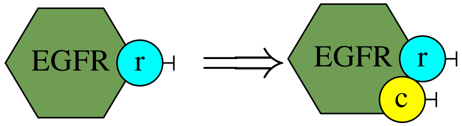
Some rules



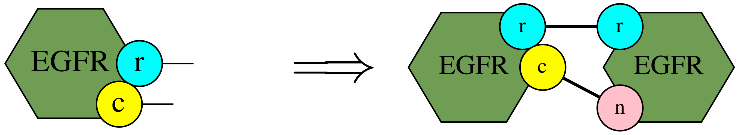
Some invariants

- 

whenever the site *l* is free, the site *r* is free as well.

- 

whenever the site *r* is free, the site *c* is free as well.

- 

whenever the sites *c* and *r* are both bound, they are bound to a single occurrence of an occurrence of the protein *EGFR*.

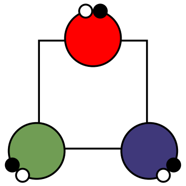
Outline

1. Context and motivations
2. Static analysis
    - Motivations
    - Case studies
    - Reachable patterns analysis
    - Specialization to orthogonal sets
    - Conclusion and perspectives
3. Flow-based model reduction (differential)
4. Flow-based model reduction (stochastic)
5. Symmetries
6. Conclusion and perspectives

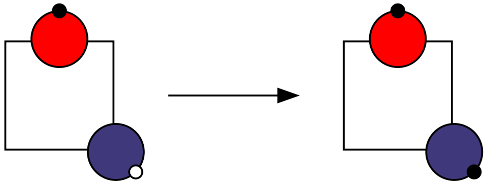
Reachability semantics

A rule stands for a multi-set of reactions.

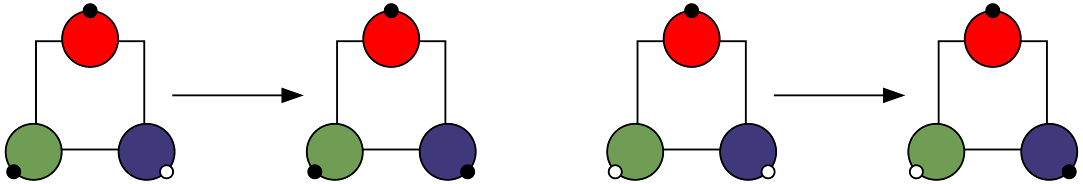
For instance, considering the protein:



the rule



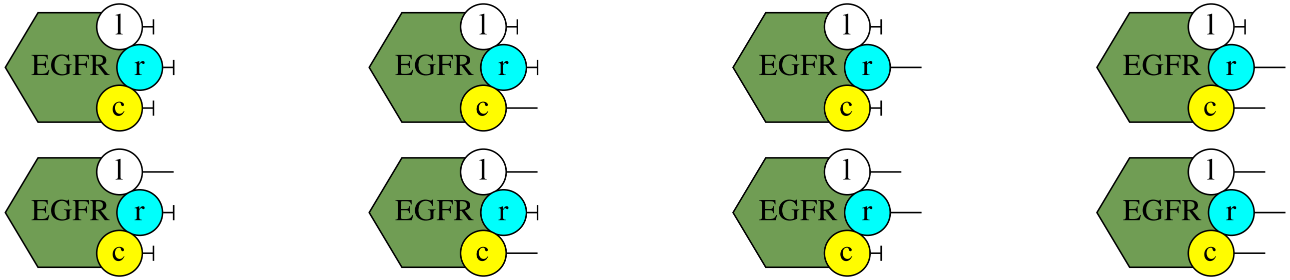
stands for both reactions:



We are interested in the set of states that are reachable after applying an arbitrary number of reactions (starting from an initial state).

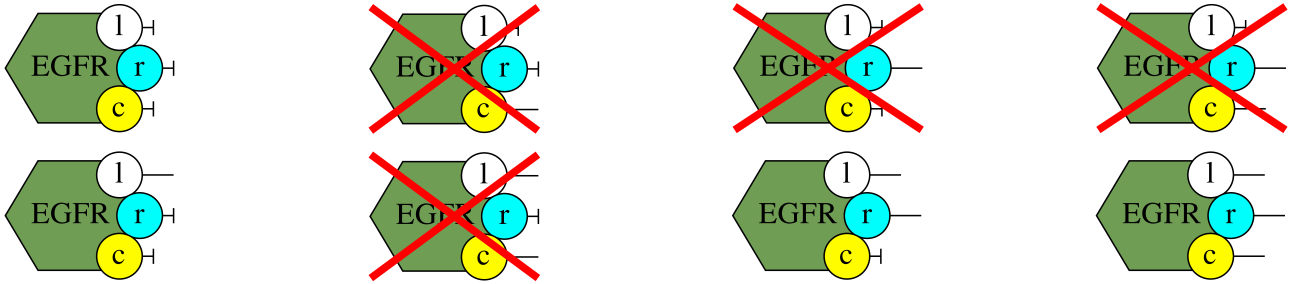
Abstract domain

Among a set of patterns,



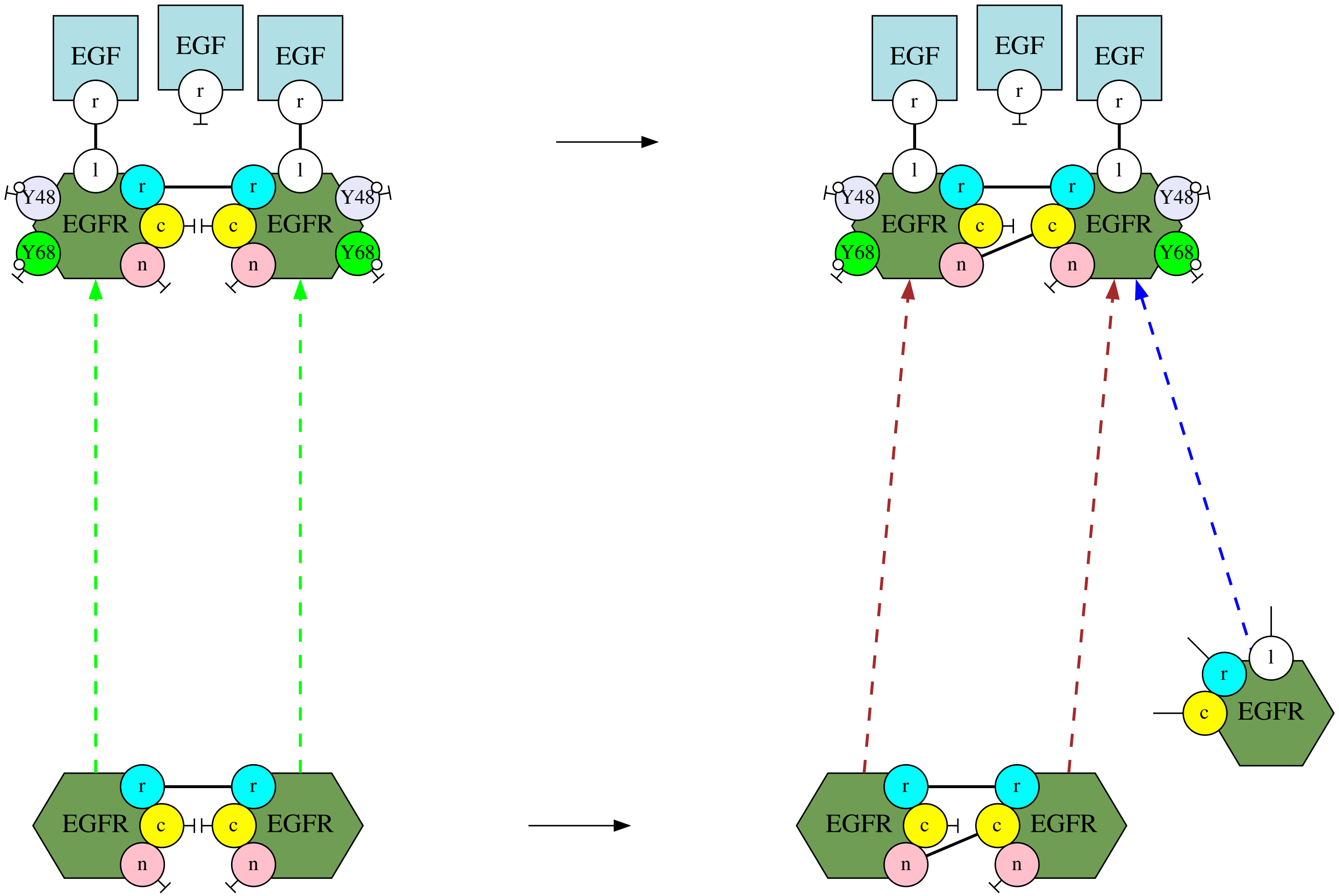
Abstract domain

Among a set of patterns,

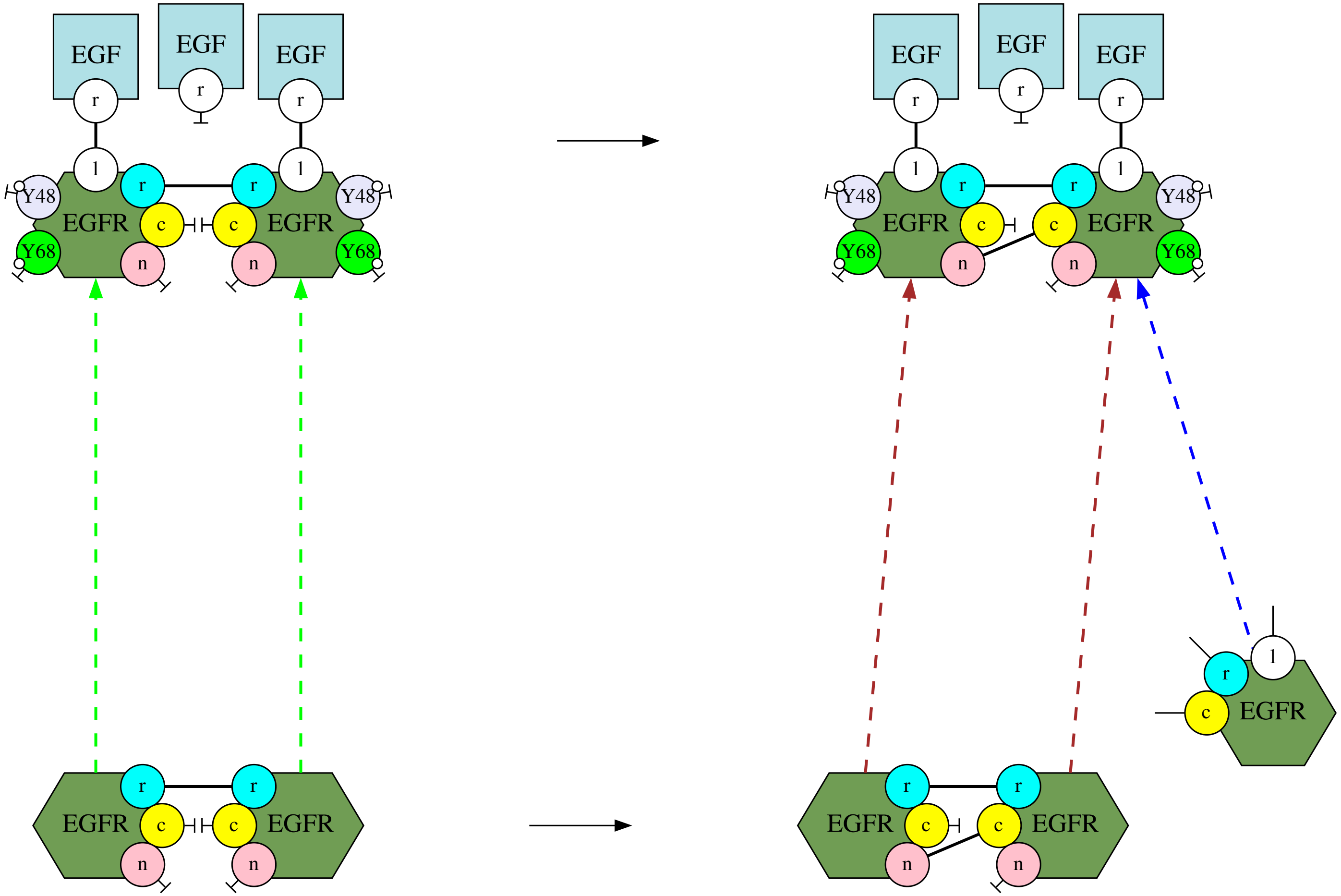


we would like to compute which ones may occur in a reachable state.
(*the initial state is given in the specification.*)

Most precise abstract transition

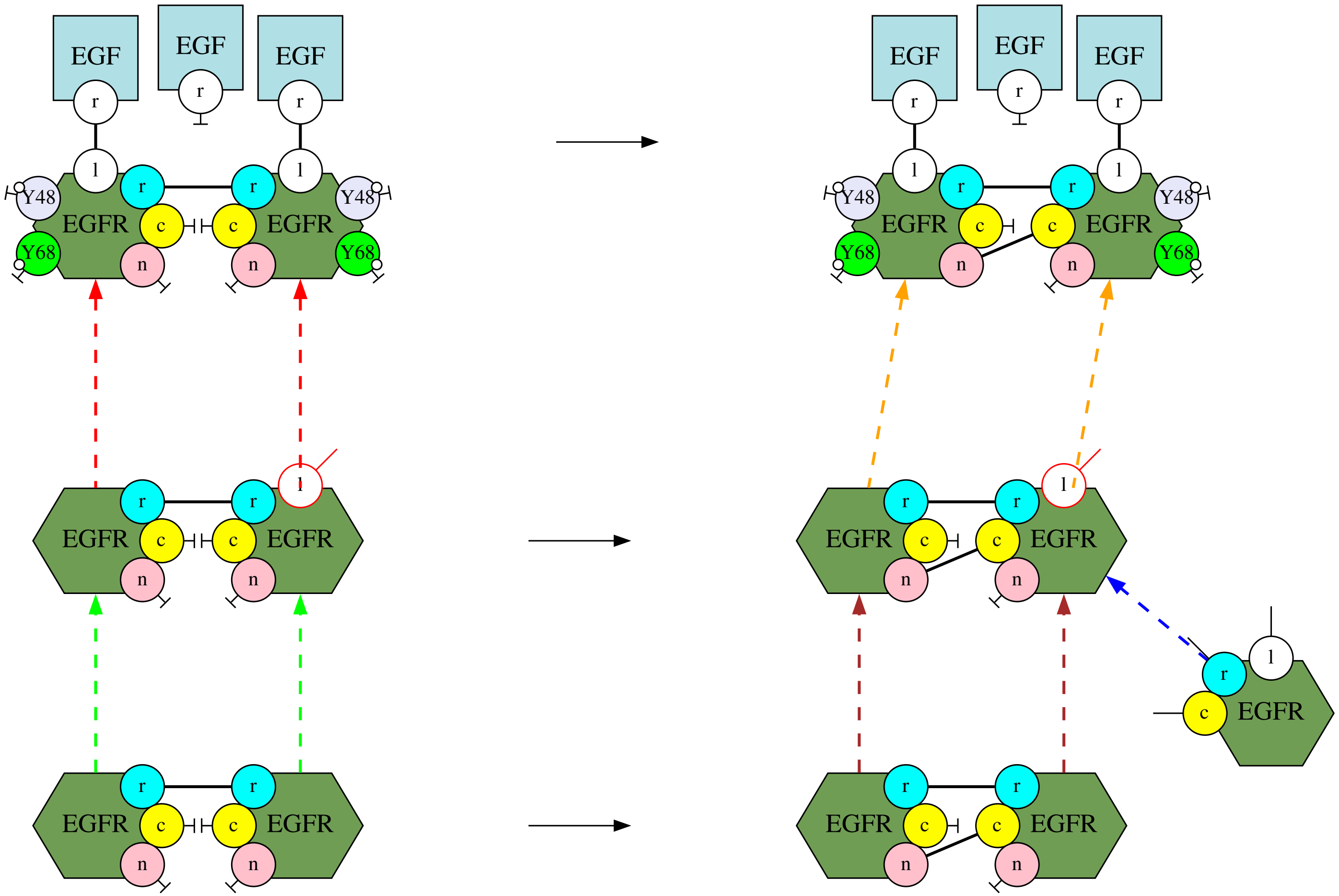


Most precise abstract transition

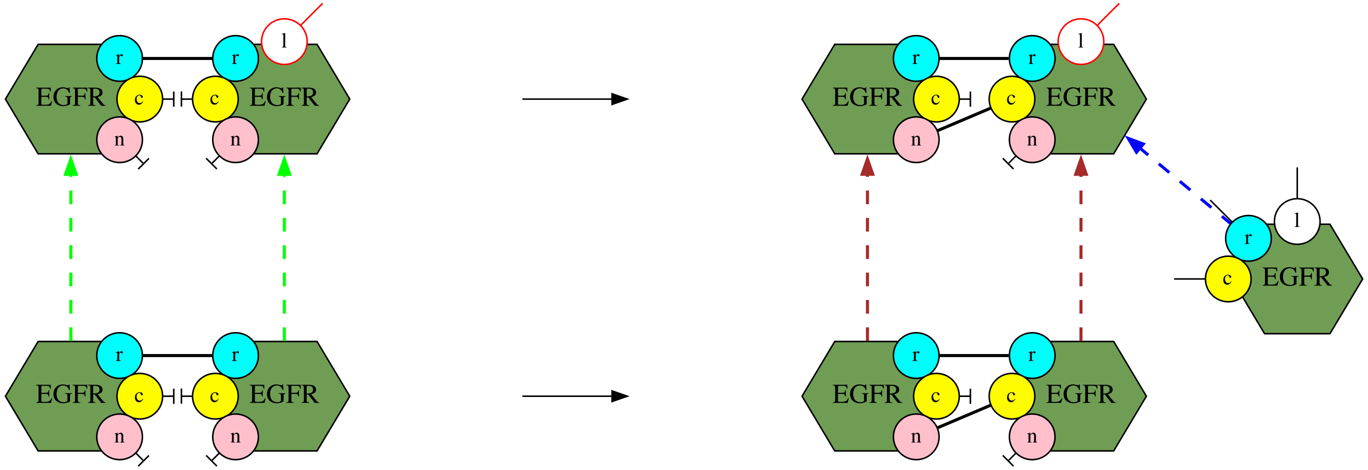


when the left hand side of the transition (top) contains no forbidden pattern.

Simplified abstract transition



Simplified abstract transition

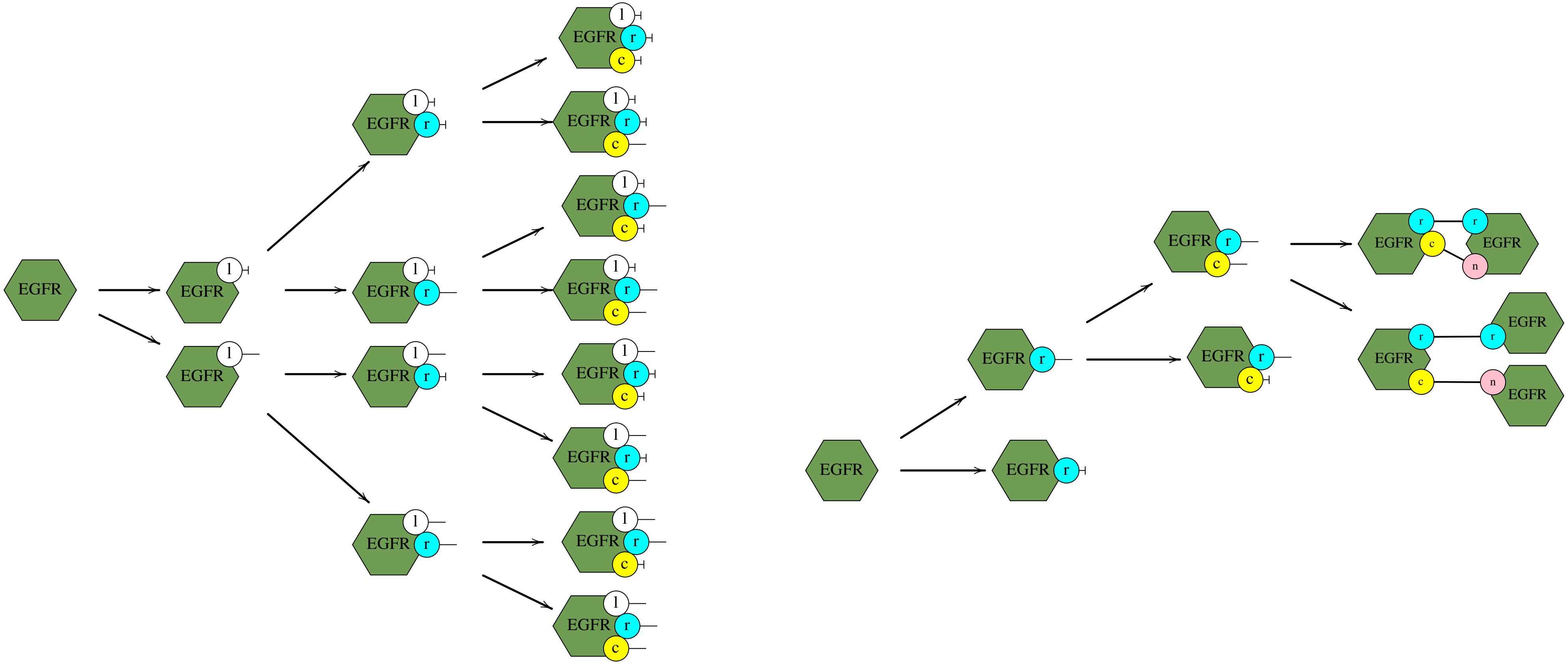


We are left to check whether the left hand side of the refined rule is reachable given the forbidden patterns.

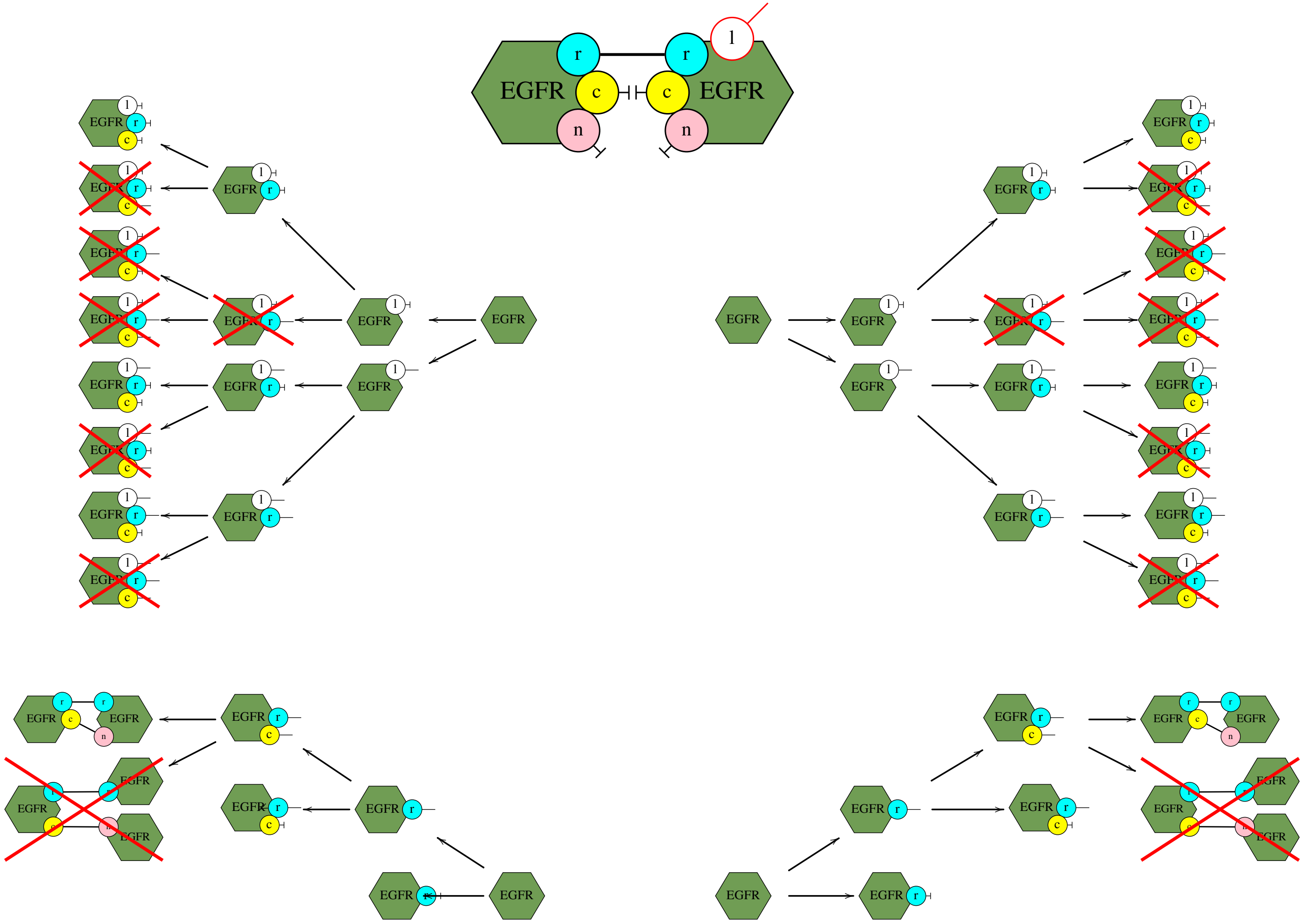
Outline

1. Context and motivations
2. Static analysis
    - Motivations
    - Case studies
    - Reachable patterns analysis
    - Specialization to orthogonal sets
    - Conclusion and perspectives
3. Flow-based model reduction (differential)
4. Flow-based model reduction (stochastic)
5. Symmetries
6. Conclusion and perspectives

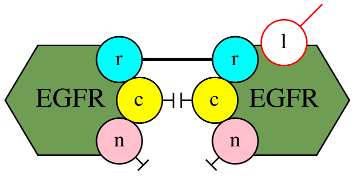
Orthogonal sets of patterns

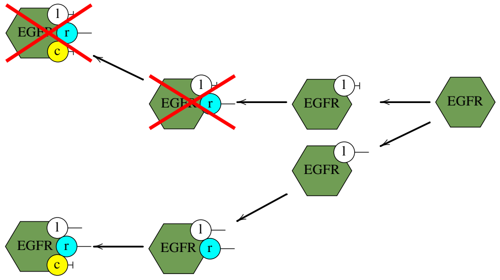


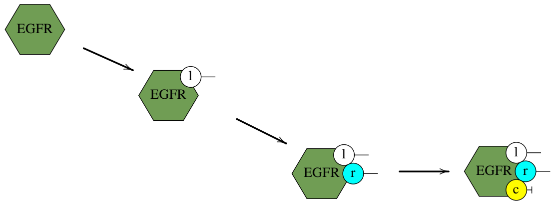
Compatibility checking procedure

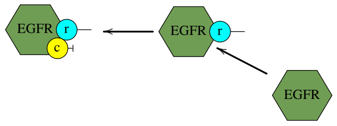


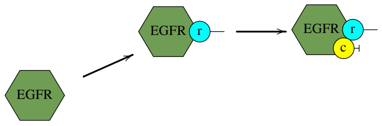
Compatibility checking procedure



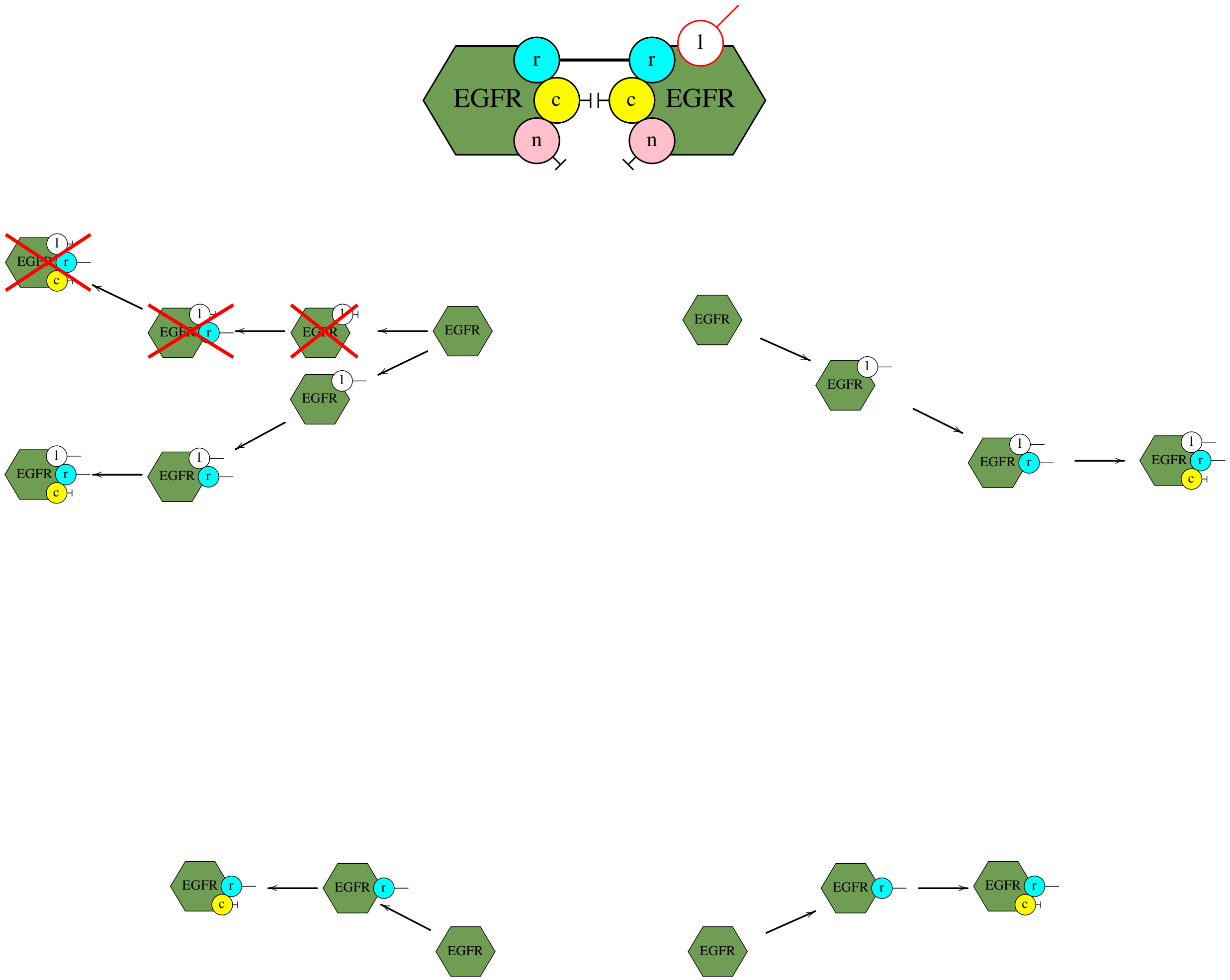




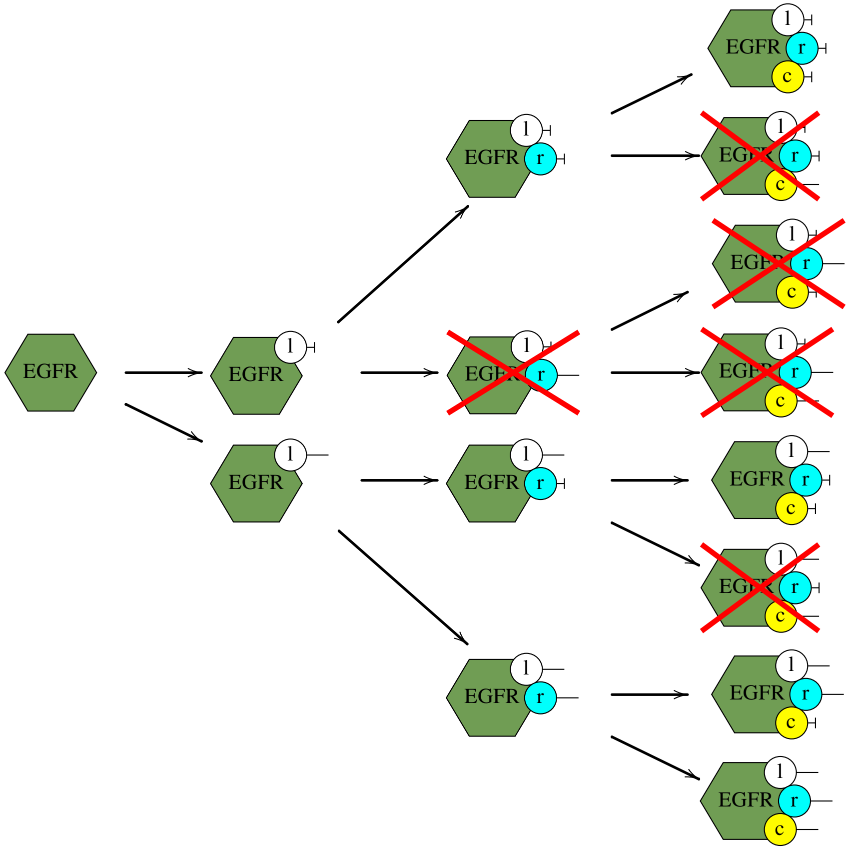


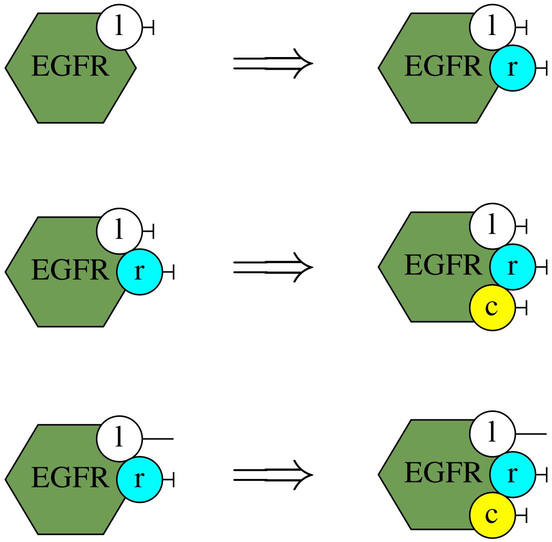


Compatibility checking procedure



Visualization





Outline

1. Context and motivations
2. Static analysis
    - Motivations
    - Case studies
    - Reachable patterns analysis
    - Specialization to orthogonal sets
    - Conclusion and perspectives
3. Flow-based model reduction (differential)
4. Flow-based model reduction (stochastic)
5. Symmetries
6. Conclusion and perspectives

Benchmarks

model	rules	inferred constraints	detected dead rules	analysis time (seconds)
repressilator	42	0	0	0.005
fceri_fyn_trimer	362	4	36	0.301
fceri_fyn_lyn_745	40	4	2	0.021
egfr	20	9	0	0.010
egfr, erk, mapk, ras	69	7	0	0.046
machine	220	13	7	0.405
ensemble	233	26	0	0.364
korkut (2017/01/17)	12896	0	874	24
korkut (2017/02/06)	5750	0	884	57
TGF (2018/04/19)	292	13	0	0.625
BigWnt (2017/03/22)	1486	14	12	8.74

on a MacBook Pro - Intel Core i7-6567U (3.3 GHz)

Conclusion

- Mutual induction between several orthogonal sets;
- Automatic parameterization (by one-pass inspection of the rules);
- Accurate analysis (on the models we were given);
- Efficient analysis

  but not enough to analyse big models in a text editor.

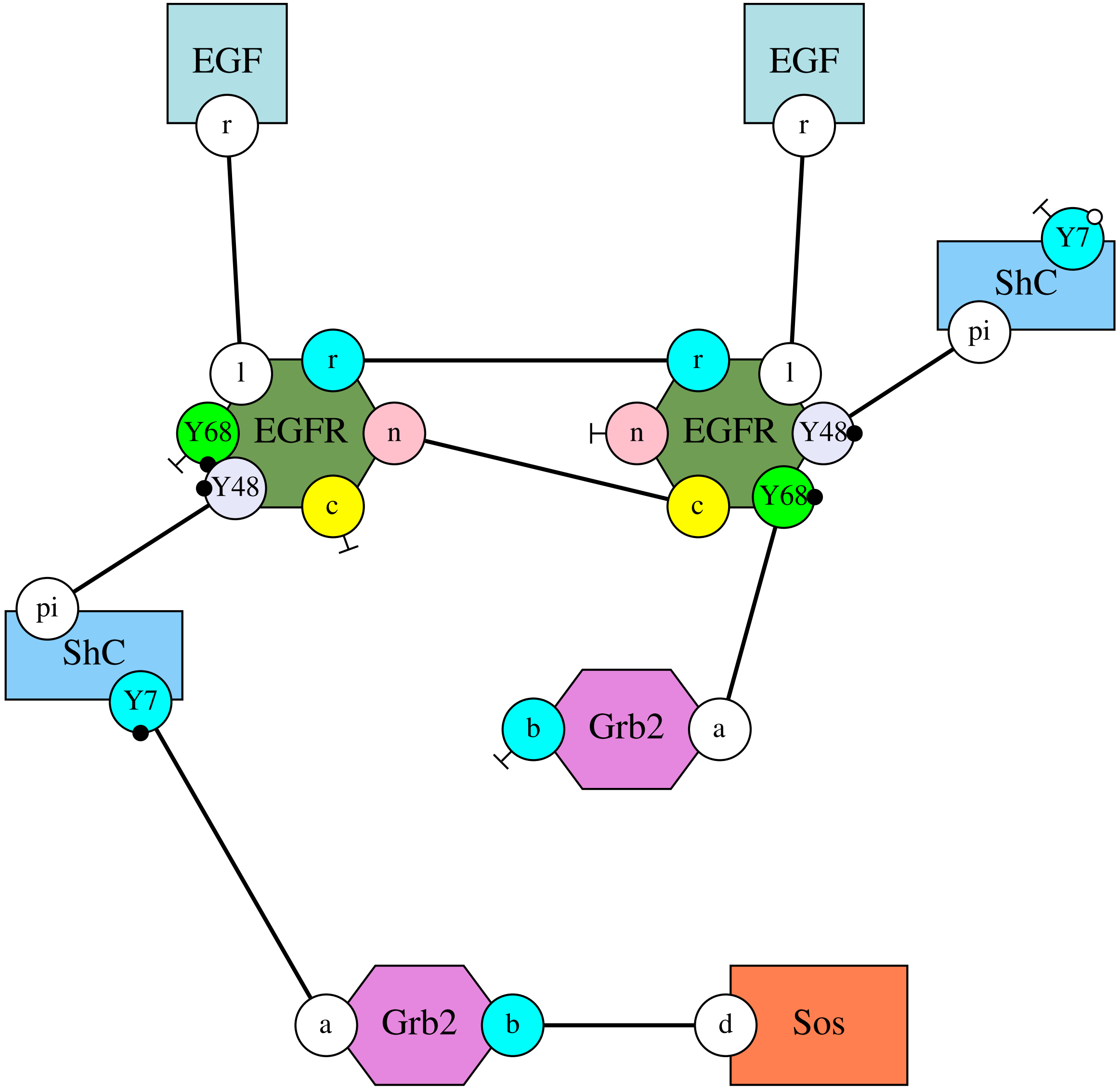
Perspectives

- Analysis of families of models;
- Incremental analysis.

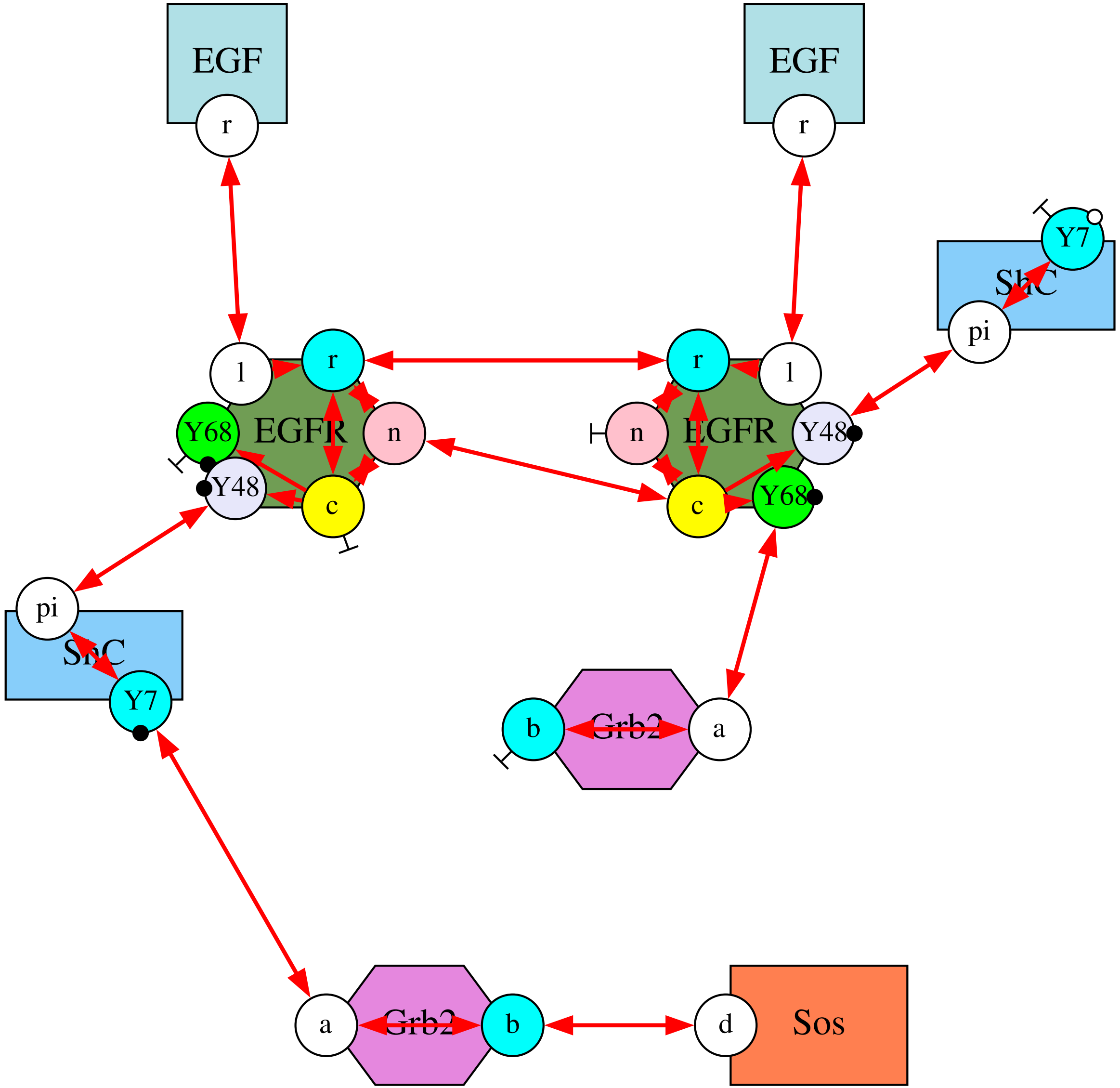
Outline

1. Context and motivations
2. Static analysis
3. Flow-based model reduction (differential)
    - Intuitions
    - Case studies
    - Abstraction of the information flow
    - Model reduction
    - Conclusion and perspectives
4. Flow-based model reduction (stochastic)
5. Symmetries
6. Conclusion and perspectives

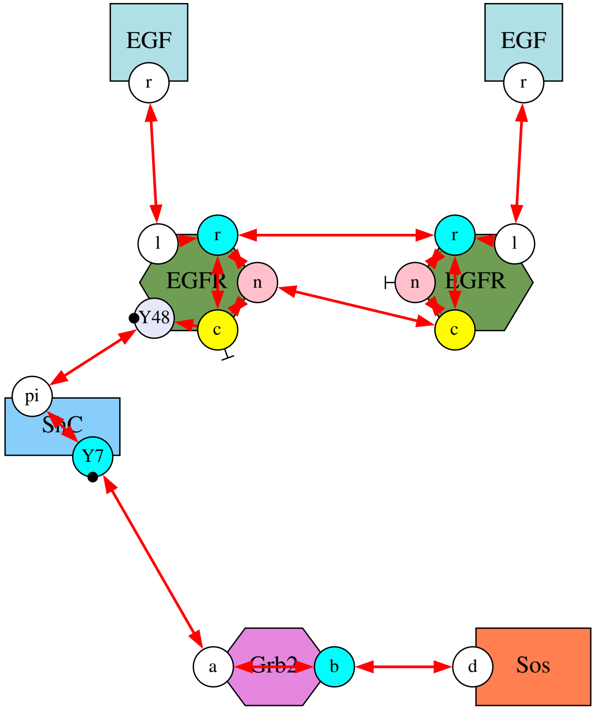
Combinatorial wall

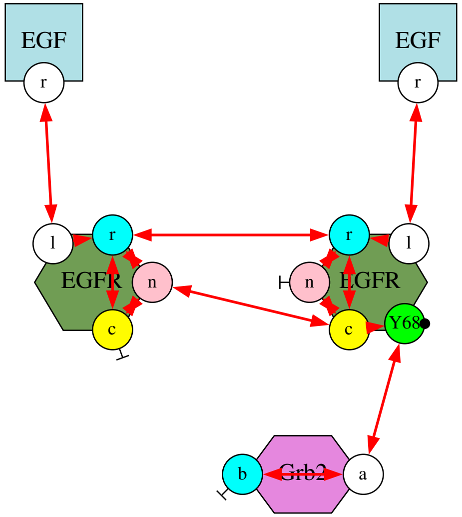


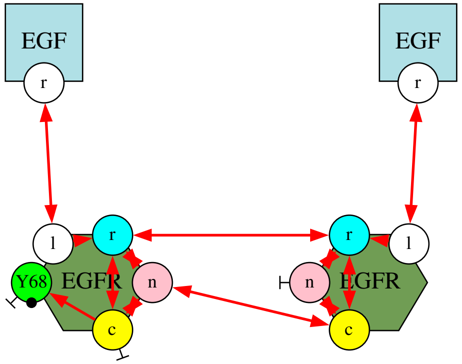
Information flow

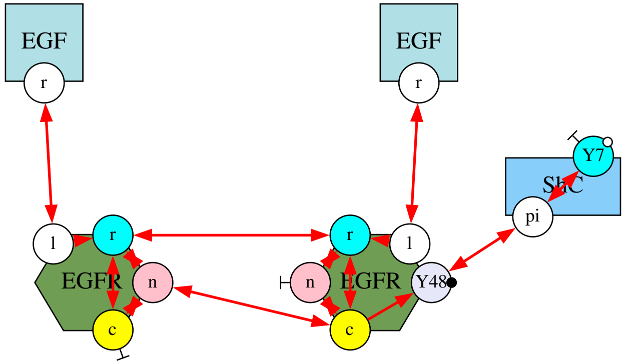


A potential breach





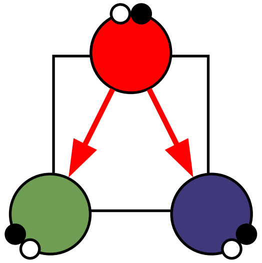


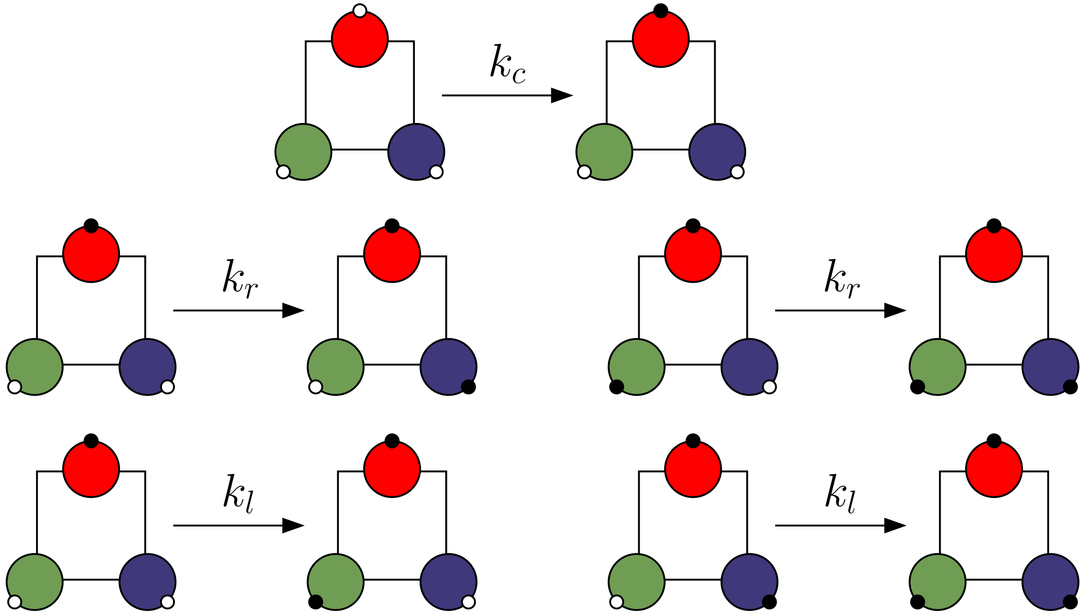


Outline

1. Context and motivations
2. Static analysis
3. Flow-based model reduction (differential)
    - Intuitions
    - Case studies
    - Abstraction of the information flow
    - Model reduction
    - Conclusion and perspectives
4. Flow-based model reduction (stochastic)
5. Symmetries
6. Conclusion and perspectives

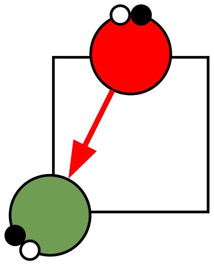
Case study





$$\left\{\begin{array}{l}
\frac{d[(u,u,u)]}{dt} = -k_c\cdot[(u,u,u)] \\
\frac{d[(u,p,u)]}{dt} = -k_l\cdot[(u,p,u)] + k_c\cdot[(u,u,u)] - k_r\cdot[(u,p,u)] \\
\frac{d[(u,p,p)]}{dt} = -k_l\cdot[(u,p,p)] + k_r\cdot[(u,p,u)] \\
\frac{d[(p,p,u)]}{dt} = k_l\cdot[(u,p,u)] - k_r\cdot[(p,p,u)] \\
\frac{d[(p,p,p)]}{dt} = k_l\cdot[(u,p,p)] + k_r\cdot[(p,p,u)]
\end{array}\right.$$

Case study

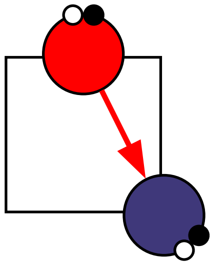


$$[(u,u,u)] = [(u,u,u)]$$

$$[(u,p,?)] \triangleq [(u,p,u)] + [(u,p,p)]$$

$$[(p,p,?)] \triangleq [(p,p,u)] + [(p,p,p)]$$

$$\left\{\begin{array}{rcl} \frac{d[(u,u,u)]}{dt} & = & -k_c\cdot[(u,u,u)] \\ \frac{d[(u,p,?)]}{dt} & = & -k_l\cdot[(u,p,?)] + k_c\cdot[(u,u,u)] \\ \frac{d[(p,p,?)]}{dt} & = & k_l\cdot[(u,p,?)] \end{array}\right.$$



$$[(u,u,u)] = [(u,u,u)]$$

$$[(?,p,u)] \triangleq [(u,p,u)] + [(p,p,u)]$$

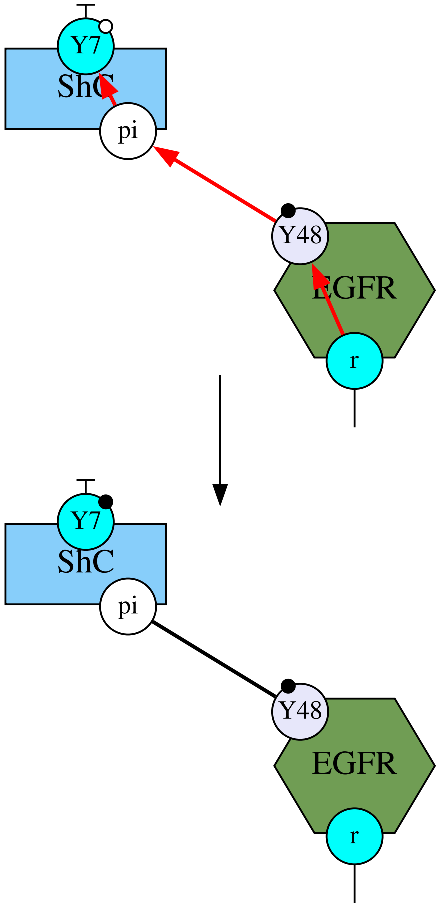
$$[(?,p,p)] \triangleq [(u,p,p)] + [(p,p,p)]$$

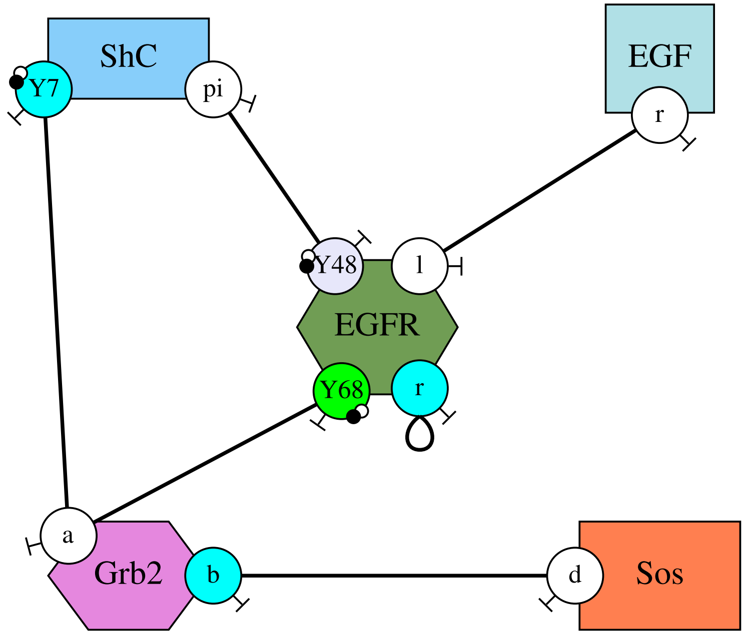
$$\left\{\begin{array}{rcl} \frac{d[(u,u,u)]}{dt} & = & -k_c\cdot[(u,u,u)] \\ \frac{d[(?,p,u)]}{dt} & = & -k_r\cdot[(?,p,u)] + k_c\cdot[(u,u,u)] \\ \frac{d[(?,p,p)]}{dt} & = & k_r\cdot[(?,p,u)] \end{array}\right.$$

Outline

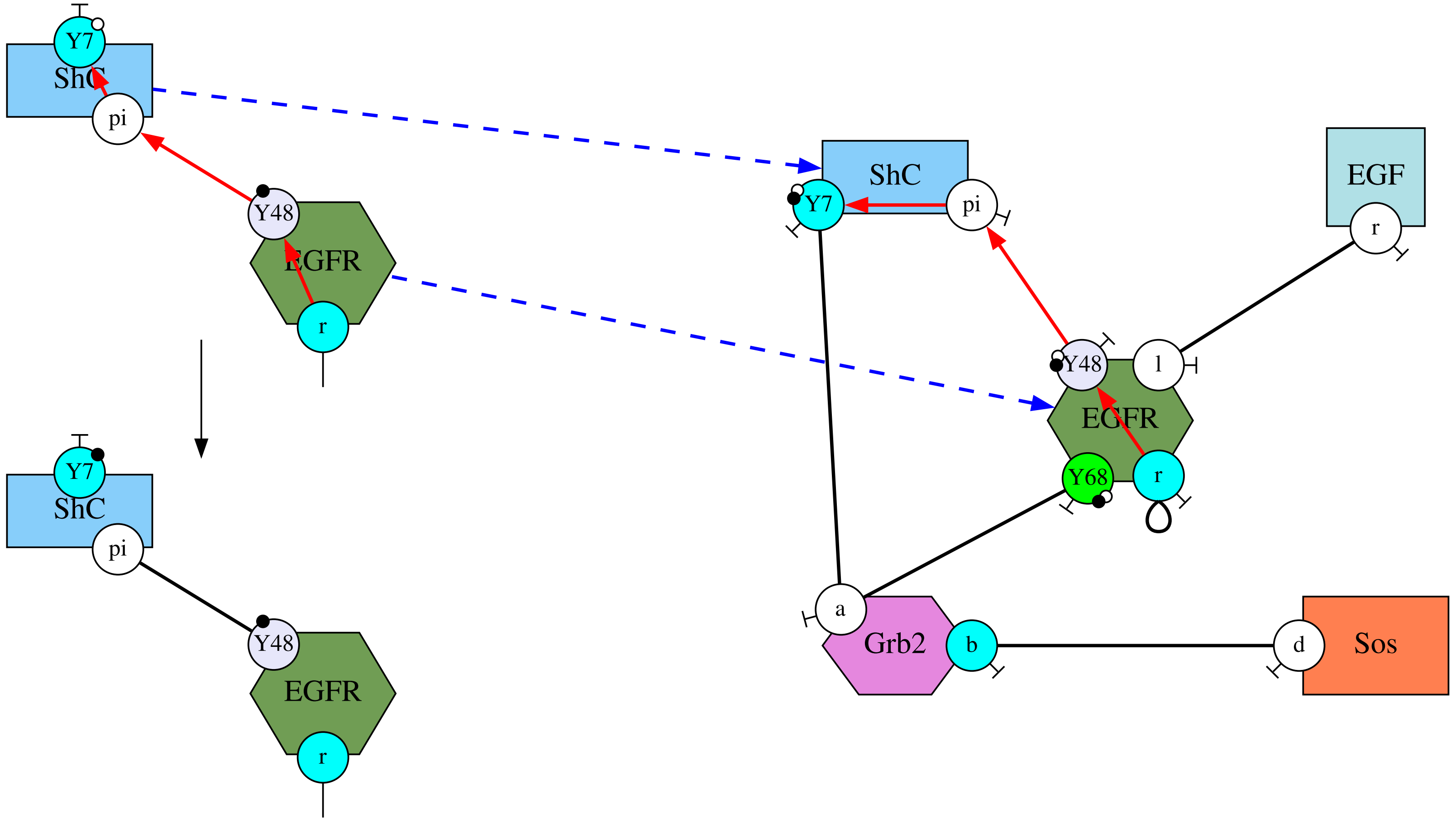
1. Context and motivations
2. Static analysis
3. Flow-based model reduction (differential)
    - Intuitions
    - Case studies
    - Abstraction of the information flow
    - Model reduction
    - Conclusion and perspectives
4. Flow-based model reduction (stochastic)
5. Symmetries
6. Conclusion and perspectives

Approximation of the flow of information

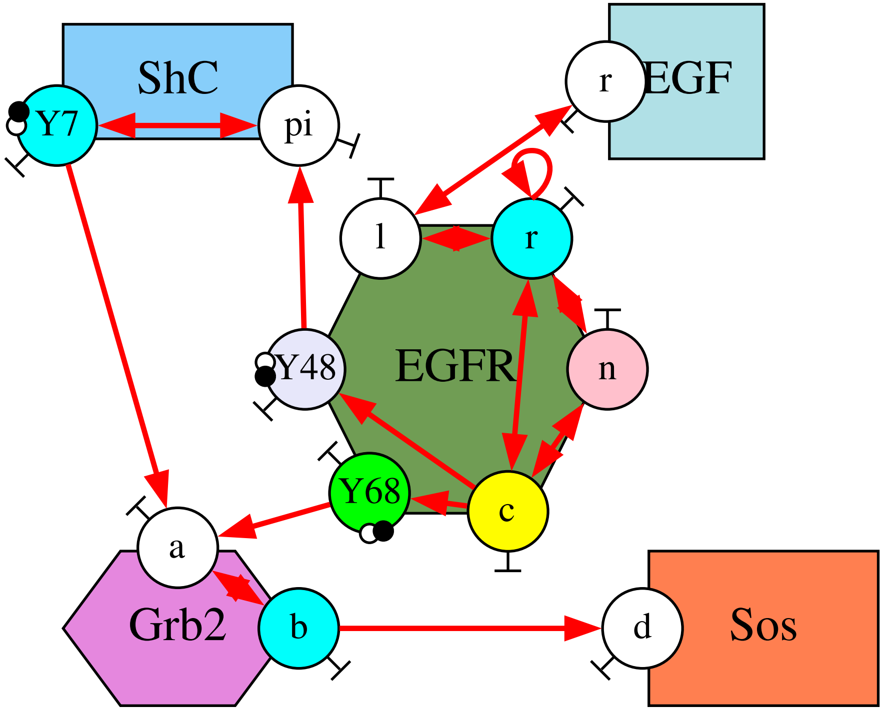




Approximation of the flow of information



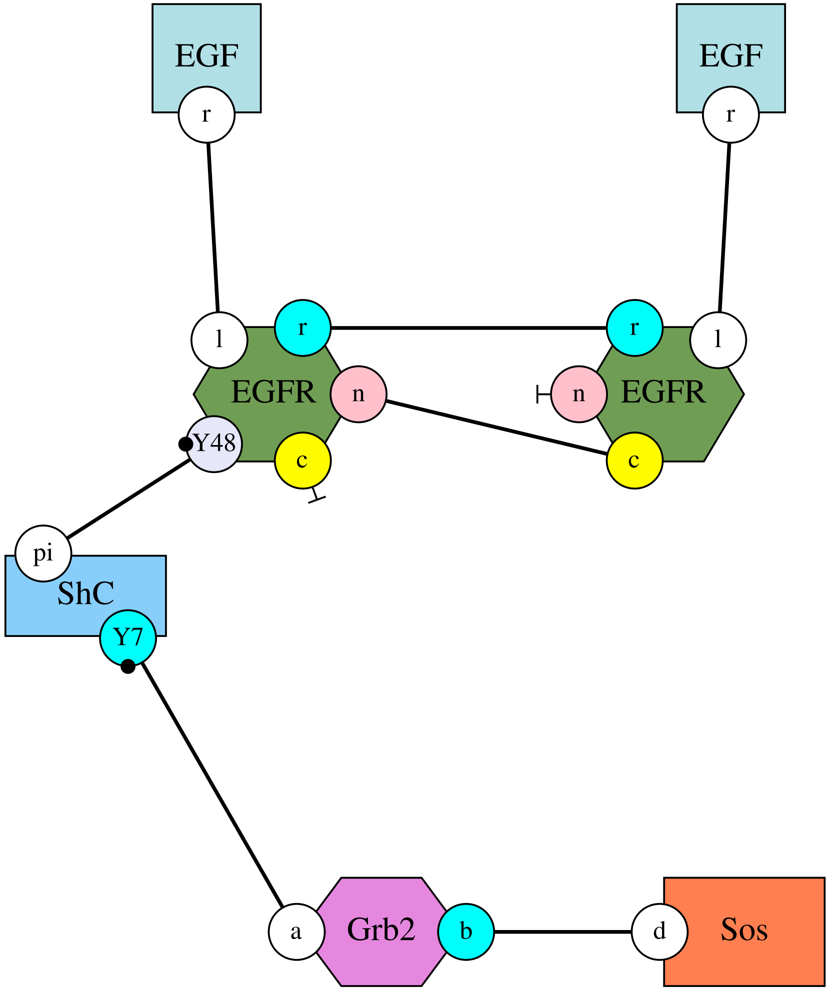
Annotated interaction map

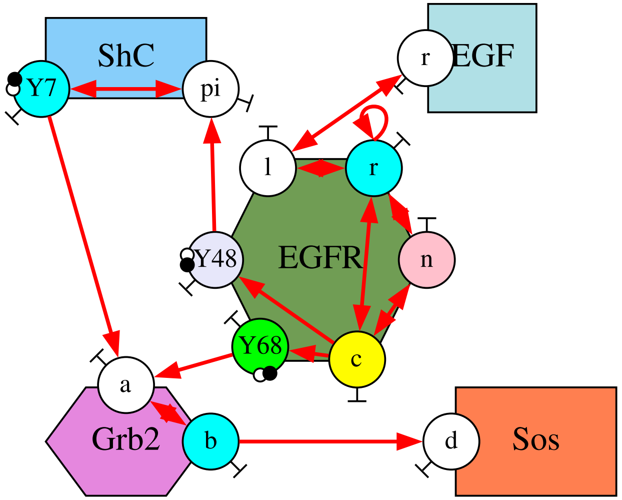


Outline

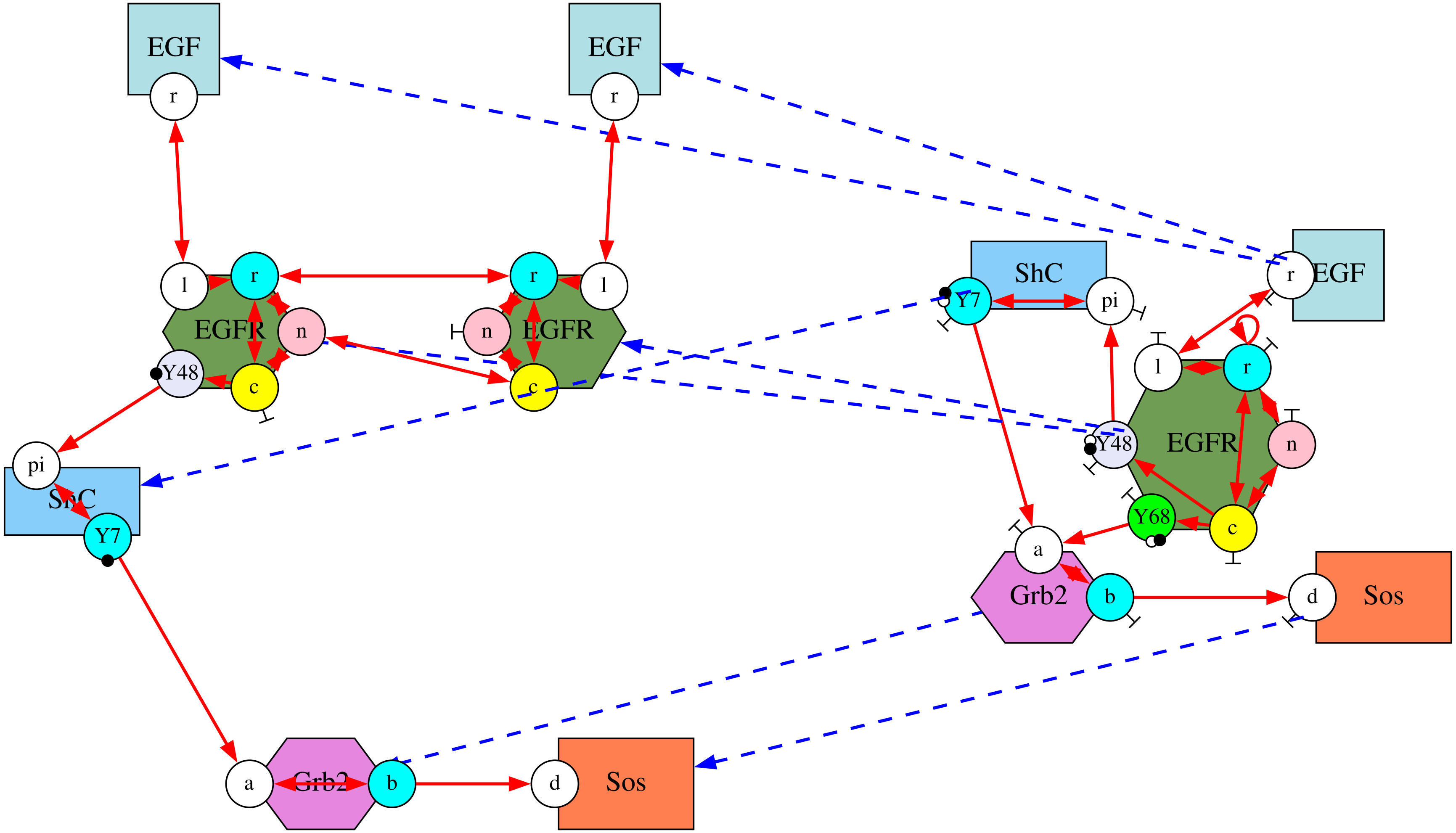
1. Context and motivations
2. Static analysis
3. Flow-based model reduction (differential)
    - Intuitions
    - Case studies
    - Abstraction of the information flow
    - Model reduction
    - Conclusion and perspectives
4. Flow-based model reduction (stochastic)
5. Symmetries
6. Conclusion and perspectives

Pattern annotation

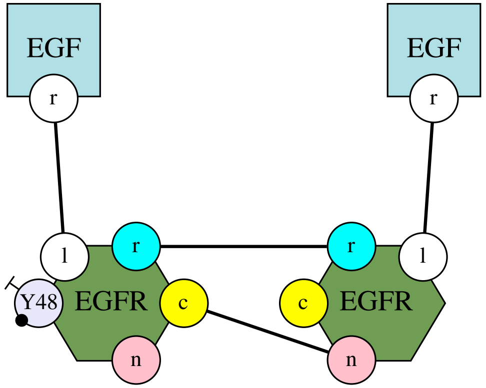


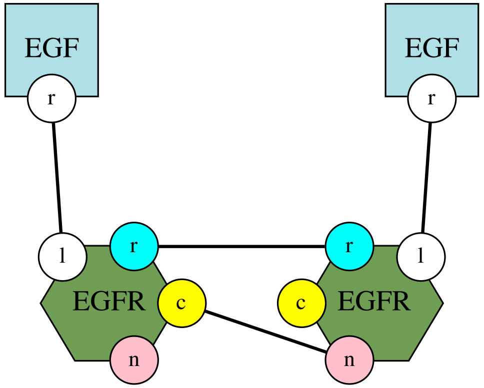


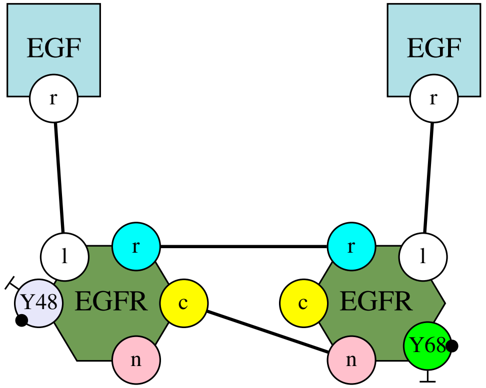
Pattern annotation

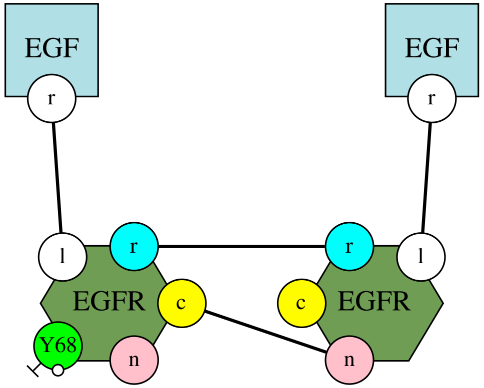


Which patterns to keep as variables?

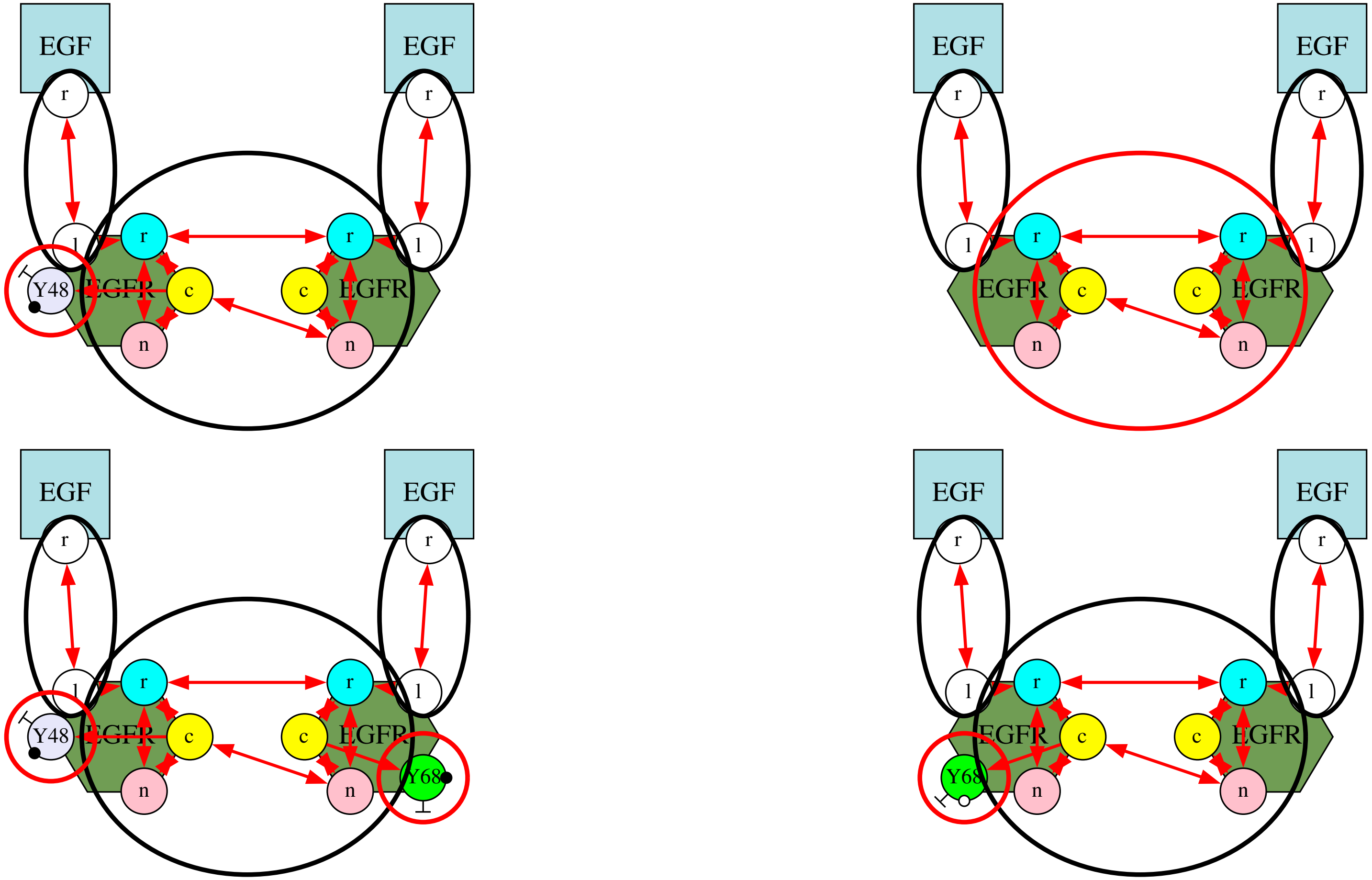






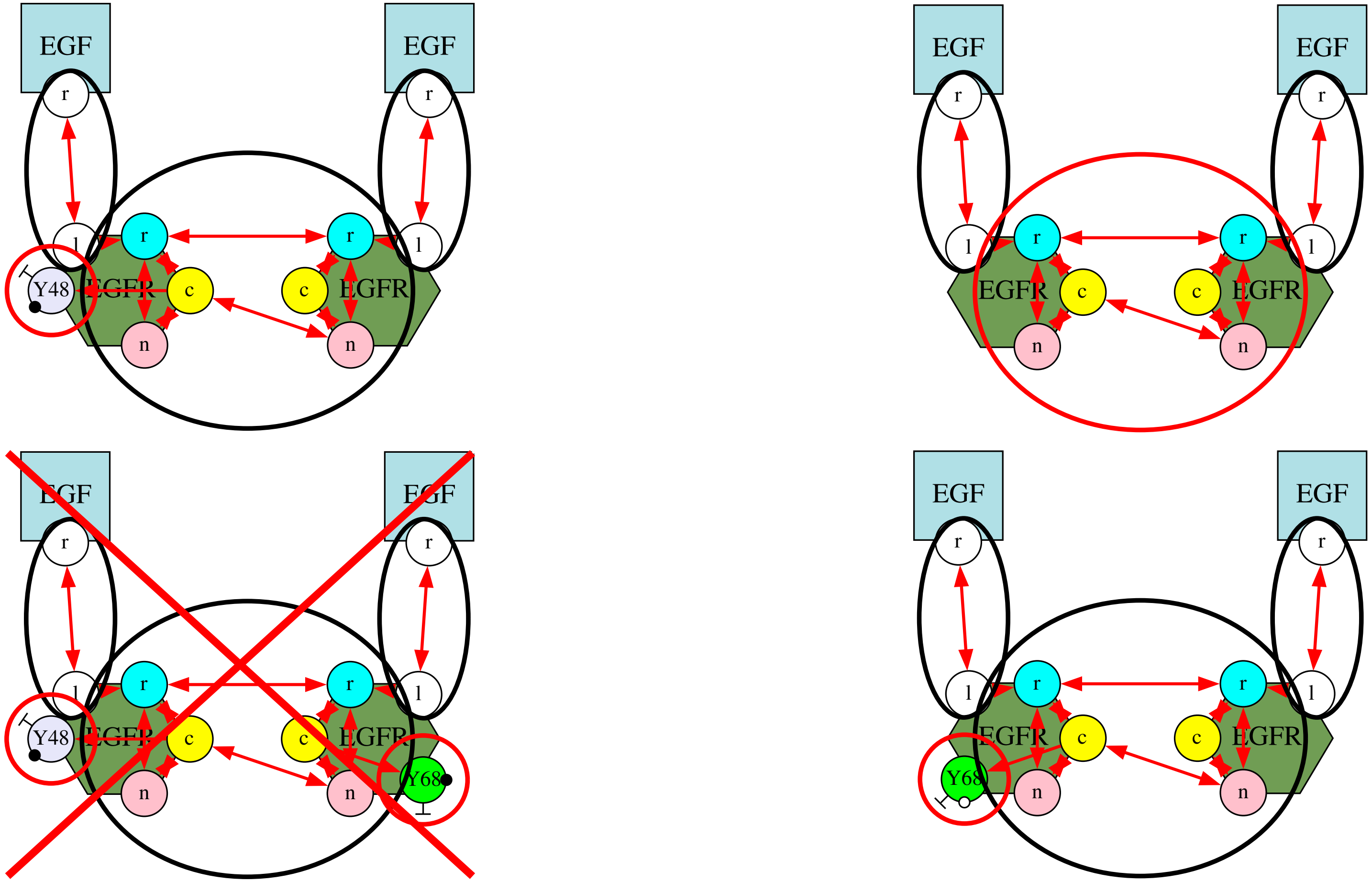


Which patterns to keep as variables?



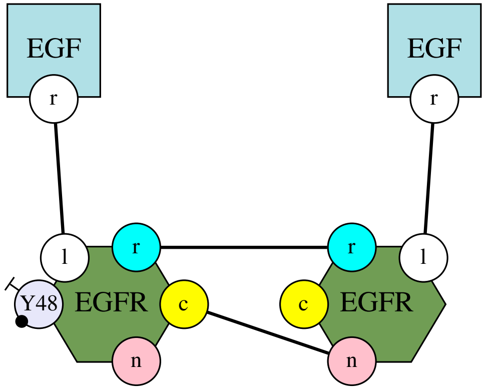
Which patterns to keep as variables?

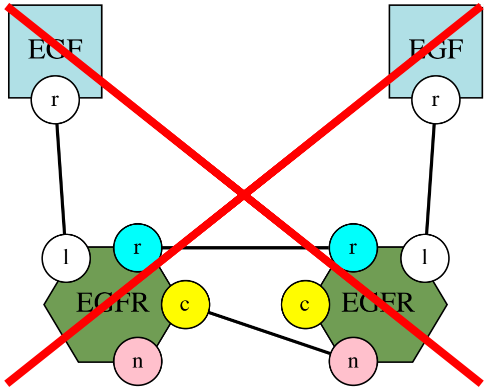
Prefragments: only one terminal strongly connected component

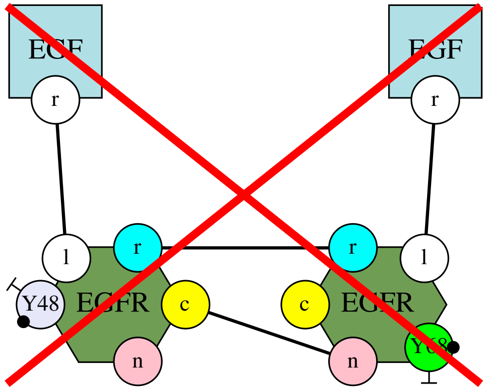


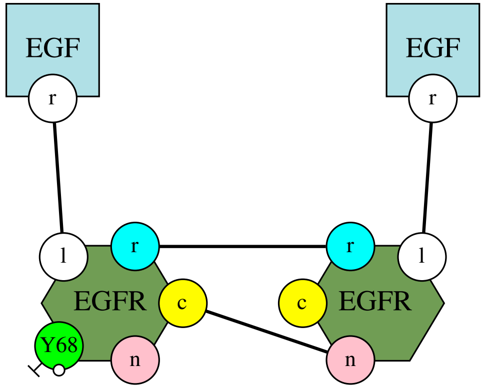
Which patterns to keep as variables?

Fragments: maximal (embedding order) prefragments

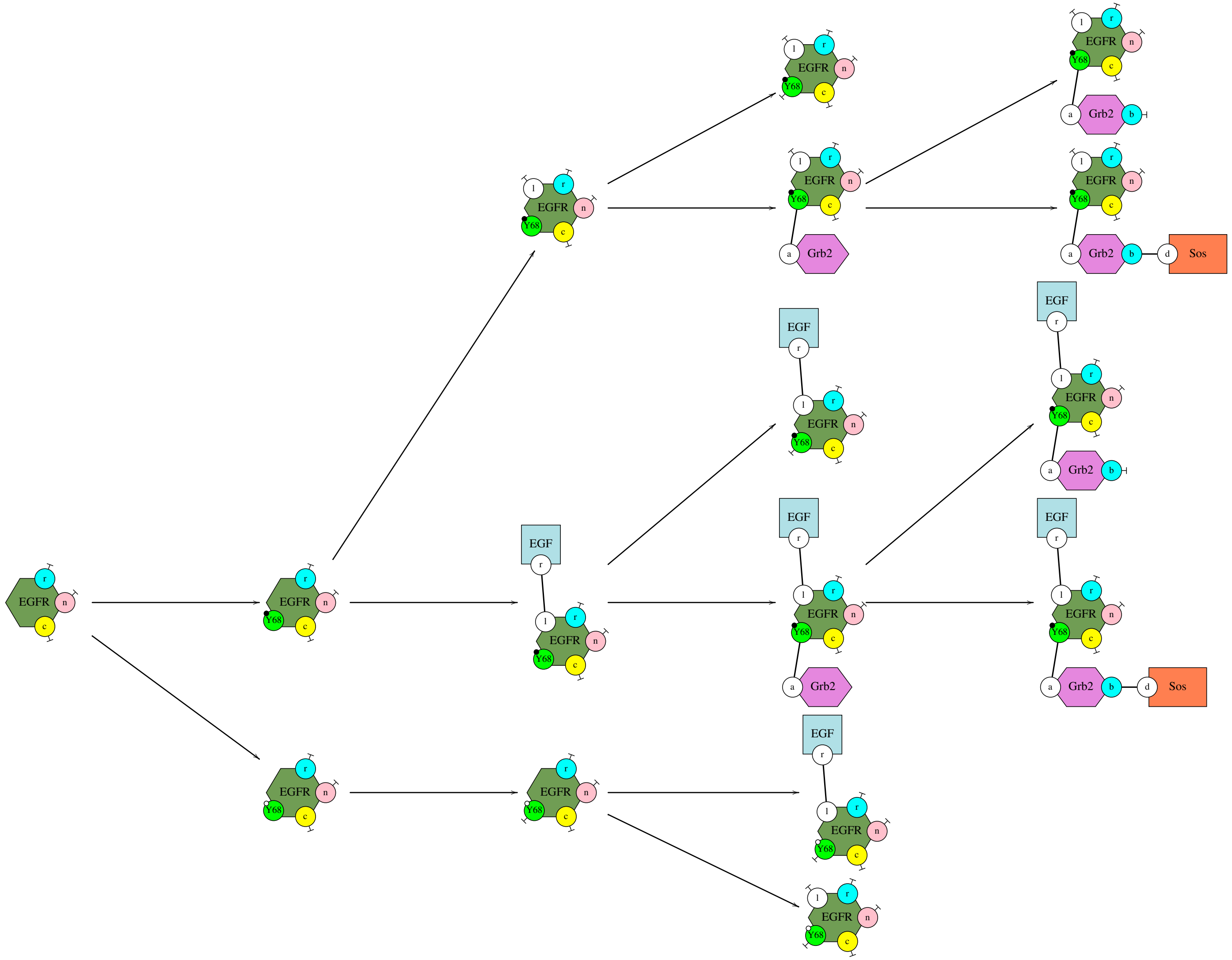




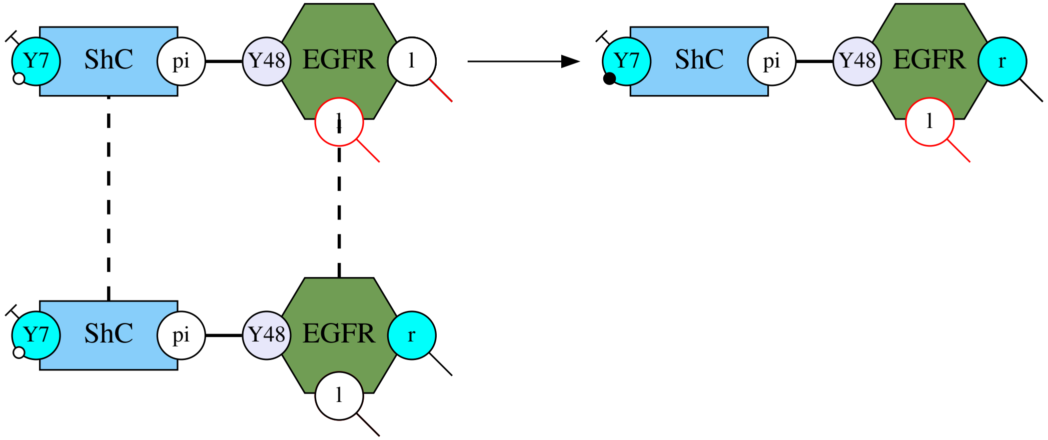




Quantity of a prefragement



Fragments consumption



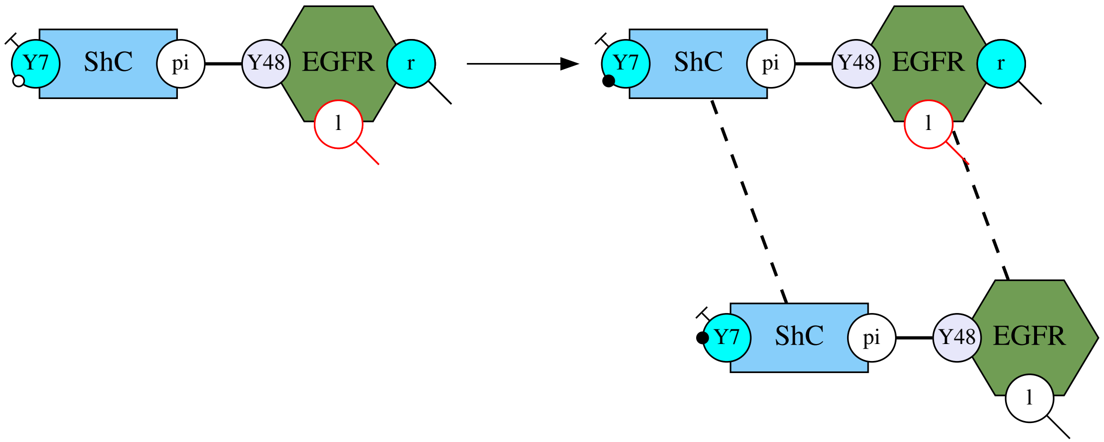
For each fragment F, for each rule:

$$\textit{r}: \ C_1, \ldots, C_n \rightarrow \textit{rhs} \quad k$$

and for each occurrence of a connected component C_j in the fragment F:

$$\frac{d[F]}{dt} \overset{-}{=} \frac{k \cdot [F] \cdot \prod_{i \neq j} [C_i]}{\mathsf{SYM}[C_1, \ldots, C_n] \cdot \mathsf{SYM}[F]}.$$

Fragments production



For each overlap between a fragment and the right hand side of a rule:

$$r \ : C_1, \ldots, C_m \rightarrow \textit{rigth hand side} \quad k,$$

we have the following contribution:

$$\frac{d[F]}{dt} \overset{+}{=} \frac{k \cdot \prod_i [C'_i]}{\mathsf{SYM}[C_1, \ldots, C_m] \cdot \mathsf{SYM}[F]}.$$

where $C'_1, \ldots, C'_n$ is the left hand side of the refined rule.

Outline

1. Context and motivations
2. Static analysis
3. Flow-based model reduction (differential)
    - Intuitions
    - Case studies
    - Abstraction of the information flow
    - Model reduction
    - Conclusion and perspectives
4. Flow-based model reduction (stochastic)
5. Symmetries
6. Conclusion and perspectives

Benchmarks

model	number of variables	number of fragments	generation time
egfr (simplified)	356	38	0.3 s.
egfr	1232	238	0.2 s.
egfr, erk, mapk, ras	$\sim 2.10^{19}$	$\sim 2.10^{5}$	180 s.

on a MacBook Pro Intel Core i7-6567U (3.3 GHz)).

Conclusion

We propose an exact model reduction framework :

- relying on the structure of biological complexes;
- formally proven with respect to the concrete semantics as opposed to [Borisov *et al.* 2005, Conzelman (2006,2007)];
- no need to enumerate biological complexes, nor the reactions as opposed to [Tribastone *et al.* (2015-2023)];
- able to break the combinatorial complexity on large examples;
- context-sensitivity of the approximation of information flow can be tuned. [Ferdinanda Camporesi PhD (2017)].

Perspectives

- How to tune context-sensitivity automatically?
- Approximate reduction.

Outline

1. Context and motivations
2. Static analysis
3. Flow-based model reduction (differential)
4. Flow-based model reduction (stochastic)
    - Master equation
    - Case studies
    - Abstraction of the information flow
    - Bisimulations
    - Conclusion and perspectives
5. Symmetries
6. Conclusion and perspectives

Master equation

$P_t(q^\star)$ denotes the probability that the system is in state $q^\star$ at time t.

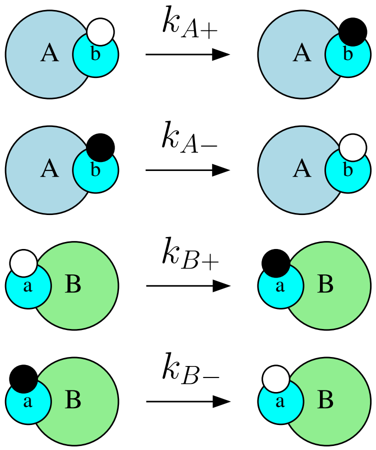
It is defined by the following equation:

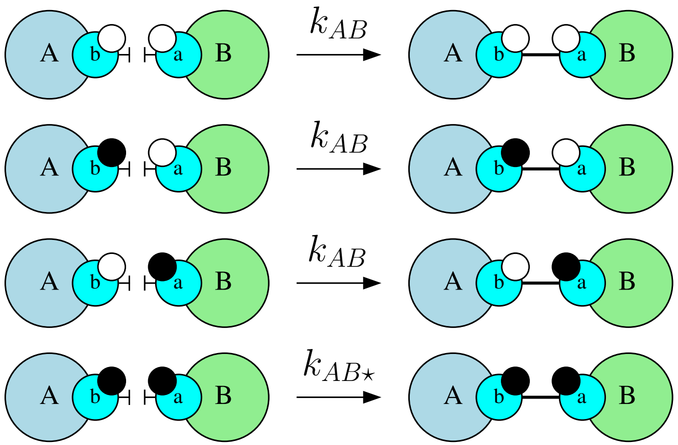
$$\frac{\mathsf{d}\,P_t(q^\star)}{\mathsf{d}t} = \left(\sum_{q \xrightarrow{\lambda} q^\star} \lambda \cdot P_t(q)\right) - \left(\sum_{q^\star \xrightarrow{\lambda} q'} \lambda \cdot P_t(q^\star)\right).$$

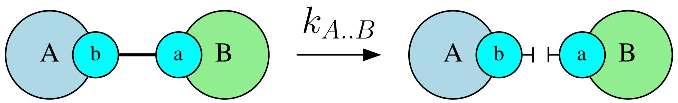
Outline

1. Context and motivations
2. Static analysis
3. Flow-based model reduction (differential)
4. Flow-based model reduction (stochastic)
    - Master equation
    - Case studies
    - Abstraction of the information flow
    - Bisimulations
    - Conclusion and perspectives
5. Symmetries
6. Conclusion and perspectives

First case study

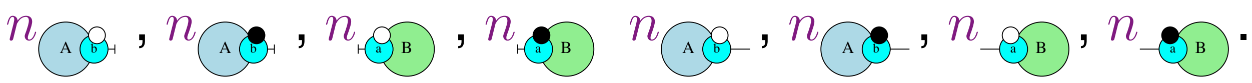






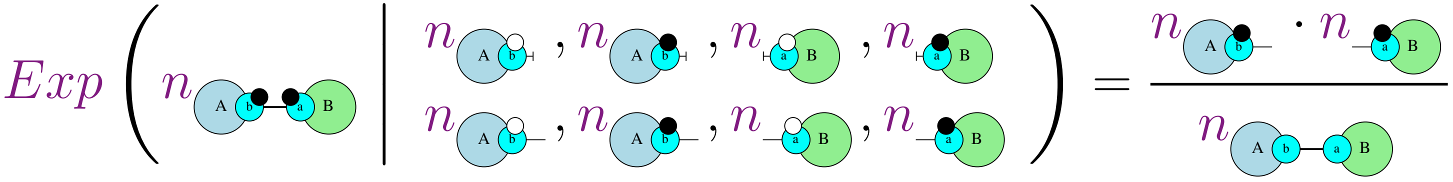
First case study

We want to abstract a state by:

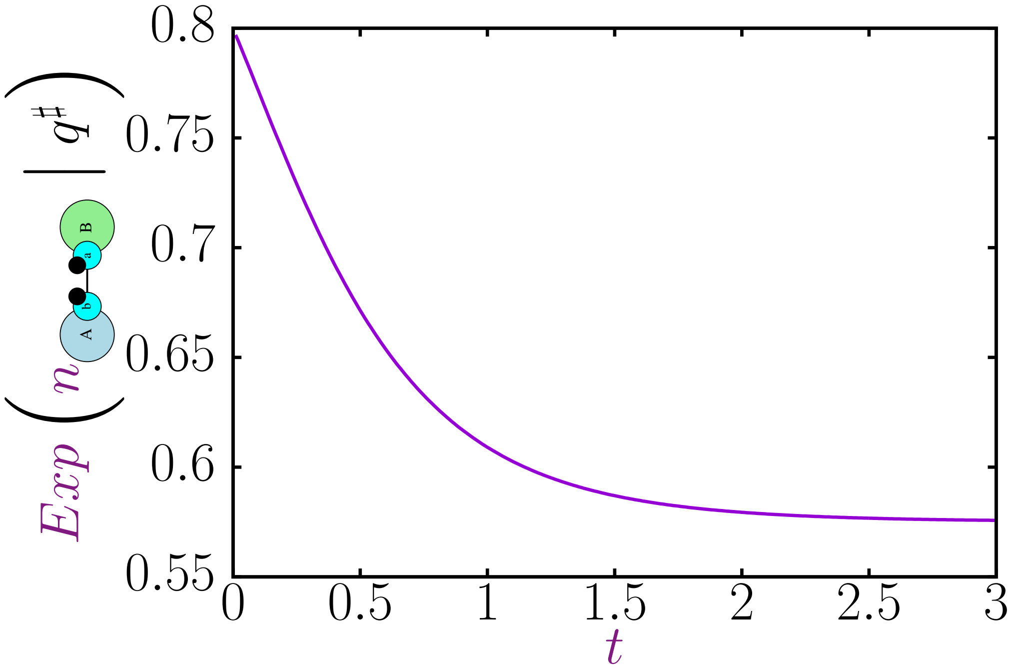


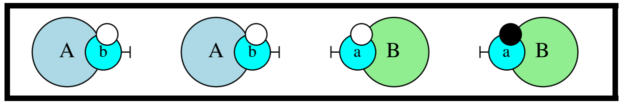
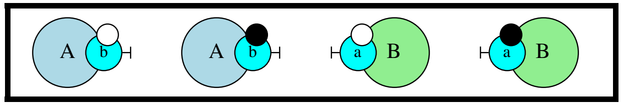
To simulate dissociation, we need to know the correlation between the states of both proteins in dimer occurrences.

When $k_{AB} = k_{AB\star}$, we have (for well chosen initial distributions):

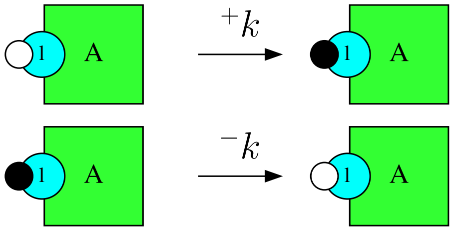


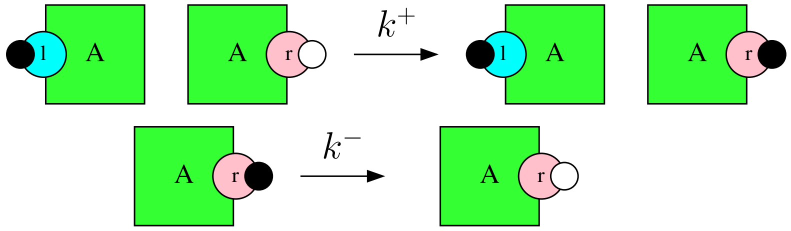
First case study: correlation



- starting in the state: ;
- with the rates $k_{A+} = k_{A-} = k_{B+} = k_{B-} = k_{AB} = k_{A..B} = 1$ and $k_{AB\star} = 10$;
- and with the abstract state $q^\sharp$ equal to: .

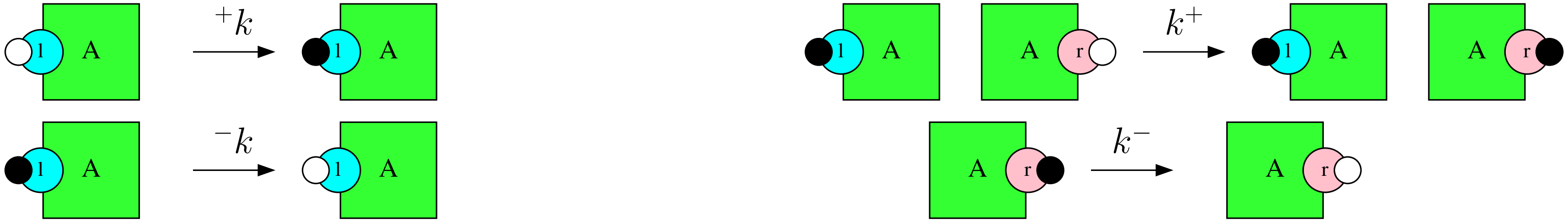
Second case study



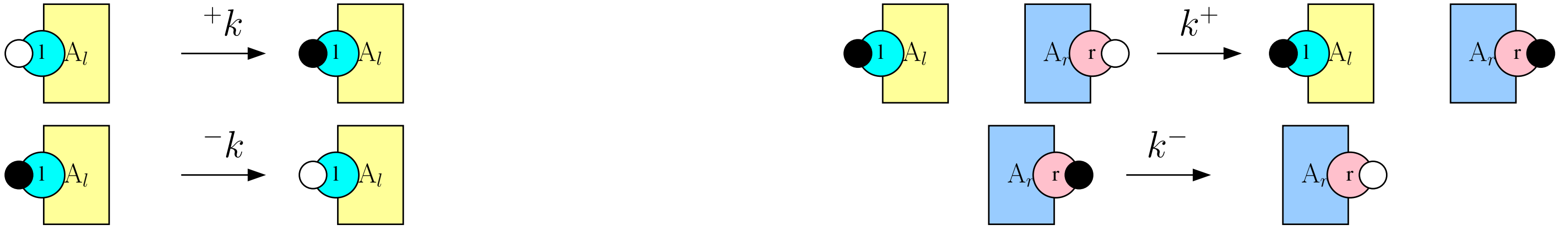


Second case study

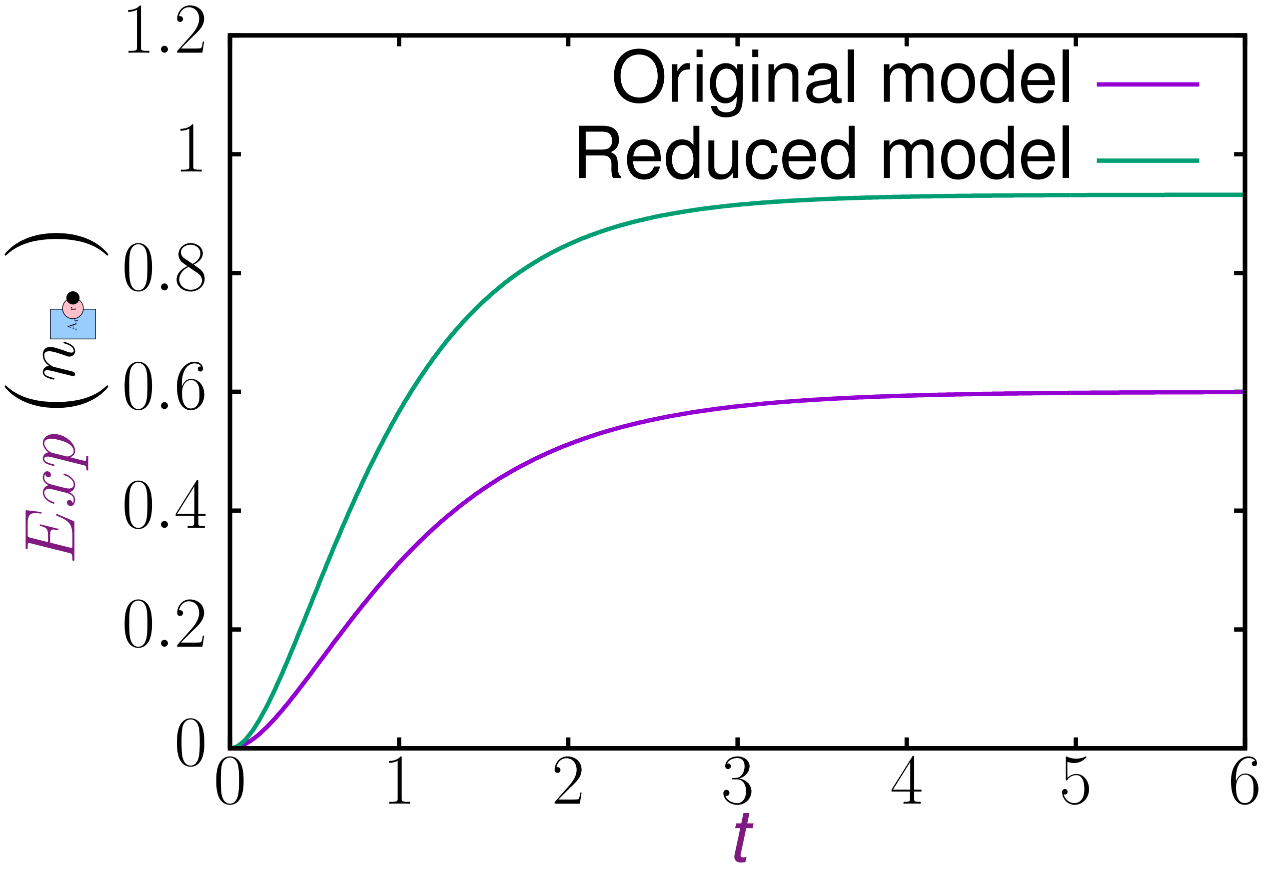
Initial model:



Reduced model:



Second case study: distant control



with two occurrences of the protein, all rates equal to 1.

Outline

1. Context and motivations
2. Static analysis
3. Flow-based model reduction (differential)
4. Flow-based model reduction (stochastic)
    - Master equation
    - Case studies
    - Abstraction of the information flow
    - Bisimulations
    - Conclusion and perspectives
5. Symmetries
6. Conclusion and perspectives

Stochastic fragments

- Issues:
  - correlations cannot be discarded
    when bonds are released,
    when modifying overlapping regions between fragments,
  - occurrences of proteins may affect each others even if not in the same rule connected component;
- What is left:
  - cut each occurrence of proteins into equivalence classes ;
  - two sites occurring in a same rule shall belong to the same equivalence class.

Outline

1. Context and motivations
2. Static analysis
3. Flow-based model reduction (differential)
4. Flow-based model reduction (stochastic)
    - Master equation
    - Case studies
    - Abstraction of the information flow
    - Bisimulations
    - Conclusion and perspectives
5. Symmetries
6. Conclusion and perspectives

Properties

The reduction induces a back-and-forth bisimulation:

- Forward:

  The behaviors of two equivalent states are the same with respect to equivalence classes.

- Backward:

  If the probability of every two equivalent states is inversely proportional to their numbers of automorphisms, it remains this way all along the execution of the model.

Outline

1. Context and motivations
2. Static analysis
3. Flow-based model reduction (differential)
4. Flow-based model reduction (stochastic)
    - Master equation
    - Case studies
    - Abstraction of the information flow
    - Bisimulations
    - Conclusion and perspectives
5. Symmetries
6. Conclusion and perspectives

Benchmarks

Model	early EGF	EGF/Insulin	SFB
Number of mollecular species	356	2899	$\sim 2.10^{19}$
Number of fragments (ODEs semantics)	38	208	$\sim 2.10^{5}$
Number of fragments (CTMC semantics)	356	618	$\sim 2.10^{19}$

Conclusion

- We use information flow to infer back-and-forth bisimulations ;
- Formally sound — but not very useful in practice ;
- Better understanding of the various semantics
  and why it is difficult to reduce them exactly.

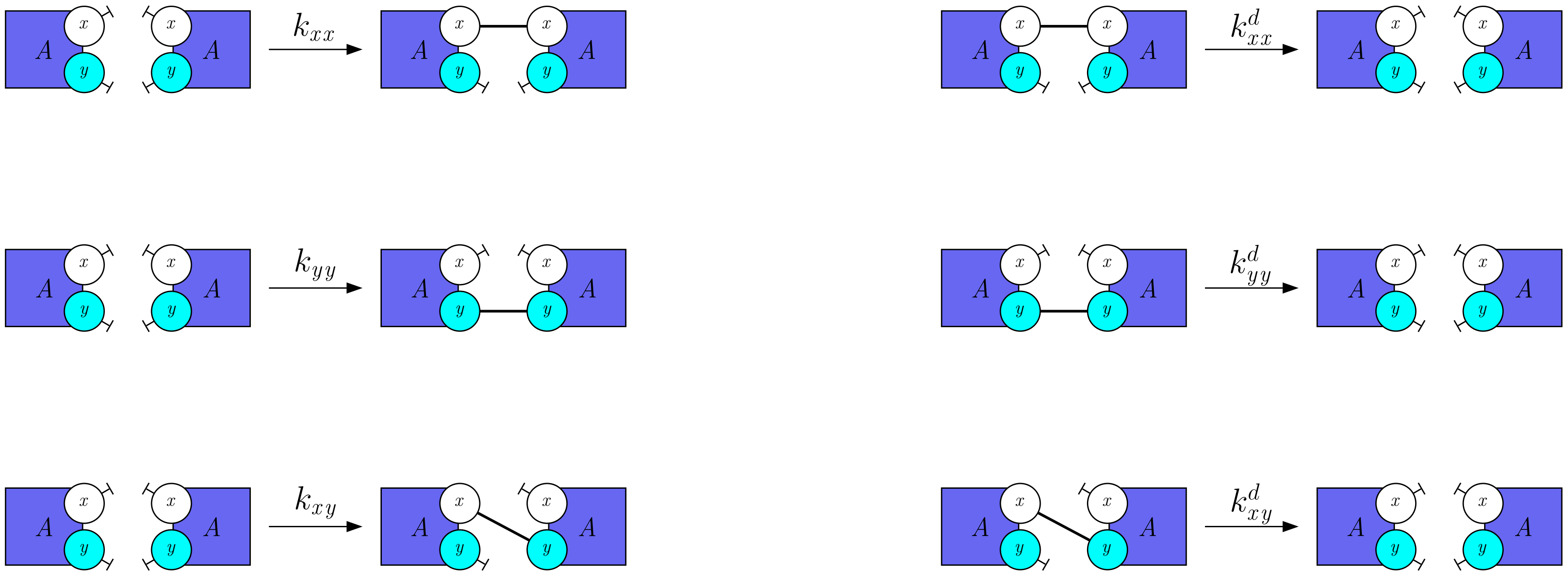
Perspectives

- Bisimulation metrics [Panagaden (1999-), Ferns (2011-2012)]
- Conservative approximate reductions.

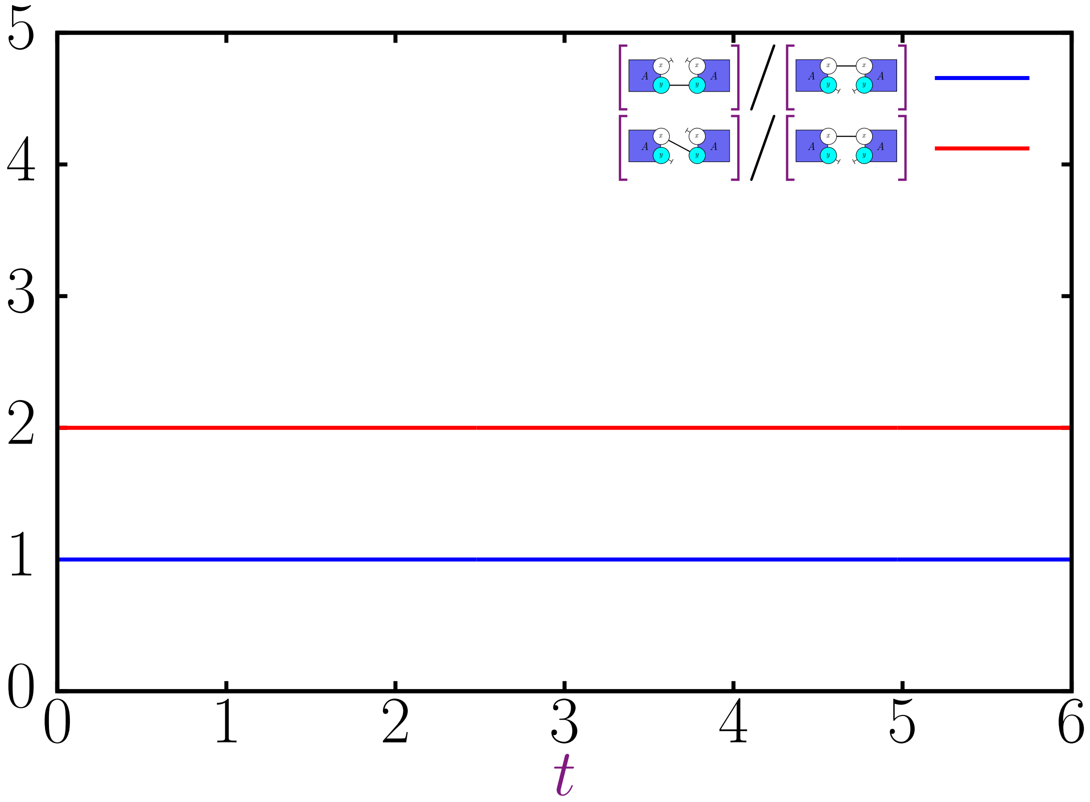
Outline

1. Context and motivations
2. Static analysis
3. Flow-based model reduction (differential)
4. Flow-based model reduction (stochastic)
5. Symmetries
    - Case studies
    - Fondamental property
    - Symmetry actions
    - Bisimulations
    - Conclusion and perspectives
6. Conclusion and perspectives

Case study

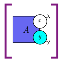
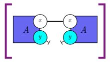
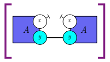
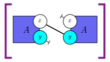


Quotients between concentrations

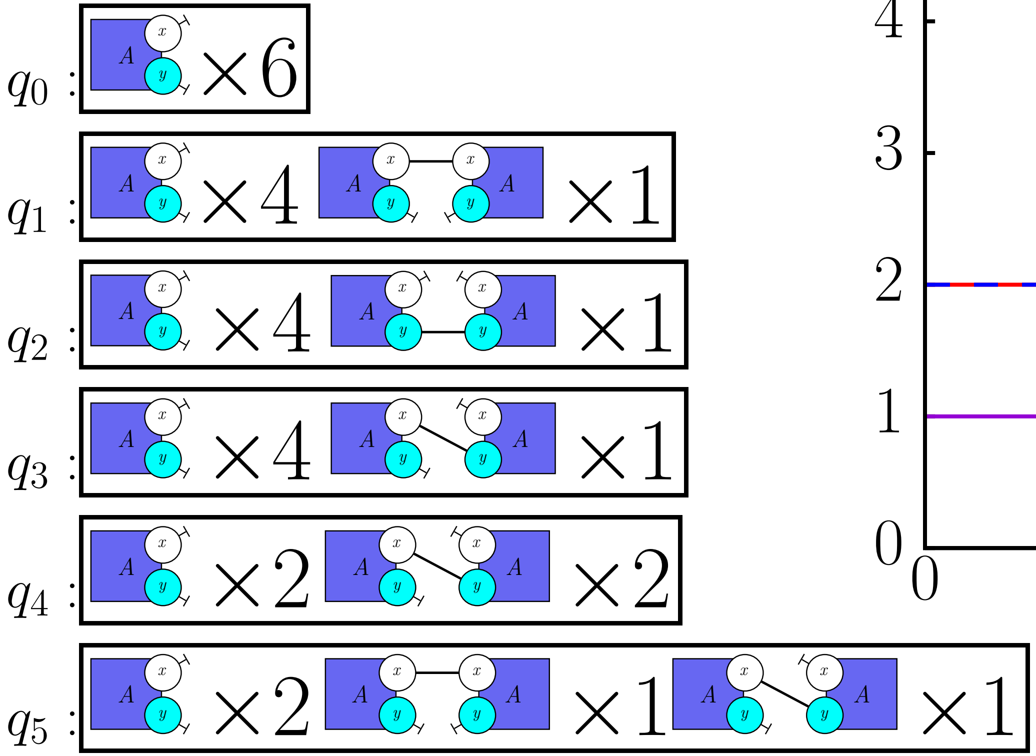


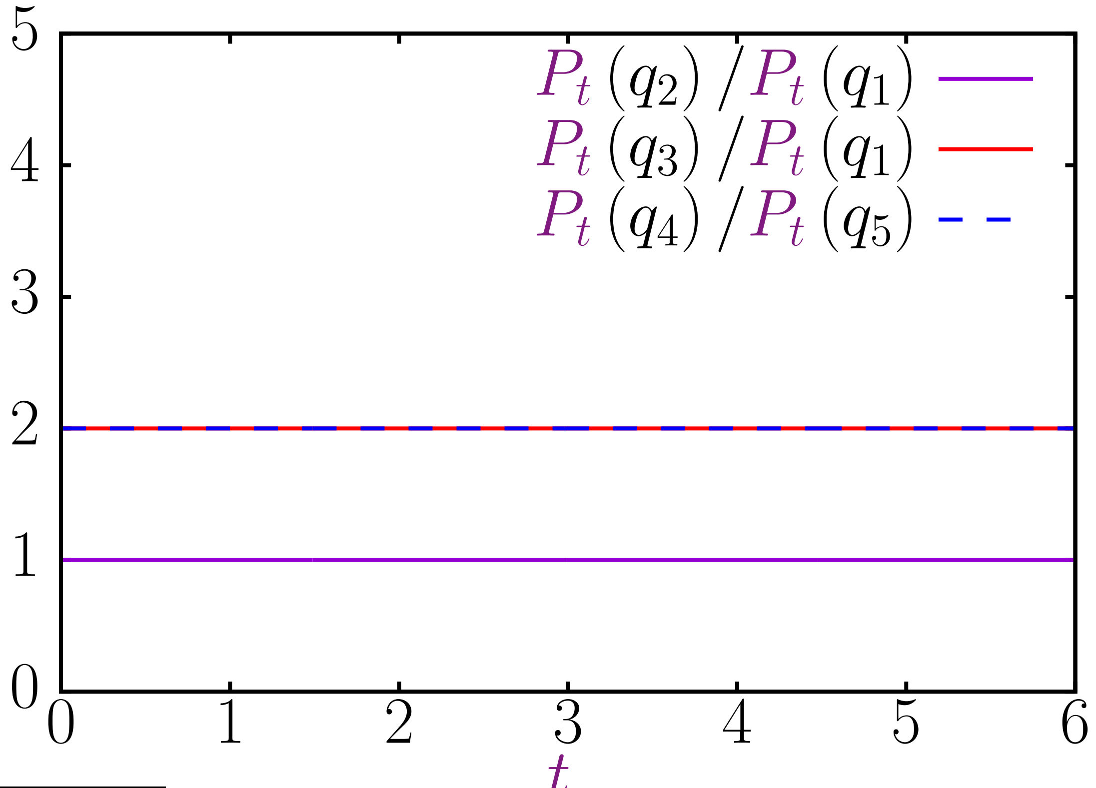
with:

$$k_{xx} = k_{yy} = 0.5,\ k^d_{xy} = 2,\ k_{xy} = k^d_{xx} = k^d_{yy} = 1$$

 $_0 = 6$,  $_0 =$  $_0 =$  $_0 = 0$

Quotients between state probabilities

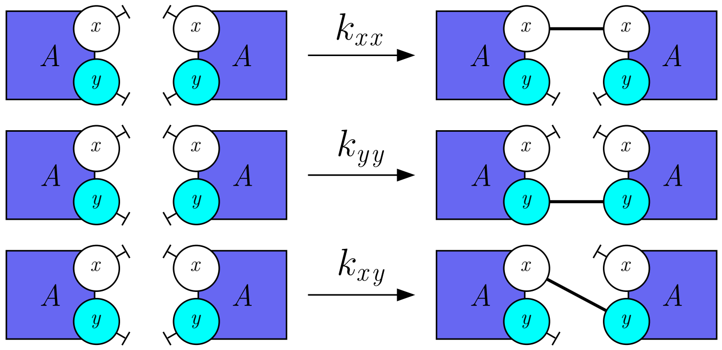


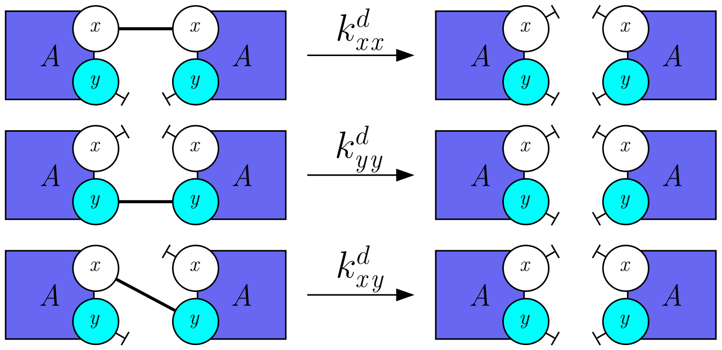


with: $\begin{cases} k_{xx} = k_{yy} = 0.5,\ \ k^d_{xy} = 2 \\ k_{xy} = k^d_{xx} = k^d_{yy} = 1 \\ P_0(q_0) = 1 \end{cases}$

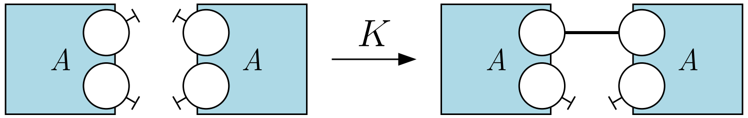
Initial model

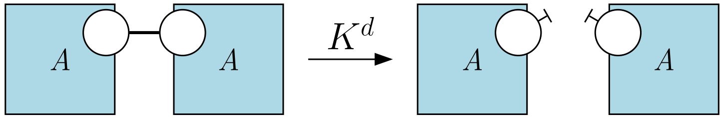
Initial model



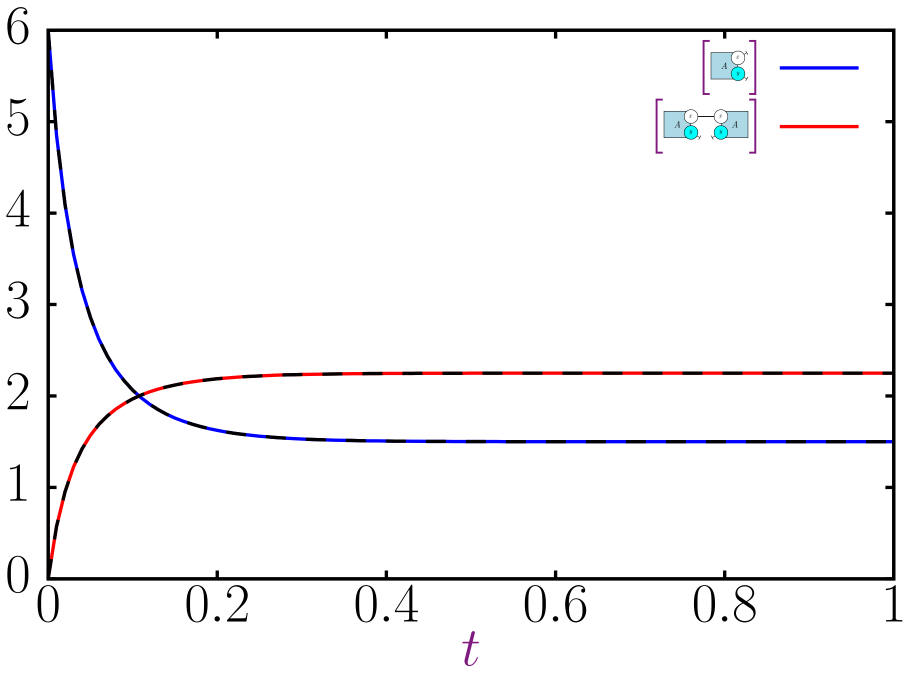


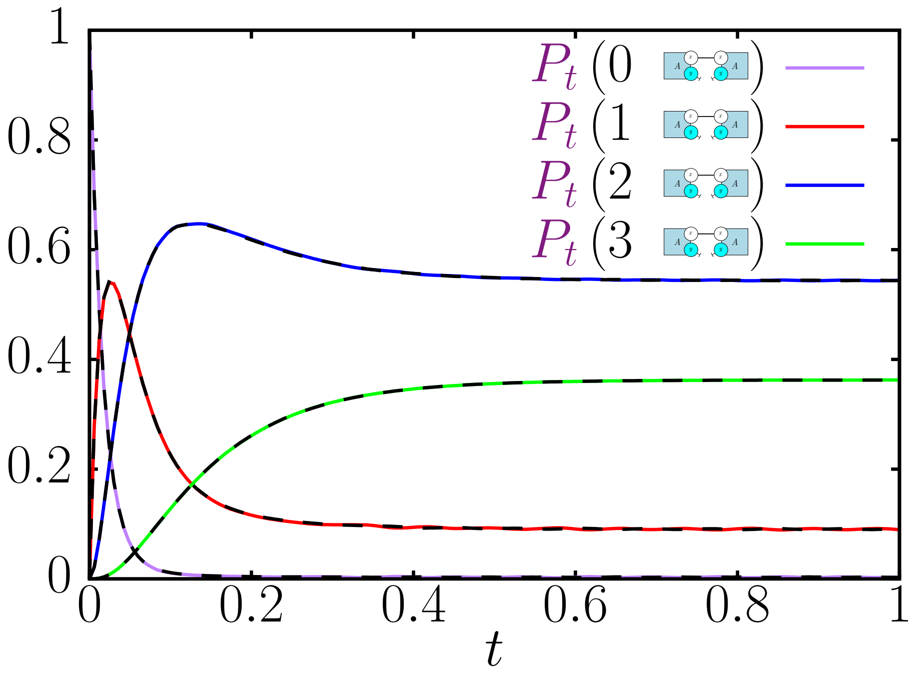
Reduced model





Macrostate distribution





with:

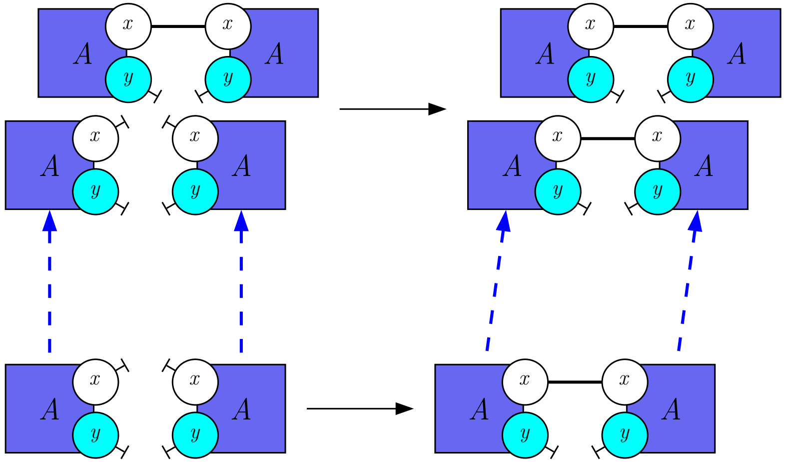
- $k_{xx} = k_{yy} = 0.5$, $k^d_{xy} = 2$, $k_{xy} = k^d_{xx} = k^d_{yy} = 1$;
- $K = 2$, $K^d = 1$,

(continuous lines: initial model; dashed lines: reduced model)

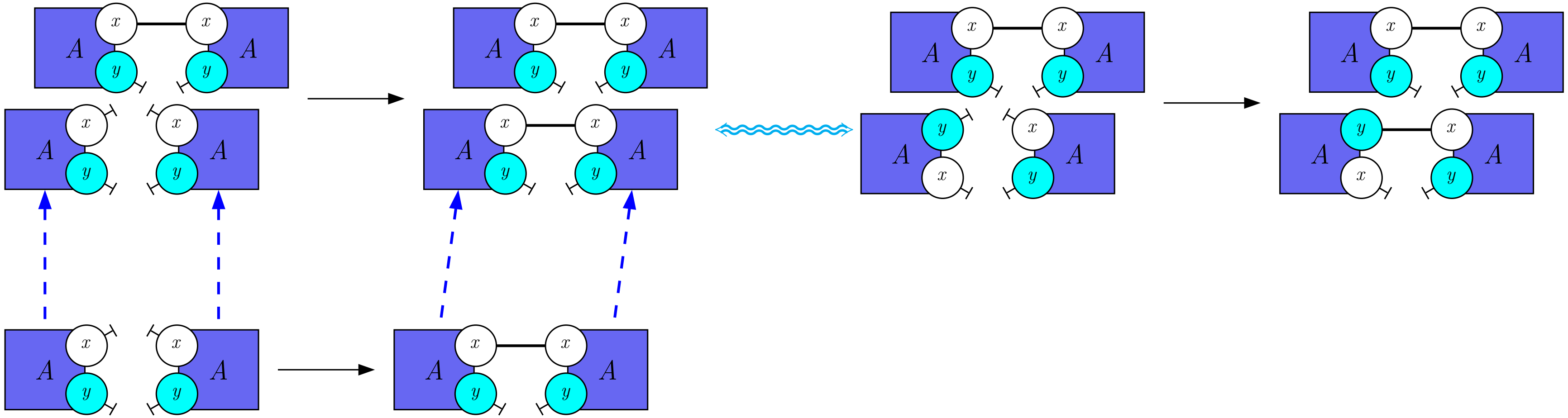
Outline

1. Context and motivations
2. Static analysis
3. Flow-based model reduction (differential)
4. Flow-based model reduction (stochastic)
5. Symmetries
    - Case studies
    - Fondamental property
    - Symmetry actions
    - Bisimulations
    - Conclusion and perspectives
6. Conclusion and perspectives

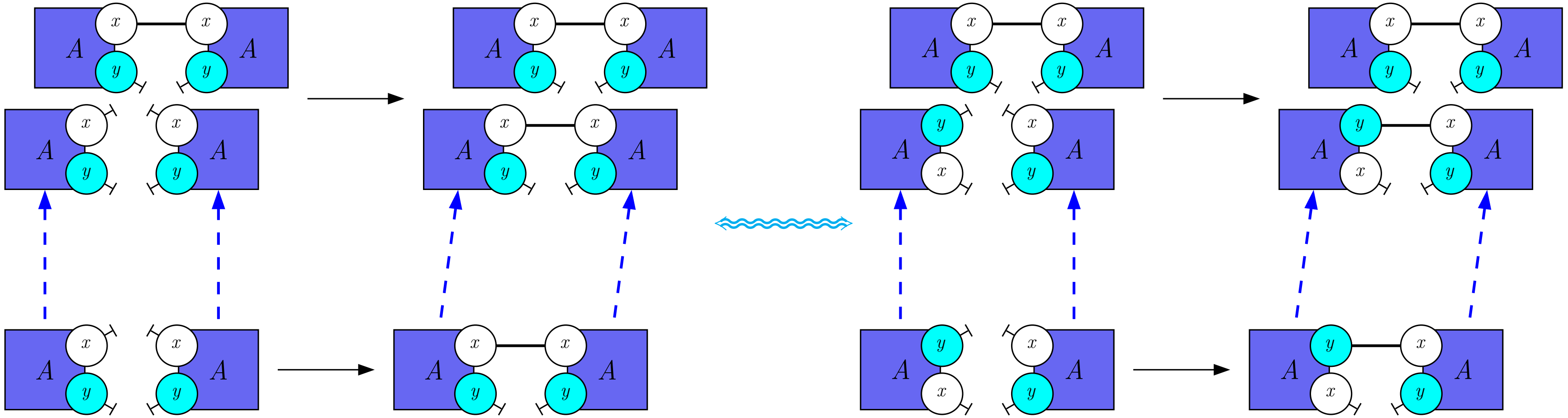
Symmetric of a transition step



Symmetric of a transition step



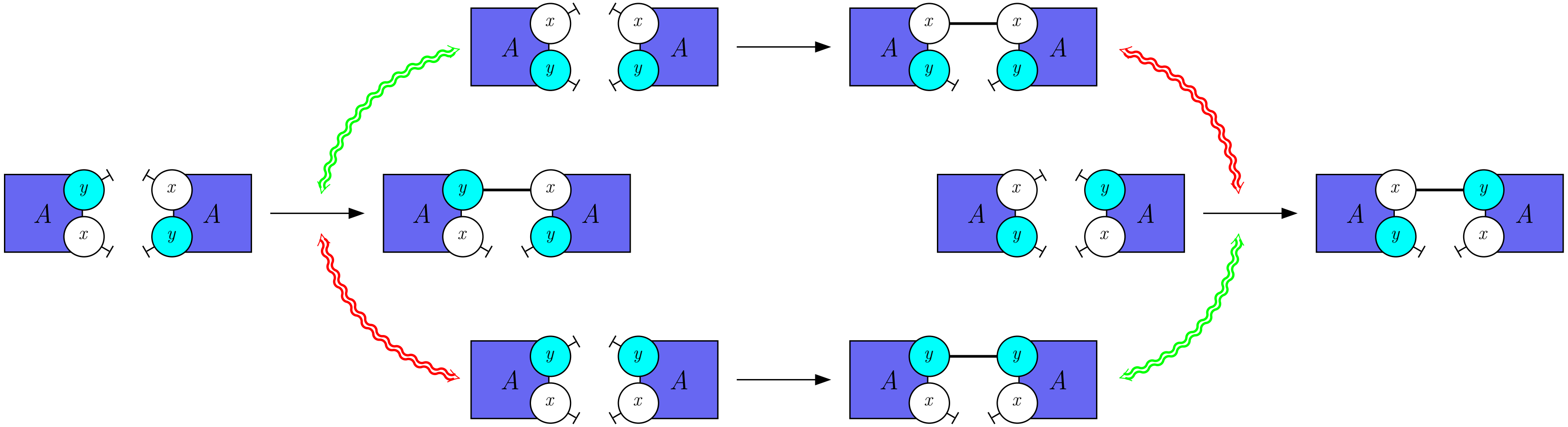
Symmetric of a transition step



Outline

1. Context and motivations
2. Static analysis
3. Flow-based model reduction (differential)
4. Flow-based model reduction (stochastic)
5. Symmetries
    - Case studies
    - Fondamental property
    - Symmetry actions
    - Bisimulations
    - Conclusion and perspectives
6. Conclusion and perspectives

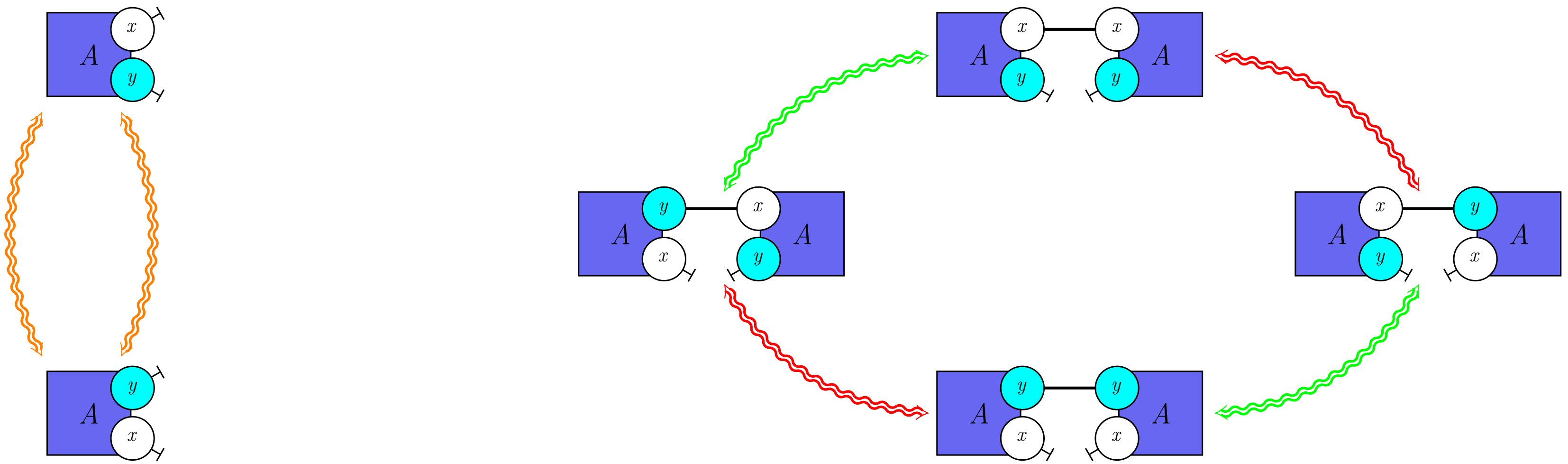
Symmetric sets of rules



A set of rules is symmetric if the rates of two symmetric rules are proportional to their numbers of occurrences in the corresponding orbit.

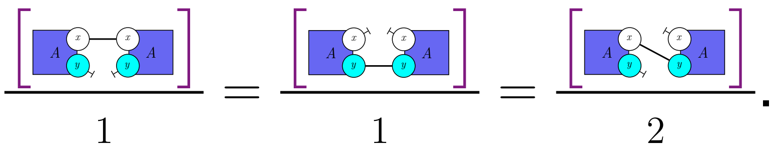
Here: $\frac{k_{xx}}{1} = \frac{k_{yy}}{1} = \frac{k_{xy}}{2}$.

Symmetric continuous states

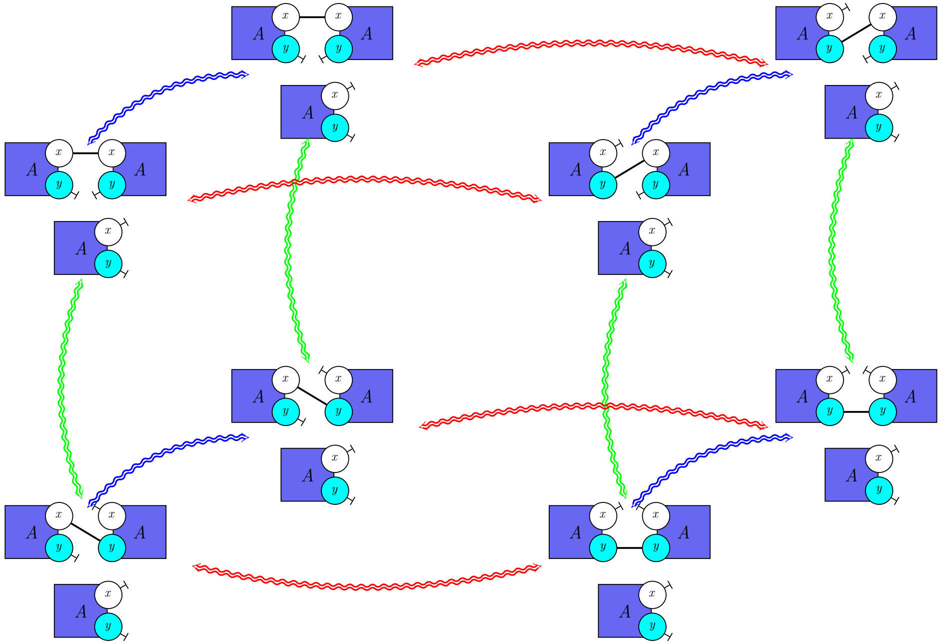


A continuous state is symmetric if the quantity of two symmetric biological complexes are proportional to their numbers of occurrences in the corresponding orbit.

Here:



Symmetric distributions of states



A continuous state is symmetric if the quantity of two symmetric biological complexes are proportional to their numbers of occurrences in the corresponding orbit.

Here proportional to their numbers of occurrences of asymmetric dimers.

Outline

1. Context and motivations
2. Static analysis
3. Flow-based model reduction (differential)
4. Flow-based model reduction (stochastic)
5. Symmetries
    - Case studies
    - Fondamental property
    - Symmetry actions
    - Bisimulations
    - Conclusion and perspectives
6. Conclusion and perspectives

Properties

Permutations of sites induce back-and-forth bisimulations:

- Forward:

  The semantics can be computed directly on symmetric classes of states (resp. distributions of states).

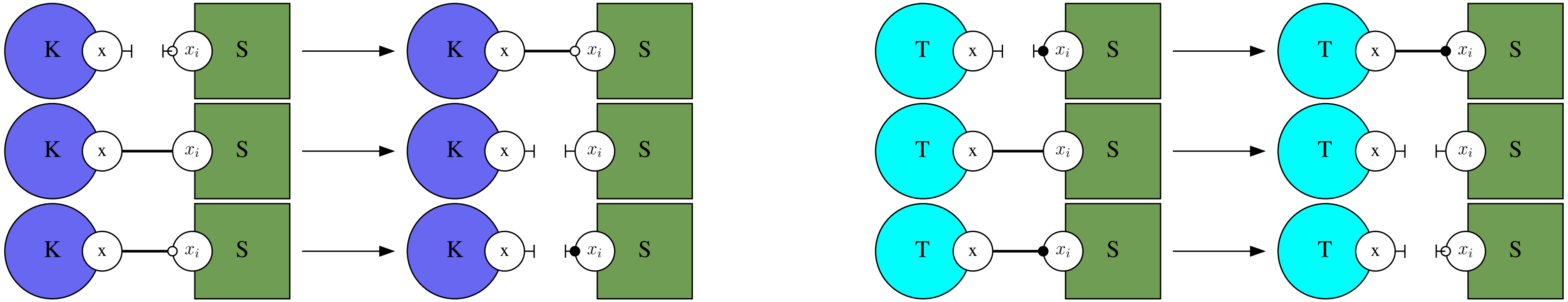
- Backward:

  Starting in a symmetric state (resp. distribution of states), the state (resp. distribution of states) is symmetric at any time.

Outline

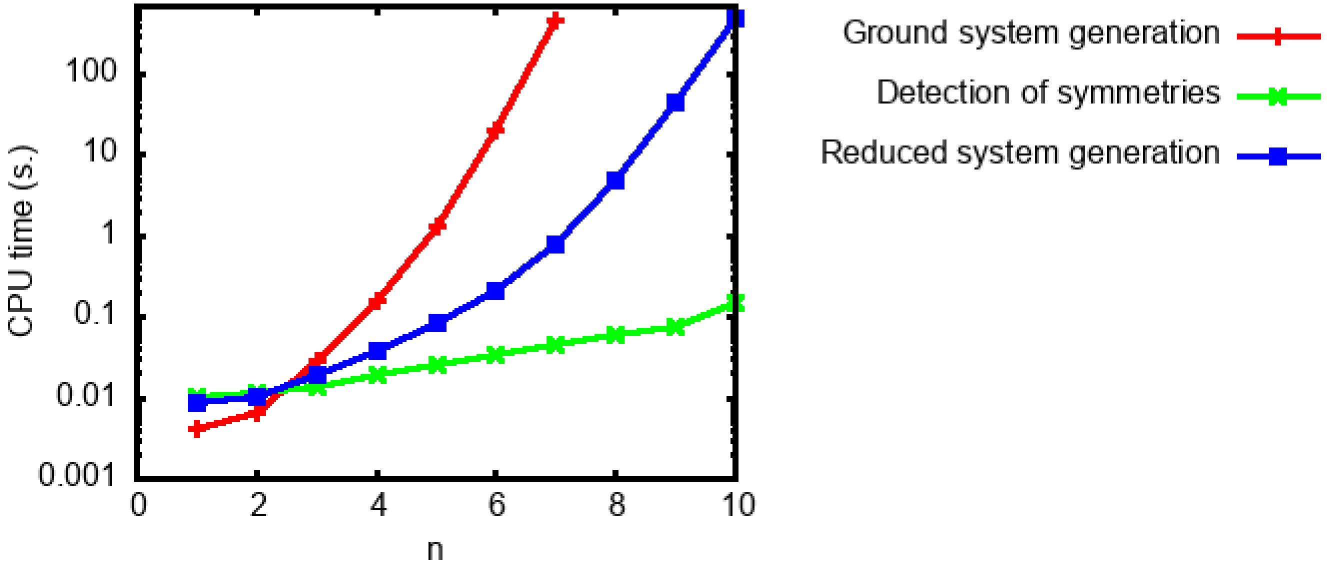
1. Context and motivations
2. Static analysis
3. Flow-based model reduction (differential)
4. Flow-based model reduction (stochastic)
5. Symmetries
    - Case studies
    - Fondamental property
    - Symmetry actions
    - Bisimulations
    - Conclusion and perspectives
6. Conclusion and perspectives

Performance: parametric example



- Number of sites in S: n
- Number of reactions (initial model): $6{\cdot}n{\cdot}4^{n-1}$
- Number of reactions (reduced model): $n{\cdot}(n+1)\cdot(n+2)$

Performance: computation time



on a MacBook Pro - Intel Core i7-6567U (3.3 GHz)

Conclusion

- Model reduction based on site permutations
  - it scales on large models
- Categorical framework
  - subgroups of symmetries;
  - necessary conditions to induce (forward and/or backward) bisimulations

Perspectives

- Investigate specific subgroups of symmetries
  (Chemistry, contextual symmetries)
- Automatic parameterization of context-sensitivity.

Outline

1. Context and motivations
2. Static analysis
3. Flow-based model reduction (differential)
4. Flow-based model reduction (stochastic)
5. Symmetries
6. Conclusion and perspectives

Conclusion

- An efficient and precise static analysis.
  - To improve the confidence in models.
  - Based on the detection of structural properties of reachable states.

- Exact model reductions
  - No need to generate the set of biological complexes;
    1. Relying on the topology of information flow;
       Differential case: efficient on models of signaling pathways;
       Stochastic case: no reduction in most of the cases.
    2. Relying on groups of symmetries
       One-to-one relations between reactions and transitions.

Perspectives on static analysis

- Analyze families of models (meta-languages, scripts);
- Incremental analysis.

Perspectives on exact model reduction

- Provide heuristic for tuning context sensitivity automatically;
- Investigate more (sub)groups of symmetries.

Perspectives on approximate model reduction

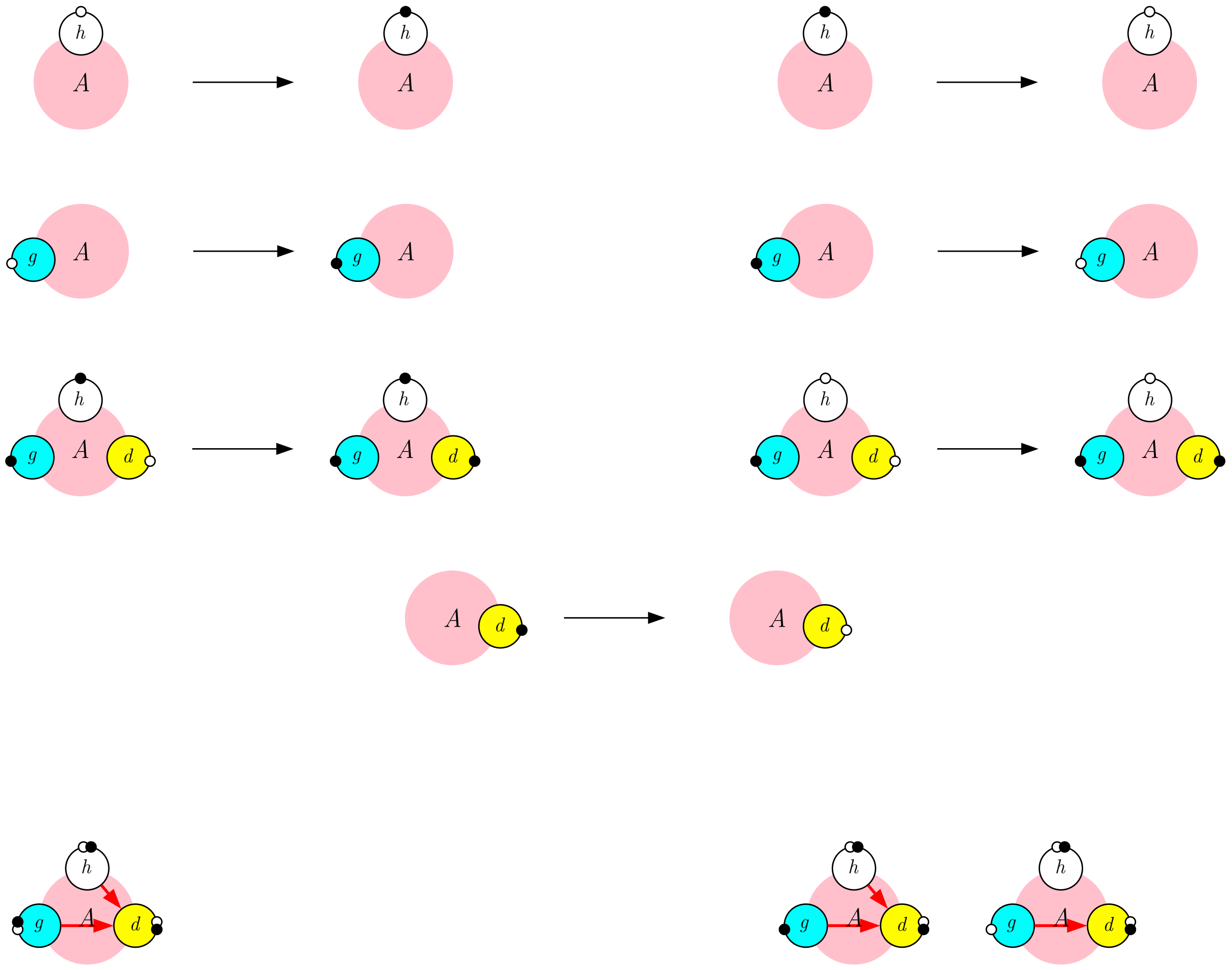
Design conservative numerical methods for approximate model reduction

- the choice of fragments is not imposed by the analysis;
- with interval bounds computed *a posteriori*;
- to be declined for differential, stochastic, hybrid semantics.

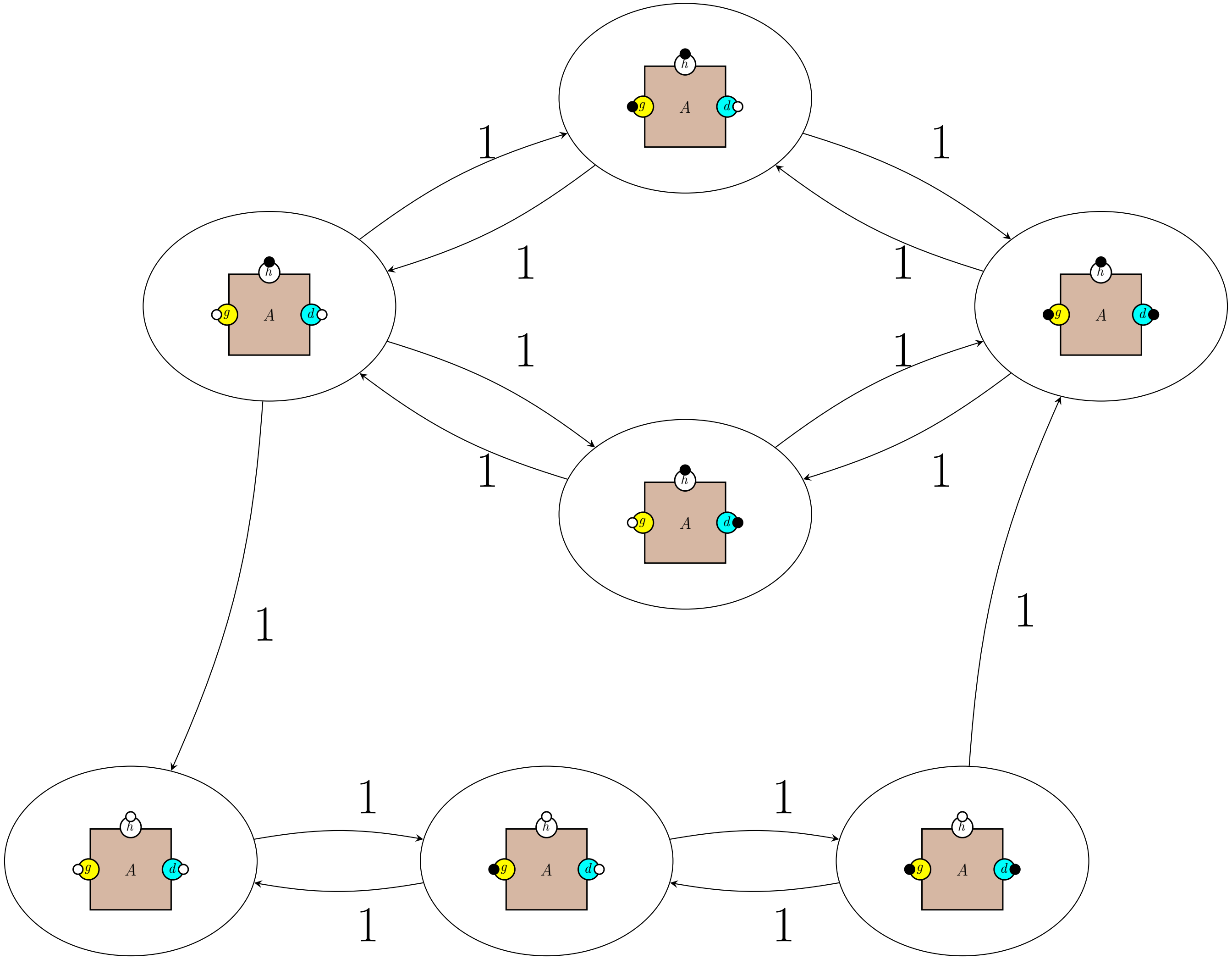
A toolkit to recast and unify existing approaches and propose new ones:

- early results on tropicalization [Beica *et al.* 2020];
- early results on finite expansions (models of polymers);
- tropical equilibration;
- flow of information;
- …

Contextual flow



Contextual symmetries



Homogeneous permutations