Hybrid Modeling and Analysis of Biological Networks

Ashish Tiwari

Tiwari@csl.sri.com

Computer Science Laboratory

SRI International

Menlo Park CA 94025

http://www.csl.sri.com/~tiwari

Collaborators: Claire Tomlin, Ronojoy Ghosh, Adam Arkin, Denise Wolf, Patrick Lincoln, Alessandro Abate

Systems Biology

- Tremendous amount of experimental data accumulated over the years,
- Which has been used by biologists to build mental models of biological processes
- But such models have not been formally specified or computationally analyzed
- Goal: Develop models of biological processes and tools to play with the models
- So that wet lab experiments can be replaced by faster and less risky computational ones

Analysis is not the only challenge

Outline

- Three case studies:
  - Delta-Notch lateral inhibition
  - Sporulation initiation in *B. Subtilis*
  - Human blood glucose metabolism
- For each case study:
  - Biology
  - Formal Model
  - Analysis technique and results

Information Metabolism

- Cells are highly responsive to specific chemicals in its environment
  Cells receive, process, and respond to information from the env.

Signal → Reception → Transduction → Response(s)

- About half of 25 largest protein families encoded by human genome deal with information processing
- Signal transduction pathways: sense and process the external stimuli

Information metabolism = Signal transduction + Response

Signal Transduction

- Membrane-bound receptor protein senses external signalling molecules, ligand, by binding to them
- Causes the structure of (the intracellular domain of) the receptor to alter
- This causes activation of protein kinases: enzymes that transfer phosphoryl group from ATP to proteins, thus activating the protein
- Protein phosphatases can undo this by removing the Phosphoryl group, thus terminating the signalling process
- Errors can lead to cancer

Caveat: There are exceptions to everything, but above is a common scenario.

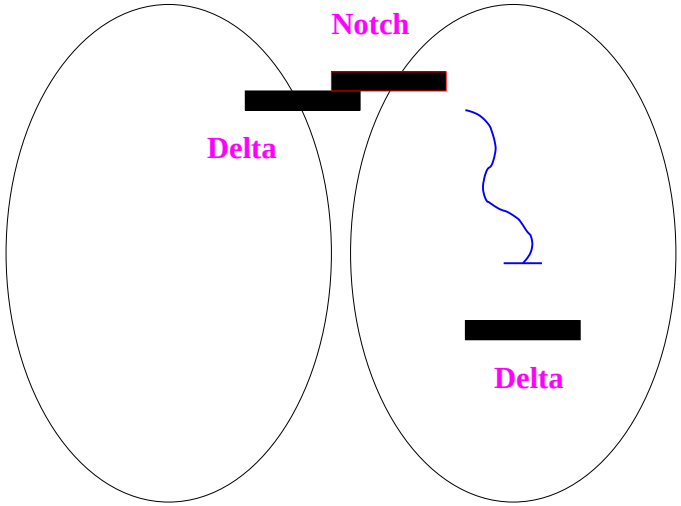
Response

- Usually via regulation of gene expression
- Rate of synthesis of proteins changes 1000-fold in bacteria in response to env. changes
- Differences in gene expression cause different cell types in multicellular organisms (e.g. muscle and nerve), even though they contain exactly the same DNA
- Gene expression = transcription + translation
- Transcription is regulated by proteins that bind to specific DNA sites (promoter regions)

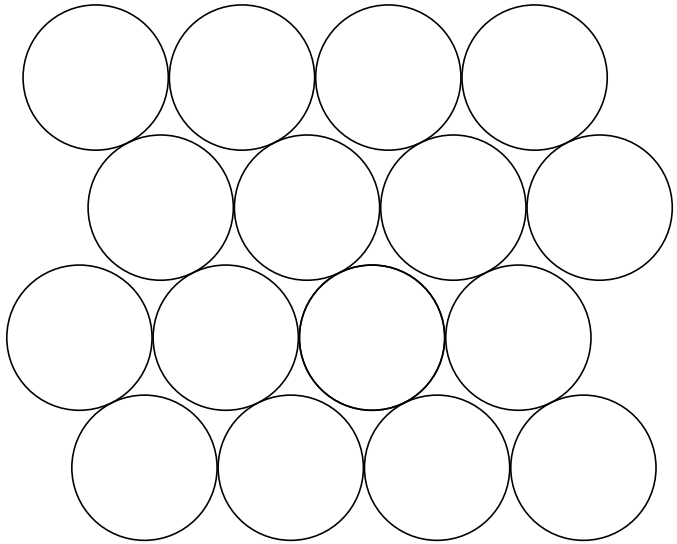
Delta-Notch Lateral Inhibition

Implicated in cell differentiation

External Signal	:	External Delta	:	Binds to receptor Notch
Sensor	:	Notch	:	transmembrance receptor protein
Response	:	Internal Delta	:	Notch inhibits Delta
			:	Delta is also a transmembrane protein

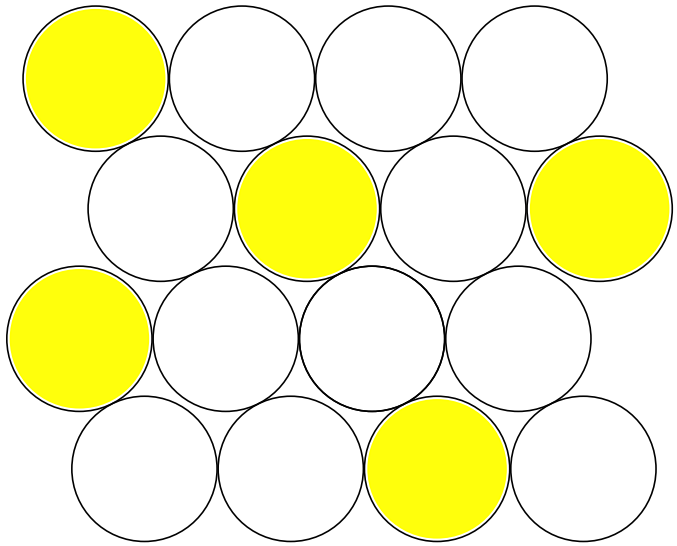


Delta-Notch: Array of cells



Causes pattern formation across many biological species

Salt-and-Pepper pattern in South African claw-toed frog's epidermal layer

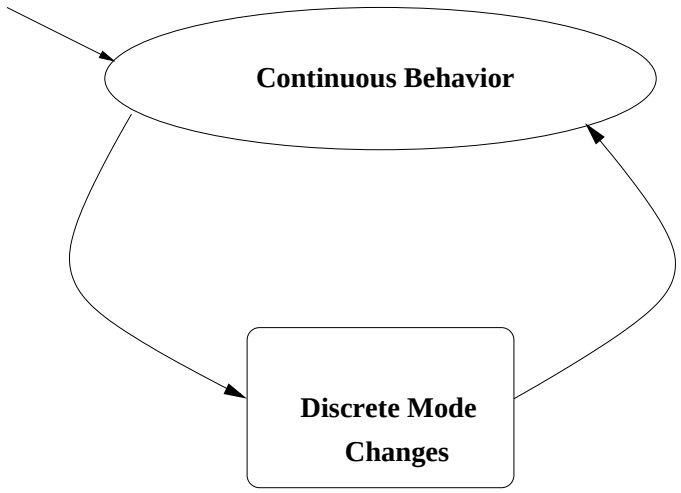


Colored cells have differentiated into ciliated cells (high Delta, low Notch) while rest as epidermal cells (low Delta, high Notch)

Modeling

Formal models:

- Continuous dynamical systems (Traditional Sciences)
- Discrete state transition systems (Computer Science)
- Hybrid systems: Continuous and discrete components



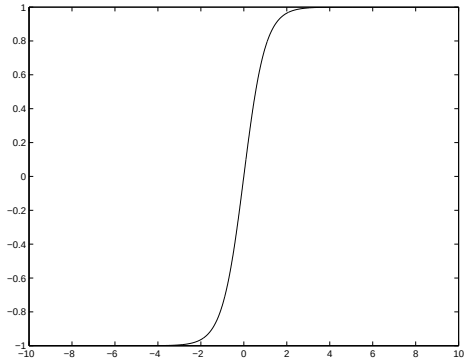
Hybrid Systems

- Formal models that combine differential equations with discrete boolean logic
- Natural for modeling
  - embedded systems
  - software controlled systems
  - multi-modal dynamical systems
- Matlab supports modeling and simulation via Simulink and Stateflow

Traditional → Hybrid Model: I

Trying to build models for the experimental data:

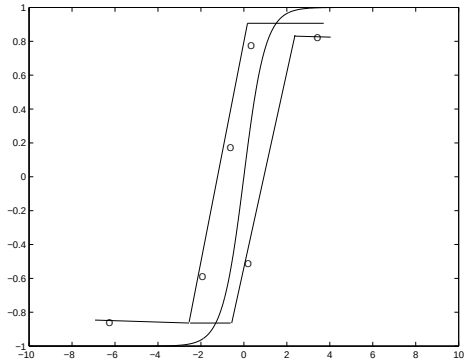
Experimental Data ⟶ Model — Modeling Formalism



↗ Continuous

Data Points

↘ Hybrid



(Levels of Abstractions)

Traditional → Hybrid Model: II

Dynamics resulting from participation of small number of molecules is discrete

Transcription: There are only a pair of genes in a cell, and few mRNAs in a cell

Genes being "on" or "off" can be seen as a discrete switch

Sigmoidal functions is one way to model behavior

Discrete step function is another way

Delta-Notch: A Hybrid Model for One Cell

v_D, v_N : concentration of Delta and Notch in a cell

delta is on : $v_N < threshold_2$

notch is on : External Delta concentration $> threshold_1$

So, a cell can be in four modes.

delta is "on" and *notch* is "off" :

$$\begin{aligned} dv_D/dt &= R_D - \lambda_D v_D \\ dv_N/dt &= -\lambda_N v_N \end{aligned}$$

Composing these hybrid models, we can get models of $2, 4, 8, \ldots$ cells

Delta-Notch: Analysis Results

Challenges in the analysis:

- Unknown parameters
- Intercellular interaction

In isolation, for given parameters, easy to prove that the cell is bistable:

- if external Delta is high, then Notch is high, Delta is low
- if external Delta is low, then Notch is low, Delta is high

Delta-Notch: Analysis Results

Using a symbolic approach, we can show that the above result holds for any set of parameter values within certain (symbolic) bounds

The multiple cell configuration can also be analyzed :

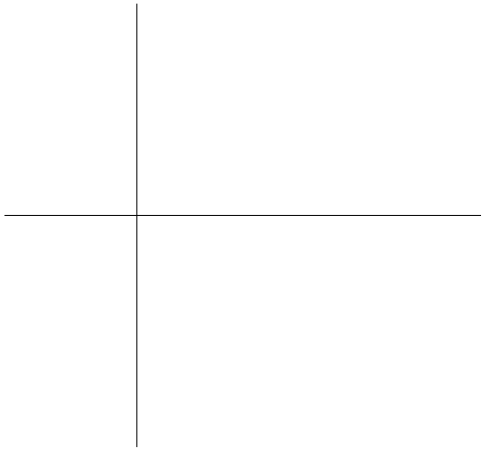
If a cell differentiates (high Delta), then none of its neighbours can differentiate

Symbolic Systems Biology: Analysis Approach

The approch to analyzing hybrid system models uses an abstraction based on partitioning the space

$$\begin{aligned} dv_D/dt &= -v_D \mid 1 - v_D \\ dv_N/dt &= -v_N \mid 1 - v_N \end{aligned}$$

Concrete state space: $\Re^2$



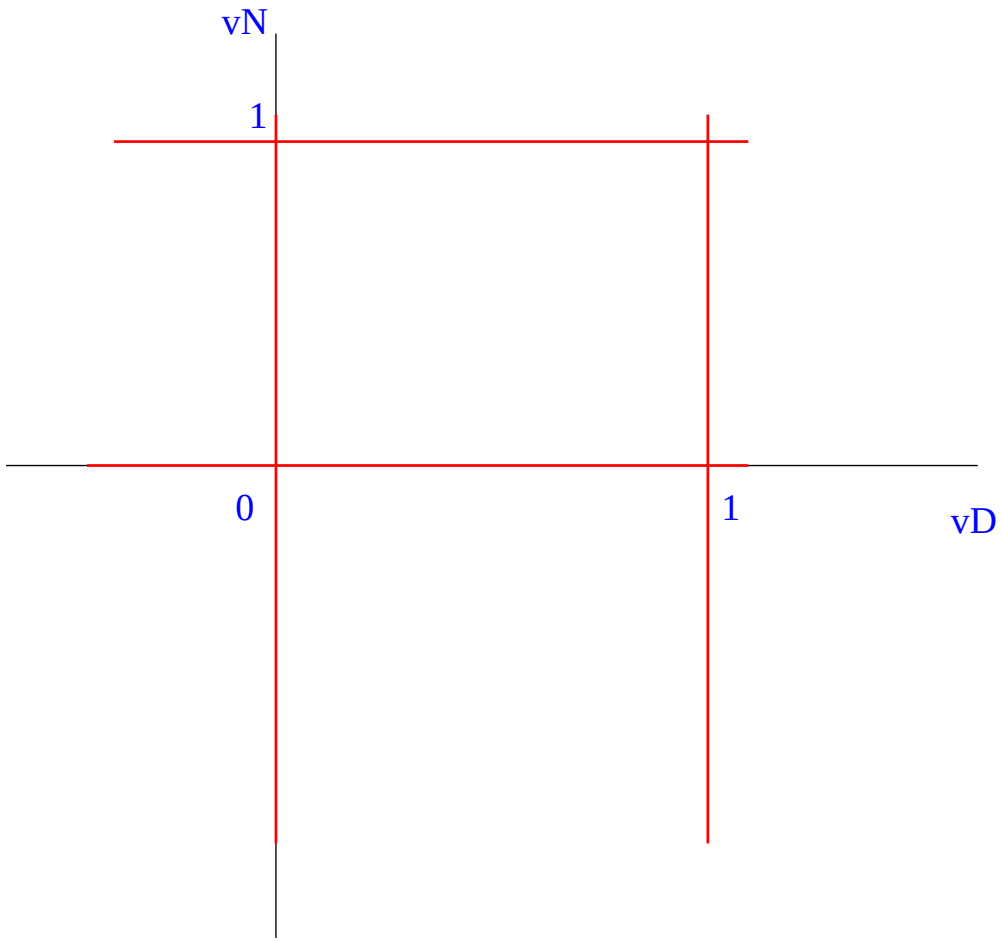
Abstraction Algorithm: Choosing Polynomials

- Initial Polynomials of Interest: v_D, v_N
- To track their progress (increasing, decreasing, constant), I need (the signs of): $\dot{v_D}$ and $\dot{v_N}$ in *all* modes. Thus, we get

$$-v_D, -v_N, 1 - v_D, 1 - v_N.$$

- To track their progress (increasing, decreasing, constant), I need (the signs of): $1 - \dot{v_D}$ and $1 - \dot{v_N}$ in *all* modes. But, this we have already.

Partitioning: Choosing Polynomials



Partitioning: Choosing More Polynomials

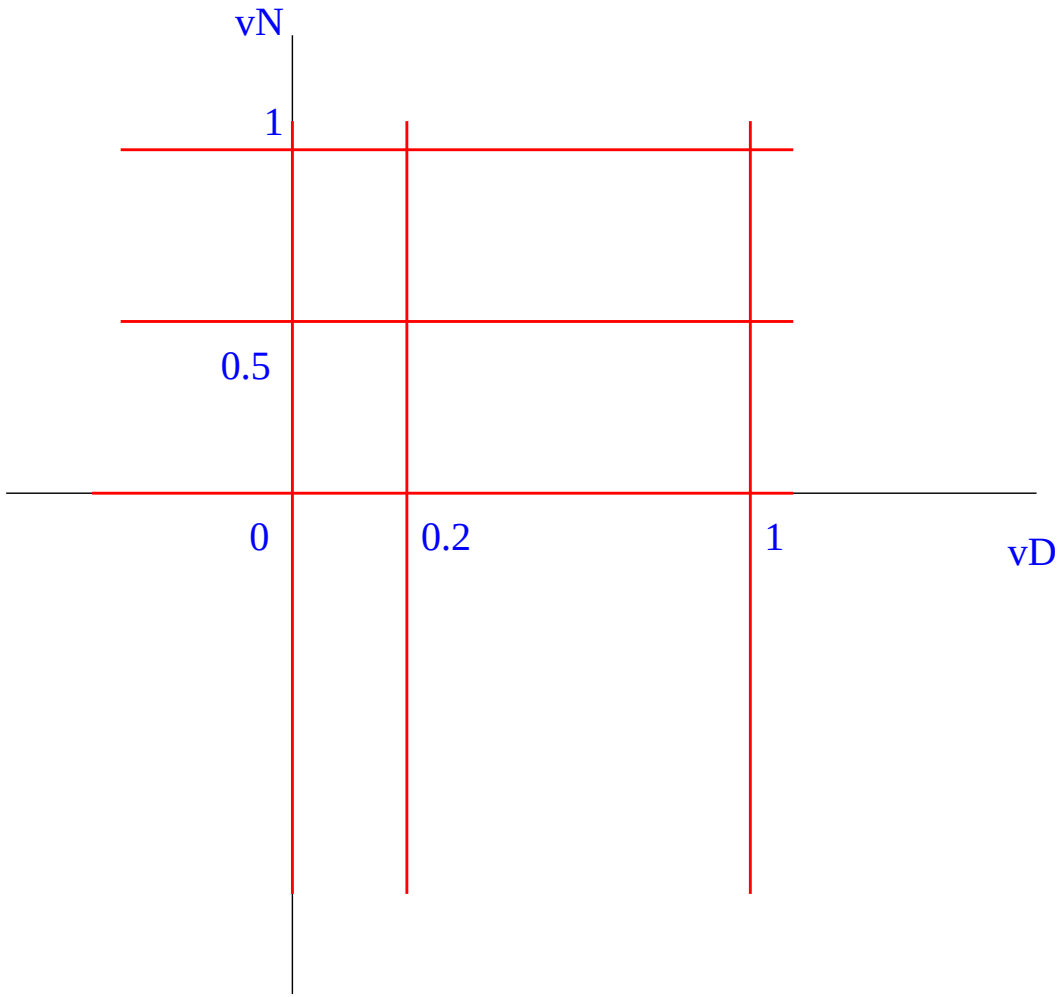
delta is on : $v_N < 0.5$

notch is on : External Delta concentration > 0.2

- Discrete mode switch conditions:
  We need to know when either $v_N < 0.5$ or $u_N > 0.2$ changes.
- To trace this, we need $(-\dot{v_N} + 0.5)$ and $(\dot{u_N} - 0.2)$.
- Now, the sign of $(-\dot{v_N} + 0.5)$, in all modes, is known from the signs of the already computed polynomials.
- The derivative of $u_N - 0.2$ is 0.
- We also include $v_D - 0.2$ in the set.

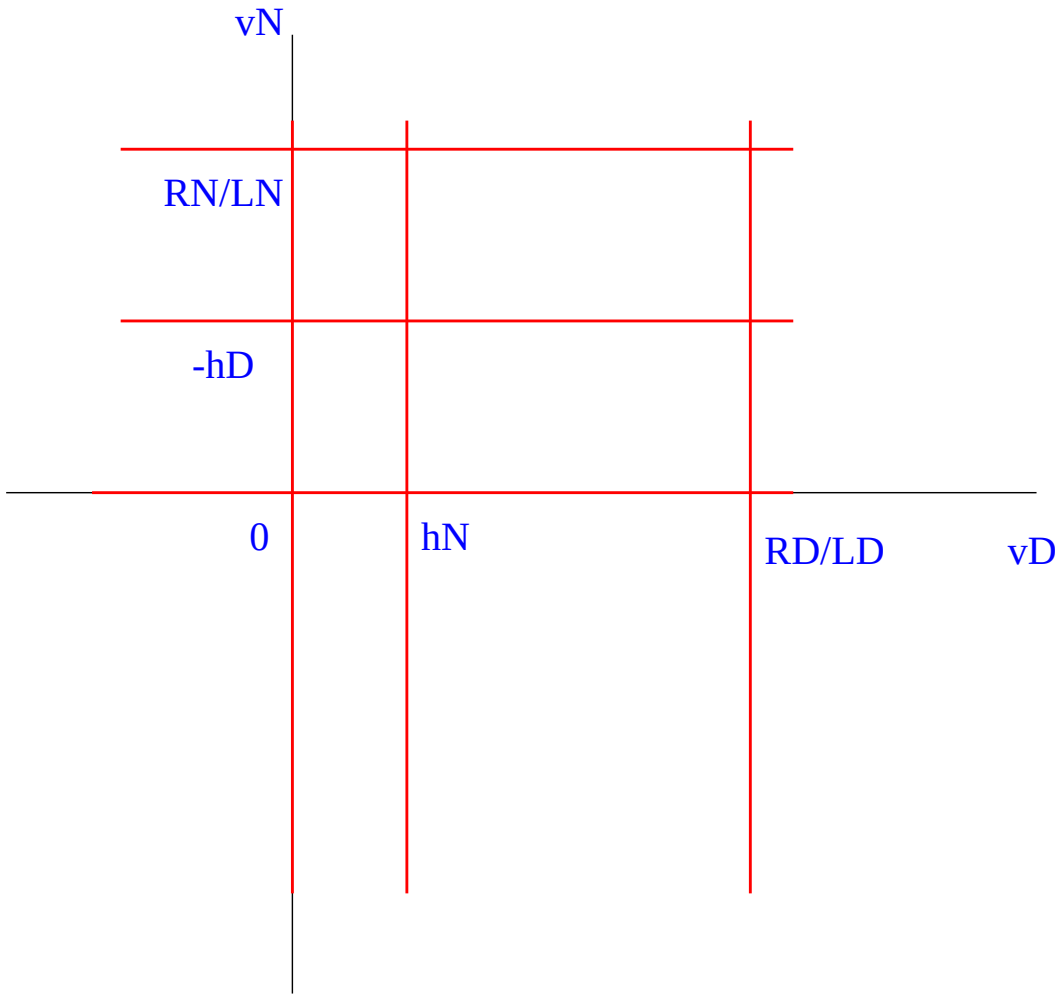
Two new polynomials: $v_D - 0.2$ and $-v_N + 0.5$.

Partitioning: Choosing More Polynomials



Partitioning: Choosing Polynomials

Assuming $\lambda_D > 0,\ \lambda_N > 0,\ R_D > 0,\ R_N > 0,\ h_D < 0,\ h_N > 0$:



Abstracting Continuous Dynamics

For each mode $l \in \mathbf{Q}$:

if q_{pi}, q_{pj} are abstract variables s.t. $\dot{p_i} = p_j$ in mode l, then apply rules of the form:

- if $q_{pi} = pos$ and $q_{pj} = pos$, then new value q'_{pi} is pos.
- if $q_{pi} = pos$ and $q_{pj} = zero$, then new value q'_{pi} is pos.
- if $q_{pi} = pos$ and $q_{pj} = neg$, then new value q'_{pi} is either pos or $zero$.
- $\ldots$

If $q_{p0}, q_{p1}, q_{p2}, \ldots, q_{pn}$ is s.t. $\dot{p_i} = p_{i+1}$ in mode l, then

- if $q_{p0} = q_{p1} = q_{p2}, \ldots q_{pn-1} = zero$ and $\dot{q_{pn}} = pos$, then new values $q'_{p0} = q'_{p1} = q'_{p2}, \ldots q'_{pn-1} = pos$.

Mapping

Each abstract state corresponds to a region where the computed polynomials are sign-invariant. We know the signs of polynomials of interest and signs of some of its higher-order derivatives.

$$
\begin{array}{lcl}
g0 & : & R_N + (-1) * v_N * \lambda_N \\
g1 & : & R_D + (-1) * v_D * \lambda_D \\
g2 & : & v_D \\
g6 & : & -v_N - h_D
\end{array}
\qquad
\begin{array}{lcl}
g3 & : & v_N \\
g4 & : & u_N - h_N \\
g5 & : & v_D - h_N
\end{array}
$$

For example, sign($\dot{g3}$) in modes when " notch high" is equal to sign(g0). sign($\dot{g0}$) in the same mode is equal to $-$sign($g0$).

Abstracting Continuous Dynamics

delta low AND notch high AND

(g4 = pos OR g4 = zero) AND g6 = neg $\longrightarrow$

g3' IN IF g3 = pos THEN IF g0 = pos OR g0 = zero THEN {pos}
ELSE {pos, zero} ENDIF
ELSIF g3 = neg THEN IF g0 = neg OR g0 = zero THEN {neg}
ELSE {neg, zero} ENDIF
ELSE IF g0 = pos THEN {pos}
ELSIF g0 = neg THEN {neg} ELSE {zero} ENDIF
ENDIF;

g0' IN . . .

g1' IN . . .

⋮

Abstracting Discrete Transitions

Discrete Transition: $(q, \psi(X), q', New(X))$, where

- q, q': modes,
- $\psi(X)$: enabling condition, and
- $New(X)$: assignments to continuous variables.

Abstract Discrete Transition: $((q, \phi_1), (q', \phi_2))$ if

- The formula $\phi_1 \wedge \psi$ is satisfiable and
- The formula $\exists X^o : \psi(X^o) \wedge X = New(X^o) \wedge \phi_2(X)$ is satisfiable.

Abstracting Discrete Transitions

Current Mode	Condition	New Mode
not (delta high AND notch low)	$-v_N \geq h_D$ and $u_N < h_N$	delta high and notch low

is abstracted to:

g4 = neg AND (g6 = pos OR g6 = zero)
AND NOT(delta high AND notch low) $\longrightarrow$

notch ' = low

delta ' = high

where $g4$ maps to $u_N - h_N$ and $g6$ maps to $-v_N - h_D$.

Discharging Proof Obligations

Using a sound, but incomplete, procedure for reals
(Ref: Tiwari:CSL05)

This gives an automated abstraction algorithm for hybrid automata that are specified using only expressions in the theory of real closed fields.

We can also detect infeasible abstract states to generate a global invariant of the resulting abstract model.

Model Checking Results

If the system ever reaches a state " delta high AND notch low" then the system continues to remain in that state subsequently forever.

$$G(\text{ delta high AND notch low} \Rightarrow G(\text{ delta high AND notch low }))$$

Under additional "fairness" assumptions: The system always eventually *reaches* one of these two equilibrium states.

$$GF((\text{ delta high AND notch low}) \text{ OR } (\text{ delta low AND notch high}))$$

Two Cell Cluster

A two cell complex is created by composing two single cells.

Certain variables (which are not "local" to the module) need to be renamed to avoid conflicts. E.g. the names $delta$ and $notch$ are renamed.

Communication is captured by renaming variables to the same name. E.g. $g4$ of one cell and $g5$ of the other are renamed to the same variable.

```
twocells: MODULE =
LOCAL vd1, vd2 IN
( (RENAME g4 TO vd2, g5 TO vd1 IN cell)
  []
  (OUTPUT delta2, notch2 IN
  (RENAME g4 TO vd1, g5 TO vd2, delta TO delta2, notch TO notch2 IN cell)) );
```

Two and Four Cells: Results

Model checking shows that the states where one cell has high Delta and low Notch concentration, while the other has low Delta and high Notch concentration, are stable states for a two cell complex.

The cell fate is determined by whether initially $v_{D\,1} < v_{D\,2} \ \wedge \ v_{N\,1} > v_{N\,2}$

With four cells, there are three stable states:

- Cells 1 and 4 differentiate
- Only cell 2 differentiates
- Only cell 3 differentiates

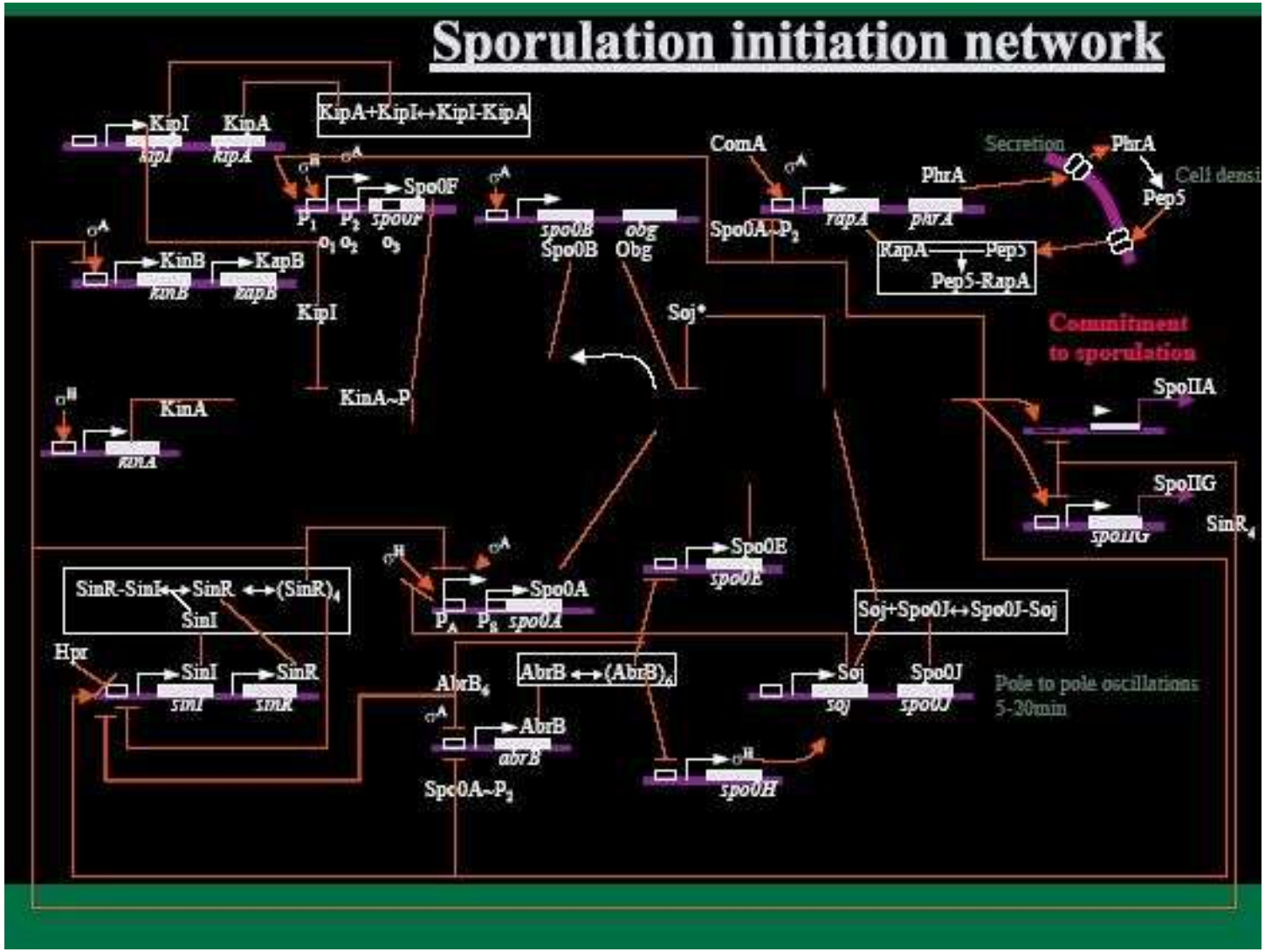
Sporulation Initiation

B. Subtilis is an *Anthrax*-like bacteria

- Shows a variety of different responses when stressed
  A complex stress response network has been proposed
  motility, degradative enzyme synthesis, competence, sporulation
- Sporulation is one possible response
- Decision to sporulate is a big one for the bacteria
- Cell undergoes several transformations after commitment

How and when does *B.Subtilis* commit to sporulation?

The Biologists View



Understanding the Network

Decision to sporulate regulated by Spo0AP

Spo0A obtains P through phosphorelay:
(Protein Kinase) : Spo0FP : Spo0BP : Spo0AP

Regulation by phosphatases: Spo0E, RapA

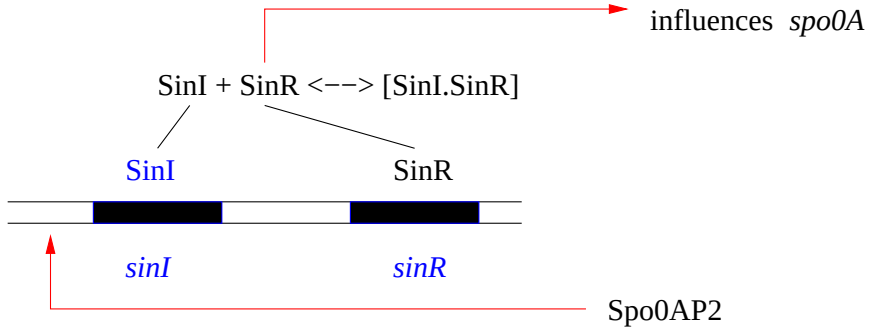
Quorum Sensing: RapA dephosphorates Spo0FP; however, under high cell-density, Pep5i binds to RapA

Stress Sensor: KinA phosphorolates Spo0F, but KinI inhibits this. However, under stress, KipA binds to KipI

Global Switch: SinR represses transcription of *spo0A*, SinI binds to SinR; Spo0AP promotes *sinI* transcription, SinR represses it

The SinIR component

Example of a "component": SinI-SinR global bistable switch



- Tetrameric SinR represses *spo0A*
- SinI inactivates SinR repression by binding to it
- *sinI* expression is controlled by Spo0A-P, SinR, Hpr, and AbrB

$$
\begin{aligned}
dSinI/dt &= \Delta_I - \lambda_I SinI - k SinI\, SinR \\
dSinR/dt &= \Delta_R - \lambda_R SinR - k SinI\, SinR \\
\Delta_I &= \texttt{if}(Spo0AP = high\ \texttt{and}\ SinR = low) \\
&\quad \texttt{then}\ 1\ \texttt{elsif}\ (\ldots)\ \texttt{else}\ (\ldots)
\end{aligned}
$$

SinI-SinR Components

In isolation, assuming simple dynamics for $Spo0AP$, the SinI-SinR component exhibits bistability

Either
(1) $Spo0AP$ is high and $SinR$ is low, or
(2) $Spo0AP$ is low and $SinR$ is high

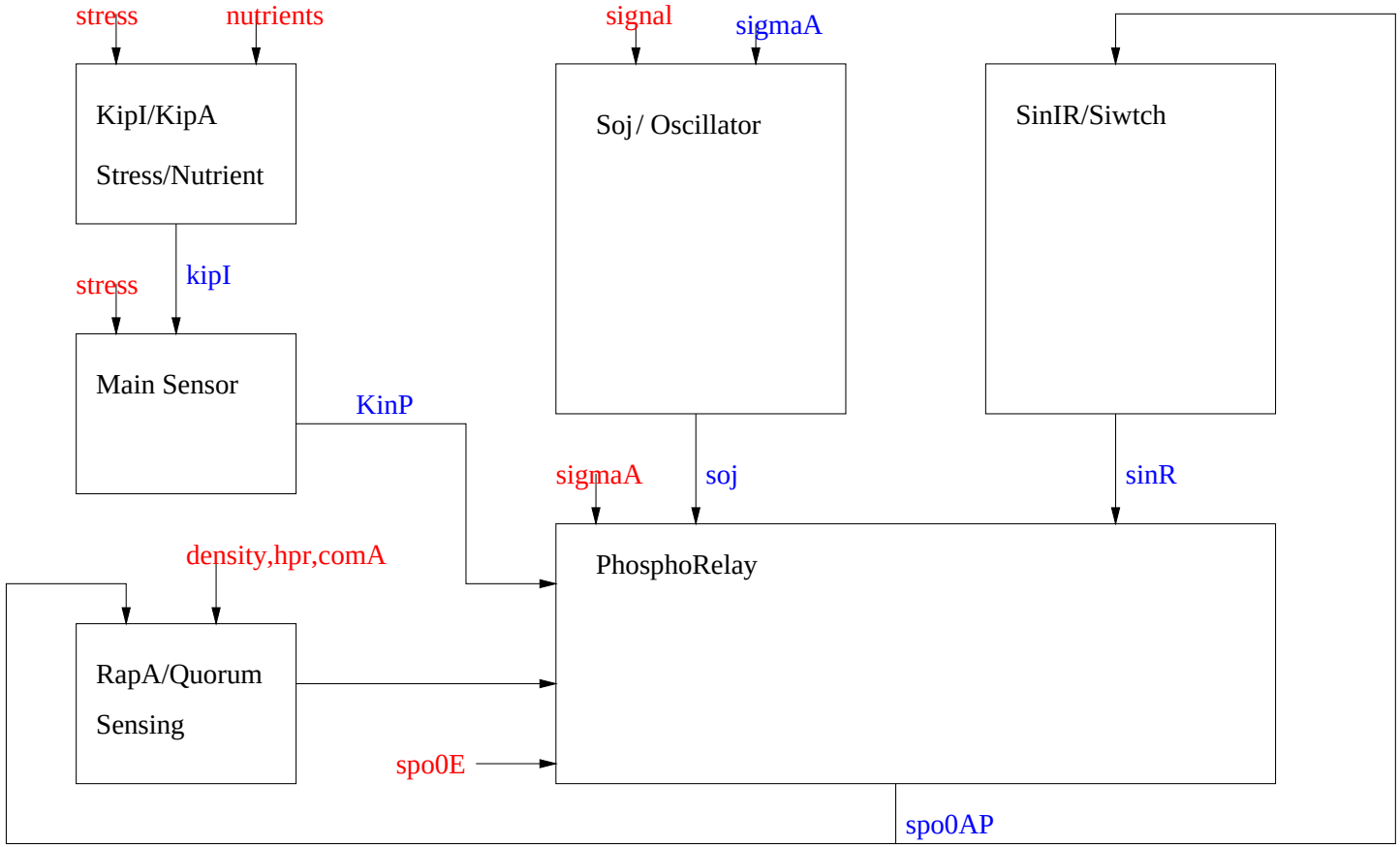
But only for certain values of the parameters

These are derived as constraints using a refinement paradigm

$$\Delta_I, \Delta_R > 0$$
$$\lambda_I = \lambda_R > 0$$

Our View

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



Analysis: B.Sub I

Model Refinement: Unknown parameters need to be constrained to observe desired stable states

Given that $SinR4$ low and $Spo0AP$ medium is a stable state

A sufficient condition is that flow field vectors point inwards in each polygonal face: (box invariance)

Equivalent to a quantified formula, which is equivalent to

$$Vbm - 3 * Vba/2 - Vbmp + Vbap > 0$$

where

$Vbm, Vbmp$: forward and reverse rate for phosphate transfer from Spo0F to Spo0B

$Vba, Vbap$: forward and reverse rate for phosphate transfer from Spo0B to Spo0A

Analysis: B.Sub II

Using our abstraction + model-checking approach, we check the abstract model for stability properties

Observation: The set of states with SinR4 low and Spo0AP high cannot be a stable region.

Why?: SinR4 low, and Spo0AP is high,
implies: transcription of Spo0A is switched off
implies: concentration of Spo0A drops
implies: drop in the concentration of Spo0AP
implies: system out of the stable region

The model checkers produces this behavior

Analysis: B.Sub III

Incorrect transcriptional control logic:

Under the initial logic proposed for SinI-SinR operon, observed that system does not reach a high Spo0AP and low SinR4 state

Forced us (and our collaborating biologists) to change the logic

Analysis: B.Sub IV

SinR4 low and Spo0AP medium still not stable!

- the cell produces Spo0FP, Spo0BP, and consequently some Spo0AP (because of baseline Spo0A synthesis);
- eventually, Spo0AP goes high and AbrB gets low;
- this causes SinR4 to go low;
- but Spo0AP reverts back to low; and SigmaH goes high;
- now, AbrB becomes high, SinI drops, and SinR4 becomes high.

Reason: Stochastic and noise bahavior that is captured by the HybridAbstractor.

Reach stable region as a “transient”

Had to constrain the response rates of AbrB, Hpr, and SigmaH further

Human Blood Glucose Metabolism

- Glucose concentration remains in a narrow range throughout the day
- Glucose turnover is approx. 2 mg/kg/min (in 70 kg adult)
- Plasma glucose concentration balance between
  - intake (glucose absorption from the gut)
  - tissue utilization (glycolysis, other pathways, glycogen synthesis)
  - endogenuous production (glycogenolysis, gluconeogenesis)
- Controlled mainly by hormones: insulin and glucagon

Glucose Homeostasis Controllers

Insulin and anti-insulin hormone, glucagon,

- Secreted by pancreatic islets of Langerhans (β- and α-cells)
- In response to glucose levels
- Insulin works by promoting
  - uptake of glucose into tissues,
  - glycogen synthesis
  - intracellular glucose metabolism

  Glucagon:
  - glycogen $\rightarrow$ glucose
  - glucose synthesis

Modeling Human Blood Glucose Metabolism

Different level of details

- Organ/tissue-level compartmental model
- Intracellular signaling and response model

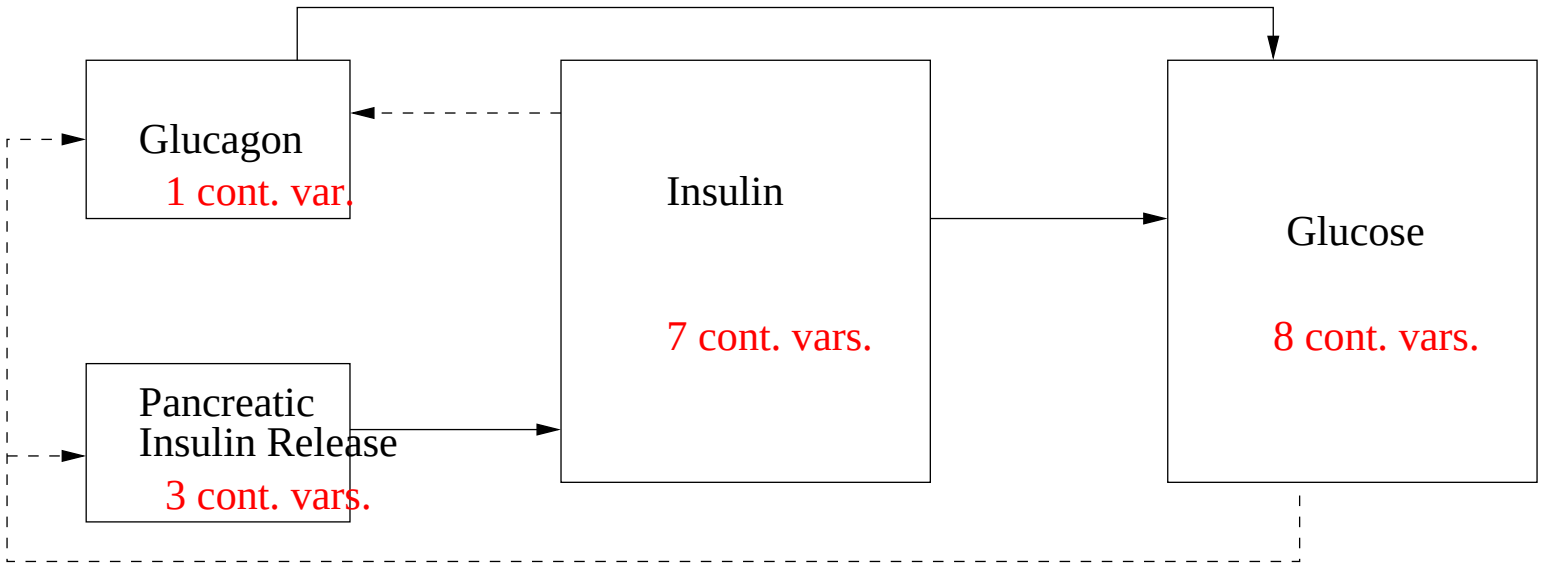
Organ-level compartmental model:

Inputs: Meals, Exercises, Injected Insulin,

Outputs: Glucose and Insulin concentration in different organs

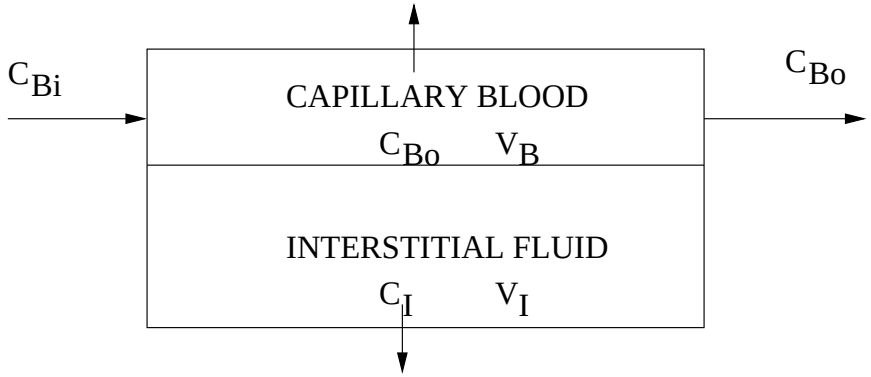
Model: Human Blood Glucose Metabolism

A module each for Glucose, Insulin, Glucagon, and Pancreas Insulin Response.



Dynamical System Model

Modeling blood 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

Modeling convection, diffusion, and metabolic sink.

Discrete Components in the Model

$\mathrm{r_{KGE}}$: kidney glucose excretion rate

Continuous Model:

$$\mathrm{r_{KGE}} = \begin{cases} 71 + 71\tanh[0.11(G_K - 460)] & 0 \leq G_K \leq 460\mathrm{mg/dl} \\ -330 + 0.872G_K & G_K \geq 460\mathrm{mg/dl} \end{cases}$$

Discrete Model:

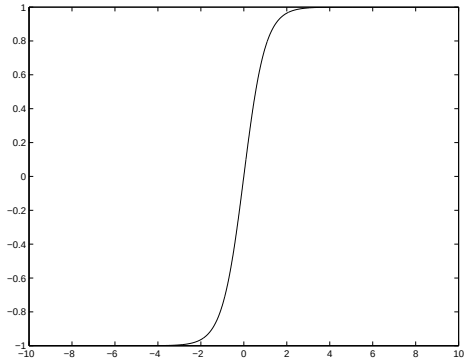
$$\mathrm{r_{KGE}} = \begin{cases} 0 & 0 \leq G_K \leq 460\mathrm{mg/dl} \\ -330 + 0.872G_K & G_K \geq 460\mathrm{mg/dl} \end{cases}$$

Faulty modes in biology indeed behave like discrete switches

Discrete Components of Model

As conservative models of the experimental data

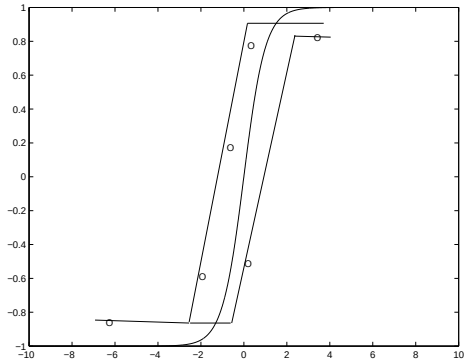
Experimental Data $\longrightarrow$ Model — Modeling Formalism



Continuous

Data Points

Hybrid



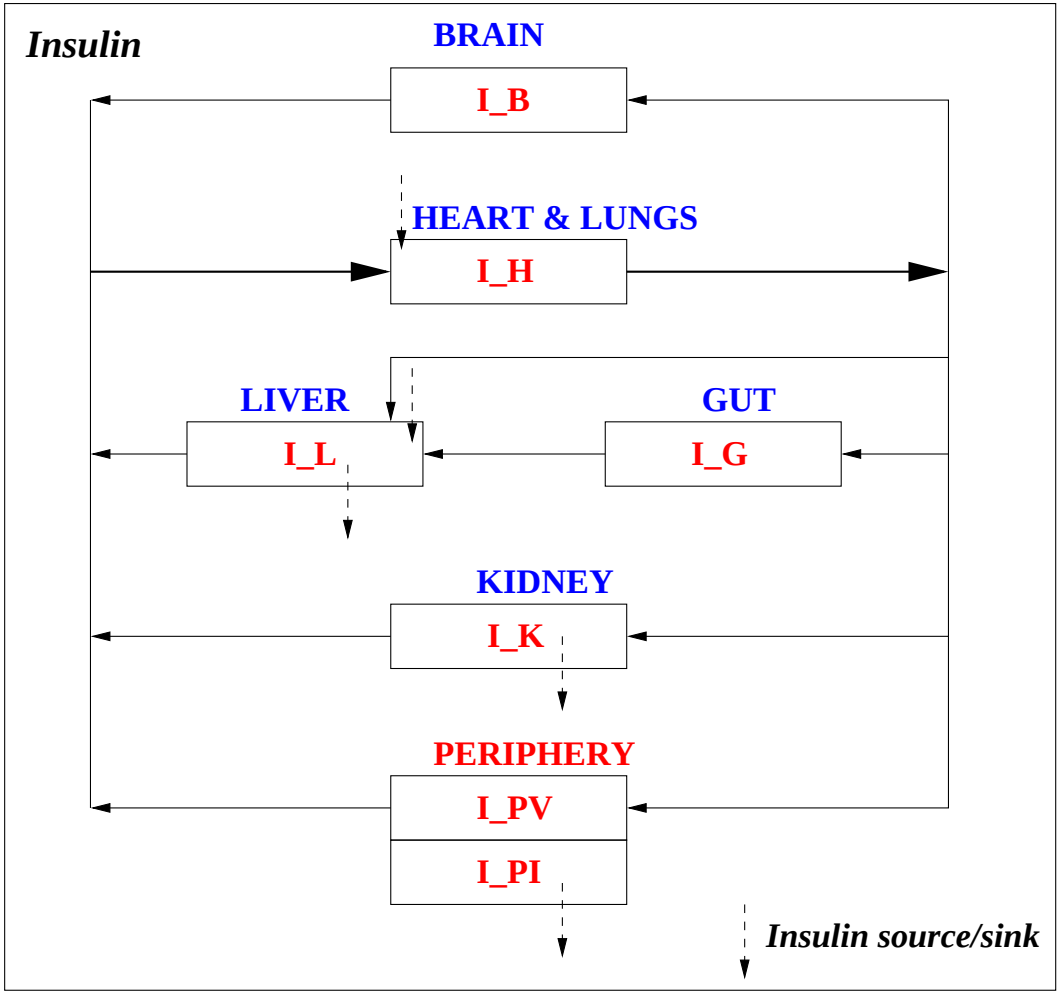
(Levels of Abstractions)

Modeling

Human blood glucose metabolism:

- The body is divided into six compartments: brain, heart and lungs, gut, liver, kidney, periphery
- Each compartment is modeled using the generic framework (interstitial space (diffusion) neglected often)
- Mostly linear dynamics except
  - Pancreas Insulin Response (PIR)
  - Certain metabolic uptake and sinks (all modeled using sigmoidal functions)
- Type-I diabetic patient modeled by eliminating PIR
- Treatment via insulin injection: extra input term in the equations for heart and lung compartment

The Insulin Component



Analysis: Glucose I

Observations:

- Glucagon concentration changes very slowly
- If change in environment, first insulin concentrations stabilize, followed by glucose concentrations
- If insulin conc. drop, then glucose conc. increase, and vice-versa.
- Insulin module can be analyzed in isolation

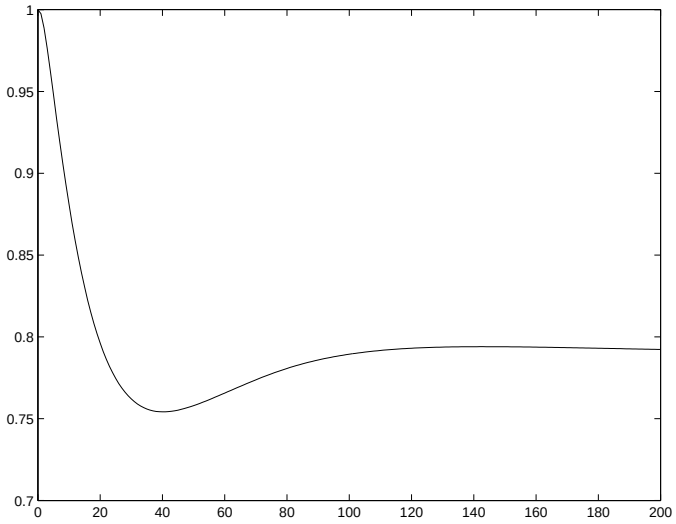
Insulin Module:

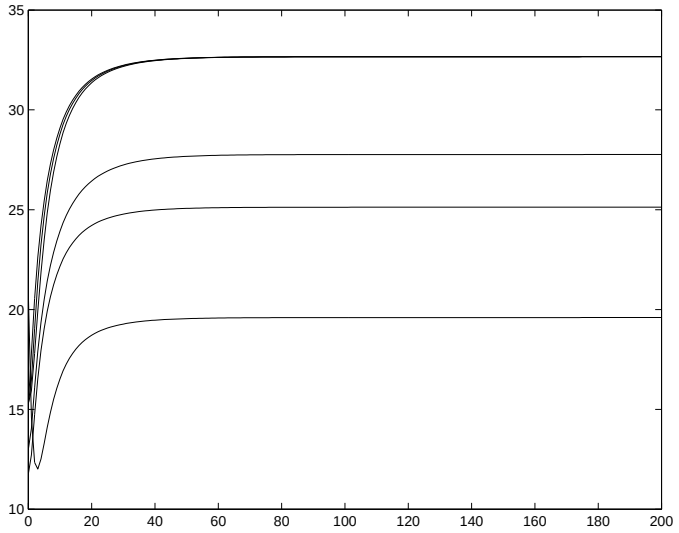
Input: Insulin injection

Output: Insulin concentrations in various compartments

Analysis: Glucose II

If the rate of insulin injection is changed, say from 20 units/sec to 25 units/sec, what is the maximum ranges of fluctuation in the concentrations of insulin in various compartments.





Approximate Reachability for Linear Systems

Linear system $\dot{\vec{x}} = A\vec{x}$

Let $\vec{r}$ be an eigenvector of A^T corresponding to eigenvalue $k < 0$

Consider $f(\vec{x}) = \vec{r}^T\vec{x}$

Question: What is $\dot{f}$?

$$\dot{f} \;=\; \vec{r}^T\vec{x} \;=\; \vec{r}^T A\vec{x} \;=\; k\vec{r}^T\vec{x}$$

$\therefore$ in all reachable states, f lies between 0 and $f(\vec{(0)})$

This gives an over-approximation of the reach set

This analysis can be extended to complex eigenvalues too

Analysis: Glucose III

Using approximate reachability computation for linear systems techniques, we can automatically compute:

Insulin Concentration	stable1	stable2	Reachable Range
Brain	26	33	26–38
Heart	26	33	26–35
Gut	26	33	25–34
Liver	14	18	8–19
Kidney	23	29	23–34
Peri. Vascul.	20	20	20–29
Peri. Interst.	2.2	2.8	1–3

Analysis: Glucose IV

- The worst case insulin concentrations can be used to determine ranges for the glucose concentrations
- To verify that blood glucose concentration remains within 80–120 mg/dl.
- This can be used to design or verify controllers of pumps for automatic insulin injection
- This analysis is tolerant to minor changes in parameter values

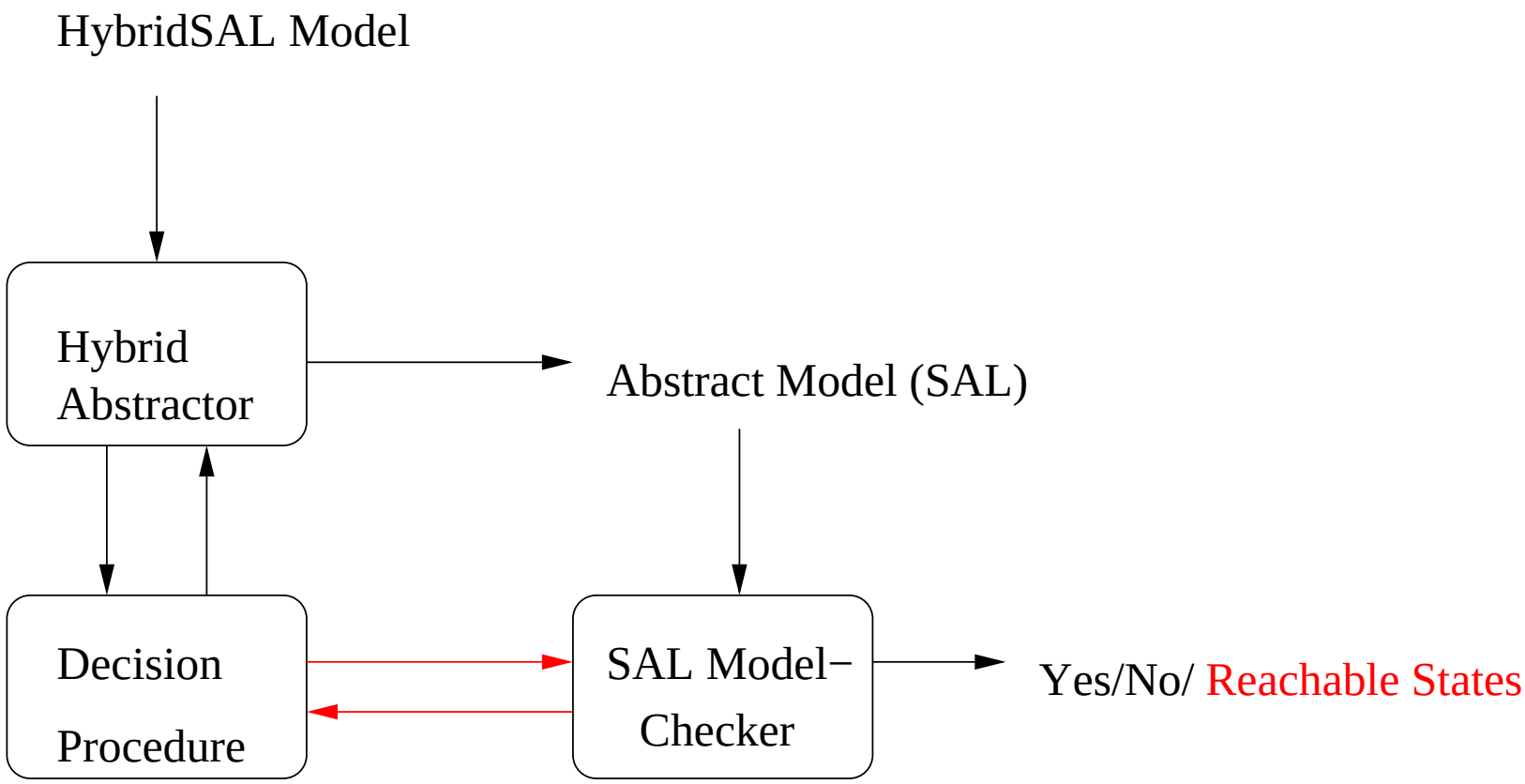
The HybridSal Abstractor

- 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

Summary

- Modeling biological systems is challenging: what level of abstraction, what question do you want to ask
- Hybrid systems provides a rich language for modeling such systems
- Unknown parameters and noise need to be handled
- Automated analysis tools can help refine models and suggest experiments
  - Models are robust: parameters, noise
  - Techniques based on abstraction well-suited
  - Abstraction verifies robustness!

Summary: Analysis Techniques

- Qualitative abstraction: for creating sound discrete abstractions of hybrid systems
- Approximating reach sets using structural analysis of the differential equations:
  linear algebra, algebraic and differential geometry
- Decision procedure for reals:
  - Abstraction based theorem proving: *Theorem provers should abstract*
  - Sound, but incomplete hierarchies of decision procedures based on *abstraction, refinement*
- Model reduction
- Model refinement