

# MPRI 2.19 Biochemical Programming

Rule-based Modeling

Conservative approximation of polymers

Jérôme Feret

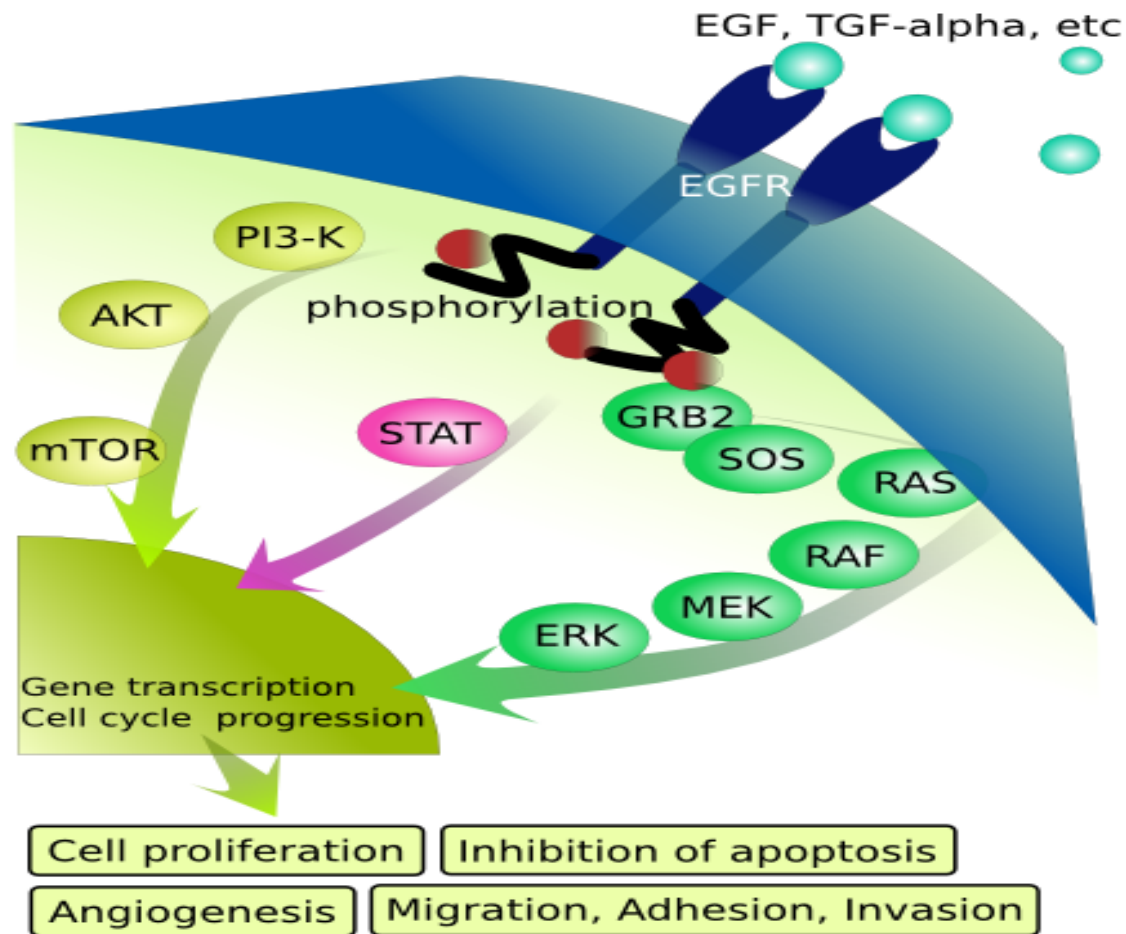
Antique, Inria Paris



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

Tuesday, the 11th of January, 2022

# Signaling Pathways



Eikuch, 2007

# Challenges for computer science

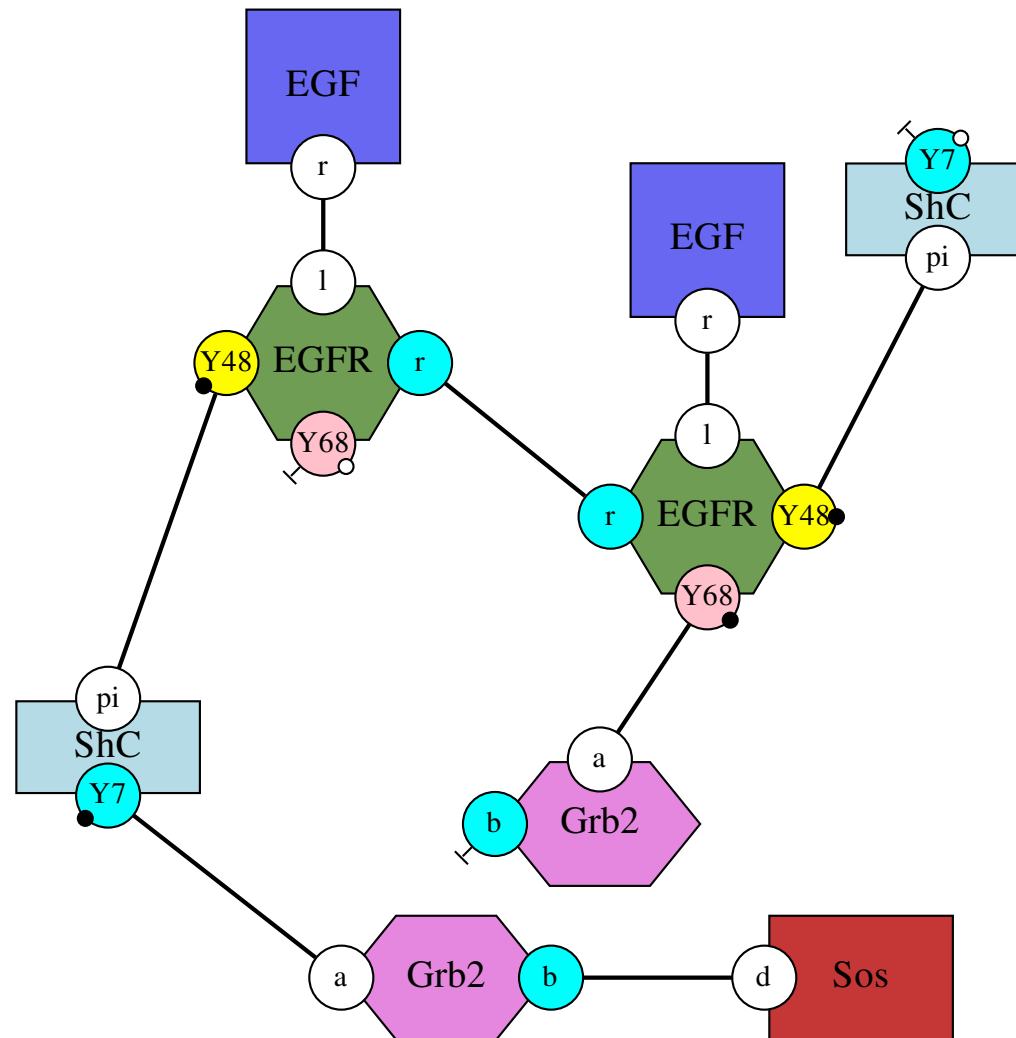
## 1. Break down combinatorial complexity

- Occurrences of proteins can form huge instances of chemical species;
- There may be many (or even infinitely many) different kinds of chemical species.

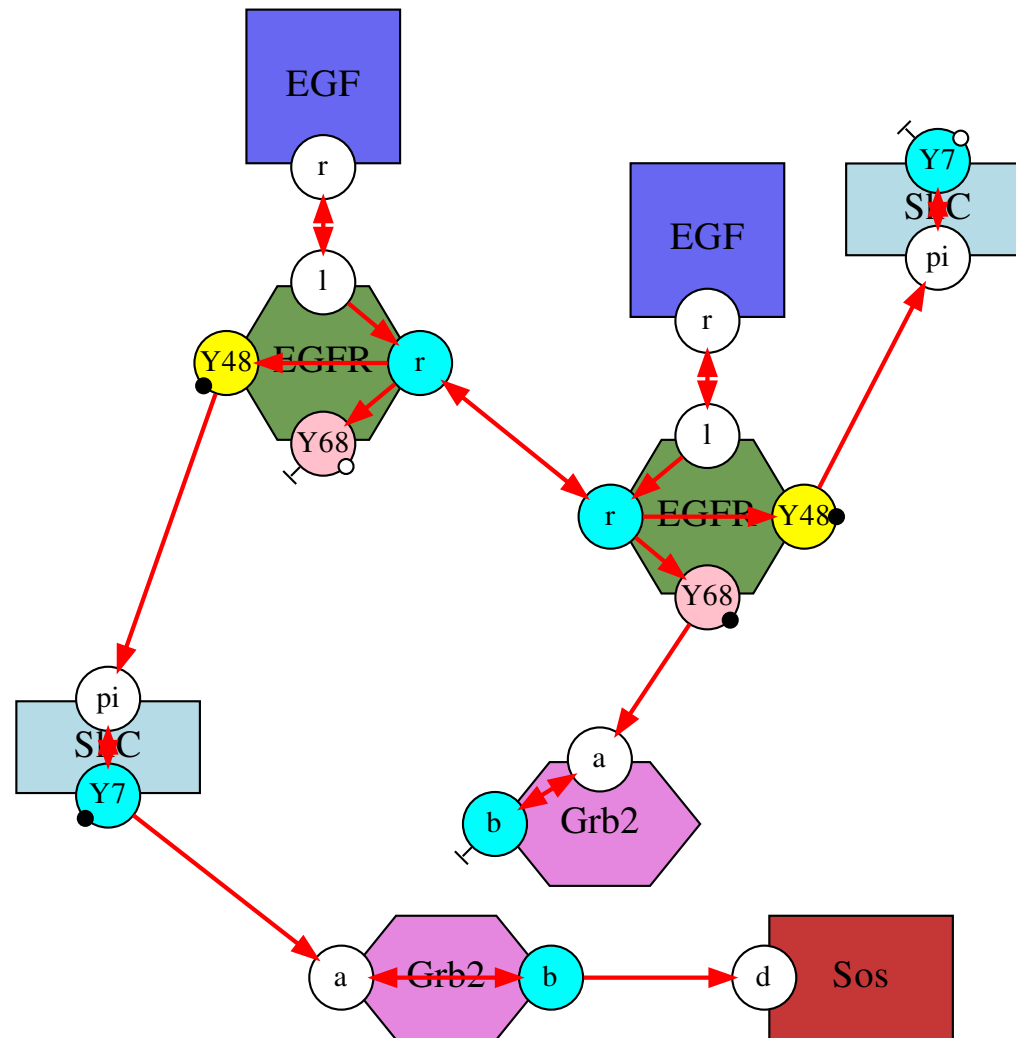
## 2. Understand how collective behaviors emerge from individual interactions.

- races for shared-resources;
- sequestration effects;
- separation of time- and concentration- scales;
- non-linear feedback loops.

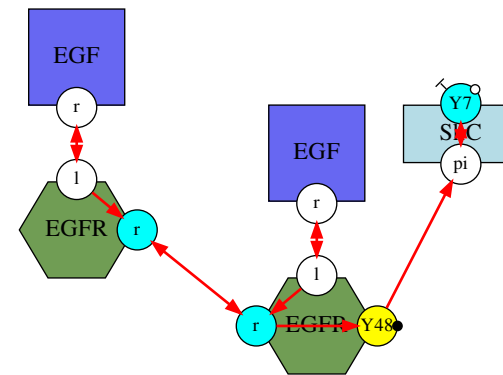
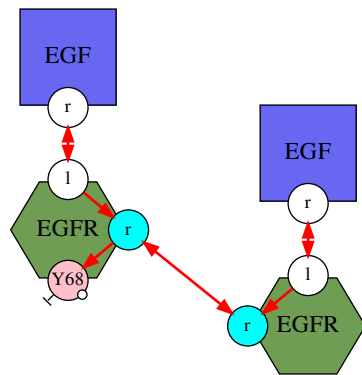
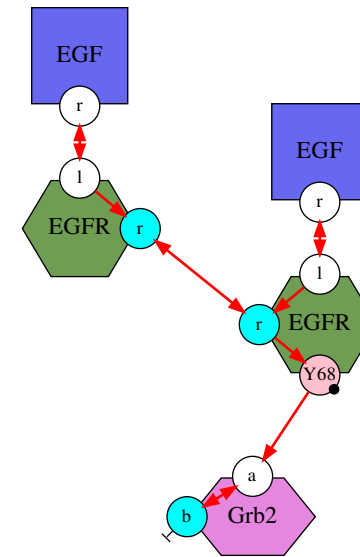
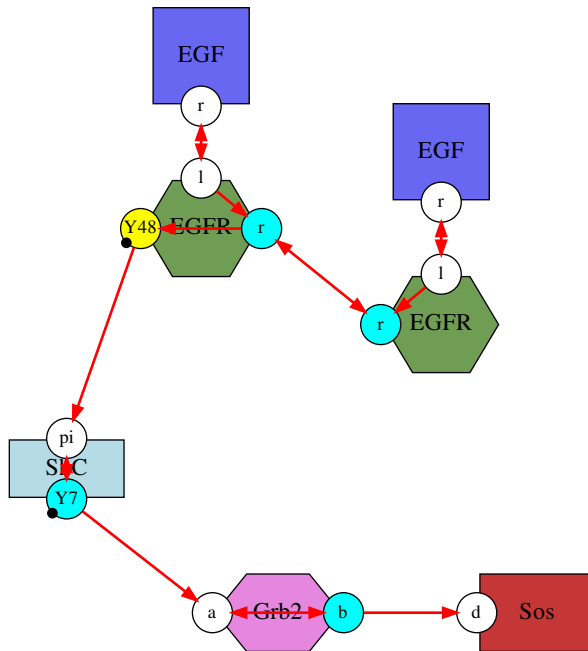
# Model reduction



# Model reduction



# Model reduction



# Exact model reduction

- We can derive an ode to describe the exact evolution of the concentration of patterns of interest.
- The choice of patterns is fixed by the analysis.

The trade-off between complexity and accuracy is imposed by the framework.

# On the menu today

## Conservative model reduction

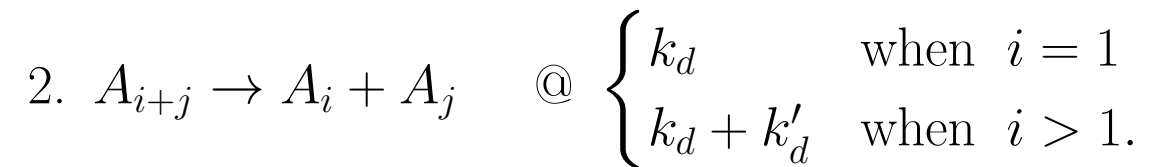
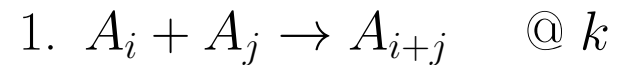
1. Motivating example
2. Evolution systems
3. Box approximation
4. Symbolic reasoning
5. Conclusion



# An example with polymers

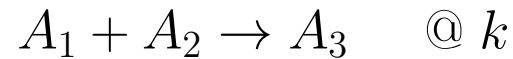
We denote by  $A_n$  a chain of  $n$  proteins.

We consider the following reactions (for  $i, j \geq 1$ ):



# Principle of mass action

The following reaction:



has the following contribution:

$$\begin{cases} \frac{d[A_1]}{dt} = -k \cdot [A_1] \cdot [A_2] \\ \frac{d[A_2]}{dt} = -k \cdot [A_1] \cdot [A_2] \\ \frac{d[A_3]}{dt} = k \cdot [A_1] \cdot [A_2]. \end{cases}$$

# (Infinite) system of odes

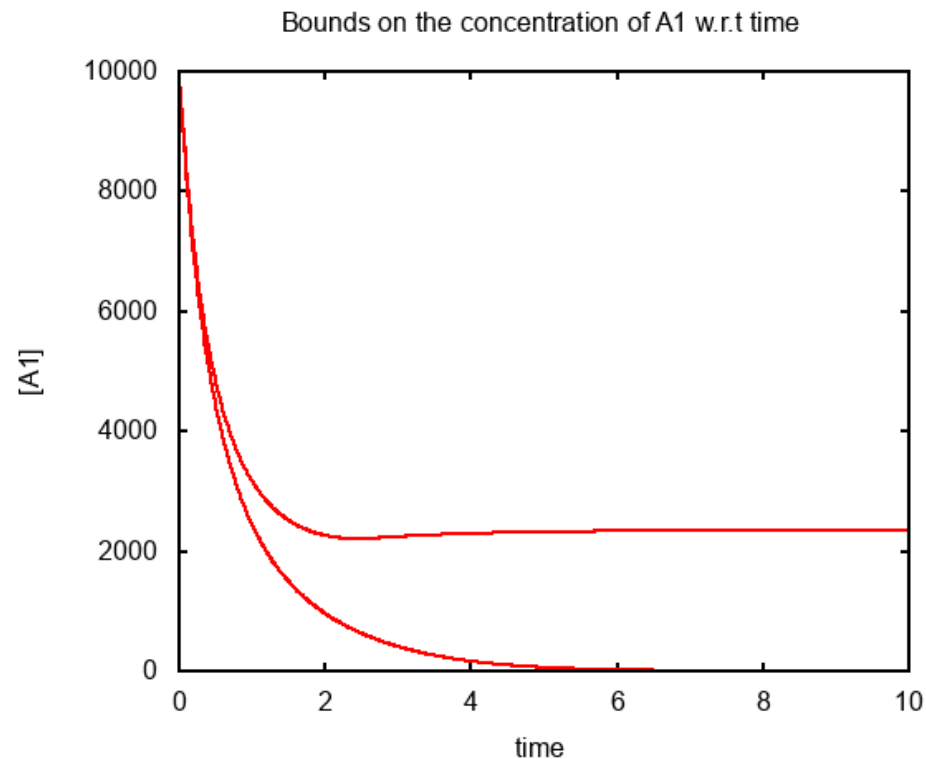
$$\frac{d[A_n]}{dt} = t_1^+(n) + t_2^+(n) + t_3^+(n) - t_1^-(n) - t_2^-(n) - t_3^-(n)$$

where:

$$\begin{aligned} t_1^+(n) &\triangleq k \cdot \sum_{i+j=n} [A_i] \cdot [A_j]; \\ t_2^+(n) &\triangleq 2 \cdot k_d \cdot \sum_{i=n+1}^{+\infty} [A_i]; \\ t_3^+(n) &\triangleq \begin{cases} k'_d \cdot \sum_{i=3}^{+\infty} [A_i] & \text{if } n = 1, \\ k'_d \cdot \sum_{i=n}^{+\infty} ([A_{i+1}] + [A_{i+2}]) & \text{if } n \geq 2; \end{cases} \\ t_1^-(n) &\triangleq 2 \cdot k \cdot [A_n] \cdot \sum_{i=1}^{+\infty} [A_i]; \\ t_2^-(n) &\triangleq k_d \cdot (n-1) \cdot [A_n]; \\ t_3^-(n) &\triangleq \begin{cases} k'_d \cdot (n-2) \cdot [A_n] & \text{if } n \geq 3, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

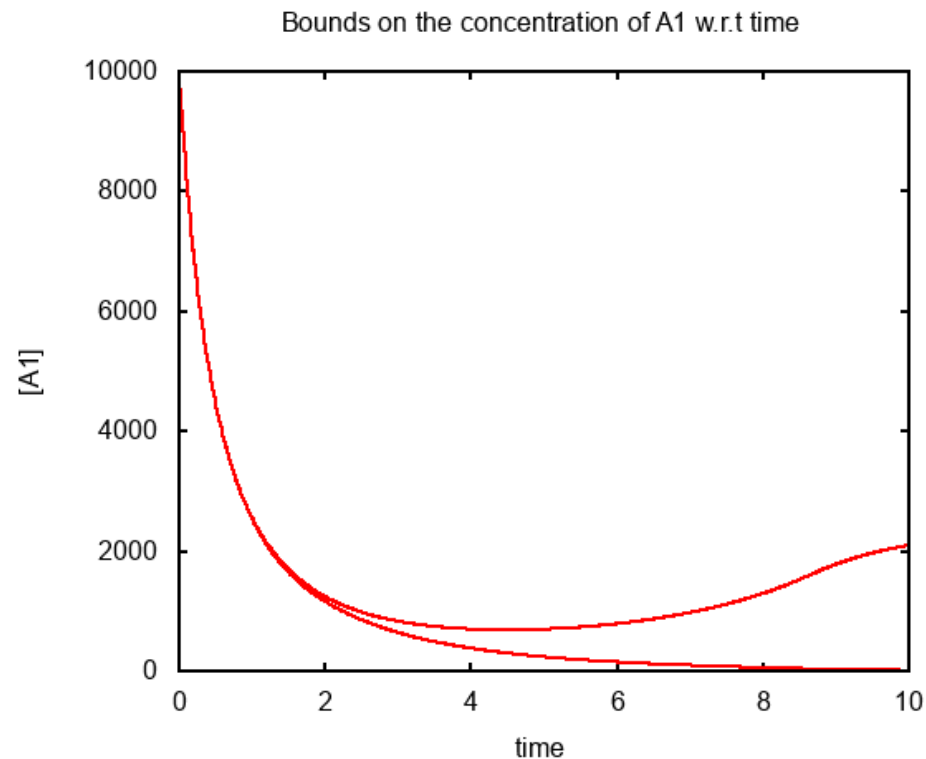
with the side condition:  $\sum_{n \in \mathbb{N}} n \cdot [A_n] < +\infty$ .

# Our goal: Bounding the concentration of $A_1$



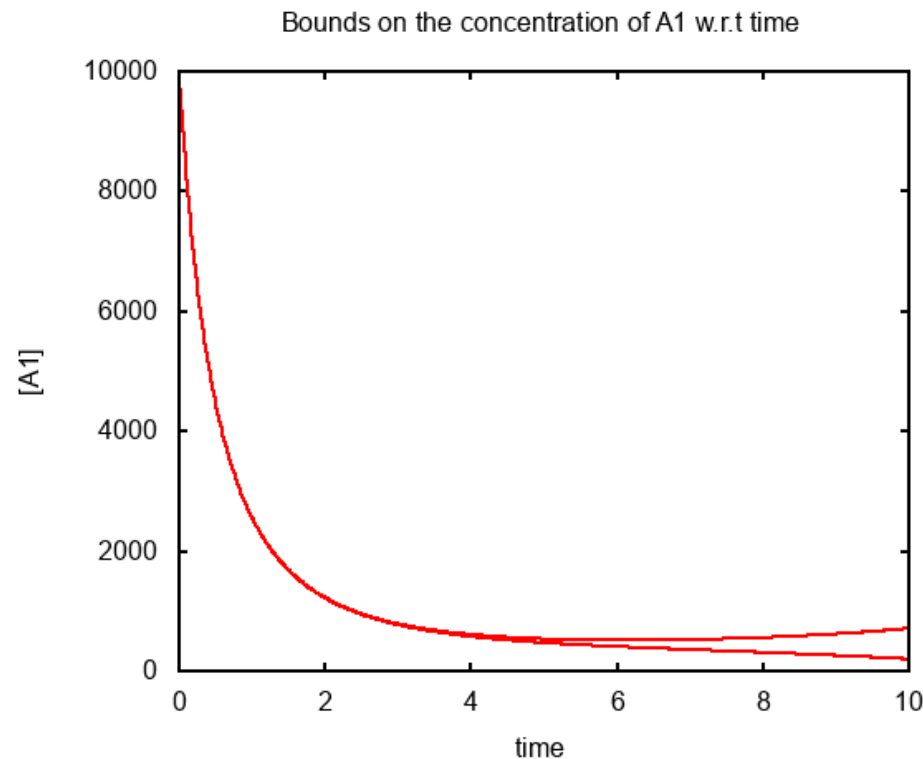
Obtained thanks to an odes of 18 variables.  
(with parameters  $[A_1]_0 = 10000$ ,  $k = 10^{-4}$ ,  $k_d = 10^{-2}$ , and  $k'_d = 10^{-1}$ ).

# Our goal: Bounding the concentration of $A_1$



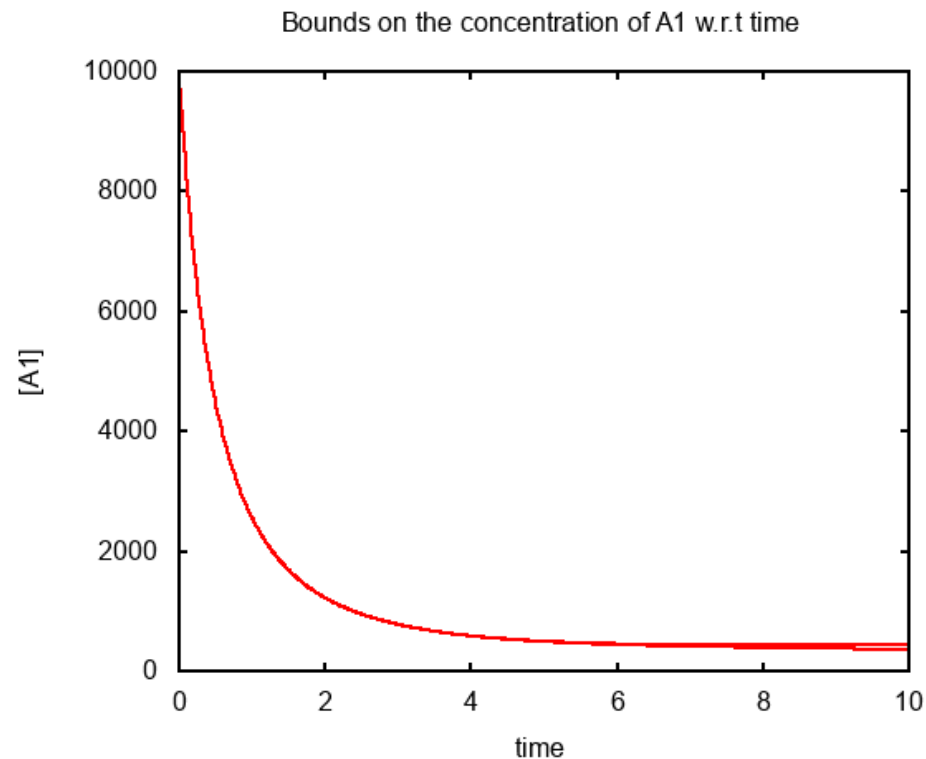
Obtained thanks to an odes of 36 variables.  
(with parameters  $[A_1]_0 = 10000$ ,  $k = 10^{-4}$ ,  $k_d = 10^{-2}$ , and  $k'_d = 10^{-1}$ ).

# Our goal: Bounding the concentration of $A_1$



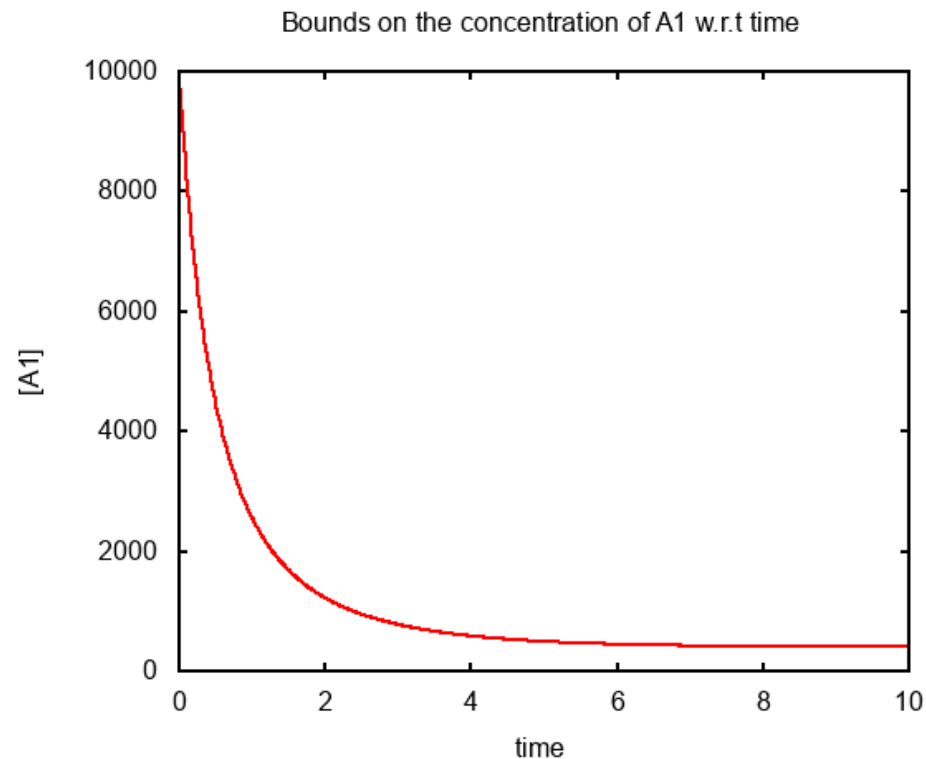
Obtained thanks to an odes of 54 variables.  
(with parameters  $[A_1]_0 = 10000$ ,  $k = 10^{-4}$ ,  $k_d = 10^{-2}$ , and  $k'_d = 10^{-1}$ ).

# Our goal: Bounding the concentration of $A_1$



Obtained thanks to an odes of 72 variables.  
(with parameters  $[A_1]_0 = 10000$ ,  $k = 10^{-4}$ ,  $k_d = 10^{-2}$ , and  $k'_d = 10^{-1}$ ).

# Our goal: Bounding the concentration of $A_1$



Obtained thanks to an odes of 90 variables.  
(with parameters  $[A_1]_0 = 10000$ ,  $k = 10^{-4}$ ,  $k_d = 10^{-2}$ , and  $k'_d = 10^{-1}$ ).



# Approach

1. Use a **high level language** to:
  - (a) describe the model;
  - (b) show the existence and uniqueness of the solution;
  - (c) **reason symbolically** about some differentiable auxiliary variables:
    - express their derivatives,
    - infer inequalities among these derivatives.
2. Use **box approximation** to define a system of odes with two variables per auxiliary variable (one for the lower bound, one for the upper bound)  
(error bounds are computed *a posteriori*).

# Main advantages

- Numerical approximations relax constraints on the choice of auxiliary variables.
- The choice of auxiliary variables can be driven by various methods:
  - truncation (this talk);
  - tropicalization (see Andreea Beica's PhD);
  - flow of information(a unifying framework to combine these methods)
- The approach provides error bounds.

# On the menu today

## Conservative model reduction

1. Motivating example
2. Evolution systems
3. Box approximation
4. Symbolic reasoning
5. Conclusion

# Definition

In a Banach space, a system defined as:

$$\frac{dX}{dt} = F(X, t) + G(X, t)$$

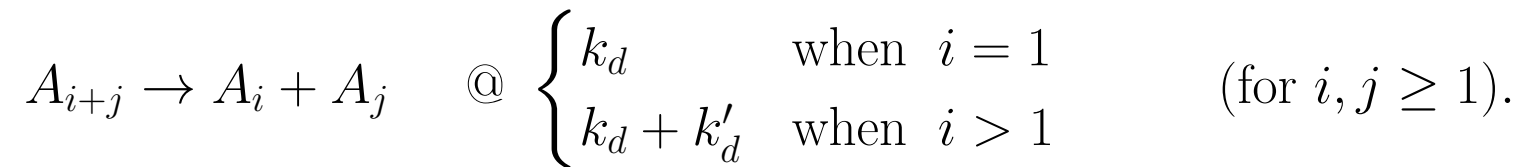
where:

1.  $F$  is linear and triangular;
2.  $G$  is Lipschitz on every bounded set;

is called an evolution system.

# Case study

- Define the norm of a state as  $\sum n \cdot |A_n|$ .
- Define  $F$  as the contribution of the unbinding reactions:



- Define  $G$  as the contribution of the binding reactions:



# Properties

Evolutions systems:

- they have **unique** maximal solutions;
- their maximal solutions are **locally Lipschitz**.
- whenever a maximal solution is **not defined over  $\mathbb{R}^+$ , the norm diverges**;
- whenever  $G$  is  $\mathcal{C}_1$  and its derivative is bounded on bounded sets, maximal solutions are also  $\mathcal{C}_1$ .

[Hundertmark et al., Operator Semigroups and Dispersive Equations, 16th Internet Seminar on Evolution Equations 2013]

# On the menu today

## Conservative model reduction

1. Motivating example
2. Evolution systems
3. Box approximation
4. Symbolic reasoning
5. Conclusion

# Principle

Given a finite system of odes:

$$\frac{dX}{dt} = F(X, t).$$

Box approximation:

1. approximate the state of the system by a (hyper)-box  
(twice many variables as in the initial system)
2. associate to each (hyper)-face an expression that bounds conservatively the partial derivative of the system with respect to the corresponding variable over this (hyper)-face.

Sound whenever  $F$  is locally Lipschitz w.r.t to the state and continuous w.r.t time.

[M. Kirkilionis and S. Walcher, On comparison systems for ordinary differential equations, J. Math. Anal. Appl. 299 (2004)]



# Example: odes

Consider the following system of odes:

$$\begin{cases} \frac{dx}{dt} = y \cdot (2 - \cos(y)) - x \cdot (2 - \sin(y)) \\ \frac{dy}{dt} = x \cdot (2 - \cos(y)) - y \cdot (2 - \sin(y)) \\ x(0) = y(0) = 1 \end{cases}$$

# Example: Invariants

Consider the following system of odes:

$$\begin{cases} \frac{dx}{dt} = y \cdot (2 - \cos(y)) - x \cdot (2 - \sin(y)) \\ \frac{dy}{dt} = x \cdot (2 - \cos(y)) - y \cdot (2 - \sin(y)) \\ x(0) = y(0) = 1 \end{cases}$$

We have:

$$\begin{cases} y - 3 \cdot x \leq \frac{dx}{dt} \leq 3 \cdot y - x \\ x - 3 \cdot y \leq \frac{dy}{dt} \leq 3 \cdot x - y. \end{cases}$$

# Example:

## Box approximation

Thus, the following system of odes:

$$\begin{cases} \frac{dx}{dt} = y \cdot (2 - \cos(y)) - x \cdot (2 - \sin(y)) \\ \frac{dy}{dt} = x \cdot (2 - \cos(y)) - y \cdot (2 - \sin(y)) \\ x(0) = y(0) = 1 \end{cases}$$

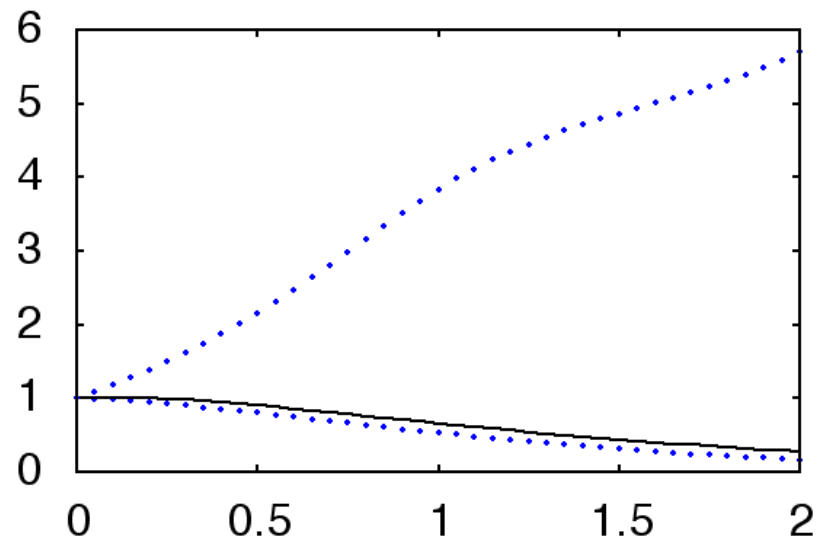
can be safely approximated by the following one:

$$\begin{cases} \frac{d\underline{x}}{dt} = \underline{y} - 3 \cdot \underline{x} \\ \frac{d\bar{x}}{dt} = 3 \cdot \bar{y} - \bar{x} \\ \frac{d\underline{y}}{dt} = \underline{x} - 3 \cdot \underline{y} \\ \frac{d\bar{y}}{dt} = 3 \cdot \bar{x} - \bar{y} \\ \underline{x}(0) = \bar{x}(0) = \underline{y}(0) = \bar{y}(0) = 1 \end{cases}$$

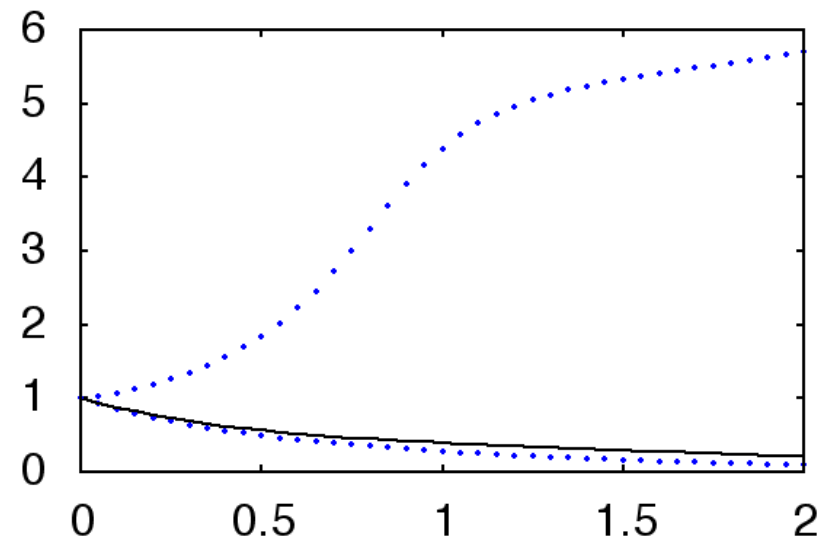
# Example:

## Numerical results

$x(t)$  with respect to  $t$



$y(t)$  with respect to  $t$

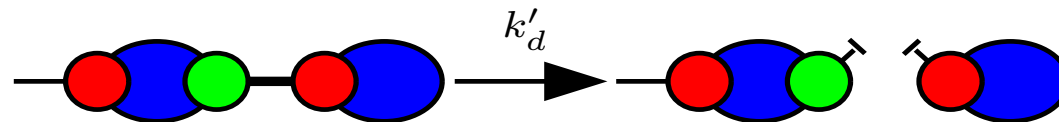
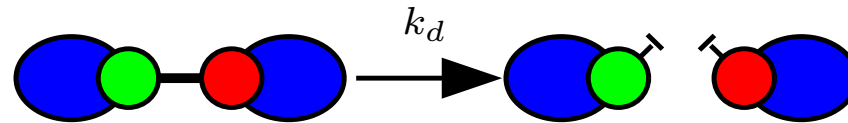
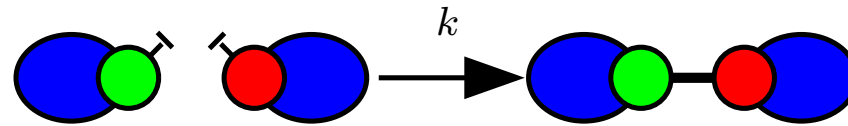


# On the menu today

## Conservative model reduction

1. Motivating example
2. Evolution systems
3. Box approximation
4. Symbolic reasoning
5. Conclusion

# Our model of polymers in Kappa



# Back to evolution systems

Let us consider both following definitions:

1. A rule is **dispersive** if it is unary and it splits its pattern into smaller ones.
2. A rule is **locally Lipschitz** if for every pattern  $P$  the number of embedding from the patterns in the lhs of the rules and the pattern  $P$  is uniformly bounded.

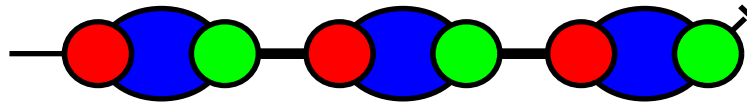
Then: A **finite** set of Kappa rules such that **every rule** is **either dispersive, or locally Lipschitz** (or both) induces **an evolution system**, for the norm defined as the overall concentration of proteins.

# Auxiliary variables

Connected patterns have both an intensional and an extensional meaning.

A connected pattern may be seen:

1. as a connected graph:



2. as a multi-set (potentially infinite) of fully specified connected graphs:

$$[-A_3\vdash] = \sum_{n \geq 4} [\vdash A_n \vdash]$$

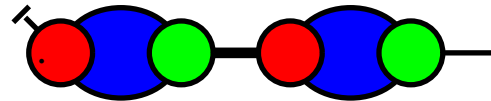
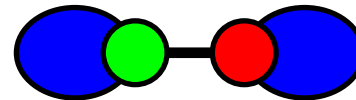


# List of connected patterns

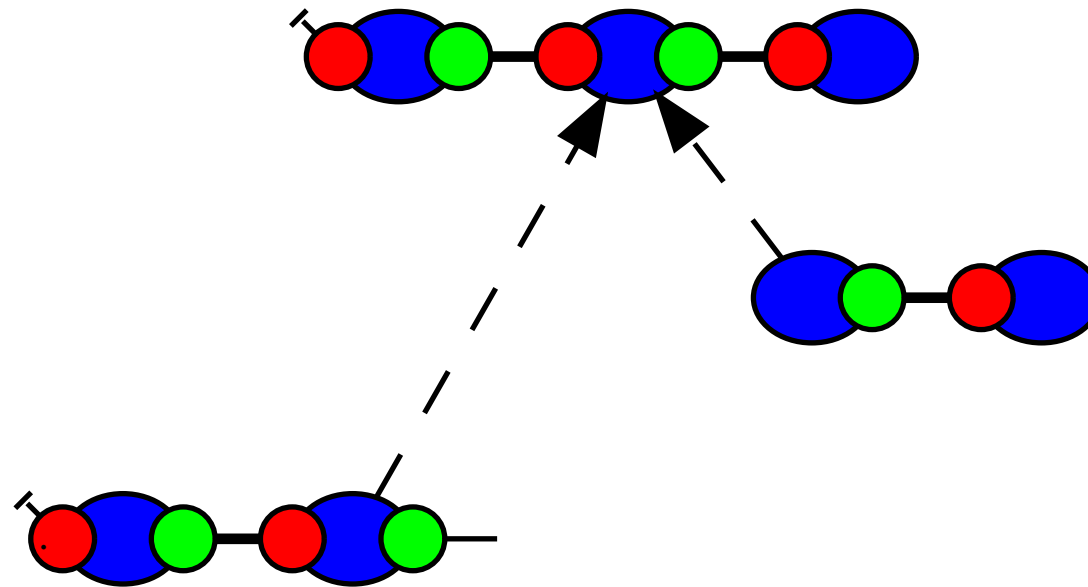
- $[\vdash A_n \dashv]$ : concentration of polymer of length  $n$ ;
- $[\vdash A_n]$ : concentration of polymer of length at least  $n$ ;
- $[A_n \dashv]$ : concentration of polymer of length at least  $n$ ;
- $[\vdash A_n -]$ : concentration of polymer of length at least  $n + 1$ ;
- $[-A_n \dashv]$ : concentration of polymer of length at least  $n + 1$ ;
- $[A_n] = \sum_{i \in \mathbb{N}} (i + 1) \cdot [\vdash A_{n+i} \dashv]$ ;
- $[A_n -] = [A_{n+1}]$ ;
- $[-A_n] = [A_{n+1}]$ ;
- $[-A_n -] = [A_{n+2}]$ .

# Overlap between connected patterns

.

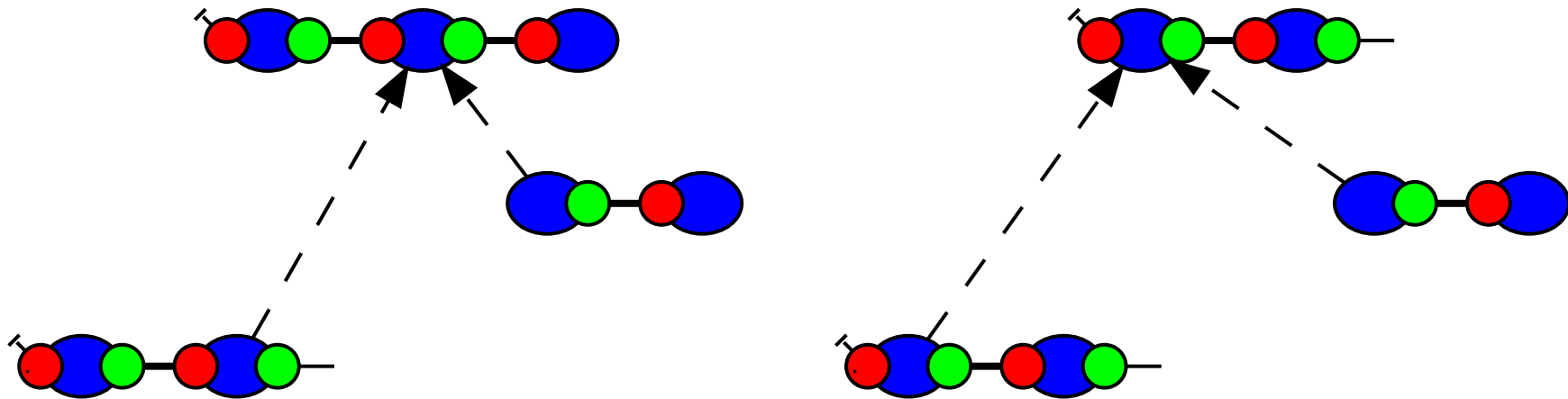


# Overlap between connected patterns

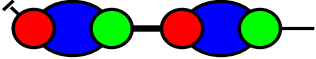
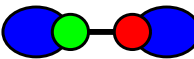




# Overlap between connected patterns

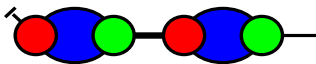
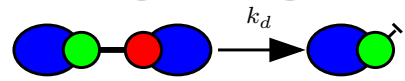
There may be several overlaps between two patterns.

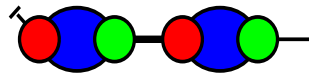


# Consumption

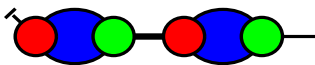
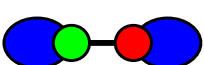
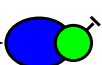
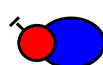
Which quantity of the pattern  is consumed due to the rule   $\xrightarrow{k_d}$    ?

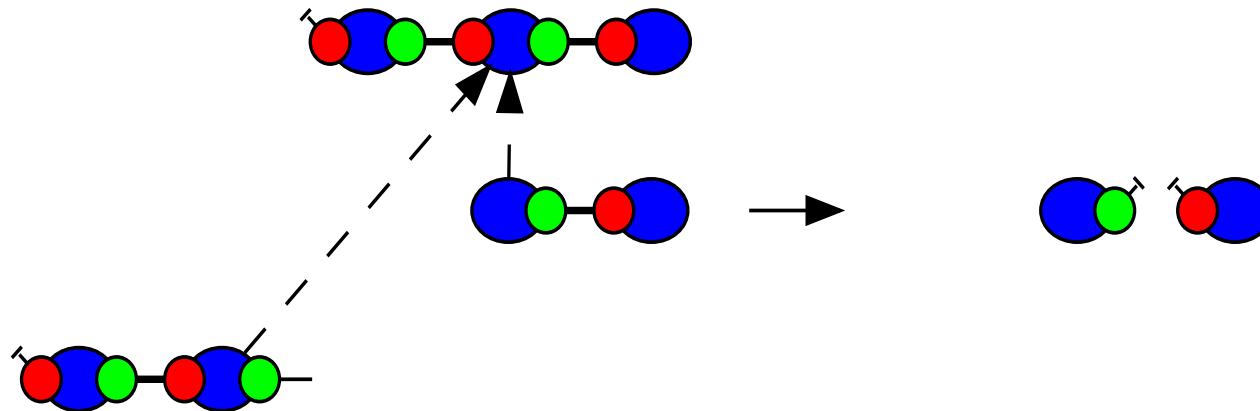
# Consumption

Which quantity of the pattern  is consumed due to the rule ?

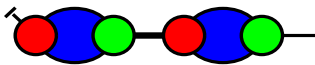
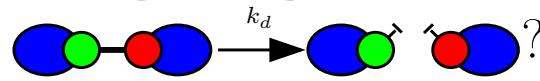


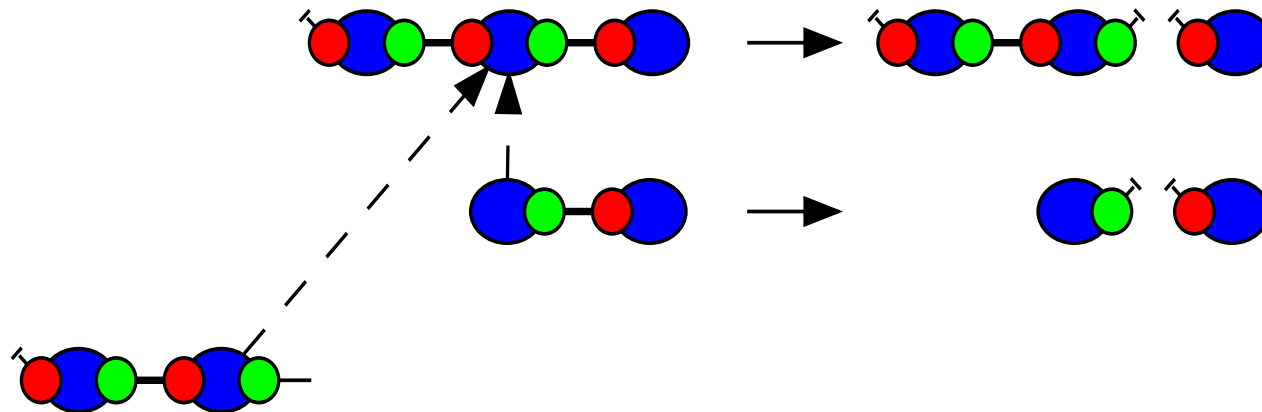
# Consumption

Which quantity of the pattern  is consumed due to the rule   $\xrightarrow{k_d}$    ?



# Consumption

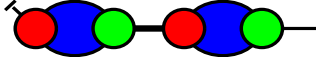
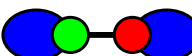
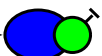
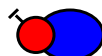
Which quantity of the pattern  is consumed due to the rule ?

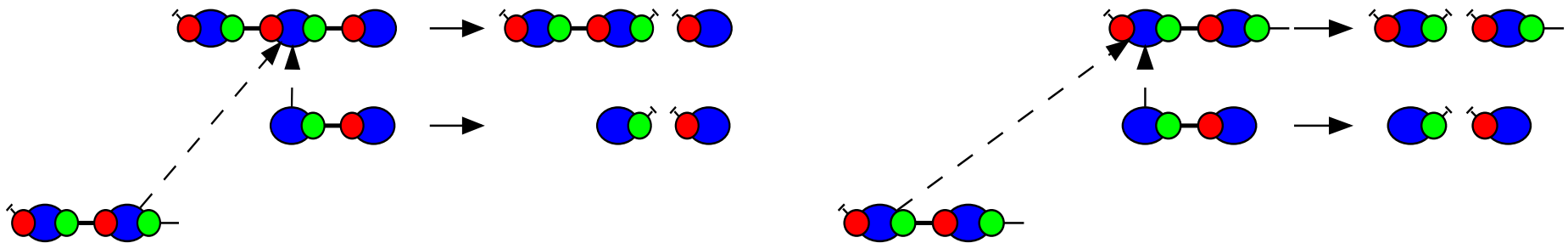


$$k_d \cdot [\vdash A_3]$$



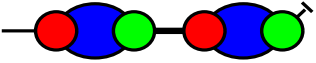
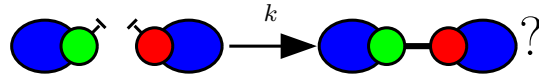
# Consumption

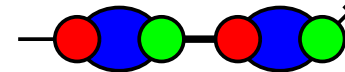
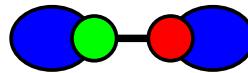
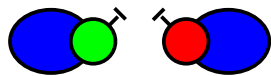
Which quantity of the pattern  is consumed due to the rule   $\xrightarrow{k_d}$   ?



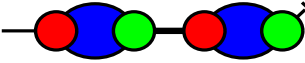
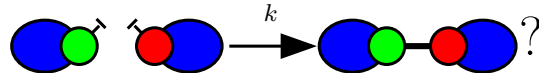
$$k_d \cdot ([\vdash A_3] + [\vdash A_2 -]).$$

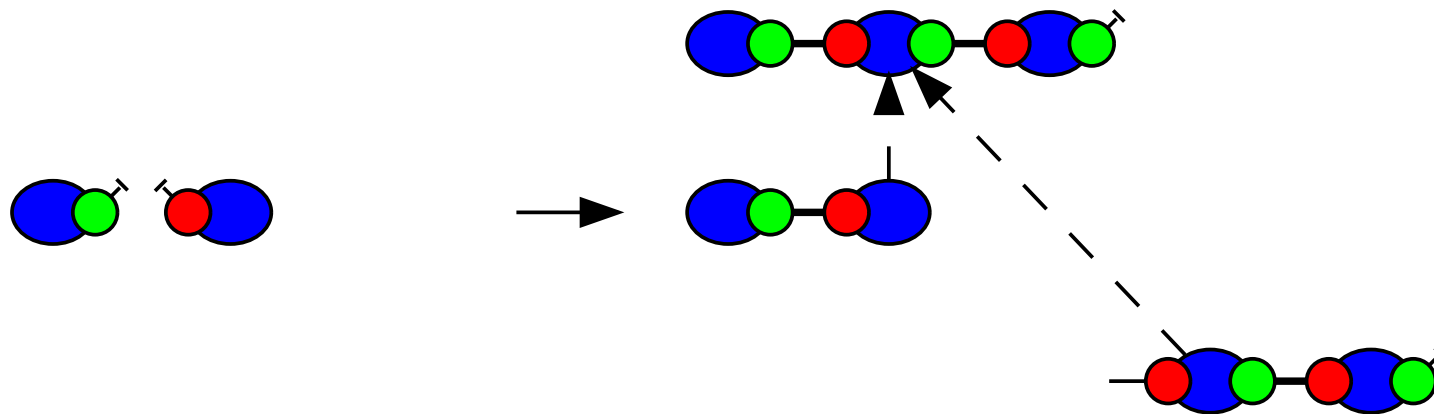
# Production

Which quantity of the pattern  is produced due to the rule ?

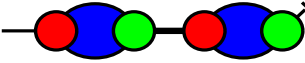
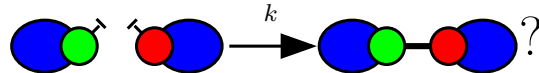


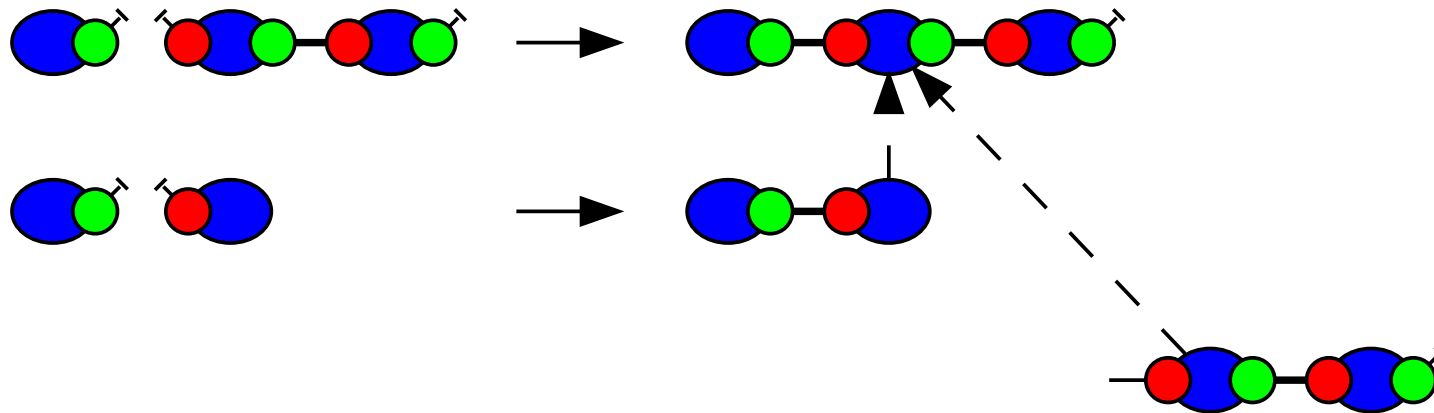
# Production

Which quantity of the pattern  is produced due to the rule .



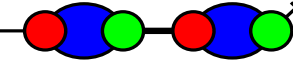
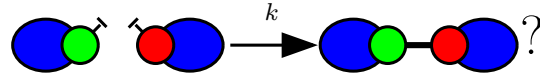
# Production

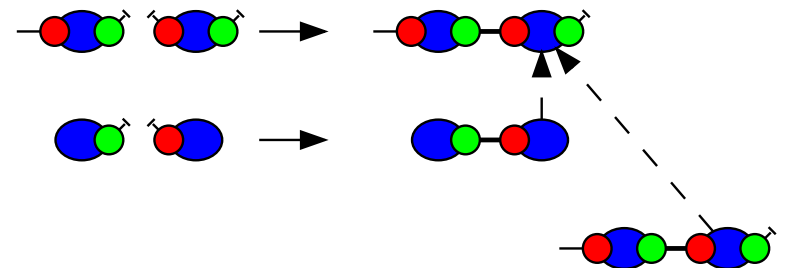
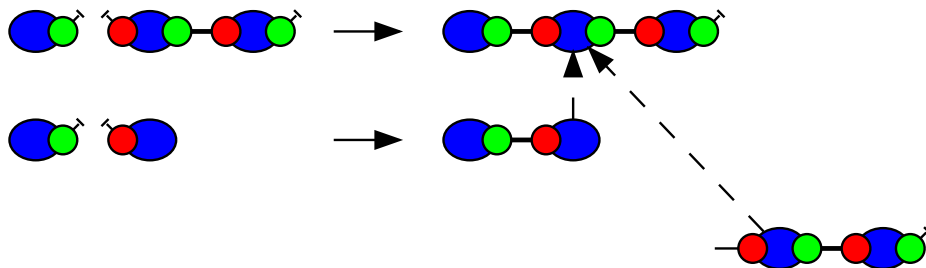
Which quantity of the pattern  is produced due to the rule ?



$$k \cdot [A_1 \dashv] \cdot [\vdash A_2 \dashv].$$

# Production

Which quantity of the pattern  is produced due to the rule .



$$k \cdot ([A_1 \dashv] \cdot [\vdash A_2 \dashv] + [-A_1 \dashv] \cdot [\vdash A_1 \dashv]).$$

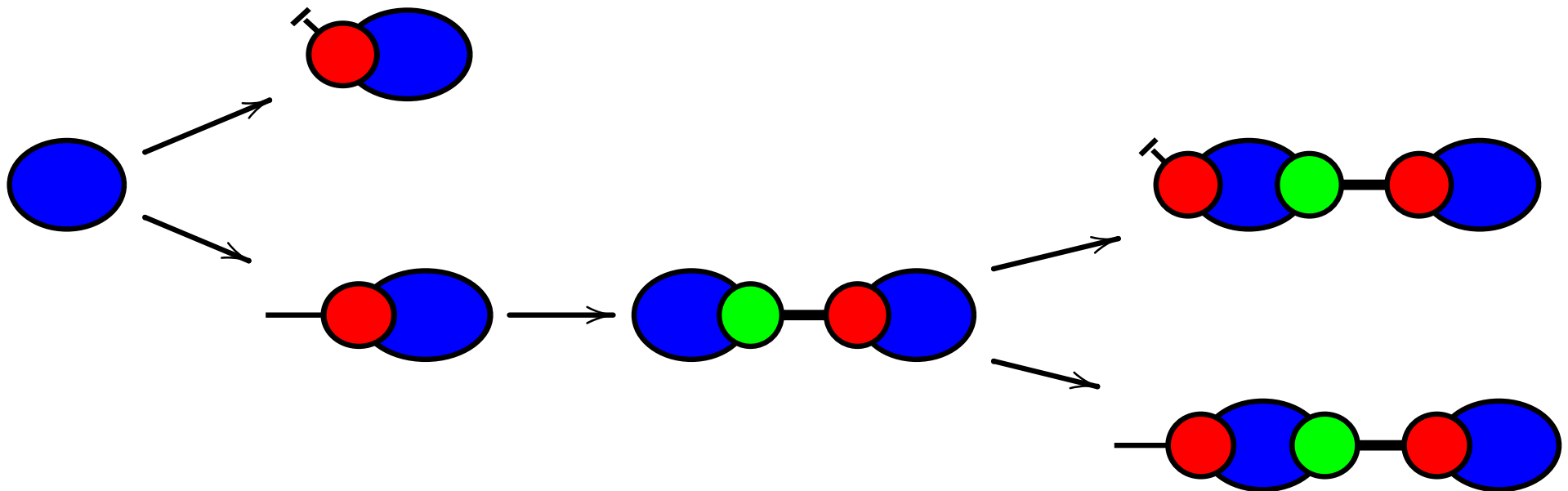
# Exact derivatives

$$\frac{d[\vdash A_n \dashv]}{dt} = \mathcal{T}_{1,1}^+(n) + \mathcal{T}_{1,2}^+(n) + \mathcal{T}_{1,3}^+(n) - \mathcal{T}_{1,1}^-(n) - \mathcal{T}_{1,2}^-(n) - \mathcal{T}_{1,3}^-(n)$$

where:

$$\begin{aligned}\mathcal{T}_{1,1}^+(n) &\triangleq k \cdot \sum_{i+j=n} [\vdash A_i \dashv] \cdot [\vdash A_j \dashv]; \\ \mathcal{T}_{1,2}^+(n) &\triangleq k_d \cdot ([\vdash A_{n+1}] + [A_{n+1} \dashv]); \\ \mathcal{T}_{1,3}^+(n) &\triangleq \begin{cases} k'_d \cdot [-A_{n+1} \dashv] & \text{if } n = 1 \\ k'_d \cdot ([-A_{n+1} \dashv] + [\vdash A_{n+1}]) & \text{if } n \geq 2; \end{cases} \\ \mathcal{T}_{1,1}^-(n) &\triangleq k \cdot [\vdash A_n \dashv] \cdot ([\vdash A_1] + [A_1 \dashv]); \\ \mathcal{T}_{1,2}^-(n) &\triangleq k_d \cdot (n - 1) \cdot [\vdash A_n \dashv]; \\ \mathcal{T}_{1,3}^-(n) &\triangleq \begin{cases} k'_d \cdot (n - 2) \cdot [\vdash A_n \dashv] & \text{if } n \geq 3 \\ 0 & \text{otherwise.} \end{cases}\end{aligned}$$

# Linear equalities among auxiliary variables



# Inequalities

$$1. \quad [\vdash A_n \dashv] \leq \frac{[A_1] - \sum_{k=1, k \neq n}^N k \cdot [\vdash A_k \dashv]}{n};$$

$$2. \quad [A_n] \leq [A_1] - \sum_{k=1}^{n-1} k \cdot [\vdash A_k \dashv];$$

$$3. \quad [A_n \dashv] \leq \frac{[A_1] - \sum_{k=1}^{n-1} k \cdot [\vdash A_k \dashv]}{n};$$

$$4. \quad [\vdash A_n] \leq \frac{[A_1] - \sum_{k=1}^{n-1} k \cdot [\vdash A_k \dashv]}{n}.$$

$$5. \quad [-A_n \diamond] = [A_{n+1} \diamond] \quad \forall \diamond \in \{\dashv, -, \varepsilon\};$$

$$6. \quad [\diamond A_n -] = [\diamond A_{n+1}] \quad \forall \diamond \in \{\dashv, -, \varepsilon\}.$$



# The odes

$$\frac{d[\vdash A_1 -]}{dt} = \underline{t_{2,1}^+} + \underline{t_{2,2}^+} + \underline{t_{2,3}^+} - \overline{t_{2,1}^-} - \overline{t_{2,2}^-} - \overline{t_{2,3}^-},$$

where:

$$\underline{t_{2,1}^+} \triangleq k \cdot [\vdash A_1 \dashv] \cdot ([\vdash A_1 \dashv] + [\vdash A_1 -]);$$

$$\underline{t_{2,2}^+} \triangleq k_d \cdot \underline{[-A_1 -]};$$

$$\underline{t_{2,3}^+} \triangleq k'_d \cdot \underline{[-A_2 -]};$$

$$\overline{t_{2,1}^-} \triangleq k \cdot \underline{[\vdash A_1 -]} \cdot \left( \min \left( \underline{[\vdash A_1 -]}, \overline{[A_1]} - \sum_{n=2}^N n \cdot \underline{[\vdash A_n \dashv]} \right) + \min \left( \overline{[-A_1 \dashv]}, \overline{[A_1]} - \sum_{n=2}^N n \cdot \underline{[\vdash A_n \dashv]} \right) \right);$$

$$\overline{t_{2,2}^-} \triangleq k_d \cdot \underline{[\vdash A_1 -]};$$

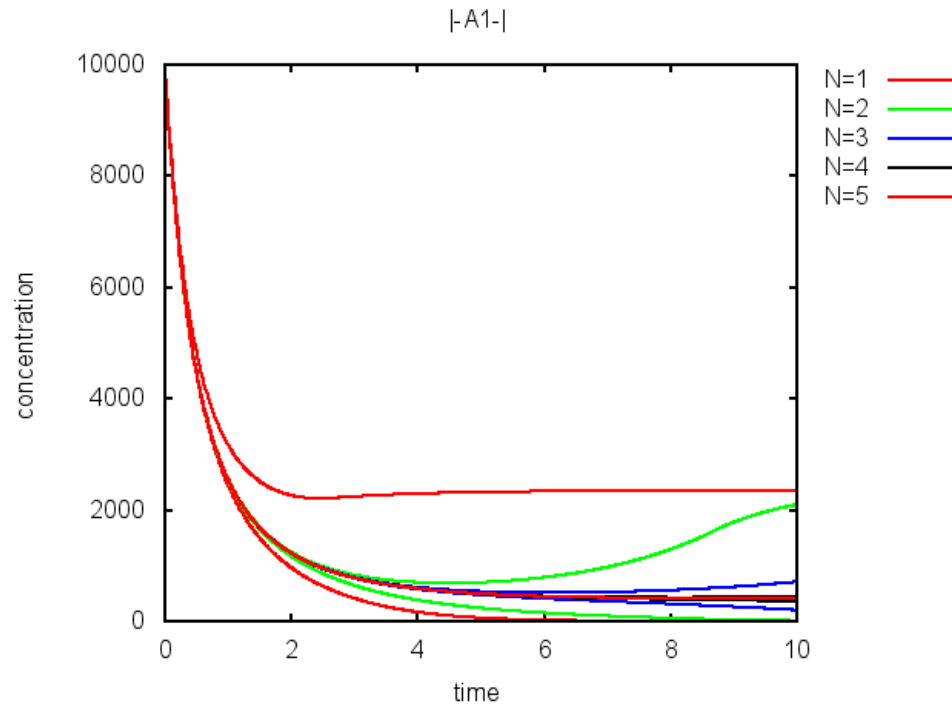
$$\overline{t_{2,3}^-} \triangleq 0.$$

# On the menu today

## Conservative model reduction

1. Motivating example
2. Evolution systems
3. Box approximation
4. Symbolic reasoning
5. Conclusion

# Numerical results



(with parameters  $[A_1]_0 = 10000$ ,  $k = 10^{-4}$ ,  $k_d = 10^{-2}$ , and  $k'_d = 10^{-1}$ ).

# Conclusion

- We can deal **models of polymers** that:
  1. are **finitely branching**;
  2. and **locally** defined.
- **High-level languages** enable to denote some infinite sums of variables and to handle them **symbolically**  
(Prove that they are differentiable, express their derivative, compare them).
- **Box approximation** can be used to derive **time-dependent bounds** on the values of some observables  
(Safe numerical bounds are computed *a posteriori* )  
(Approximation locally adapts to the state of the system)  
(Partial derivatives are considered only on the corresponding (hyper)-faces).

# Perspectives

1. Conservative approximation of the differential semantics
  - Provide a toolkit  
(box approximation, symbolic reasoning, partitioning, ...)  
to design and compose conservative approximations.
  - Recast existing methods in the light of this framework.
  - Automate the process and the choice of auxiliary variables.
2. Conservative approximation of the stochastic semantics
  - Provide analogous abstractions for stochastic simulation.
  - Design a stochastic simulator for abstract states  
(with an abstract notion of time).
3. Conservative approximation of hybrid semantics