Modeling and Analysis of Biological Systems

Ashish Tiwari

Tiwari@csl.sri.com

Computer Science Laboratory
SRI International
Menlo Park CA 94025
http://www.csl.sri.com/~tiwari

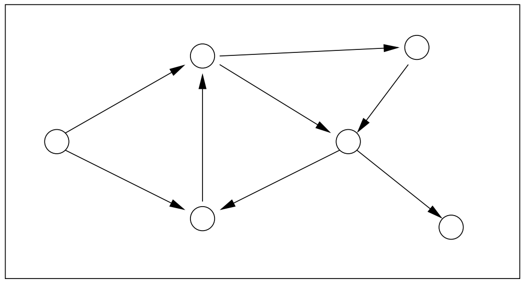
Part of the work described here is in collaboration with Alessandro Abate (Berkeley), Yu Bai (Stanford), Pat Lincoln, Merrill Knapp, Keith Laderoute, Carolyn Talcott (SRI)

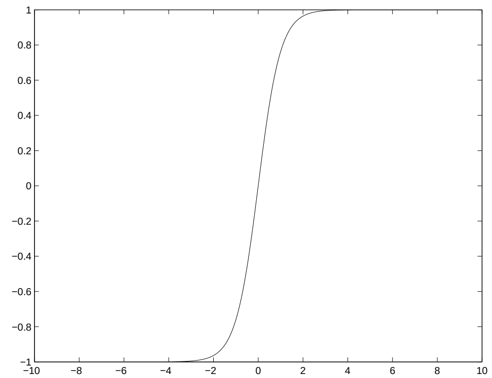
Modeling

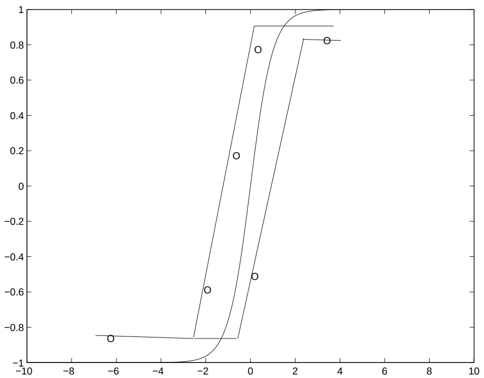
From experimental data to a formal model in some modeling formalism

Describes system at some level of abstractions

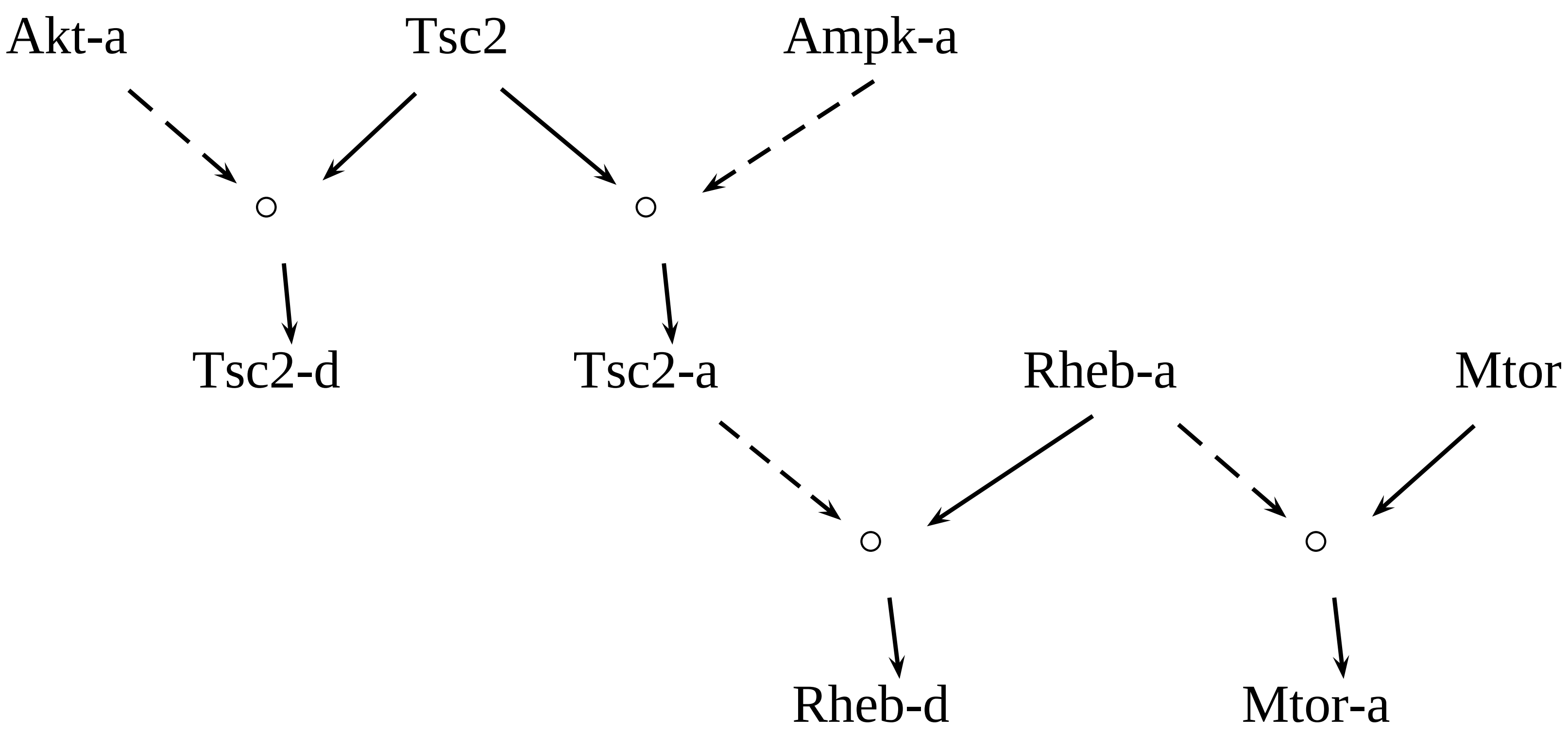
Discrete/Logical | Hybrid | Continuous



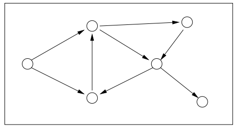




Logical Models: Example

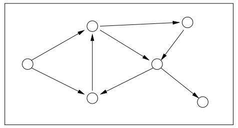


What does the diagram mean?

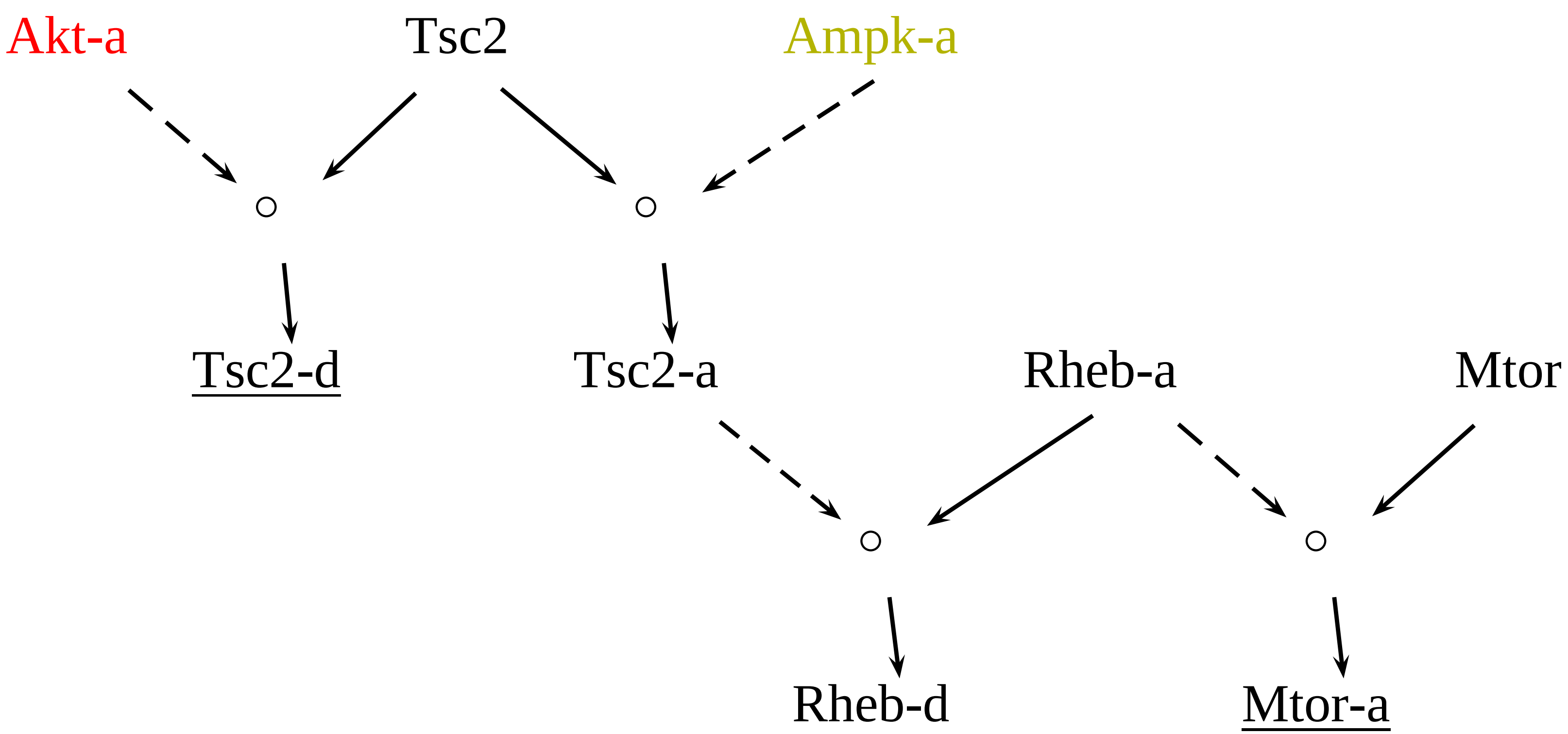


Logical Models

- Describes the process at a high level of abstraction.
- Only qualitative behavior of the system is modeled
- Interpreted as a petri-net, lack of rate information means that it is interpreted as a 1-safe petri-net
- Pathway Logic tool displays and analyzes these models
- Has been used to build models of signaling pathways in mammalian cells

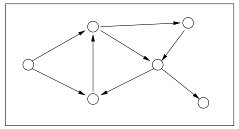


Example



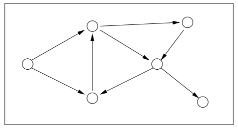
If Akt-act present, but Ampk-act is absent, then Tsc2 is deactivated and

Mtor-act is high

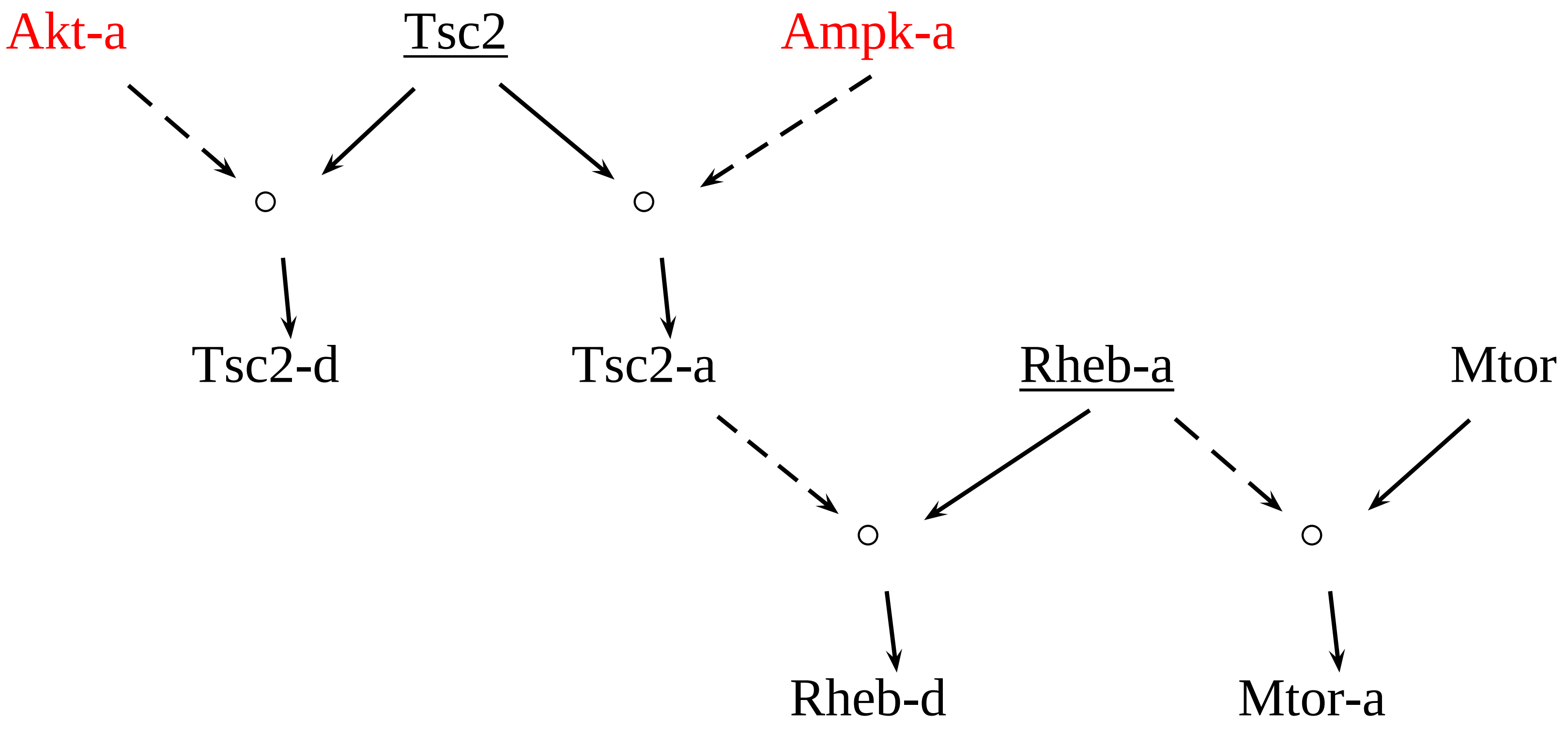


Extended Logical Models

- 2-valued interpretation gives limited information,
- Explored extensions of the 2-valued semantics by considering 3-valued qualitative interpretation of the rules in the model
- Built a tool that analyzes these networks on 3 values
- Enables analysis of networks where a resource is be shared between two competing pathways



Example



Difference in Mtor-act in the cases:

(1) Akt-a and Ampk-a are both present,

(2) Akt-a is present, but Ampk-a is not.

Stochastic Extensions of Discrete Logical Models

Limited rate information can be added in the discrete logical models in the form of transition probabilities

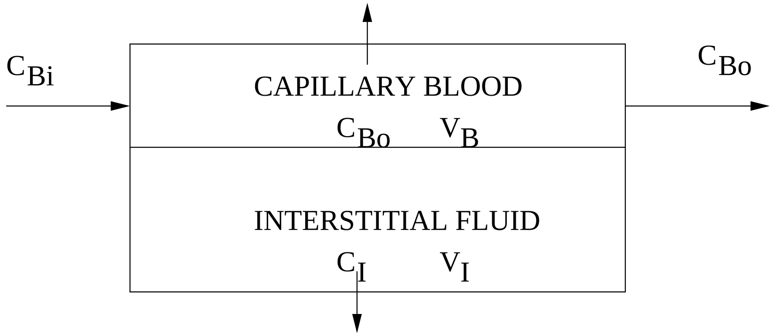
The resulting model can be simulated using a simplified variant of Gillespie's stochastic simulation algorithm (with tau-leaping)

We obtain time abstract stochastic simulations of the stochastic pathway logic models in this way

Continuous Dynamical System Models

Continuous differential equations are used classically to build models

Example. A 6-comparment PD model of glucose metabolism in human body



The mass balances for a typical physiologic compartment:

$$
\begin{aligned}
V_B \dot{C}_{Bo} &= Q_B(C_{Bi} - C_{Bo}) + PA(C_I - C_{Bo}) - r_{RBC} \\
V_I \dot{C}_I &= PA(C_{Bo} - C_I) - r_T
\end{aligned}
$$

V: volume, C: concentration, Q: flow, r: rate

Hard to determine the rate constants, still uncertainty in predicated values

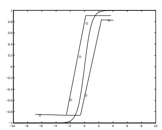
Hybrid Systems

Combines differential equations with discrete boolean logic

$$\frac{d}{dt}[RapA] \quad = \quad \underline{DRapA} - LRapA * [RapA] - k_{12} * [Pep5i] * [RapA]$$

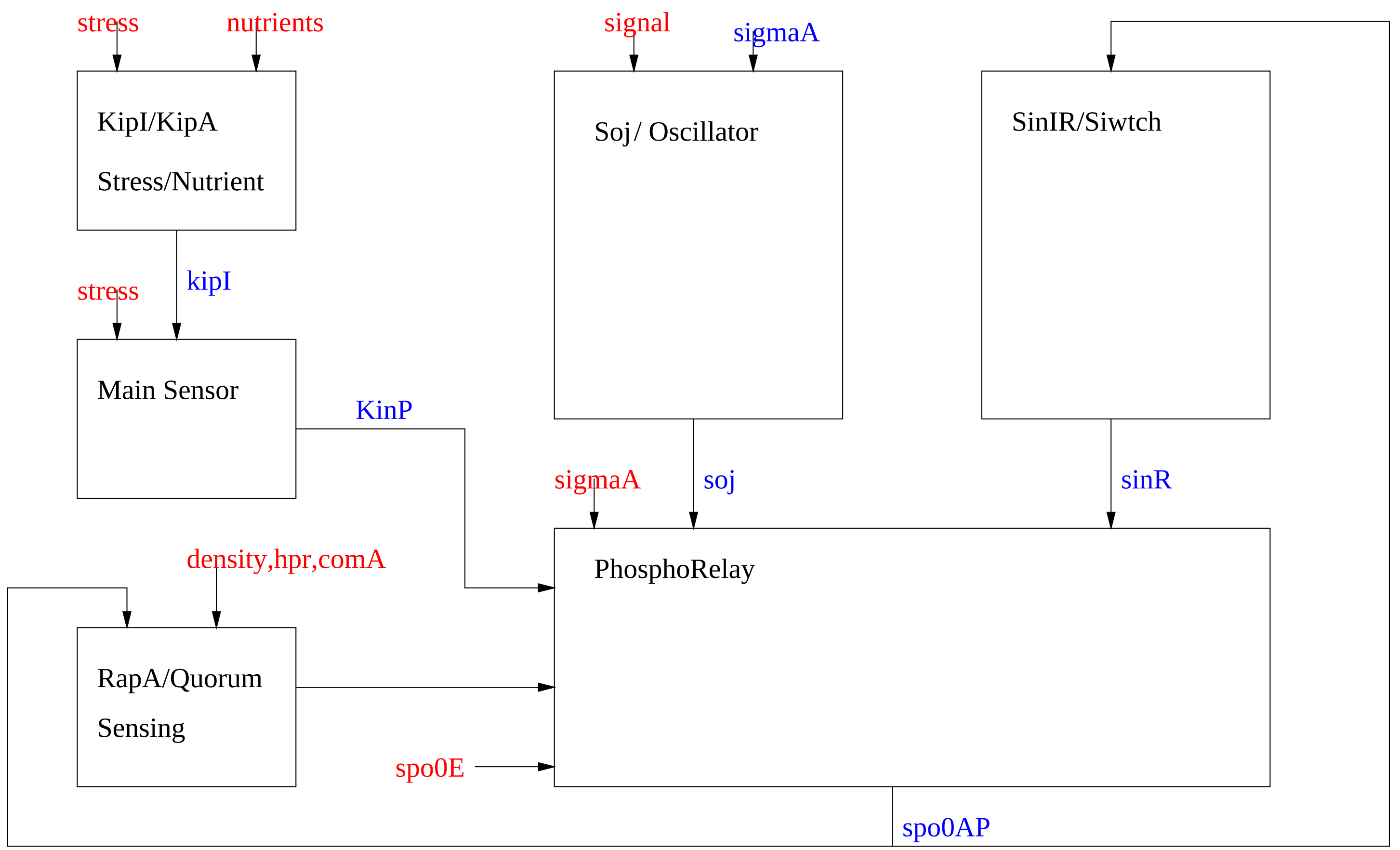
$\underline{DRapA}$ = IF $(comAP\_high?)$ THEN 1
ELSIF $(spo0AP2\_high?)$ THEN 0
ELSIF $(hpr\_high?)$ THEN 0
ELSE $1/2$ ENDIF;

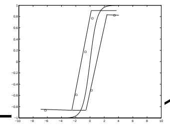
- Richer language for modeling at different levels of abstraction
- Nondeterminism allows for unknown parameter values
- Compositional development of complete model



A HybridSal Model

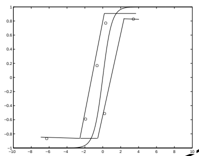
A component based view of the Sporulation Initiation Network in *B.Subtilis*





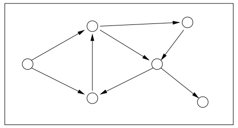
The HybridSal Abstractor

- Creates a conservative discrete approximation of the hybrid model
- The discrete abstraction has all behaviors of the original nondeterministic (partially unspecified) model
- The abstractor works compositionally and abstracts the models by abstracting its components of the model
- It can ignore certain parts of the model and focus on other parts of interest to the biologist
- It can create multiple abstract views of the same base model
- Unknown rate constants can be symbolically constrained, such as ($k_{12} > k_{21}$)

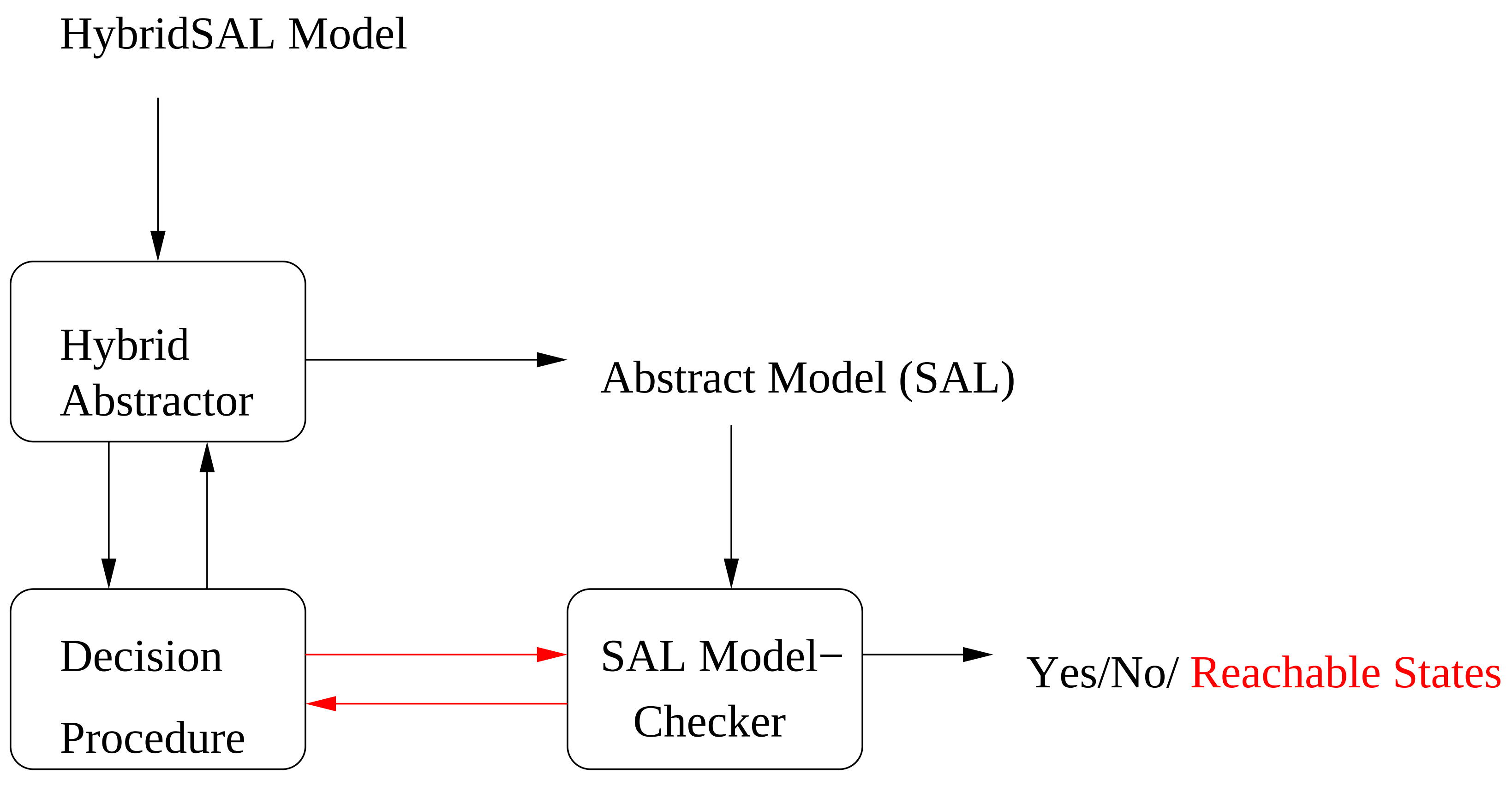


The Sal Model Checker

- The discrete abstract model is explored using a symbolic model checker
- Routinely search through state space of size 2^{100} and beyond
- Can extract interesting behaviors that the model exhibits:
  Under the given environment, can the cell go into a high $SpooAP$ state?
- Can also provably verify that certain things never happen
  It is impossible for the concentrations of proteins A and B to be high simultaneously.
  If the cell enters a particular configuration, it does not get out of it unless the environmental signals change.



The HybridSal Tool Architecture



The decision procedure for the quantifier-free theory of reals powers the HybridSal tools

Model Simplification

It is computationally difficult to analyze continuous and hybrid models with unknown parameters

Need to simplify the model without compromising much on the behaviors of the model

A new method for model simplification based on a symbolic procedure for reasoning about the real numbers:

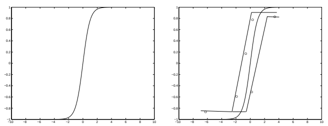
- Not all terms in an ODE are equally important. Some reactions are more influential in determining the overall behavior
- The fact that certain terms contribute little to the overall dynamics can be formally stated and proved using a symbolic reasoning engine (that decides the theory of reals).

Model Simplification: Example

Model of tetracycline resistance in bacteria (UPenn):

$$
\begin{aligned}
d[TetR]/dt &= f_1 - k_d[TetR] - k_+[Tc][TetR] + k_-[TetRTc] \\
d[TetRTc]/dt &= k_+[Tc][TetR] - k_-[TetRTc] - k_d[TetRTc] \\
d[Tc]/dt &= k_i([Tc]^0 - [Tc]) - k_p[Tc][TetA] - k_+[Tc][TetR] + \\
&\quad + k_-[TetRTc] - k_d[Tc] \\
d[TetA]/dt &= f_2 - k_d[TetA]
\end{aligned}
$$

$$
f_1, f_2 = \begin{cases} b_1 & \text{if } [TetR] > 1/c_1 \\ b_1 + k_1 & \text{otherwise} \end{cases}
$$

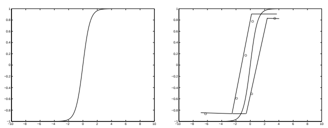


Model Simplification: Example

We simplify the above system into the following system:

$$
\begin{aligned}
d[TetR]/dt &= f_1 - k_+[Tc][TetR] + k_-[TetRTc] \\
d[TetRTc]/dt &= k_+[Tc][TetR] - k_-[TetRTc] - k_d[TetRTc] \\
d[Tc]/dt &= k_i[Tc]^0 - k_p[Tc][TetA] \\
d[TetA]/dt &= f_2 - k_d[TetA]
\end{aligned}
$$

This is a *sound* simplification: prove that the contribution of eliminated terms is much smaller than that of the few *dominating* retained terms



Model Simplification: Technique

Let $[TetR]_0, [TetRTc]_0, [Tc]_0, [TetA]_0$ denote the equilibrium concentrations. Hence, we get the formula C:

$$
\begin{aligned}
f_1 - k_d[TetR]_0 - k_+[Tc]_0[TetR]_0 + k_-[TetRTc]_0 &= 0 \ \wedge \\
k_+[Tc]_0[TetR]_0 - (k_- + k_d)[TetRTc]_0 &= 0 \ \wedge \\
k_i([Tc]^0 - [Tc]_0) - k_p[Tc]_0[TetA]_0 - k_+[Tc]_0[TetR]_0 + & \\
+k_-[TetRTc]_0 - k_d[Tc]_0 &= 0 \ \wedge \\
f_2 - k_d[TetA]_0 &= 0
\end{aligned}
$$

We prove the following:

$$
\begin{aligned}
C &\Rightarrow 10k_d[TetR]_0 < f_2 \\
C &\Rightarrow (10k_i[Tc]_0 < k_i[Tc]^0 \ \wedge \ 10k_+[Tc]_0[TetR]_0 < k_p[Tc]_0[TetA]_0 \ \wedge \\
& \quad 10k_-[TetRTc] < k_i[Tc]^0 \ \wedge \ 10k_d[Tc]_0 < k_i[Tc]^0)
\end{aligned}
$$

Model Simplification: Example

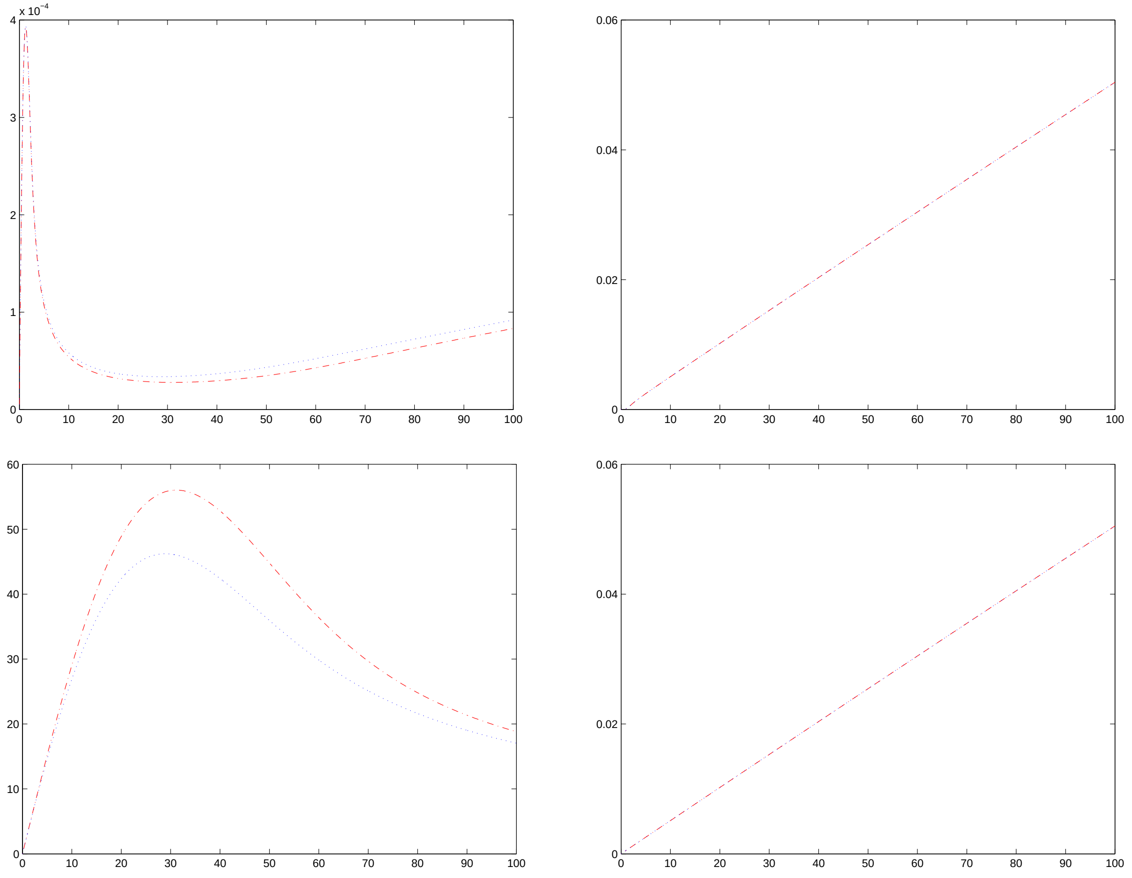


Table 1: Simulation plots for 100 time steps of the original model (red) and simplified model (blue) for each of the four species.

Model Simplification: Example

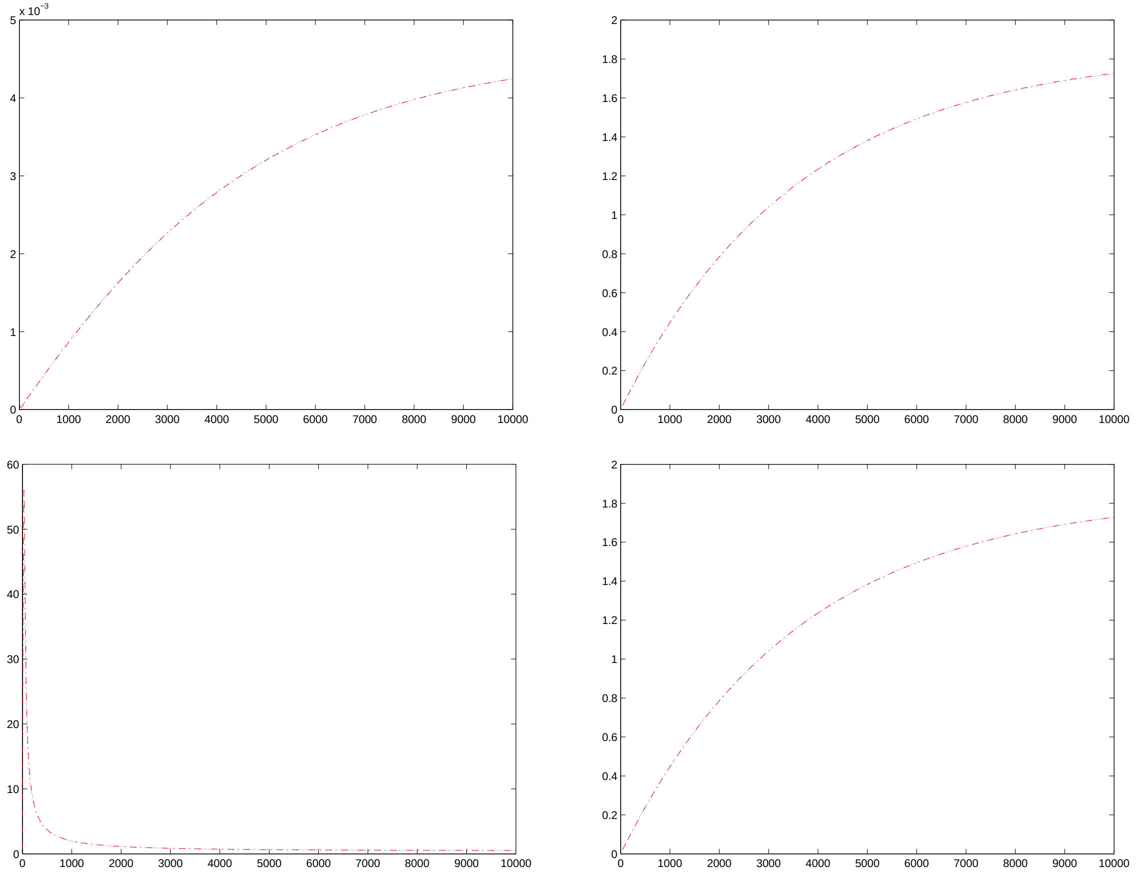


Table 2: Simulation plots for 10000 time steps of the original model (red) and simplified model (blue) for each of the four species.

Symbolic Procedure for the Reals

The core technology we use in verifying model simplification steps and performing hybrid abstraction:

Given a set of nonlinear equations and inequalities:

$$\begin{aligned} p &\approx 0, \qquad & p &\in P \\ q &> 0, \qquad & q &\in Q \\ r &\geq 0, \qquad & r &\in R \end{aligned}$$

where $P, Q, R \subset \mathbb{Q}[\vec{x}]$ are sets of polynomials over $\vec{x}$

Is the above set satisfiable over the reals?

Generalized Simplex for Nonlinear Constraints

If all polynomials are linear, then Simplex LP solver can be used

- Introduce slack variables s.t. all inequality constraints are of the form $v > 0$, or $w \geq 0$

$$P = 0, \quad Q > 0, \quad R \geq 0 \quad \mapsto$$
$$\underline{P = 0}, \quad \underline{Q - \vec{v} = 0}, \quad \underline{R - \vec{w} = 0}, \quad \vec{v} > 0,\ \vec{w} \geq 0$$

- Search for a polynomial p s.t.

$$\underline{P = 0} \quad \Rightarrow \quad p \approx 0$$
$$\vec{v} > 0,\ \vec{w} \geq 0 \quad \Rightarrow \quad p > 0$$

- To search for p, compute the Gröbner basis for P using different possible orderings (pivot)

Example: Nonlinear Constraint Solving

Consider $E = \{x^3 = x,\ x > 2\}$.

$$\frac{\dfrac{\dfrac{x^3 - x \approx 0, \qquad x - v - 2 \approx 0}{(v+2)^3 - (v+2) \approx 0, \qquad x - v - 2 \approx 0}}{(v+2)(v+1)(v+3) \approx 0, \quad x - v - 2 \approx 0}}{\bot}$$

Computing GB and projecting it onto the slack variables discovers the witness p for unsatisfiability

A. Tiwari, "An algebraic approach for the unsatisfiability of nonlinear constraints." In *Computer Science Logic, CSL 2005*, Vol 3634 of LNCS, pp 248–262, Springer.

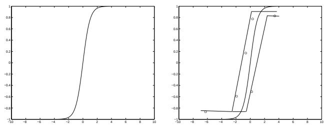
A General Principle for Physical Systems?

The analysis of models of various biological systems shows that

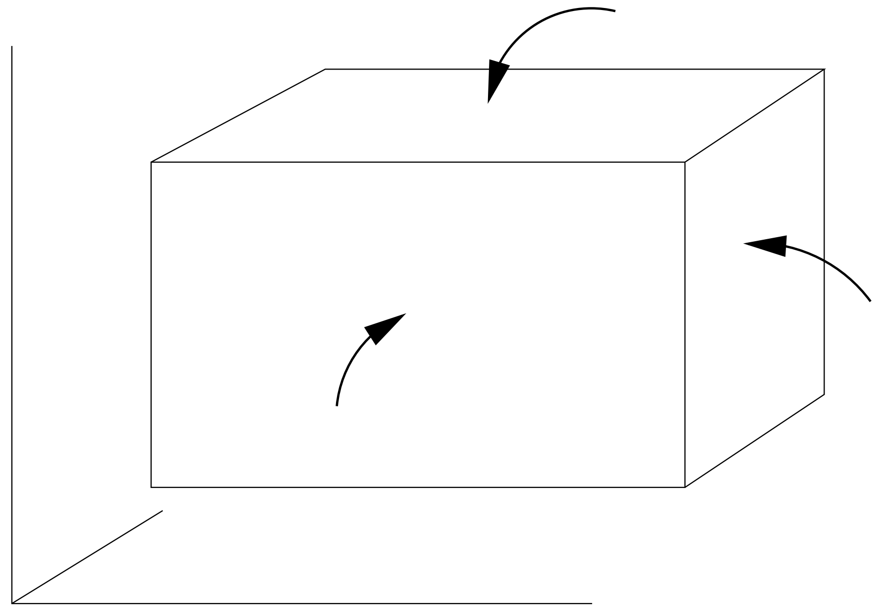
whenever the system is stable, it is also stable in a sense stronger than asymptotic stability (in case of linear systems)

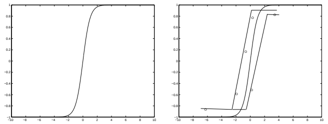
We have called this notion box stability

Box stability. A system is box stable around an equilibrium point $\vec{x}_0$ if there is a rectangular box $\vec{l} \leq \vec{x} \leq \vec{u}$ containing $\vec{x}_0$ s.t. the vector field points inwards on all surfaces of the box.



Box stability: Illustration





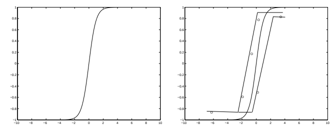
Box stability: Examples

Examples of systems that were found to be box stable:

- Glucose and insulin metabolism in human body
- Tetracycline resistance in bacteria (UPenn)
- Regulation of induction in the *lac* operon in *E.Coli* (UPenn)
- B. subtilis sporulation initiation network (LBL)
- Delta-Notch intercellular signaling mechanism (Stanford)

Example that was not box-stable:

cardiovascular model of heart (oscillatory)

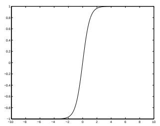


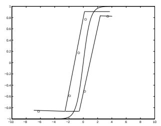
Box stability: Why?

- Computationally more tractable
- Helps in safety verification of systems
- It has a very natural interpretation
- Implies asymptotic stability in the linear case

Implications of the general observation

- Yields new computational methods to analyze linear and nonlinear systems
- As a principle, allows generation of constraints on unknown parameters
- Gives new techniques for defining senstivity and analyzing it





The BioSal Summary