MFPS XXVI

Special Session on Systems Biology

Internal coarse-graining of molecular systems*

[PNAS'09]

Jérôme Feret

Laboratoire d'Informatique de l'École Normale Supérieure

INRIA, ÉNS, CNRS

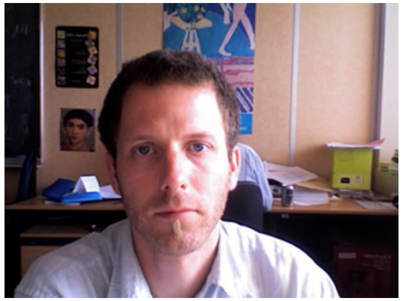
* This research has been done during my Post-Doc at Harvard Medical School

Mai 8, 2010

Joint-work with...

Walter Fontana
Harvard Medical School

Vincent Danos
Edinburgh



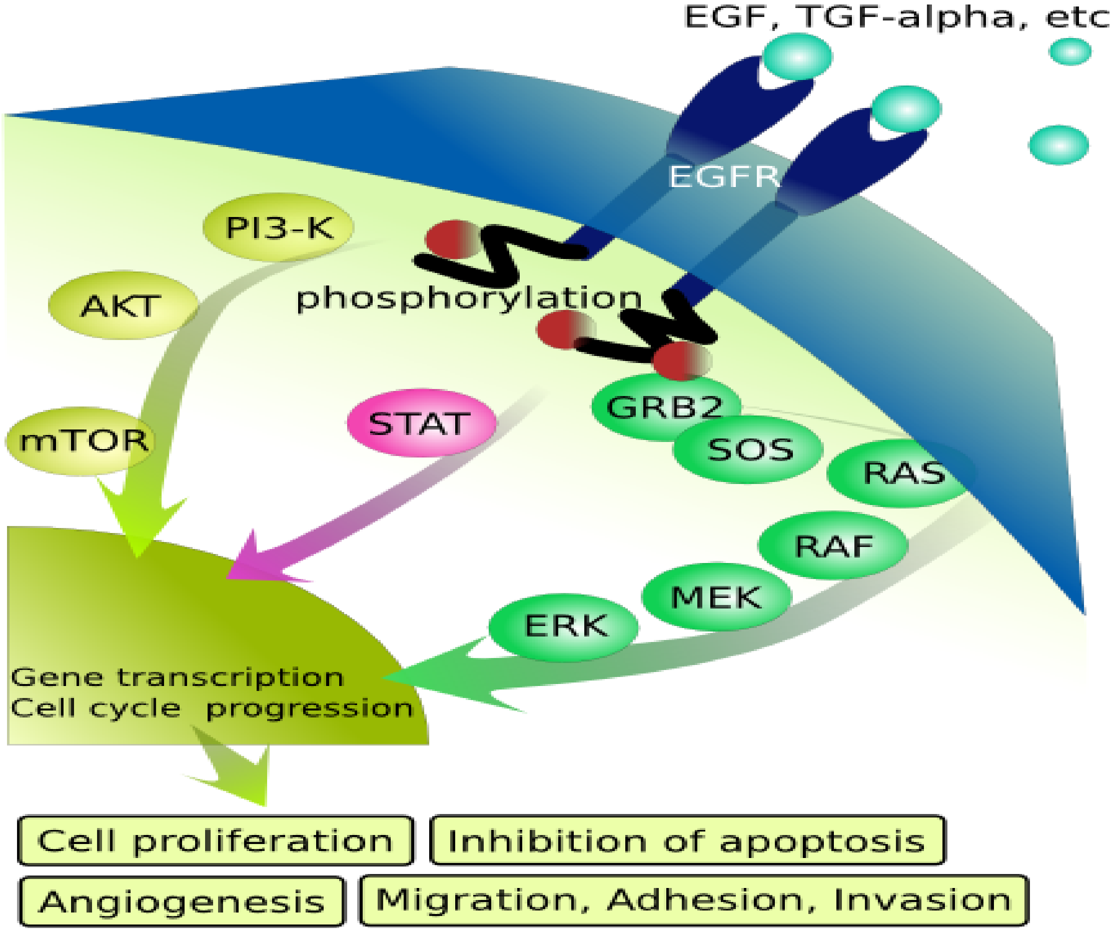
Russ Harmer
Paris VII
(Harvard Medical School)

Jean Krivine
Paris VII

Overview

1. Context and motivations
2. Handmade ODEs
3. Abstract interpretation framework
4. Instantiation
5. Conclusion

Signalling Pathways



Eikuch, 2007

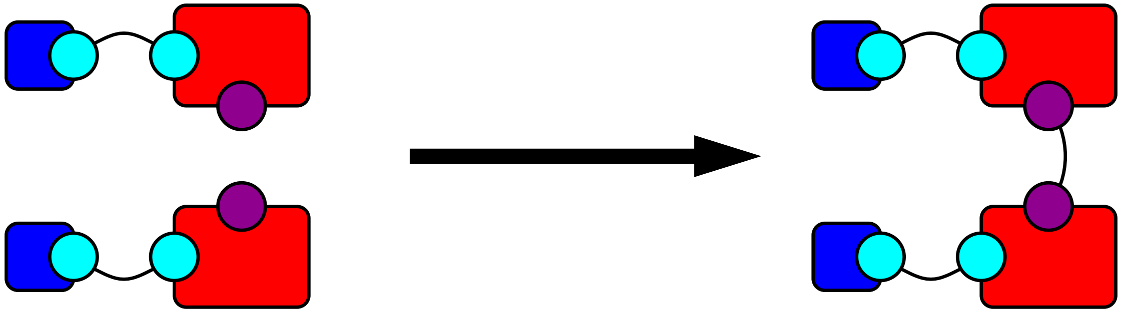
A gap between two worlds

Two levels of description:

1. Databases of proteins interactions in natural language
   + documented and detailed description
   + transparent description
   – cannot be interpreted
2. ODE-based models
   + can be integrated
   – opaque modelling process, models can hardly be modified
   – there are also some scalability issues.

Rule-based approach

We use site graph rewrite systems

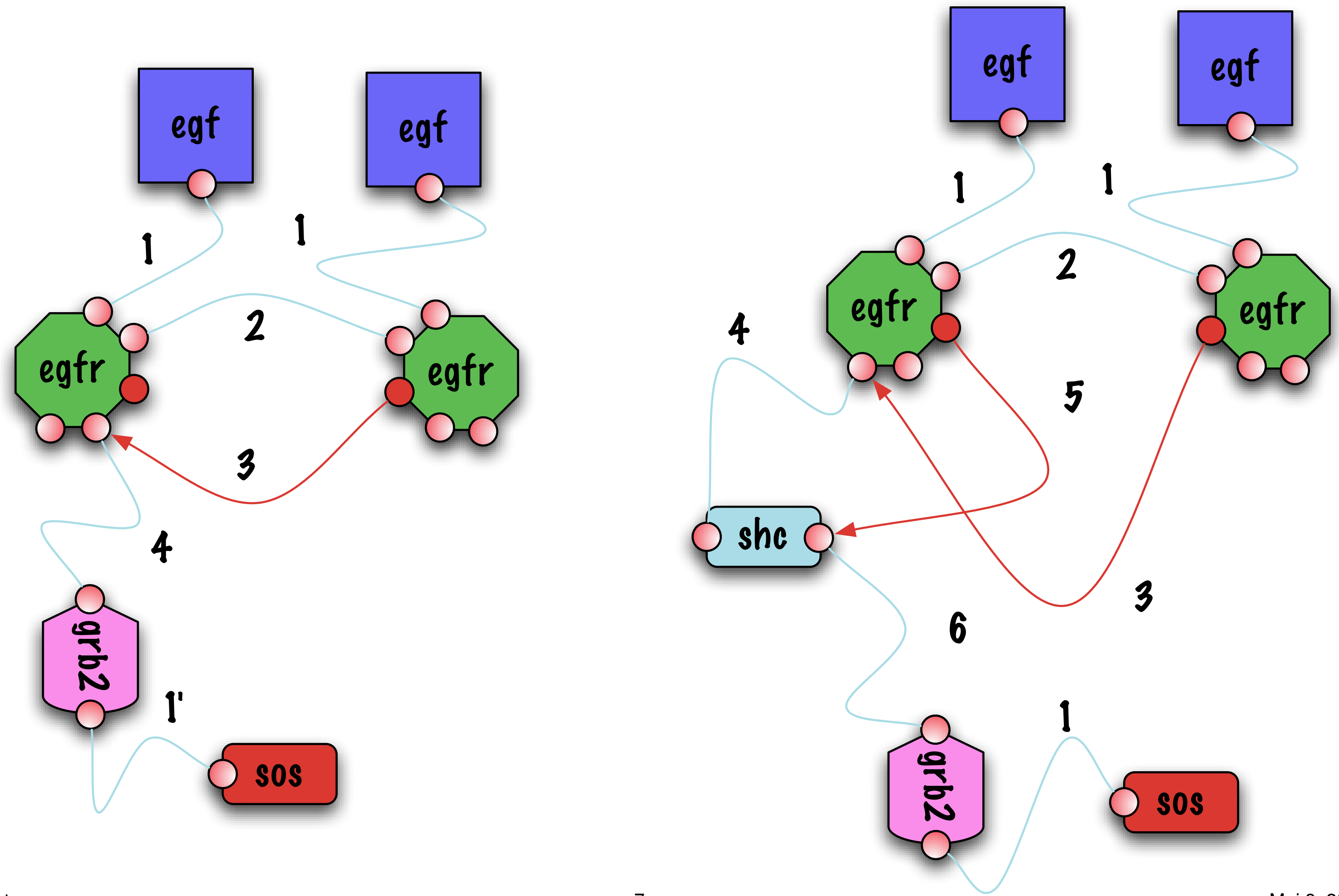


1. The description level matches with both
    - the observation level
    - and the intervention level

   of the biologist.

   We can tune the model easily.
2. Model description is very compact.
3. Quantitative semantics can be defined.

A breach in the combinatorial wall ?



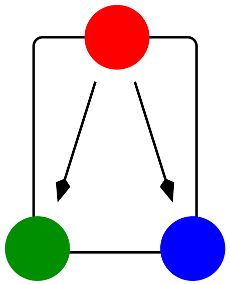
Overview

1. Context and motivations
2. Handmade ODEs
3. Abstract interpretation framework
4. Instantiation
5. Conclusion

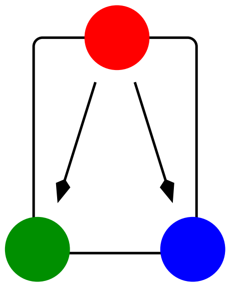
Overview

1. Context and motivations
2. Handmade ODEs
3. Abstract interpretation framework
4. Instantiation
5. Conclusion

Case study: A protein with a switch

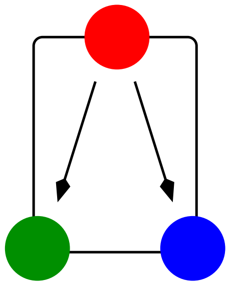


Case study: A protein with a switch



$$
\begin{array}{lcll}
(u,u,u) & \longrightarrow & (u,p,u) & k^c \\
(u,p,u) & \longrightarrow & (p,p,u) & k^l \\
(u,p,p) & \longrightarrow & (p,p,p) & k^l \\
(u,p,u) & \longrightarrow & (u,p,p) & k^r \\
(p,p,u) & \longrightarrow & (p,p,p) & k^r
\end{array}
$$

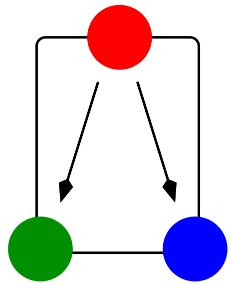
Case study: A protein with a switch



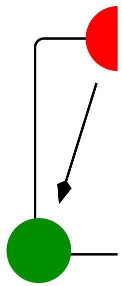
$$
\begin{array}{lcll}
(u,u,u) & \longrightarrow & (u,p,u) & k^c \\
(u,p,u) & \longrightarrow & (p,p,u) & k^l \\
(u,p,p) & \longrightarrow & (p,p,p) & k^l \\
(u,p,u) & \longrightarrow & (u,p,p) & k^r \\
(p,p,u) & \longrightarrow & (p,p,p) & k^r
\end{array}
$$

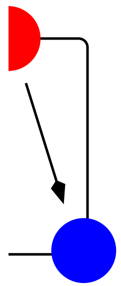
$$
\begin{cases}
\frac{d[(u,u,u)]}{dt} = -k^c \cdot [(u,u,u)] \\
\frac{d[(u,p,u)]}{dt} = -k^l \cdot [(u,p,u)] + k^c \cdot [(u,u,u)] - k^r \cdot [(u,p,u)] \\
\frac{d[(u,p,p)]}{dt} = -k^l \cdot [(u,p,p)] + k^r \cdot [(u,p,u)] \\
\frac{d[(p,p,u)]}{dt} = k^l \cdot [(u,p,u)] - k^r \cdot [(p,p,u)] \\
\frac{d[(p,p,p)]}{dt} = k^l \cdot [(u,p,p)] + k^r \cdot [(p,p,u)]
\end{cases}
$$

Case study: Two subsystems

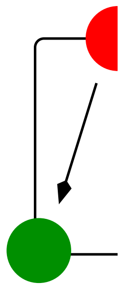


Case study: Two subsystems



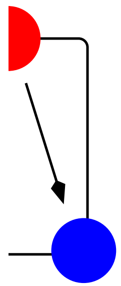


Case study: Two subsystems



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

$$\begin{cases} \frac{d[(u,u,u)]}{dt} = -k^c \cdot [(u,u,u)] \\ \frac{d[(u,p,?)]}{dt} = -k^l \cdot [(u,p,?)] + k^c \cdot [(u,u,u)] \\ \frac{d[(p,p,?)]}{dt} = k^l \cdot [(u,p,?)] \end{cases}$$



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

$$\begin{cases} \frac{d[(u,u,u)]}{dt} = -k^c \cdot [(u,u,u)] \\ \frac{d[(?,p,u)]}{dt} = -k^r \cdot [(?,p,u)] + k^c \cdot [(u,u,u)] \\ \frac{d[(?,p,p)]}{dt} = k^r \cdot [(?,p,u)] \end{cases}$$

Case study: Dependence index

We introduce:

$$[(?,p,?)] \stackrel{\Delta}{=} [(?,p,u)] + [(?,p,p)]$$

The states of left site and right site would be independent if, and only if:

$$\frac{[(p,p,p)]}{[(p,p,?)]} = \frac{[(?,p,p)]}{[(?,p,?)]}.$$

Thus we define the dependence index as follows:

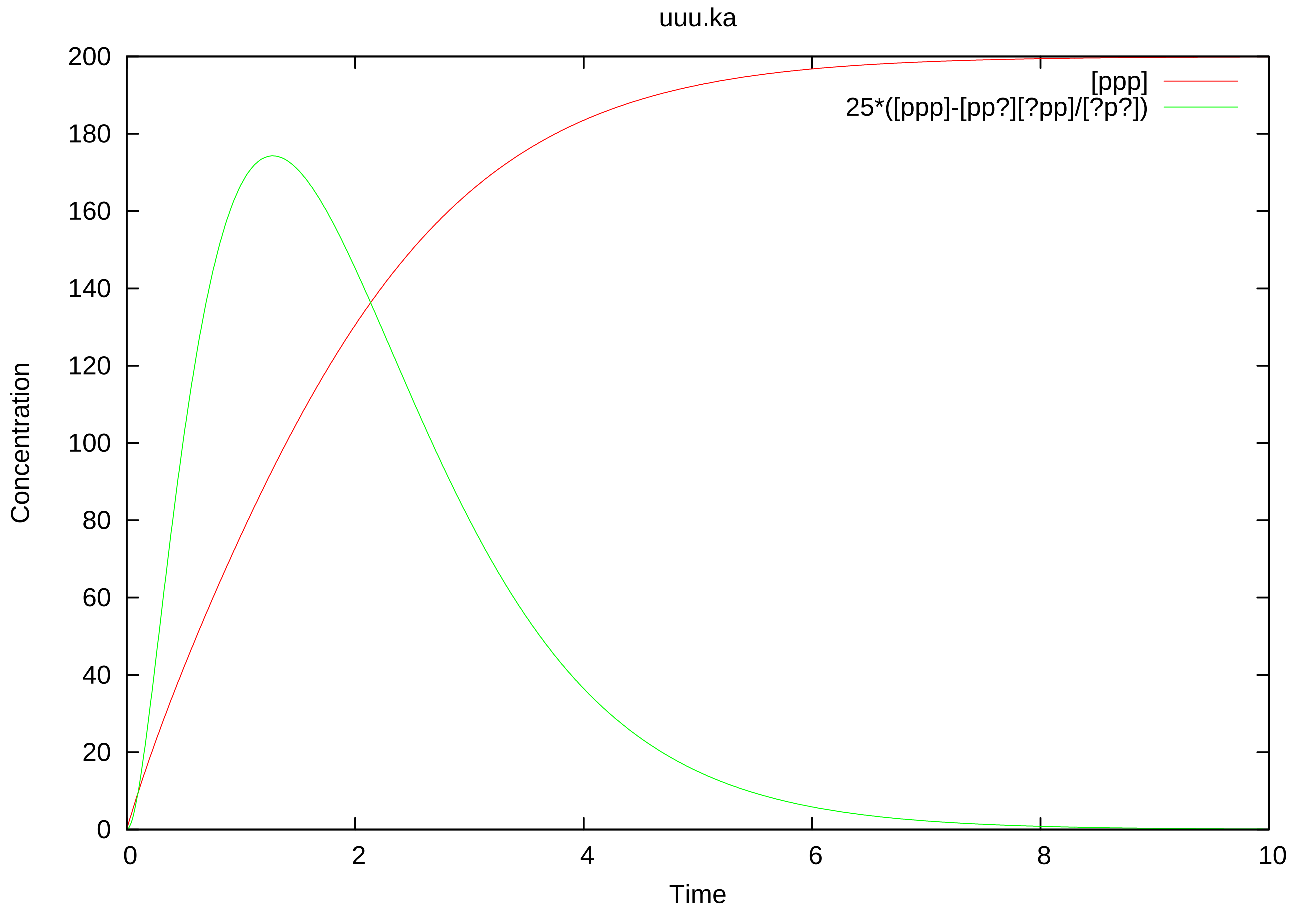
$$X \stackrel{\Delta}{=} [(p,p,p)]\cdot[(?,p,?)] - [(?,p,p)]\cdot[(p,p,?)].$$

We have (after a short computation):

$$\frac{dX}{dt} = -X \cdot \left(k^l + k^r\right) + k^c\cdot[(p,p,p)]\cdot[(u,u,u)].$$

As a consequence, the property $X = 0$ is not an invariant.

Case study: Erroneous recombination



Concentrations evolution with respect to time ($[(u,u,u)](0) = 100$).

$[(p,p,p)]$ and $25 \cdot \left([(p,p,p)] - \frac{[(p,p,?)] \cdot [(?,p,p)]}{[(?,p,?)]}\right)$

Conclusion
self-consistency

– some information is abstracted away
we cannot recover the concentration of any species;

+ it is a weak property
which is easy to ensure,
which is easy to propagate;
+ it captures the essence of the kinetics of systems.

We are going to track the correlations that are read by the system.

Overview

1. Context and motivations
2. Handmade ODEs
3. Abstract interpretation framework
4. Instantiation
5. Conclusion

Continuous differential semantics

Given $\mathcal{V}$, a finite set of variables;
and $\mathbb{F}$, a $\mathcal{C}^\infty$ mapping from $\mathcal{V} \to \mathbb{R}^+$ into $\mathcal{V} \to \mathbb{R}$.

as for instance,

- $\mathcal{V} \stackrel{\Delta}{=} \{[(u,u,u)], [(u,p,u)], [(p,p,u)], [(u,p,p)], [(p,p,p)]\}$,
- $\mathbb{F}(\rho) \stackrel{\Delta}{=} \begin{cases} [(u,u,u)] \mapsto -k^c \cdot \rho([(u,u,u)]) \\ [(u,p,u)] \mapsto -k^l \cdot \rho([(u,p,u)]) + k^c \cdot \rho([(u,u,u)]) - k^r \cdot \rho([(u,p,u)]) \\ [(u,p,p)] \mapsto -k^l \cdot \rho([(u,p,p)]) + k^r \cdot \rho([(u,p,u)]) \\ [(p,p,u)] \mapsto k^l \cdot \rho([(u,p,u)]) - k^r \cdot \rho([(p,p,u)]) \\ [(p,p,p)] \mapsto k^l \cdot \rho([(u,p,p)]) + k^r \cdot \rho([(p,p,u)]); \end{cases}$

we can define the continuous differential semantics as follows:

$$X_c : \begin{cases} (\mathcal{V} \to \mathbb{R}^+) \times \mathbb{R}^+ & \to & (\mathcal{V} \to \mathbb{R}^+) \\ (X_0, T) & \mapsto & X_0 + \int_{t=0}^{T} \mathbb{F}(X_c(X_0, t)) \cdot dt. \end{cases}$$

Abstraction

An abstraction $(\mathcal{V}^\sharp, \psi, \mathbb{F}^\sharp)$ is given by:

- $\mathcal{V}^\sharp$: a finite set of observables,
- ψ: a mapping from $\mathcal{V} \to \mathbb{R}$ into $\mathcal{V}^\sharp \to \mathbb{R}$,
- $\mathbb{F}^\sharp$: a $\mathcal{C}^\infty$ mapping from $\mathcal{V}^\sharp \to \mathbb{R}^+$ into $\mathcal{V}^\sharp \to \mathbb{R}$;

such that:

- ψ is linear with positive coefficients,
- $\mathbb{F}^\sharp$ is ψ-complete

  i.e. the following diagram commutes:

$$\begin{tikzcd} \mathcal{V} \to \mathbb{R}^+ \arrow[r, "\mathbb{F}"] \arrow[d, "\psi"'] & \mathcal{V} \to \mathbb{R} \arrow[d, "\psi"] \\ \mathcal{V}^\sharp \to \mathbb{R}^+ \arrow[r, "\mathbb{F}^\sharp"] & \mathcal{V}^\sharp \to \mathbb{R} \end{tikzcd}$$

  i.e. $\psi \circ \mathbb{F} = \mathbb{F}^\sharp \circ \psi$.

Abstraction example

- $\mathcal{V} \stackrel{\Delta}{=} \{[(u,u,u)], [(u,p,u)], [(p,p,u)], [(u,p,p)], [(p,p,p)]\}$
- $\mathbb{F}(\rho) \stackrel{\Delta}{=} \begin{cases} [(u,u,u)] \mapsto -k^c \cdot \rho([(u,u,u)]) \\ [(u,p,u)] \mapsto -k^l \cdot \rho([(u,p,u)]) + k^c \cdot \rho([(u,u,u)]) - k^r \cdot \rho([(u,p,u)]) \\ [(u,p,p)] \mapsto -k^l \cdot \rho([(u,p,p)]) + k^r \cdot \rho([(u,p,u)]) \\ \ldots \end{cases}$
- $\mathcal{V}^\sharp \stackrel{\Delta}{=} \{[(u,u,u)], [(?,p,u)], [(?,p,p)], [(u,p,?)], [(p,p,?)]\}$
- $\psi(\rho) \stackrel{\Delta}{=} \begin{cases} [(u,u,u)] \mapsto \rho([(u,u,u)]) \\ [(?,p,u)] \mapsto \rho([(u,p,u)]) + \rho([(p,p,u)]) \\ [(?,p,p)] \mapsto \rho([(u,p,p)]) + \rho([(p,p,p)]) \\ \ldots \end{cases}$
- $\mathbb{F}^\sharp(\rho^\sharp) \stackrel{\Delta}{=} \begin{cases} [(u,u,u)] \mapsto -k^c \cdot \rho^\sharp([(u,u,u)]) \\ [(?,p,u)] \mapsto -k^r \cdot \rho^\sharp([(?,p,u)]) + k^c \cdot \rho^\sharp([(u,u,u)]) \\ [(?,p,p)] \mapsto k^r \cdot \rho^\sharp([(?,p,u)]) \\ \ldots \end{cases}$

(Completeness can be checked analytically.)

Abstract continuous trajectories

Given an abstraction $(\mathcal{V}^\sharp, \psi, \mathbb{F}^\sharp)$, we have:

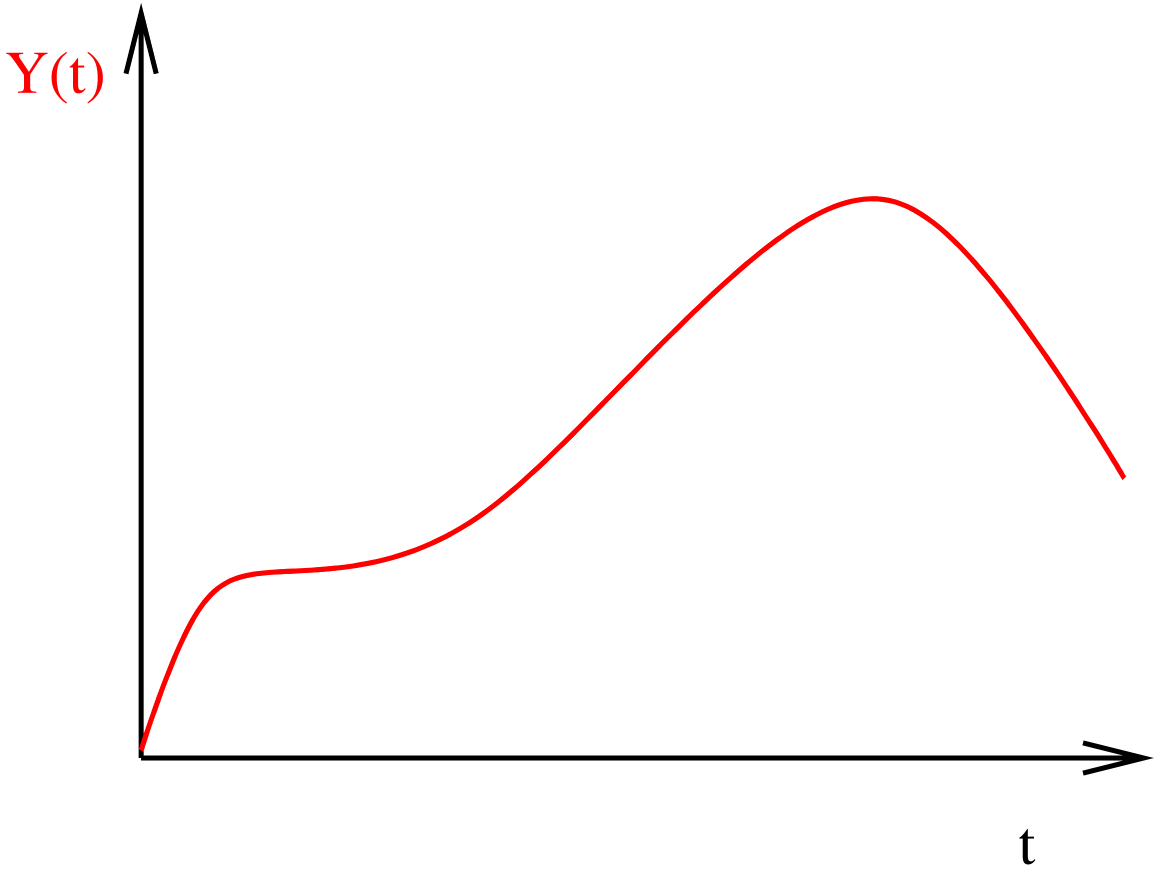
$$\begin{array}{rcll}
X_c(X_0, T) & = & X_0 + \int_{t=0}^{T} \mathbb{F}\left(X_c(X_0, t)\right) \cdot dt & \\
\psi\left(X_c(X_0, T)\right) & = & \psi\left(X_0 + \int_{t=0}^{T} \mathbb{F}\left(X_c(X_0, t)\right) \cdot dt\right) & \\
\psi\left(X_c(X_0, T)\right) & = & \psi(X_0) + \int_{t=0}^{T} [\psi \circ \mathbb{F}]\left(X_c(X_0, t)\right) \cdot dt & (\psi \text{ is linear}) \\
\psi\left(X_c(X_0, T)\right) & = & \psi(X_0) + \int_{t=0}^{T} \mathbb{F}^\sharp\left(\psi\left(X_c(X_0, t)\right)\right) \cdot dt & (\mathbb{F}^\sharp \text{ is } \psi\text{-complete})
\end{array}$$

We set $Y_0 \triangleq \psi(X_0)$ and $Y_c \triangleq \psi \circ X_c$.

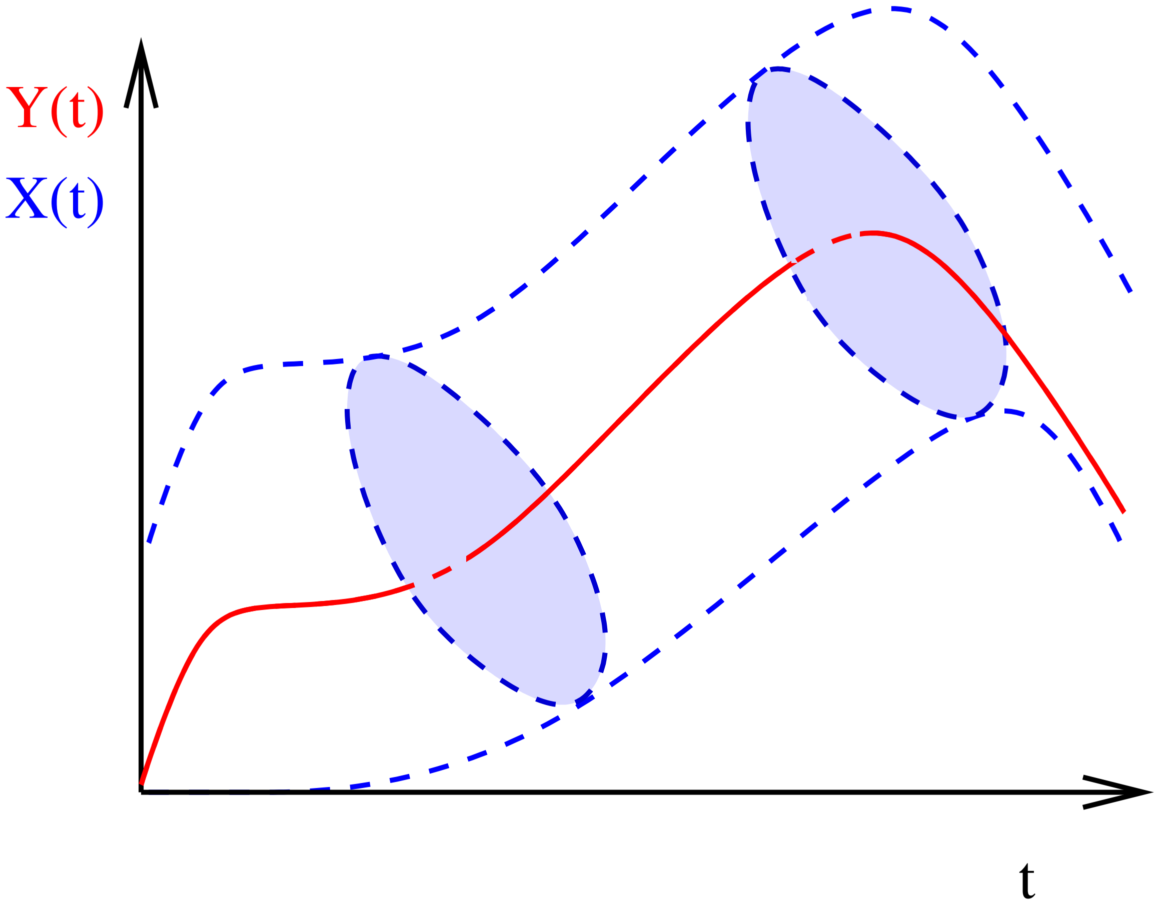
Then we have:

$$\boxed{Y_c(X_0, T) = Y_0 + \int_{t=0}^{T} \mathbb{F}^\sharp\left(Y_c(X_0, t)\right) \cdot dt}$$

Fluid trajectories



Fluid trajectories



Overview

1. Context and motivations
2. Handmade ODEs
3. Abstract interpretation framework
4. Instantiation
5. Conclusion

Differential system

Let $\mathcal{R}$ be an over-approximation of the set of reachable species.
Let us consider a rule *lhs* $\rightarrow$ *rhs* $\quad k$.

1. We write *lhs* as a multi-set $\{C_i\}$ of non empty connected components.
2. A ground instantiation of the rule *rule* is defined by a tuple $(r_i, \Phi i)$ such that $\forall i, r_i \in \mathcal{R}$ and $C_i \lhd_{\Phi i} r_i$.
3. The ground instantiation can be written as follows:
$$r_1, \ldots, r_m \rightarrow p_1, \ldots, p_n \quad k.$$
4. The activity of a ground instantiation is defined as:
$$\mathit{act}_{(r_i,\Phi i)} = \frac{k \cdot \prod [r_i]}{\sharp\{\Phi \mid \mathit{lhs} \lhd_{\Phi} \mathit{lhs}\}}.$$
5. Each ground instantiation induces the following contributions:
$$\frac{d[r_i]}{dt} \stackrel{+}{=} -\mathit{act}_{(r_i,\Phi i)}, \qquad \frac{d[p_i]}{dt} \stackrel{+}{=} \mathit{act}_{(r_i,\Phi i)}.$$

Abstract domain

We are looking for suitable pair $(\mathcal{V}^\sharp, \psi)$ (such that $\mathbb{F}^\sharp$ exists)

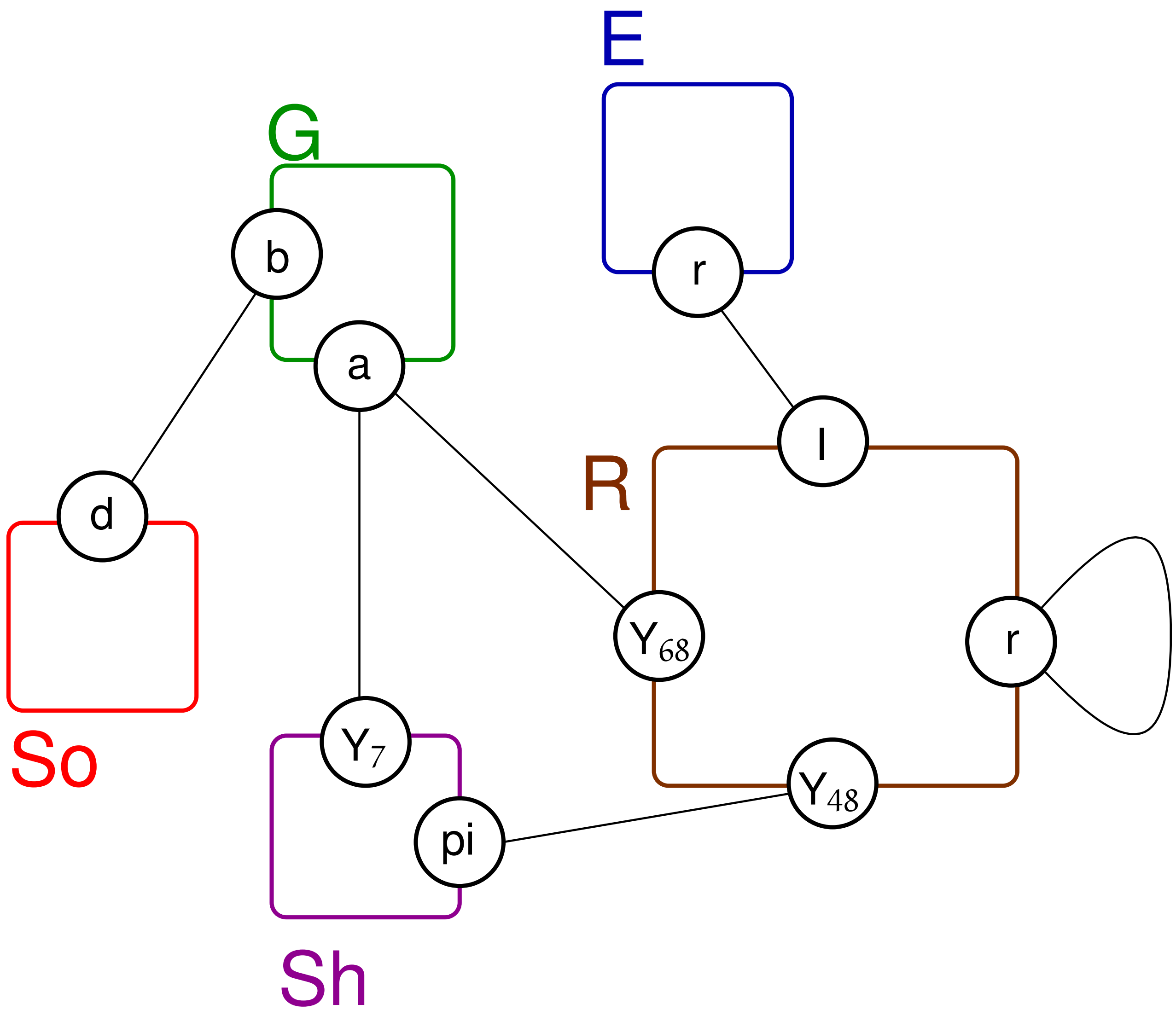
The set of linear variable changement is too big to be explored.

We introduce a specific shape on $(\mathcal{V}^\sharp, \psi)$ so as:

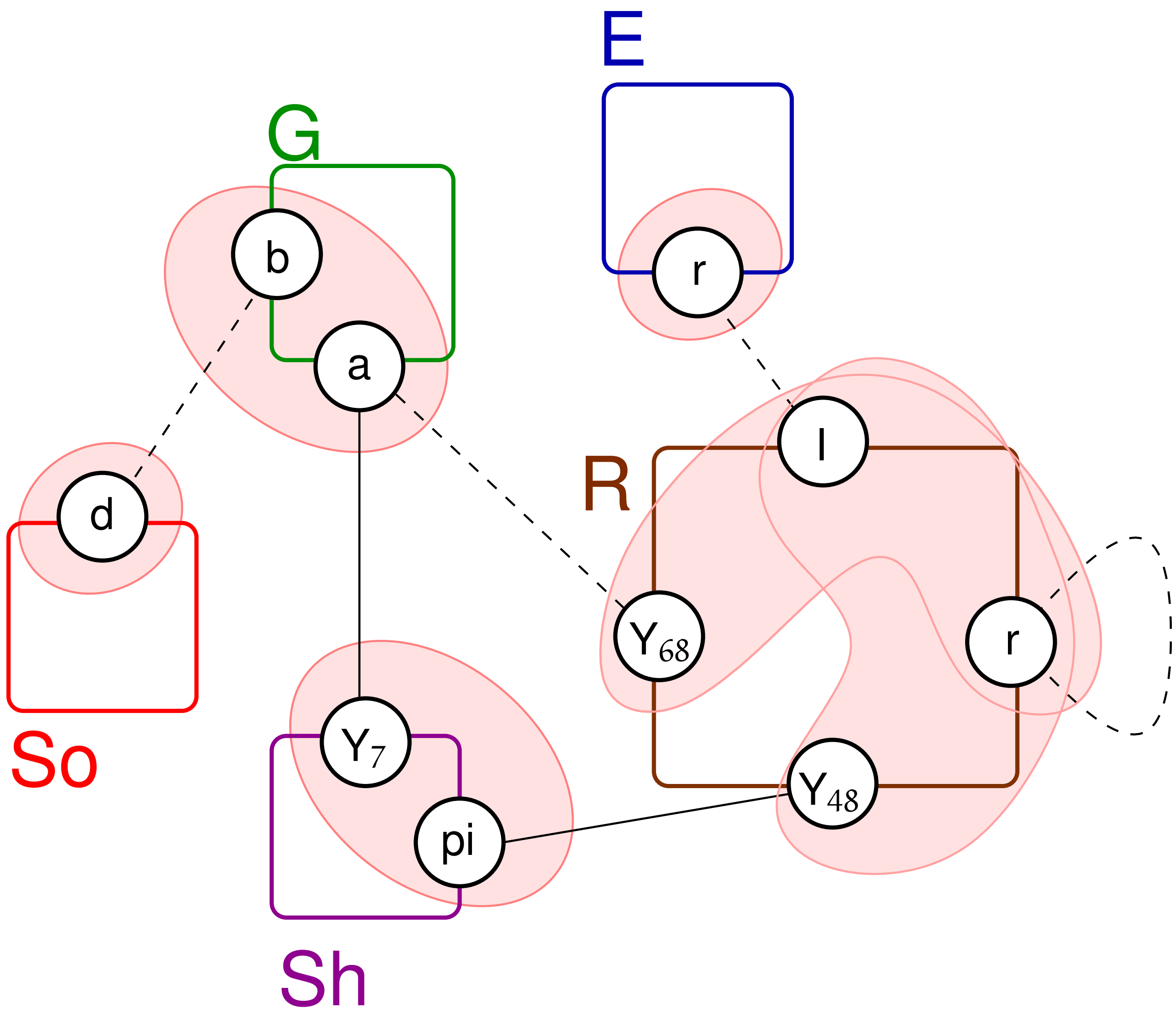
- restrict the exploration;
- drive the intuition;
- having efficient way to find suitable abstractions $(\mathcal{V}^\sharp, \psi)$ and to compute $\mathbb{F}^\sharp$.

Our choice might be not optimal, but we can live with that.

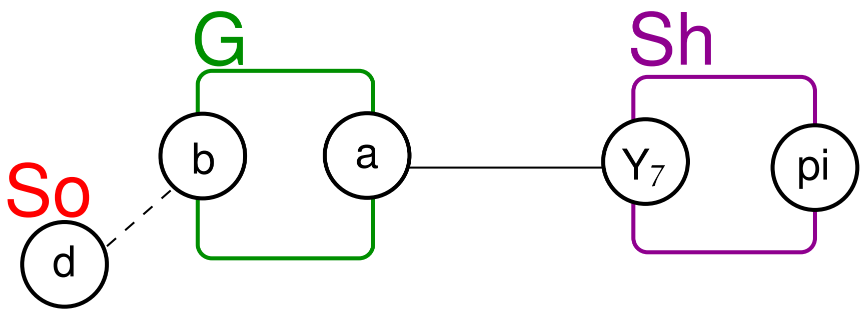
Contact map

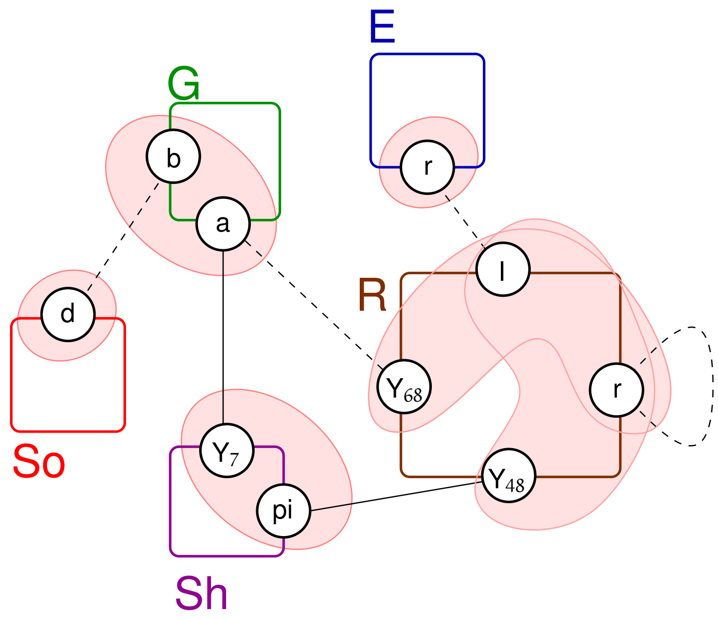


Annotated contact map



A fragment





Basic properties

The set of fragments enjoys two convenient properties:

1. Closure with respect to the operational semantics:

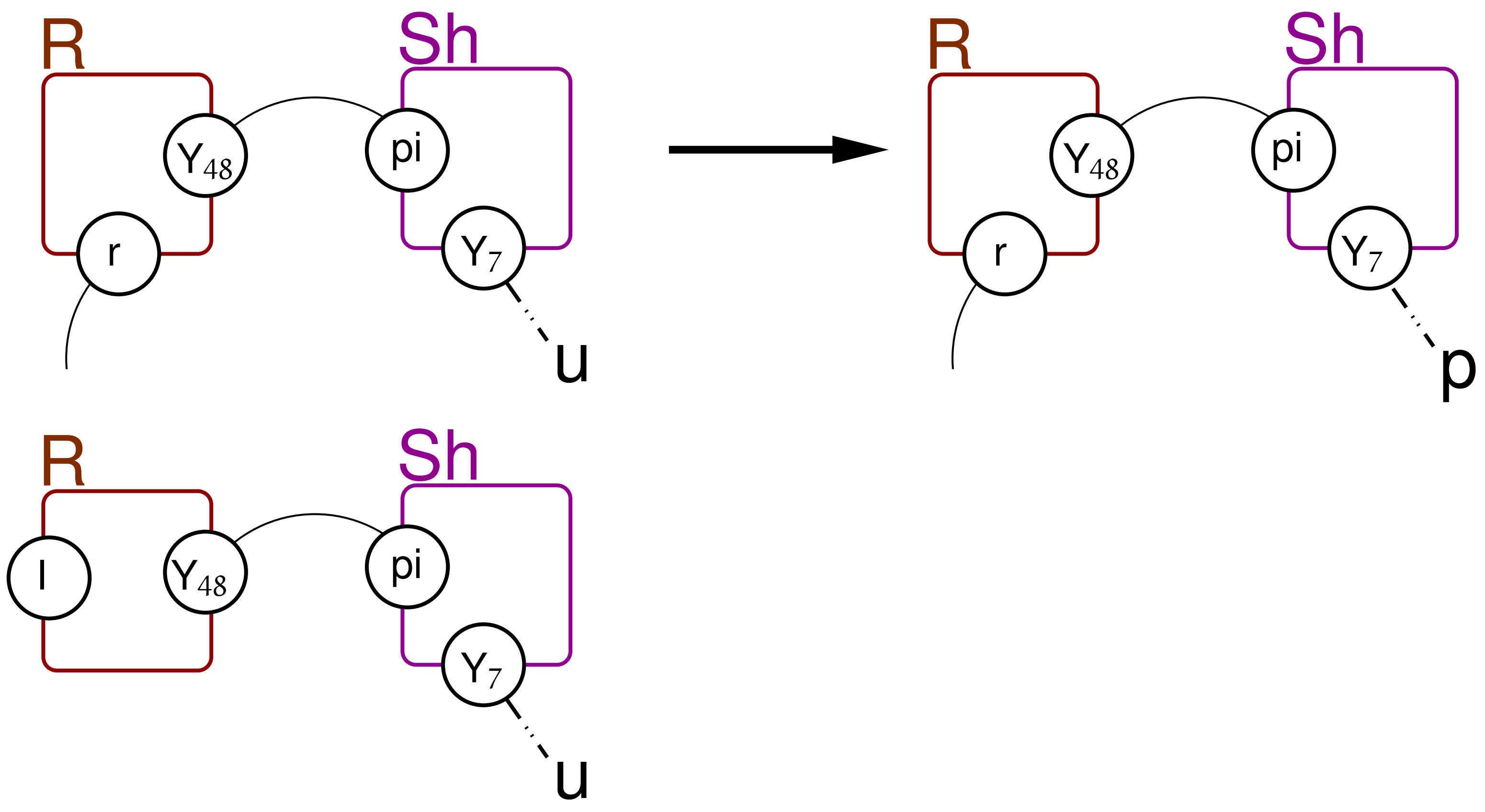
   When we apply a rule with a tuple of fragments, we get a tuple of fragments.

2. Subfragments:

   We can express the concentration of any sub-fragment as a linear combination of the concentration of some fragments.

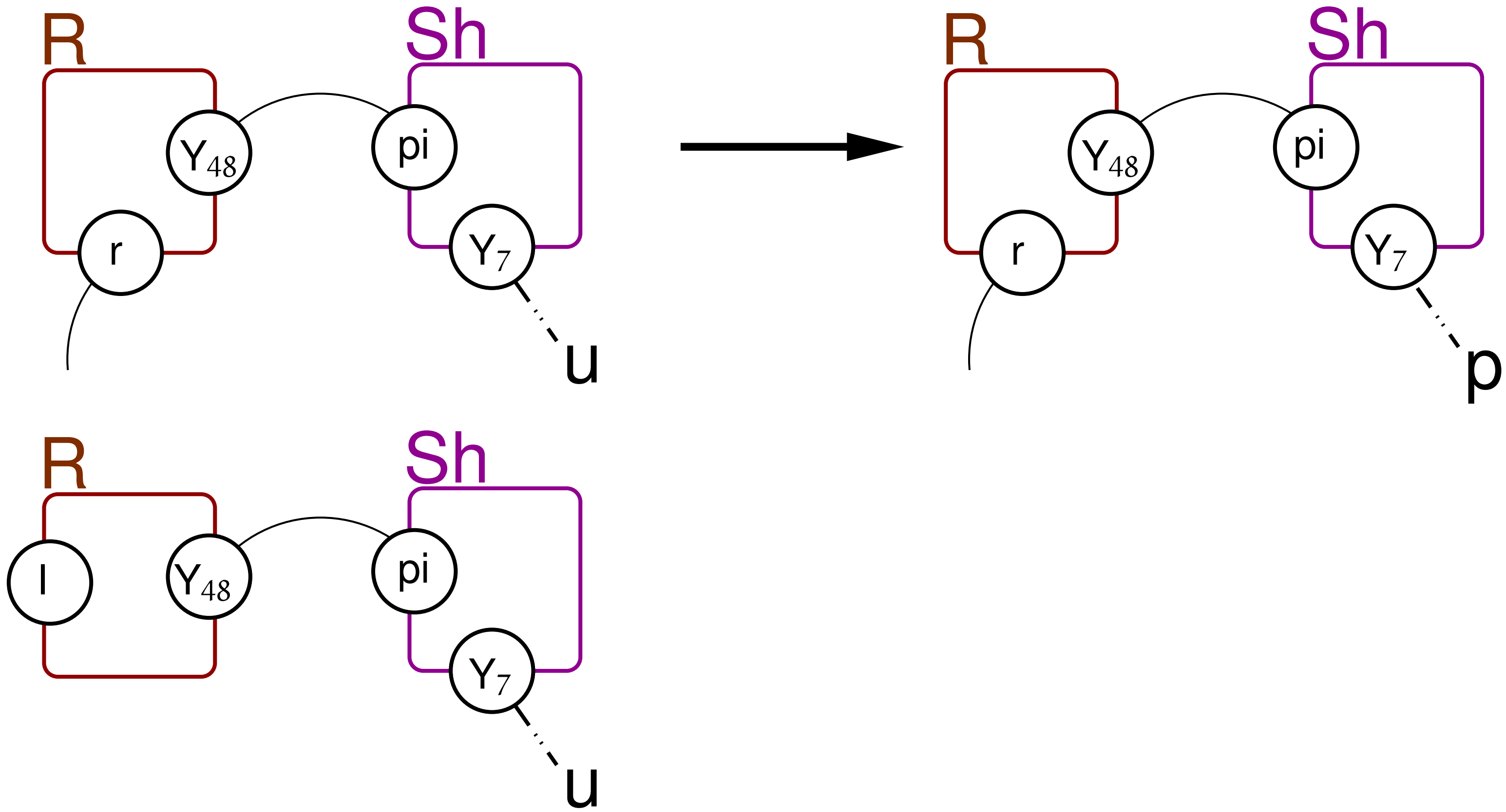
Which other properties do we need so that the function $\mathbb{F}^\sharp$ can be defined ?

Fragments consumption



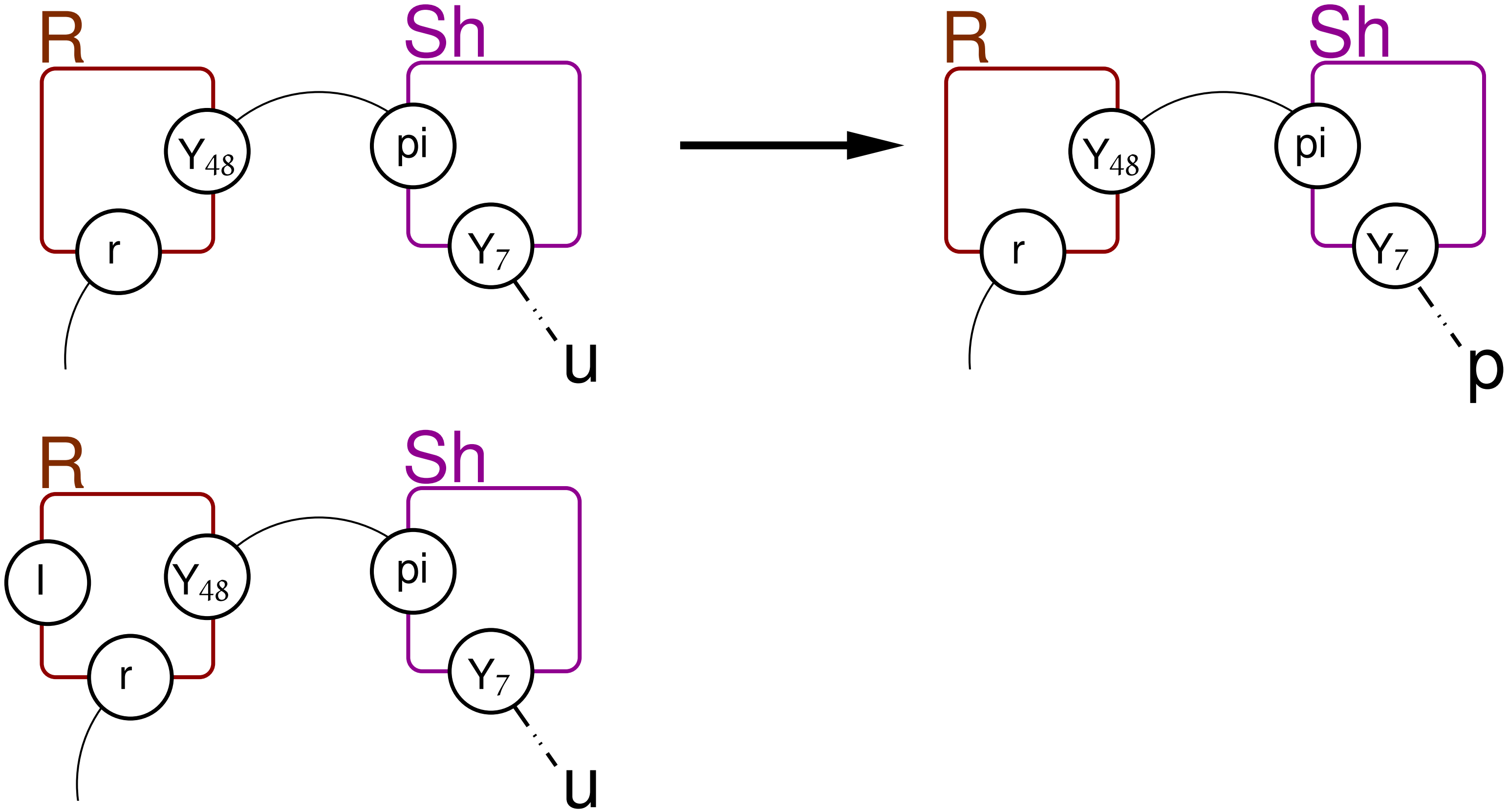
Can we express the amount (per time unit) of this fragment (bellow) concentration that is consumed by this rule (above)?

Fragments consumption



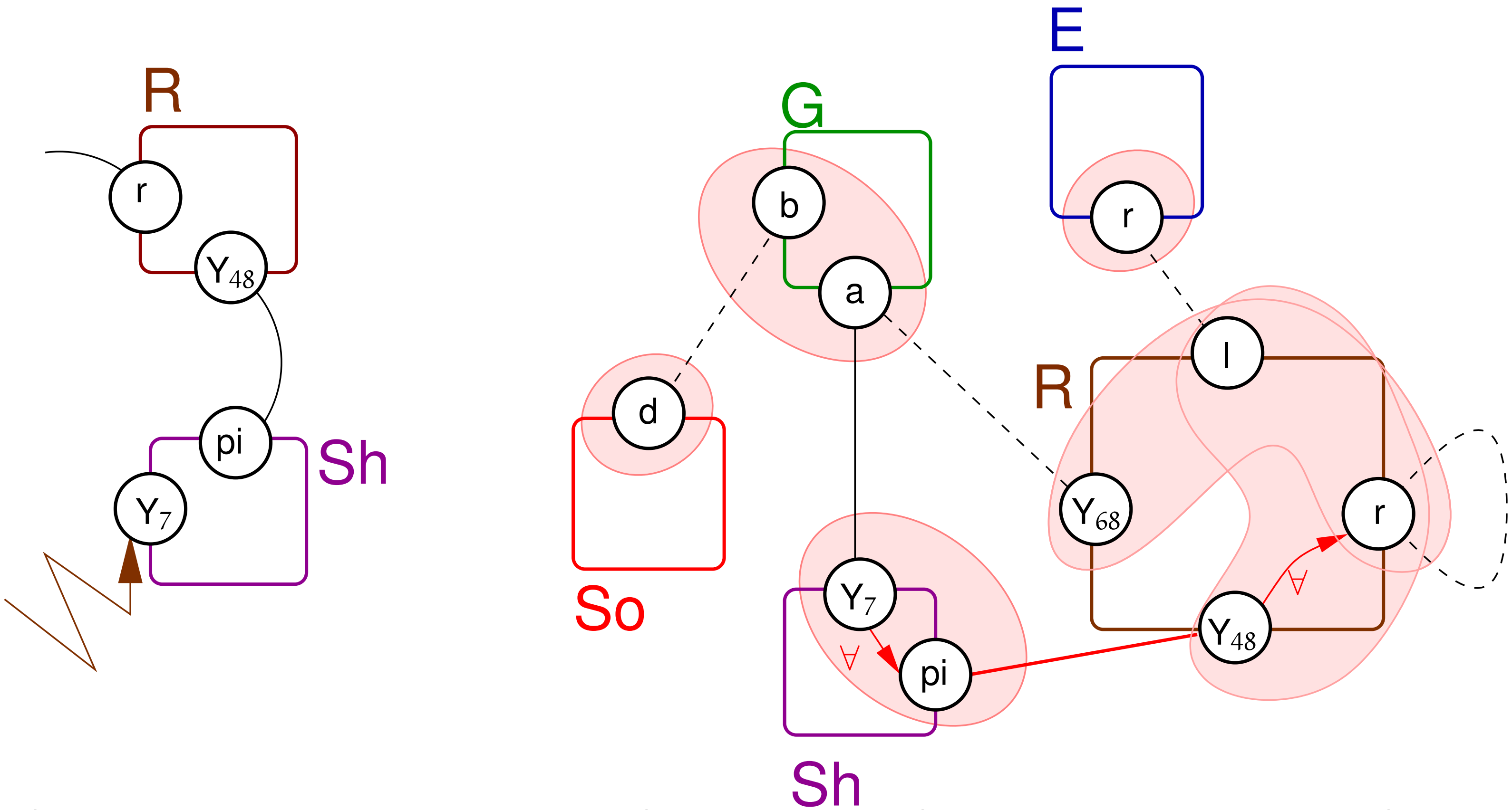
No, because we have abstracted away the correlation between the state of the site $\mathtt{r}$ and the state of the site $\mathtt{l}$.

Fragments consumption
Proper intersection



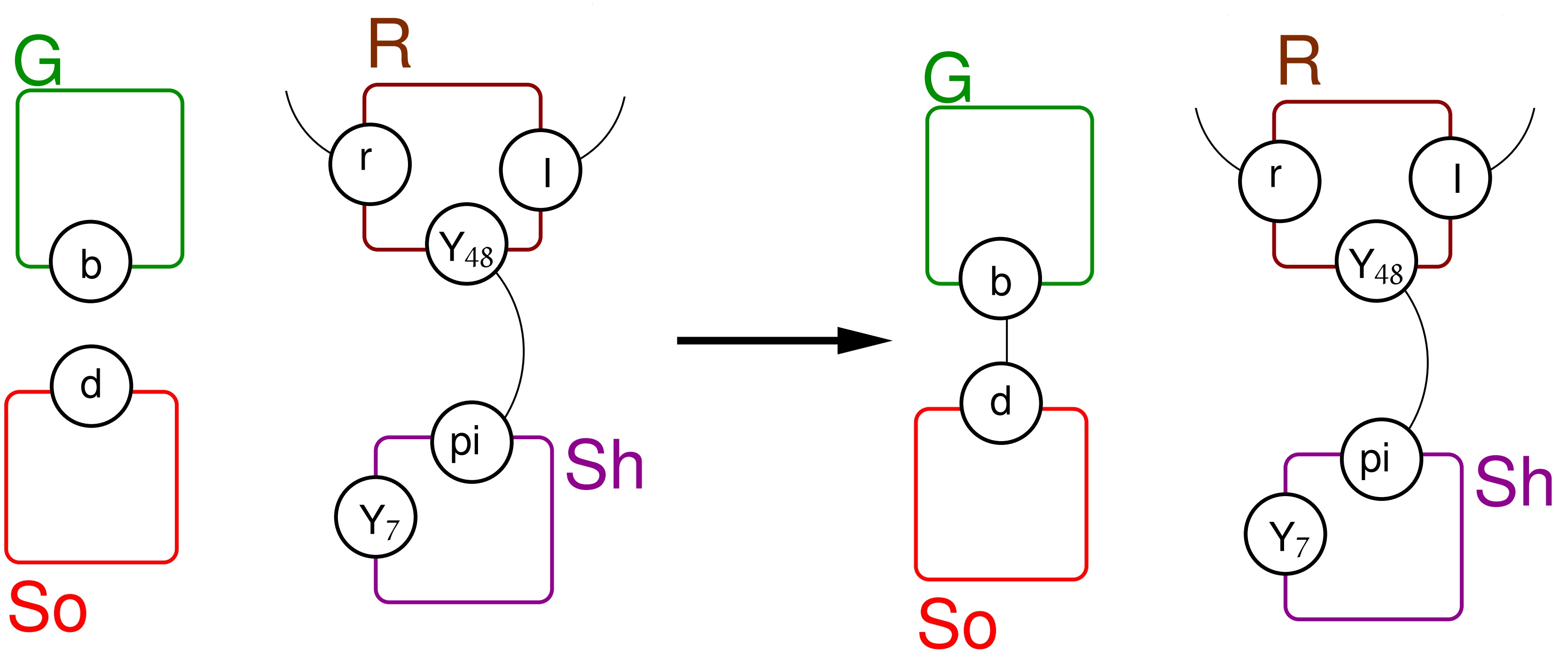
Whenever a fragment intersects a connected component of a lhs on a modified site, then the connected component must be embedded in the fragment!

Fragment consumption
Syntactic criteria



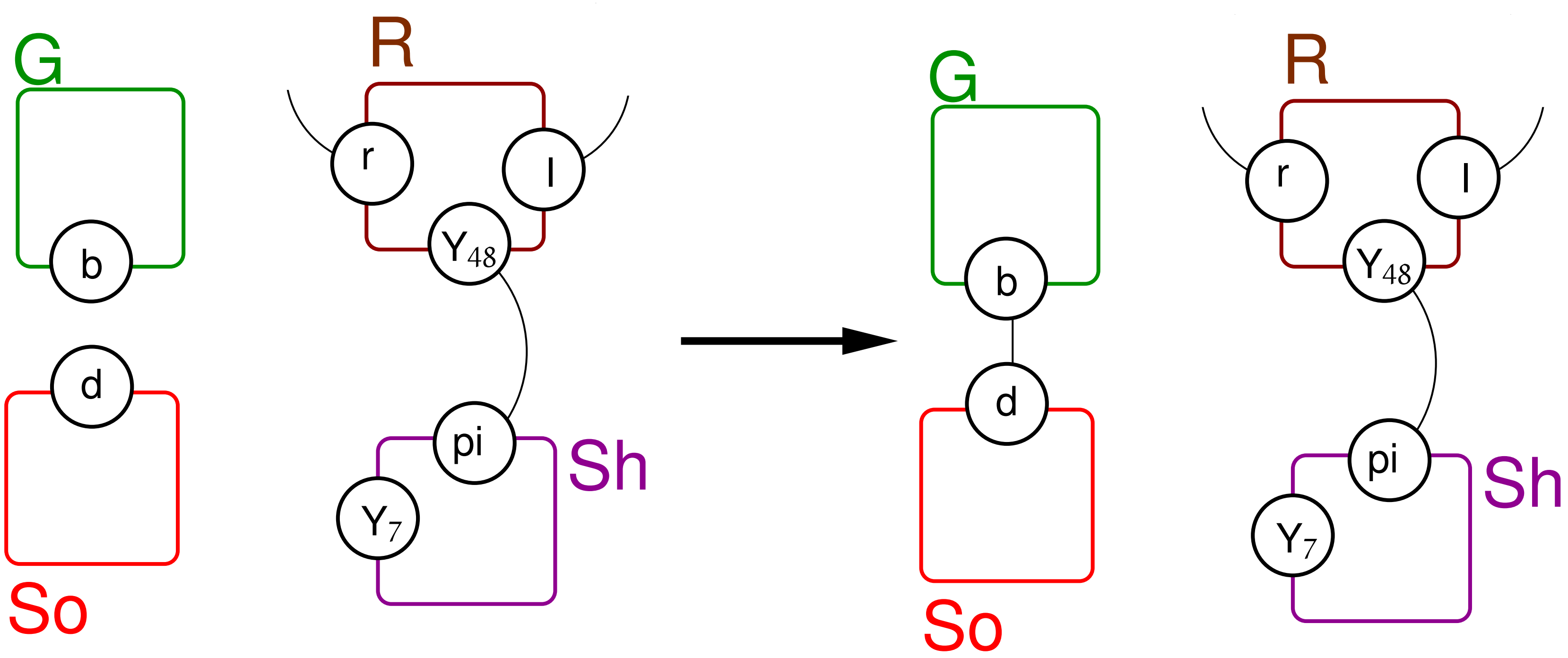
We reflect each path that stems from a modified site (in the lhs of a rule) into the annotated contact map.

Connected components



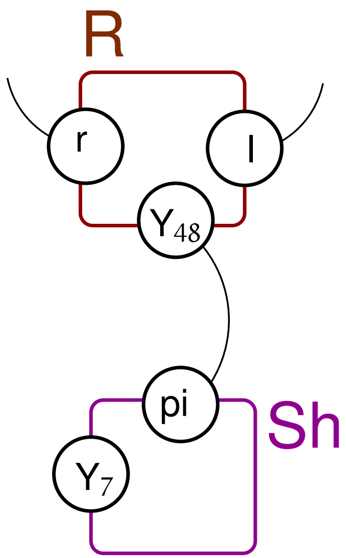
We need to express the “concentration” of any connected component of a lhs with respect to the “concentration” of fragments.

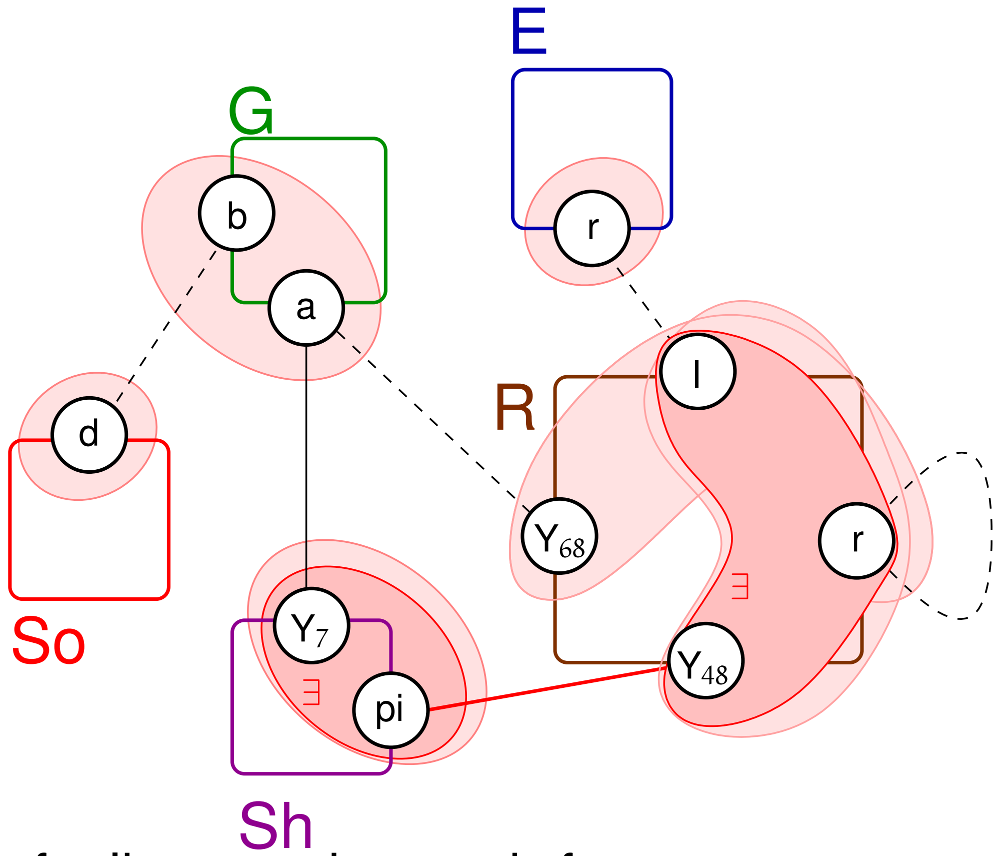
Connected components
Sub-fragment



Each connected component of a lhs must be a sub-fragment.

Connected components
Syntactic criteria





Each connected component of a lhs must be a sub-fragment.

Fragment properties

If:

- an annotated contact map satisfies the syntactic criteria,
- fragments are defined by this annotated contact map,
- we know the concentration of fragments;

then:

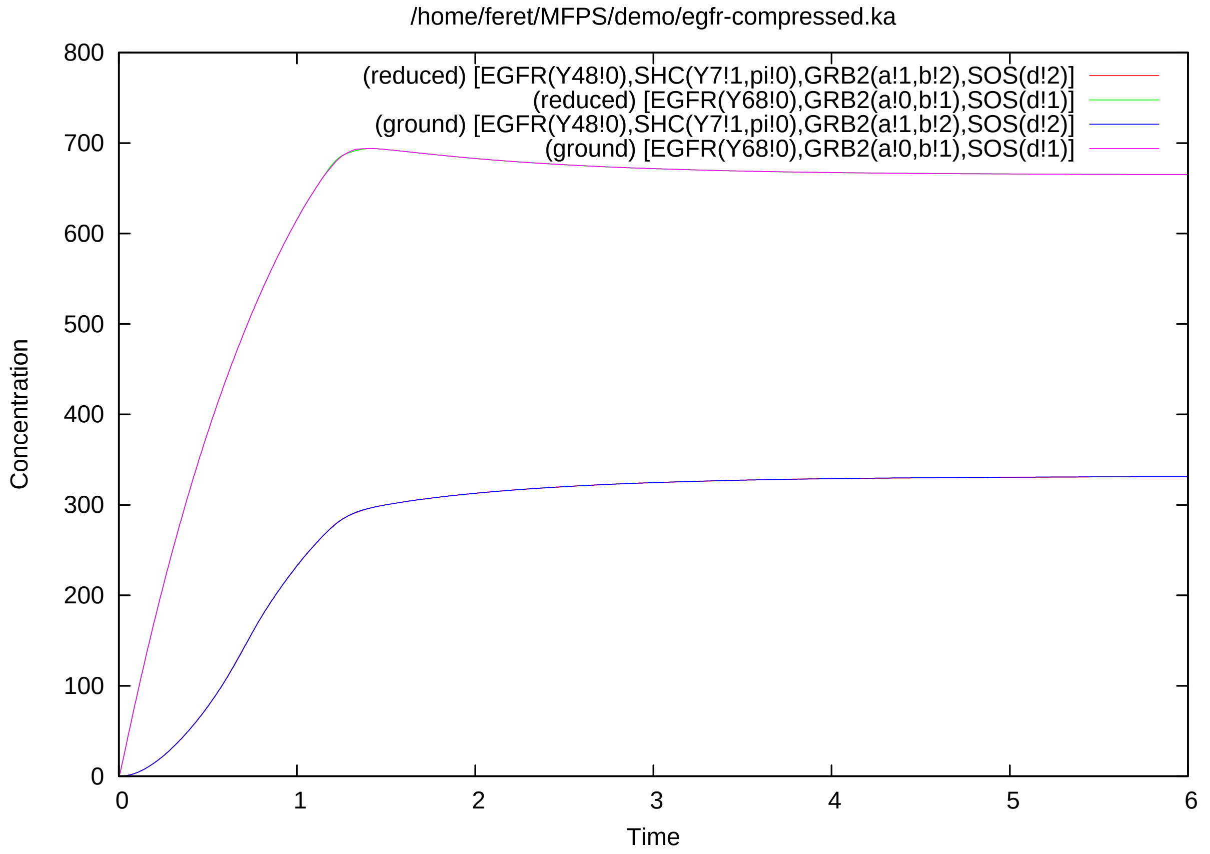
- we can express the concentration of any connected component occuring in lhss,
- we can express fragment proper consumption,
- we can express fragment proper production (eg. see the LICS'2010 paper),
- WE HAVE A CONSTRUCTIVE DEFINITION FOR $\mathbb{F}^{\sharp}$.

Overview

1. Context and motivations
2. Handmade ODEs
3. Abstract interpretation framework
4. Instantiation
5. Conclusion

Experimental results

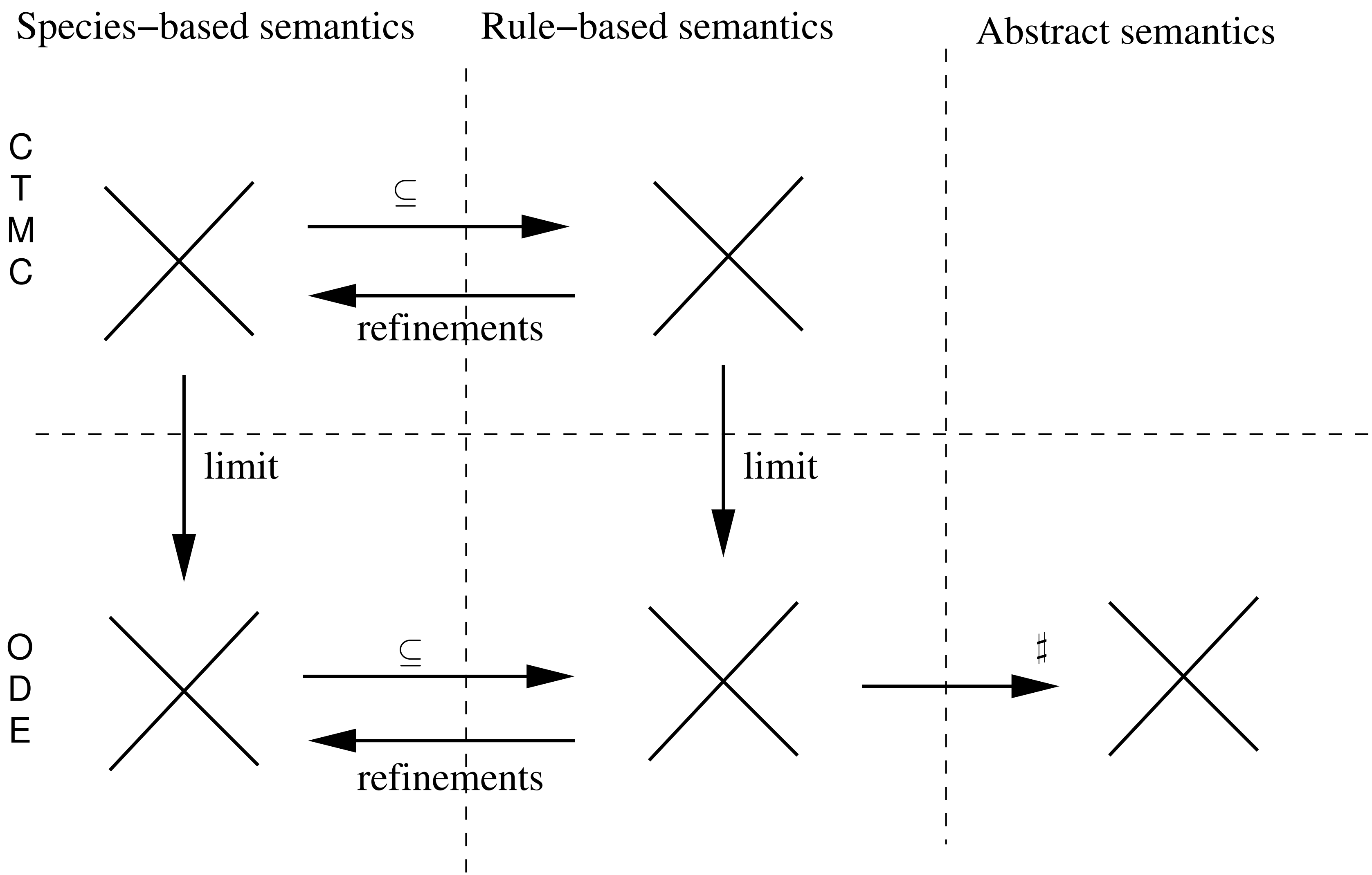
On early egfr, 356 species are simplified into 38 fragments:



Superposition of the ground and the abstract differential semantics.

On a bigger example, $\approx 2 \cdot 10^{19}$ species are simplified into $\approx 2 \cdot 10^{5}$ fragments.

Related issues I: Semantics comparisons



Related issues II:
Semantics approximations

1. ODE approximations:
   - Because of the use of annotated contact map, fragments have a homogeneous structure (or signature).
     Can we design and use heterogeneous fragments ?

   Joint work with Ferdinanda Camporesi (Bologne)
2. Stochastic semantics approximations:
   - Can we design abstraction ?
   - Find the adequate soundness criteria.

   Joint work with Tatjana Petrov and Heinz Koeppl (EPFL)

Announcements

- Call for candidates:

  If you are interested in (at least one) of these issues, there are open positions (Internships, PhD students or Post-doc fellows)...

  ANR-Chair of Excellence: AbstractCell

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

- Call for paper/participation:

  First Workshop on Static Analysis and Systems Biology (SASB 2010)

  (co-chaired with Andre Levchenko)

  13th Sept 2010, Perpignan

  http://www.di.ens.fr/sasb2010