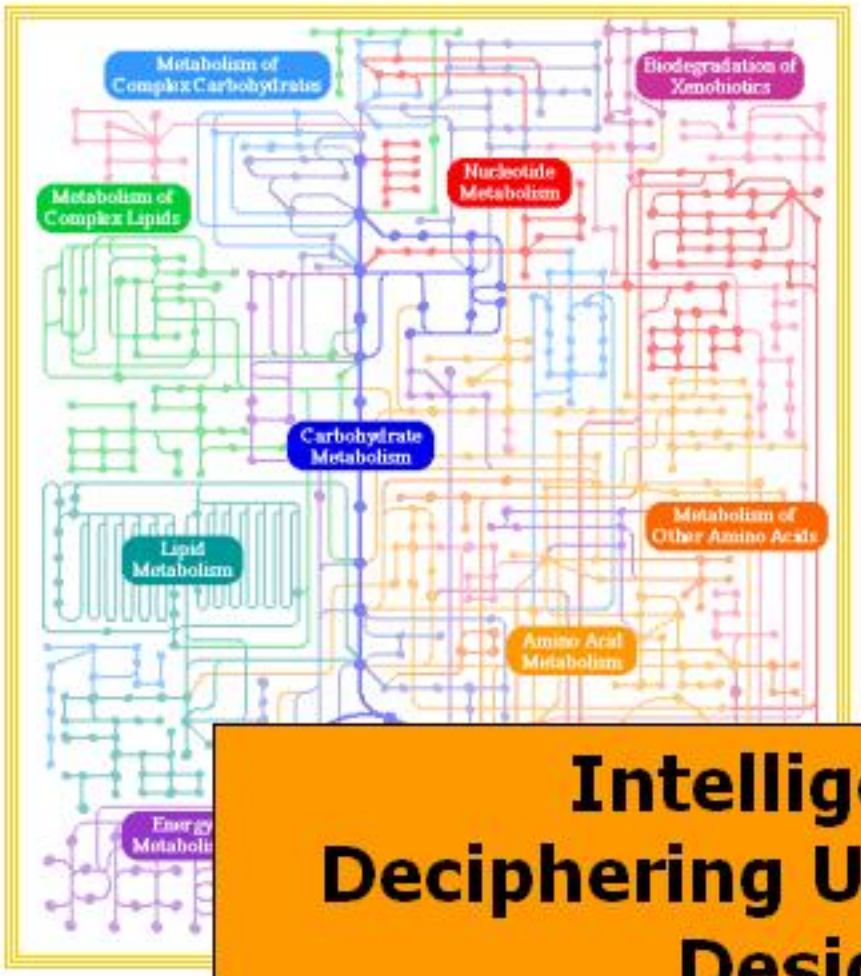


Intelligently Deciphering Unintelligible Design

Bud Mishra

Professor of Computer Science, Mathematics and Cell Biology

◇

Courant Institute, NYU School of Medicine, Tata Institute of Fundamental Research, and Mt. Sinai School of Medicine

Robert Hooke

- Robert Hooke (1635-1703) was an **experimental scientist**, **mathematician**, **architect**, and **astronomer**. Secretary of the Royal Society from 1677 to 1682, ...
- Hooke was considered the "**England's Da Vinci**" because of his wide range of interests.
- His work **Micrographia** of 1665 contained his microscopical investigations, which included the first identification of biological cells.
- In his drafts of Book II, Newton had referred to him as the most illustrious Hooke—"**Cl[arissimus] Hookius.**"
- Hooke became involved in a dispute with Isaac Newton over the priority of the discovery of the inverse square law of gravitation.

Hooke to Halley

- "[Huygen's Preface] is concerning those properties of gravity which I myself first discovered and showed to this Society and years since, which of late Mr. Newton has done me the favour to print and publish as his own inventions."

Newton to Halley

- "Now is this not very fine? Mathematicians that find out, settle & do all the business must content themselves with being nothing but dry calculators & drudges & another that does nothing but pretend & grasp at all things must carry away all the inventions...
- "**I beleive you would think him a man of a strange unsociable temper.**"

Newton to Hooke

- "If I have seen further than other men, it is because I have stood on the shoulders of giants and **you my dear Hooke, have not**."

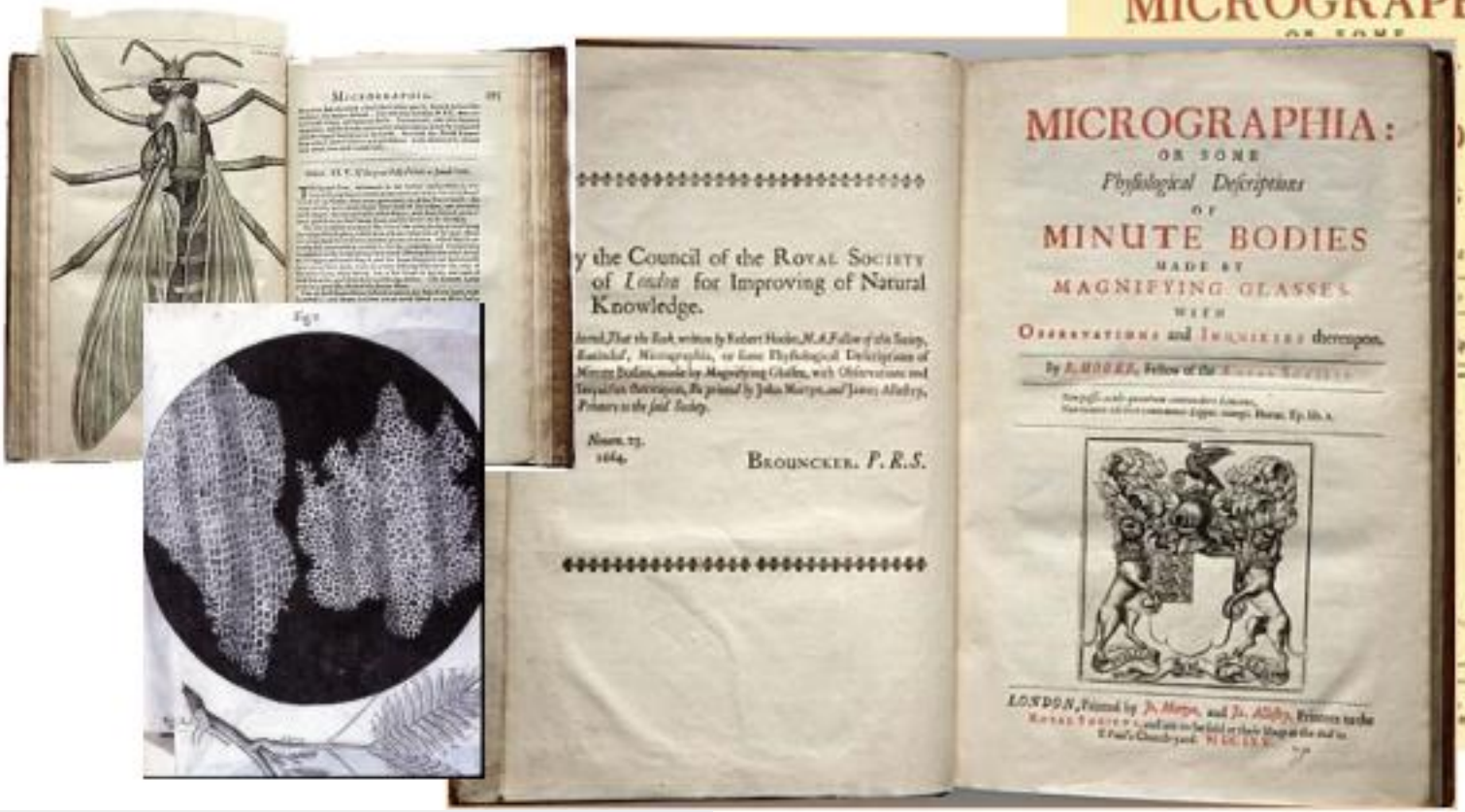
  – Newton to Hooke

Image & Logic

- The great distance between
  - a glimpsed truth and
  - a demonstrated truth
    - Christopher Wren/Alexis Claude Clairaut

Micrographia Principia

Micrographia



"The Brain & the Fancy"

- "The truth is, the science of Nature has already been too long made only a work of the brain and the fancy. It is now high time that it should return to the plainness and soundness of observations on material and obvious things."
  - Robert Hooke. (1635 - 1703), *Micrographia* 1665

Principia

PHILOSOPHIÆ
NATURALIS
PRINCIPIA
MATHEMATICA.
AUCTORE
ISAACO NEWTONO, Eq. Auratq.

Perpetuis Commentariis illustrata, communi studio
PP. Thomæ Le Seur & Francisci Jacquier
Ex Gallicanâ Minimorum Familiâ,
Matheseos Professorum.

TOMUS PRIMUS

Typis Barrillot & Filii Bibliop. & Typogr.

LAW I.

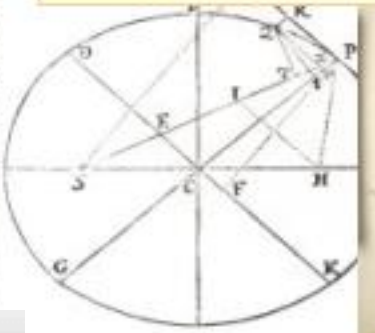
That every body perseveres in its state of resting, or of moving uniformly in a right line, as far as it is not compelled to change that state by external forces impressed upon it.

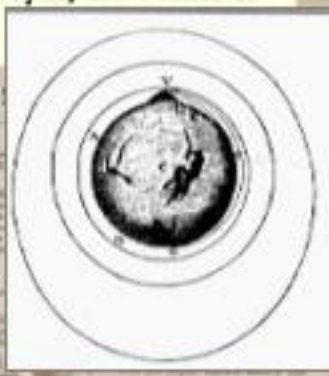
LAW II.

That the change of motion is proportional to the moving force impressed; and is produced in the direction of the right line, in which that force is impressed.

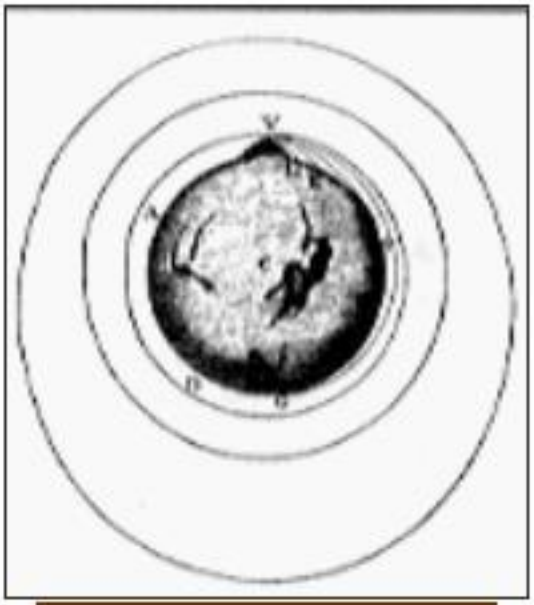
LAW III.

That reaction is always contrary and equal to action: or, that the mutual actions of two bodies upon each other are always equal, and directed to contrary parts.





"Induction & Hypothesis"



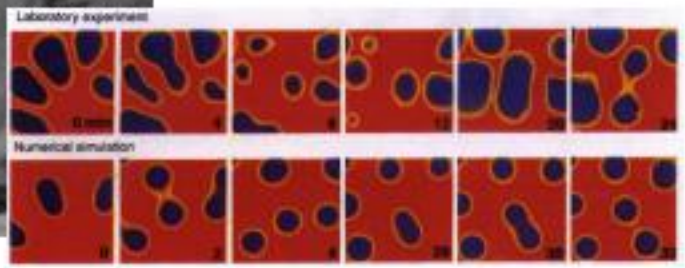
Hypotheses non fingo.
I feign no hypotheses.
Principia Mathematica.

- "Truth being uniform and always the same, it is admirable to observe how easily we are enabled to make out very abstruse and difficult matters, when once true and genuine Principles are obtained."
  - Halley, "The true Theory of the Tides, extracted from that admired Treatise of Mr. Issac Newton, Intituled, Philosophiae Naturalis Principia Mathematica," *Phil. Trans.* 226:445,447.
- This rule we must follow, that the argument of induction may not be evaded by hypotheses.

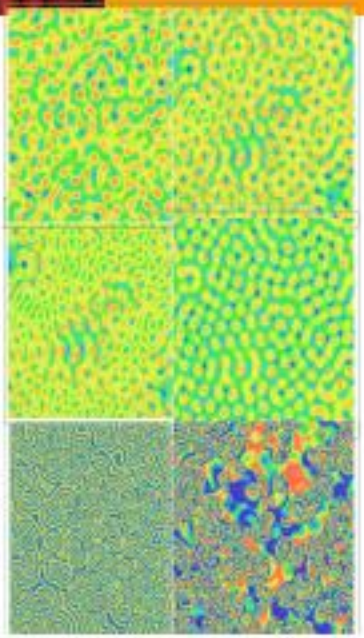
Morphogenesis

Alan Turing: 1952

- "The Chemical Basis of Morphogenesis," 1952, *Phil. Trans. Roy. Soc. of London*, Series B: Biological Sciences, **237**:37—72.
- ***A reaction-diffusion model for development.***



"A mathematical model for the growing embryo."



- A very general program for modeling embryogenesis: The `model' is "a simplification and an idealization and consequently a falsification."
- Morphogen: "is simply the kind of substance concerned in this theory..." in fact, anything that diffuses into the tissue and "somehow persuades it to develop along different lines from those which would have been followed in its absence" qualifies.

Diffusion equation

first temporal derivative: rate

$$\partial a / \partial t = D_a \nabla^2 a$$

second spatial derivative: flux

a: concentration

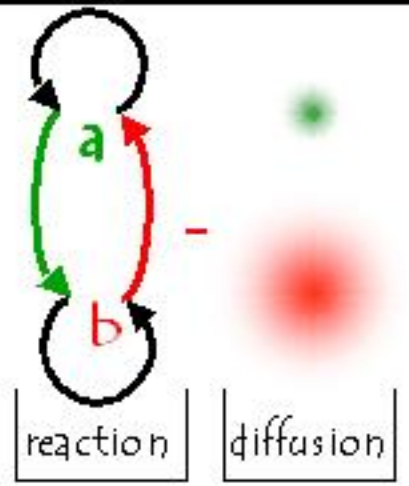
D_a: diffusion constant

Reaction-Diffusion

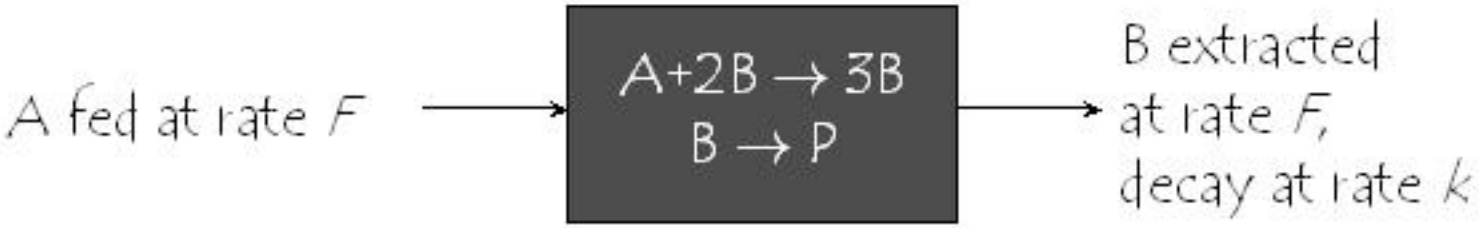
$$\partial a/\partial t = f(a,b) + D_a \nabla^2 a \qquad f(a,b) = a(b-1) - k_1$$

$$\partial b/\partial t = g(a,b) + D_b \nabla^2 b \qquad g(a,b) = -ab + k_2$$

Turing, A.M. (1952). "The chemical basis of morphogenesis." *Phil. Trans. Roy. Soc. London B* **237**: 37



Reaction-diffusion: an example



$d[A]/dt = F(1-[A])$

$d[B]/dt = -(F+k)[B]$

reaction: $-d[A]/dt = d[B]/dt = [A][B]^2$

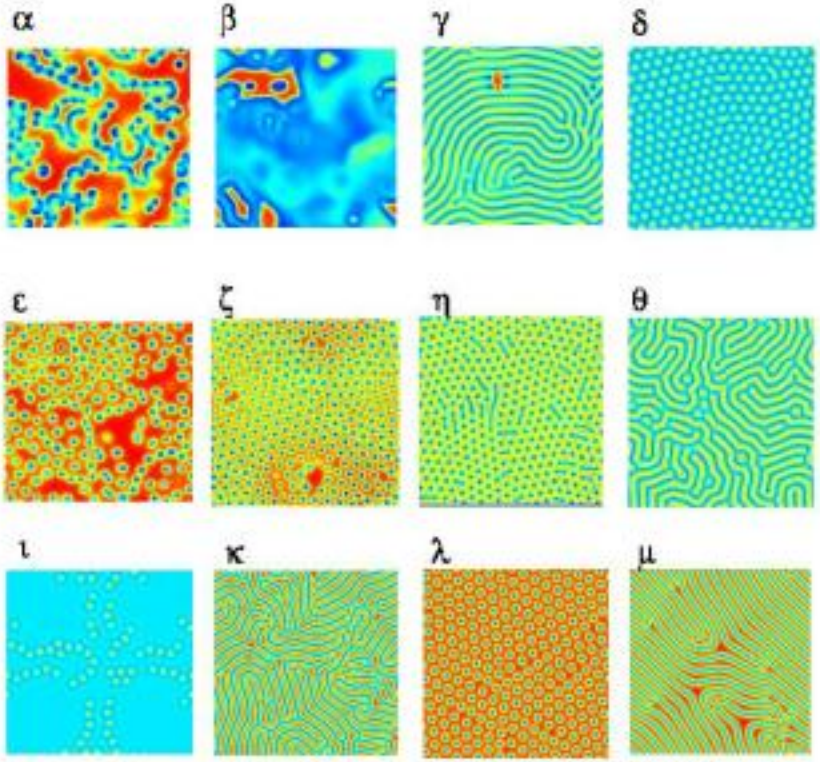
diffusion: $d[A]/dt = D_A \nabla^2[A]$; $d[B]/dt = D_B \nabla^2[B]$

$$\partial [A]/\partial t = F(1-[A]) - [A][B]^2 + D_A \nabla^2[A]$$

$$\partial [B]/\partial t = -(F+k)[B] + [A][B]^2 + D_B \nabla^2[B]$$

Pearson, J. E.: Complex patterns in simple systems. *Science* **261**, 189–192 (1993).

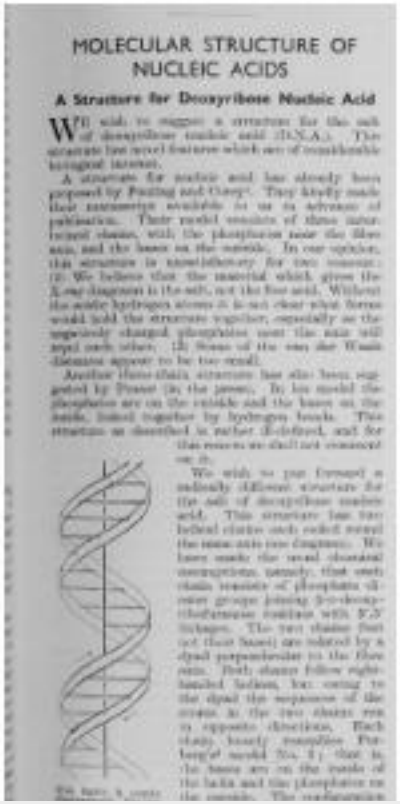
Reaction-diffusion: an example



Genes: 1952

- Since the role of genes is presumably catalytic, influencing only the rate of reactions, unless one is interested in comparison of organisms, they **"may be eliminated from the discussion..."**

Crick & Watson :1953



MOLECULAR STRUCTURE OF NUCLEIC ACIDS

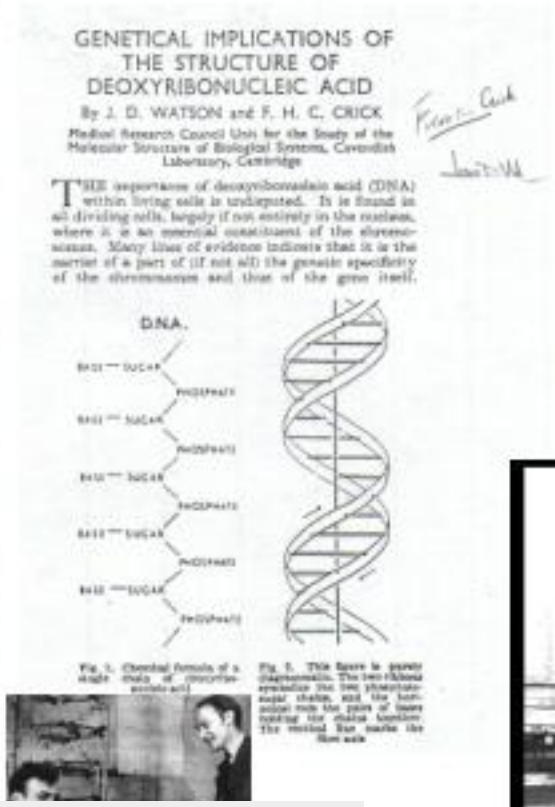
A Structure for Deoxyribose Nucleic Acid

We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagram is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3',5' linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration



GENETICAL IMPLICATIONS OF THE STRUCTURE OF DEOXYRIBONUCLEIC ACID

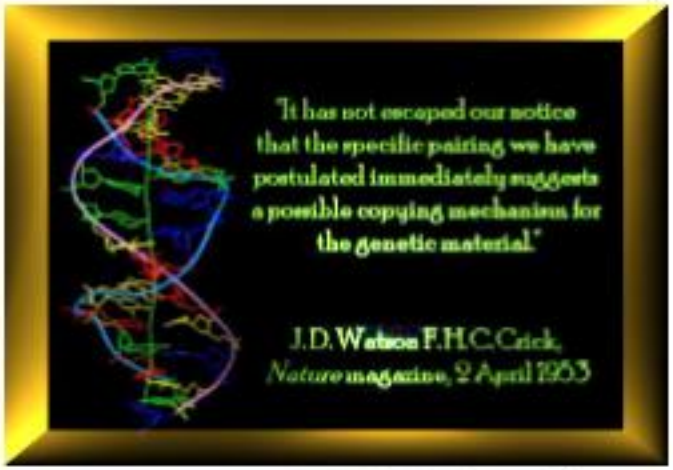
By J. D. WATSON and F. H. C. CRICK

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge

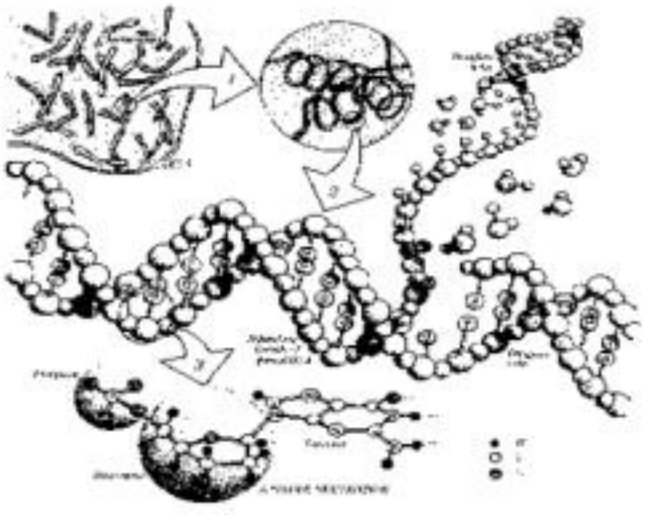
THE importance of deoxyribonucleic acid (DNA) within living cells is undisputed. It is found in all dividing cells, largely if not entirely in the nucleus, where it is an essential constituent of the chromosomes. Many lines of evidence indicate that it is the carrier of a part of (if not all) the genetic specificity of the chromosomes and thus of the gene itself.

Fig. 1. Chemical formula of a single chain of deoxyribonucleic acid

Fig. 2. This figure is purely diagrammatic. The two ribbons symbolize the two phosphate-sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis



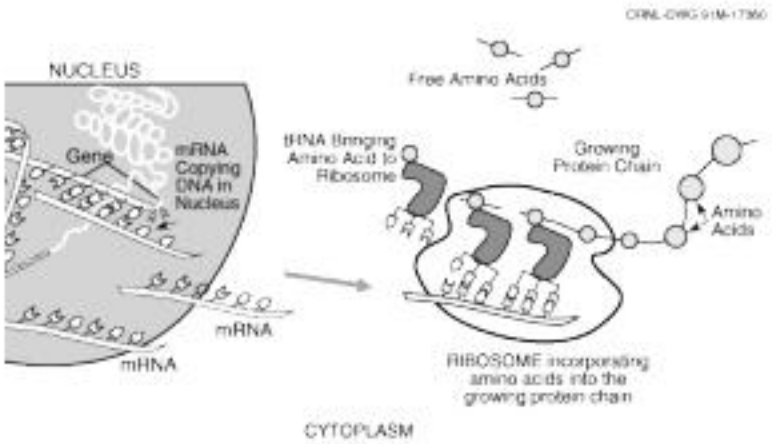
Genome



- **Genome:**
  - Hereditary information of an organism is encoded in its DNA and enclosed in a cell (unless it is a virus). All the information contained in the DNA of a single organism is its *genome*.
- DNA molecule can be thought of as a *very* long sequence of **nucleotides** or **bases**:

$$\Sigma = \{A, T, C, G\}$$

The Central Dogma



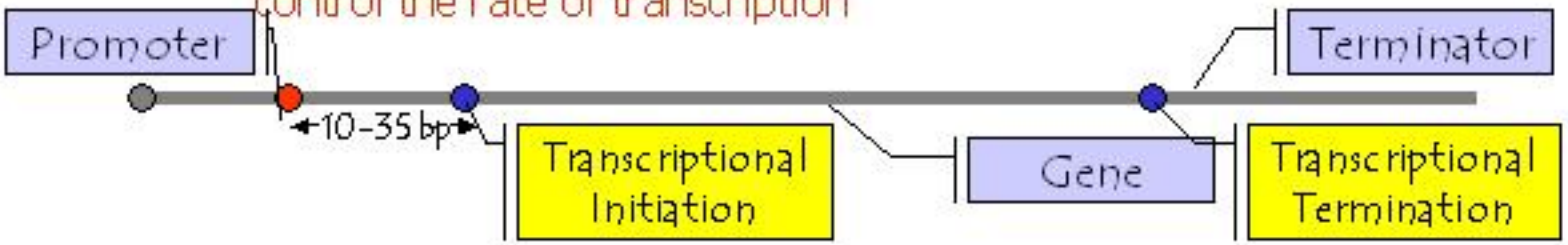
- The central dogma(due to Francis Crick in 1958) states that these information flows are all unidirectional:

"The central dogma states that once 'information' has passed into protein it cannot get out again. The transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein, may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible. Information means here the precise determination of sequence, either of bases in the nucleic acid or of amino acid residues in the protein."

DNA → Transcription → RNA → Translation → Protein

RNA, Genes and Promoters

- A specific region of DNA that determines the synthesis of proteins (through the transcription and translation) is called a **gene**
  - Originally, a gene meant something more abstract---a unit of hereditary inheritance.
  - Now a gene has been given a physical molecular existence.
- Transcription of a gene to a **messenger RNA, mRNA**, is keyed by a transcriptional activator/factor, which attaches to a **promoter** (a specific sequence adjacent to the gene).
- Regulatory sequences such as **silencers** and **enhancers** control the rate of transcription



"The Brain & the Fancy"

"Work on the mathematics of growth as opposed to the statistical description and comparison of growth, seems to me to have developed along two equally unprofitable lines... It is futile to conjure up in the imagination a system of differential equations for the purpose of accounting for facts which are not only very complex, but largely unknown,...What we require at the present time is more measurement and less theory."

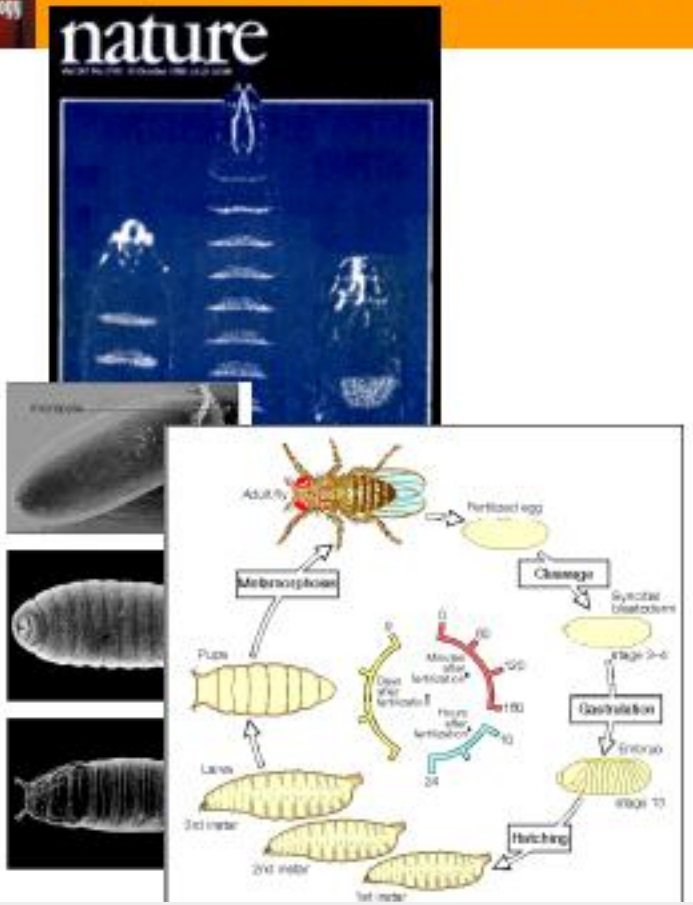
– Eric Ponder, Director, CSHL (LIBA), 1936-

"Axioms of Platitudes"

-E.B. Wilson

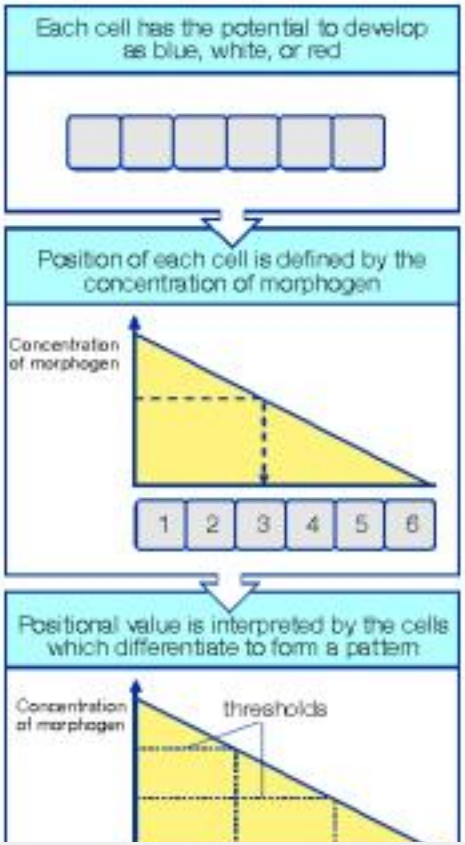
1. Science need not be mathematical.
2. Simply because a subject is mathematical it need not therefore be scientific.
3. Empirical curve fitting may be without other than classificatory significance.
4. Growth of an individual should not be confused with the growth of an aggregate (or average) of individuals.
5. Different aspects of the individual, or of the average, may have different types of growth curves.

Genes for Segmentation



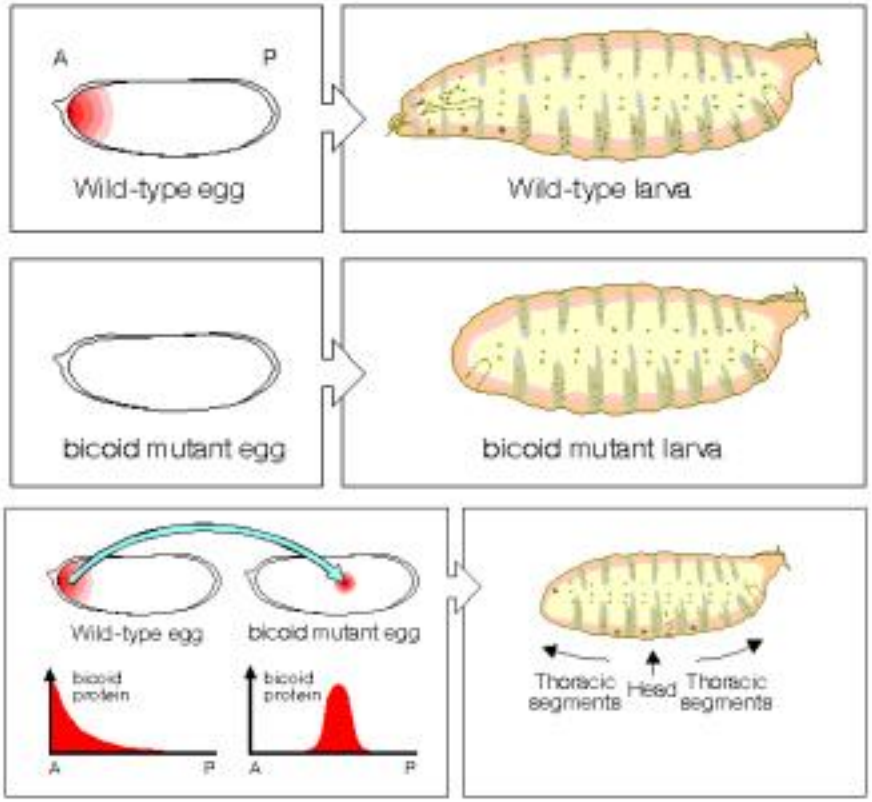
- Fertilization followed by cell division
- Pattern formation – instructions for
  - Body plan (Axes: A-P, D-V)
  - Germ layers (ecto-, meso-, endoderm)
- Cell movement - form – gastrulation
- Cell differentiation

PI: Positional Information



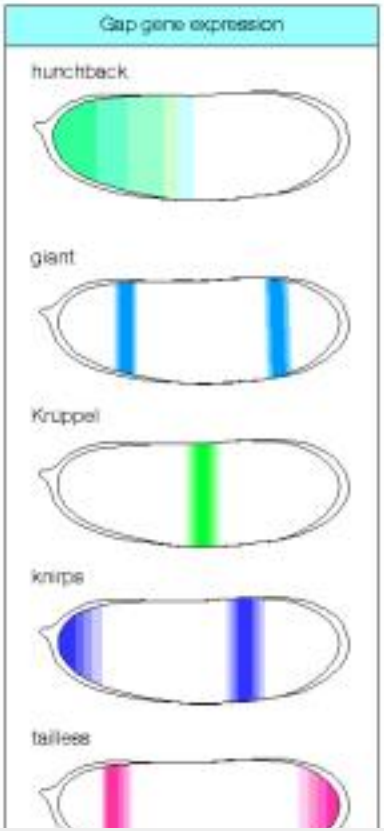
- Positional value
  - Morphogen – a substance
  - Threshold concentration
- Program for development
  - Generative rather than descriptive
- **"French-Flag Model"**

bicoid



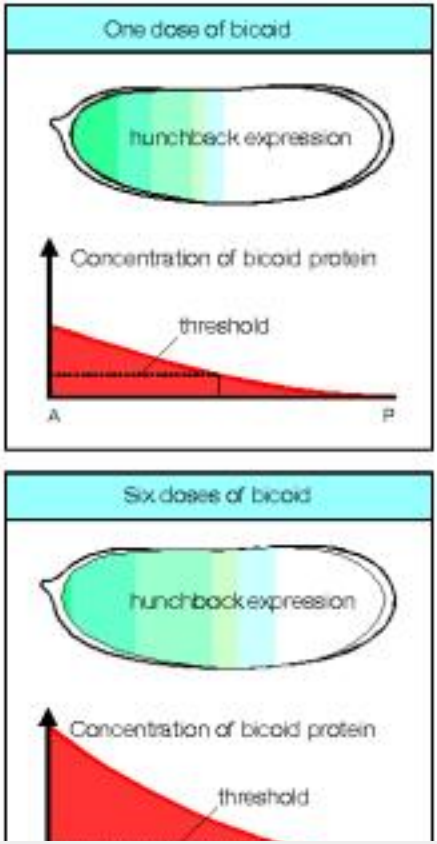
- The *bicoid* gene provides an A-P morphogen gradient

gap genes



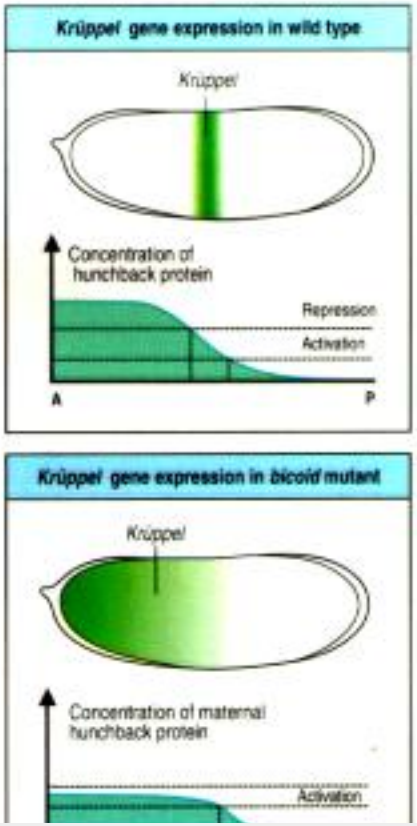
- The A-P axis is divided into broad regions by gap gene expression
- The first *zygotic* genes
- Respond to maternally-derived instructions
- Short-lived proteins, gives bell-shaped distribution from source

Transcription Factors in Cascade



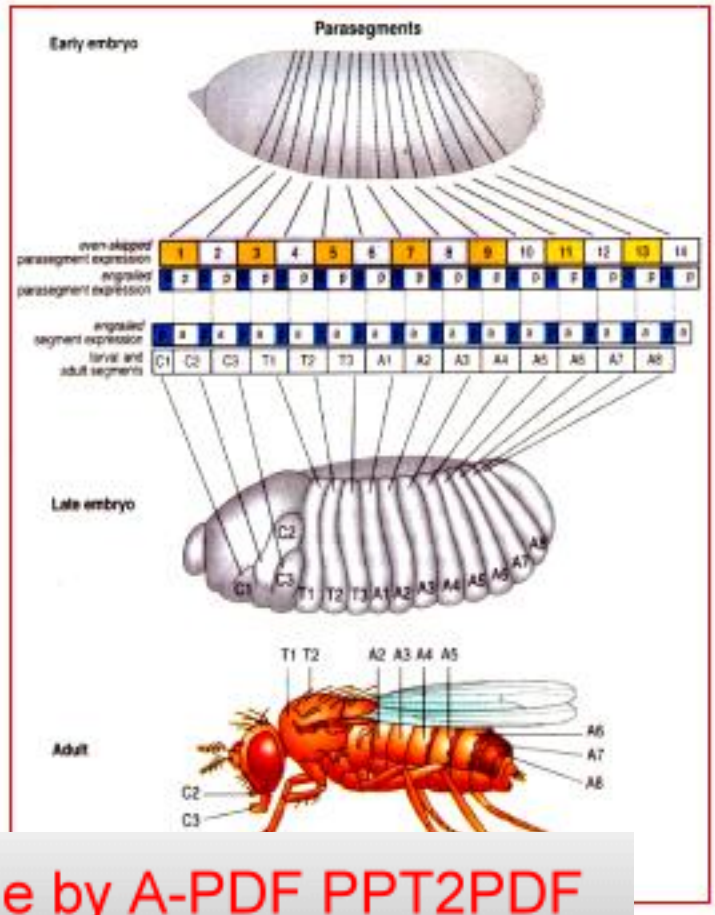
- *Hunchback (hb)* , a gap gene, responds to the dose of bicoid protein
- A concentration **above threshold** of bicoid activates the expression of *hb*
- The more *bicoid* transcripts, the further **back** *hb* expression goes

Transcription Factors in Cascade



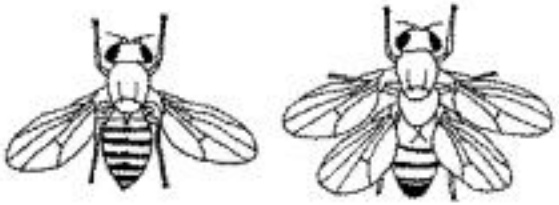
- *Krüppel (Kr)*, a gap gene, responds to the dose of hb protein
- A concentration **above minimum threshold** of hb activates the expression of *Kr*
- A concentration **above maximum threshold** of hb inactivates the expression of *Kr*

Segmentation



- Parasegments are delimited by expression of pair-rule genes in a periodic pattern
- Each is expressed in a series of 7 transverse stripes

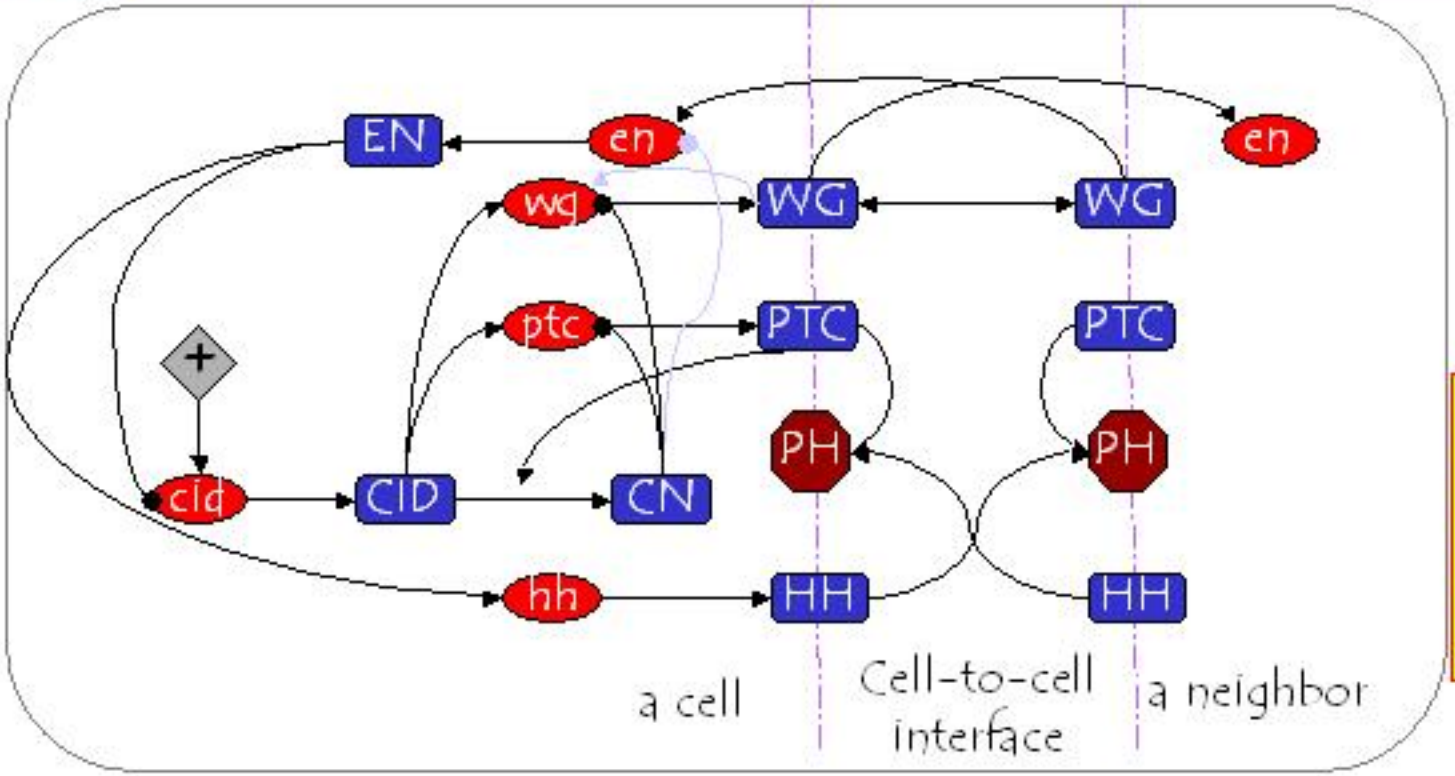
Pattern Formation



- Edward Lewis, of the California Institute of Technology
- Christiane Nuesslein-Volhard, of Germany's Max-Planck Institute
- Eric Wieschaus, at Princeton

- Each of the three were involved in the early research to find the genes controlling development of the *Drosophila* fruit fly.

The Network of Interaction



Completeness

- "We incorporated these two remedies first (light gray lines). With these links installed there are many parameter sets that enable the model to reproduce the target behavior, **so many that they can be found easily by random sampling.**"

Model Parameters

Parameter	Meaning	Realistic (General) Range	Range used for SP Model
κ	half-maximal activation coefficient	10^{-3} – 10	10^{-3} – 1
H	half-life (inverse of degradation rate)	1 – 10^{4} min. (for mRNA or protein)	5 – 100 min
ν	Hill coefficient	1 – 50 (highest measured is 35)	1 – 10
α	saturability coefficient for an enhancer	0.1 – 10	1 – 10
transfer rates	how much reaction occurs per unit time	10^{-3} – 10	10^{-3} – 1.0
transform rates	ditto; but for cleavage, phosphorylation, etc.	10^{-3} – 10	10^{-3} – 10

Complete Model

Notation: $X_{n,j+\frac{1}{2}}$ = amount of X on apposite cell face; $X_{i,T} = \sum_{j=1}^{b} X_{i,j}$; $X_{n,T} = \sum_{j=1}^{b} X_{n,j+\frac{1}{2}}$; $X_{i,b} = X_{i,j-\frac{1}{2}} + X_{i,j+\frac{1}{2}}$

a) $$\frac{d\,en_i}{d\tau} = \frac{T_r}{H_{en}}\left(\frac{EWG_{n,T}\left(1-\frac{CN_i^{\nu_{CNen}}}{\kappa_{CNen}^{\nu_{CNen}}+CN_i^{\nu_{CNen}}}\right)^{n_{WGen}}}{\kappa_{WGen}^{n_{WGen}}+EWG_{n,T}\left(1-\frac{CN_i^{\nu_{CNen}}}{\kappa_{CNen}^{\nu_{CNen}}+CN_i^{\nu_{CNen}}}\right)^{n_{WGen}}}-en_i\right)$$

b) $$\frac{d\,EN_i}{d\tau} = \frac{T_r}{H_{EN}}\left(en_i - EN_i\right)$$

c) $$\frac{d\,wg_i}{d\tau} = \frac{T_r}{H_{wg}}\left(\frac{\alpha_{CNwg}\cdot\left(\frac{CI_i\left(1-\frac{CN_i^{\nu_{CNwg}}}{\kappa_{CNwg}^{\nu_{CNwg}}+CN_i^{\nu_{CNwg}}}\right)^{\nu_{CIwg}}}{\kappa_{CIwg}^{\nu_{CIwg}}+CI_i\left(1-\frac{CN_i^{\nu_{CNwg}}}{\kappa_{CNwg}^{\nu_{CNwg}}+CN_i^{\nu_{CNwg}}}\right)^{\nu_{CIwg}}}\right)+\alpha_{WGwg}\cdot\left(\frac{IWG_i^{n_{WGwg}}}{\kappa_{WGwg}^{n_{WGwg}}+IWG_i^{n_{WGwg}}}\right)}{1+\alpha_{CIwg}\cdot\left(\frac{CI_i\left(1-\frac{CN_i^{\nu_{CNwg}}}{\kappa_{CNwg}^{\nu_{CNwg}}+CN_i^{\nu_{CNwg}}}\right)^{\nu_{CIwg}}}{\kappa_{CIwg}^{\nu_{CIwg}}+CI_i\left(1-\frac{CN_i^{\nu_{CNwg}}}{\kappa_{CNwg}^{\nu_{CNwg}}+CN_i^{\nu_{CNwg}}}\right)^{\nu_{CIwg}}}\right)+\alpha_{WGwg}\cdot\left(\frac{IWG_i^{n_{WGwg}}}{\kappa_{WGwg}^{n_{WGwg}}+IWG_i^{n_{WGwg}}}\right)}-wg_i\right)$$

d) $$\frac{d\,IWG_i}{d\tau} = \frac{T_r}{H_{IWG}}\left(wg_i - IWG_i\right) + T_r\left(r_{EnLinWG}EWG_{i,T} - r_{EnWG}IWG_i\right)$$

e) $$\frac{d\,EWG_{i,j}}{d\tau} = T_r\left(\frac{r_{EnWG}IWG_i}{6} - r_{EnLinWG}EWG_{i,j} - r_{MigInWG}EWG_{i,j} + r_{MigInWG}EWG_{n,j+\frac{1}{2}} - 2r_{LMigInWG}EWG_{i,j} + r_{LMigInWG}EWG_{i,b}\right) - \frac{T_r EWG_{i,j}}{H_{IWG}}$$

Complete Model

$$f)\ \frac{d\,ptc_i}{d\tau} = \frac{T_s}{H_{ptc}}\left(\frac{CI_i\left(1-\frac{CN_i^{\nu_{CIptc}}}{\kappa_{CIptc}^{\nu_{CIptc}}+CN_i^{\nu_{CIptc}}}\right)^{\nu_{CIptc}}}{\kappa_{CIptc}^{\nu_{CIptc}}+CI_i\left(1-\frac{CN_i^{\nu_{CIptc}}}{\kappa_{CIptc}^{\nu_{CIptc}}+CN_i^{\nu_{CIptc}}}\right)^{\nu_{CIptc}}} - ptc_i\right)$$

$$g)\ \frac{d\,PTC_{i,j}}{d\tau} = \frac{T_s}{H_{PTC}}\left(\frac{ptc_i}{6}-PTC_{i,j}\right) - T_s k_{PTCHH}[HH]_s HH_{s,j+1}\cdot PTC_{i,j} + T_s\left(r_{LMofs+PTC}PTC_{i,j+} - 2r_{LMofs+PTC}PTC_{i,j}\right)$$

$$h)\ \frac{d\,ci_i}{d\tau} = \frac{T_s}{H_{ci}}\left(\frac{B_i\left(1-\frac{EN_i^{\nu_{Bci}}}{\kappa_{ENci}^{\nu_{Bci}}+EN_i^{\nu_{Bci}}}\right)^{\nu_{Bci}}}{\kappa_{Bci}^{\nu_{Bci}}+B_i\left(1-\frac{EN_i^{\nu_{Bci}}}{\kappa_{ENci}^{\nu_{Bci}}+EN_i^{\nu_{Bci}}}\right)^{\nu_{Bci}}} - ci_i\right)$$

$$i)\ \frac{d\,CI_i}{d\tau} = \frac{T_s}{H_{CI}}\left(ci_i - CI_i\right) - T_s C_{CI} CI_i\left(\frac{PTC_{i,T}^{\nu_{PTCCI}}}{\kappa_{PTCCI}^{\nu_{PTCCI}}+PTC_{i,T}^{\nu_{PTCCI}}}\right)$$

$$j)\ \frac{d\,CN_i}{d\tau} = T_s C_{CI} CI_i\left(\frac{PTC_{i,T}^{\nu_{PTCCI}}}{\kappa_{PTCCI}^{\nu_{PTCCI}}+PTC_{i,T}^{\nu_{PTCCI}}}\right) - \frac{T_s CN_i}{H_{CI}}$$

$$k)\ \frac{d\,hh_i}{d\tau} = \frac{T_s}{H_{hh}}\left(\frac{EN_i\left(1-\frac{CN_i^{\nu_{CNhh}}}{\kappa_{CNhh}^{\nu_{CNhh}}+CN_i^{\nu_{CNhh}}}\right)^{\nu_{ENhh}}}{\kappa_{ENhh}^{\nu_{ENhh}}+EN_i\left(1-\frac{CN_i^{\nu_{CNhh}}}{\kappa_{CNhh}^{\nu_{CNhh}}+CN_i^{\nu_{CNhh}}}\right)^{\nu_{ENhh}}} - hh_i\right)$$

$$l)\ \frac{d\,HH_{i,j}}{d\tau} = \frac{T_s}{H_{HH}}\left(\frac{hh_i}{6}-HH_{i,j}\right) - T_s k_{PTCHH}[PTC]_s PTC_{s,j+1}\cdot HH_{i,j} + T_s\left(r_{LMofs+HH}HH_{i,j+} - 2r_{LMofs+HH}HH_{i,j}\right)$$

$$\frac{d\,PH_{i,j}}{\quad} = \frac{T_s PH_{i,j}}{H_{PH}}$$

Is this your final answer?

- It is not uncommon to assume certain biological problems to have achieved a cognitive finality without rigorous justification.
- Rigorous mathematical models with automated tools for reasoning, simulation, and computation can be of enormous help to uncover
  - cognitive flaws,
  - qualitative simplification or
  - overly generalized assumptions.
- Some ideal candidates for such study would include:
  - prion hypothesis
  - cell cycle machinery
  - muscle contractility
  - processes involved in cancer (cell cycle regulation, angiogenesis, DNA repair, apoptosis, cellular senescence, tissue space modeling enzymes, etc.)
  - signal transduction pathways, and many others.

Computational Systems Biology

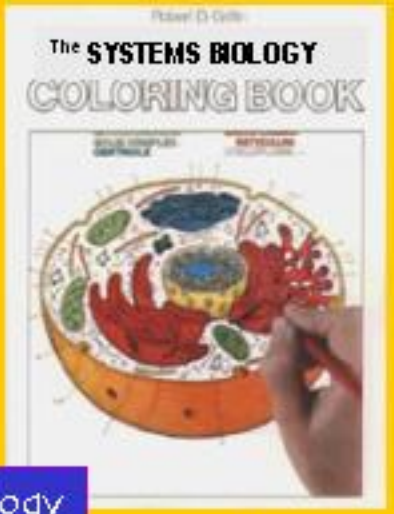
Systems Biology

Combining the mathematical rigor of numerology with the predictive power of astrology.

Numerology

Numeristan

Cyberia



Astrology

Astrostan

HOTzone

Infostan

Interpretive Biology

Integrative Biology

Computational Biology

Bioinformatics

BioSpice

Why do we need a tool?

We claim that, by drawing upon mathematical approaches developed in the context of dynamical systems, kinetic analysis, computational theory and logic, it is possible to create powerful simulation, analysis and reasoning tools for working biologists to be used in deciphering existing data, devising new experiments and ultimately, understanding functional properties of genomes, proteomes, cells, organs and organisms.

Simulate Biologists! Not Biology!!

Future Biology

Functional genomic hypothesis generation and experimentation by a robot scientist

Ross D. King[1], Kenneth E. Whelan[1], Ffion M. Jones[1], Philip G. K. Reiser[1], Christopher H. Bryant[2], Stephen H. Muggleton[3], Douglas B. Kell[4] & Stephen G. Oliver[5]

[1]Department of Computer Science, University of Wales, Aberystwyth SY23 3DB, UK
[2]School of Computing, The Robert Gordon University, Aberdeen AB10 1FR, UK
[3]Department of Computing, Imperial College, London SW7 2AZ, UK
[4]Department of Chemistry, UMIST, P.O. Box 88, Manchester M60 1QD, UK
[5]School of Biological Sciences, University of Manchester, 2.205 Stopford Building, Manchester M13 9PT, UK

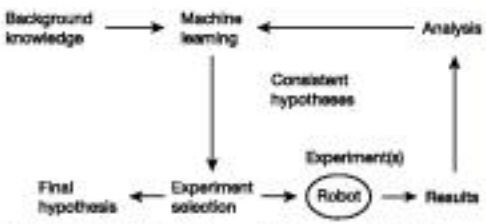
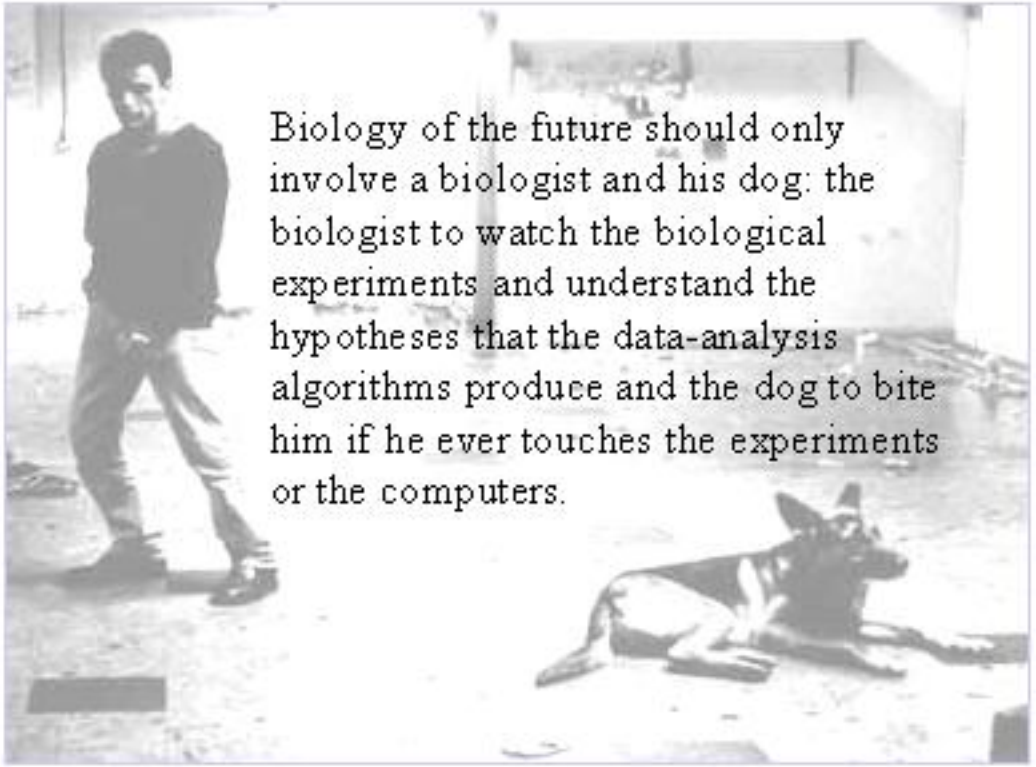


Figure 1 The Robot Scientist hypothesis-generation and experimentation loop.



Simpathica is a modular system

Canonical Form:

$$\begin{cases} \dot{X}_i = \alpha_i \prod_{j=1}^{n+m} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j^{h_{ij}} \quad i = 1 \ldots n \\ C_i(X_1(t), \ldots, X_{n+m}(t)) = \sum (\gamma_i \prod_{j=1}^{n+m} X_j^{d_{ij}}) = 0 \end{cases}$$

Characteristics:

- Predefined Modular Structure
- Automated Translation from Graphical to Mathematical Model
- Scalability

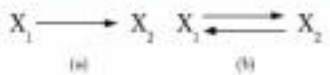


Figure 1: Representation of an unmodified and of a reversible reaction.

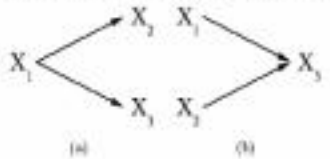


Figure 2: Representation of a divergence and of a convergence branch point (the two processes in each reaction are independent of each other).

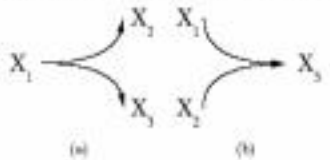


Figure 3: Representation of a single splitting reaction generating two products, X_2 and X_3, in stoichiometric proportions and of a single synthetic reaction involving two source components, X_1 and X_2 always in stoichiometric proportions.

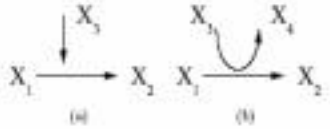
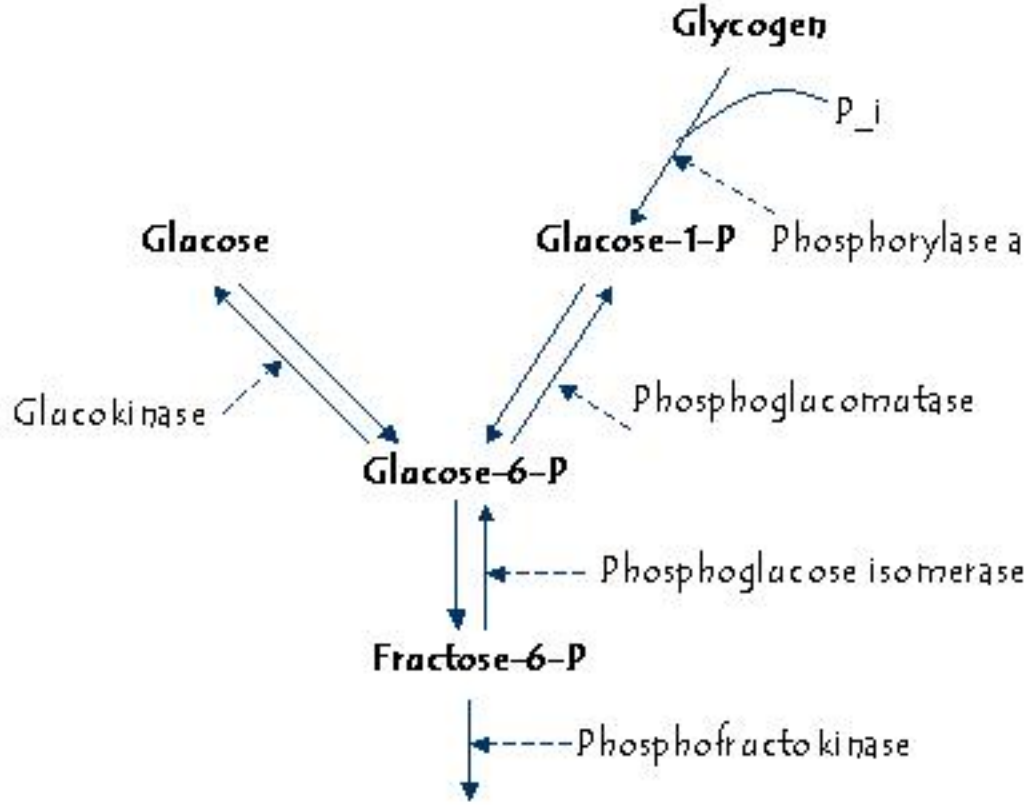


Figure 4: The conversion of X_1 into X_2 is modulated (stimulation or inhibition is represented by the sign of the arrow) by X_3. The reaction between X_1 and X_2 requires coenzyme X_3, which in the process is converted into X_4.

Glycolysis



Formal Definition of S-system

Definition 1 (S-system). *An S-system is a quadruple* $S = (DV, IV, DE, C)$ *where:*

- $DV = \{X_1, \ldots, X_n\}$ *is a finite non empty set of* dependent variables *ranging over the domains* $D_1, \ldots, D_n$, *respectively;*
- $IV = \{X_{n+1}, \ldots, X_{n+m}\}$ *is a finite set of* independent variables *ranging over the domains* $D_{n+1}, \ldots, D_{n+m}$, *respectively;*
- DE *is a set of* differential equations, *one for each dependent variable, of the form*

$$\dot{X}_i = \alpha_i \prod_{j=1}^{n+m} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j^{h_{ij}}$$

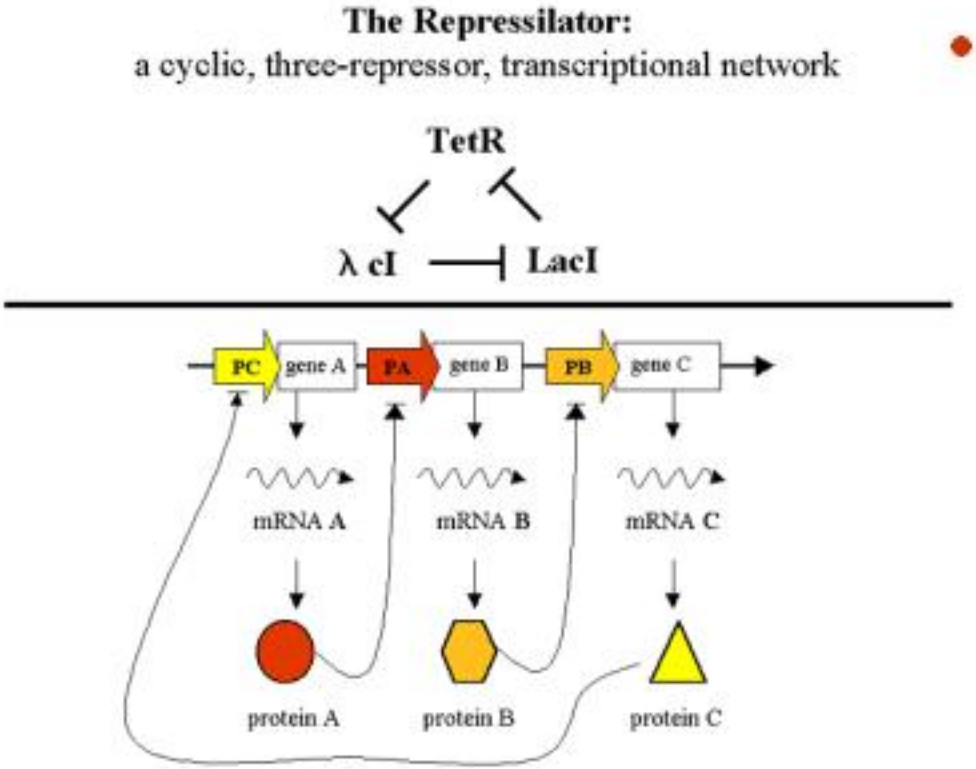
with $\alpha_i, \beta_i \geq 0$ *called* rate constants;

- C *is a set of* algebraic constraints *of the form*

$$C_j(X_1, \ldots, X_{n+m}) = \sum (\gamma_j \prod_{k=1}^{n+m} X_k^{f_{jk}}) = 0$$

with γ_j *called rate constants*

An Artificial Clock



- Three proteins:
  - LacI, tetR & λ cI
  - Arranged in a cyclic manner (logically, not necessarily physically) so that the protein product of one gene is rpressor for the next gene.

LacI→ ¬ *tetR; tetR*→ TetR

TetR→ ¬ λ *cI;* λ *cI* → λ cI

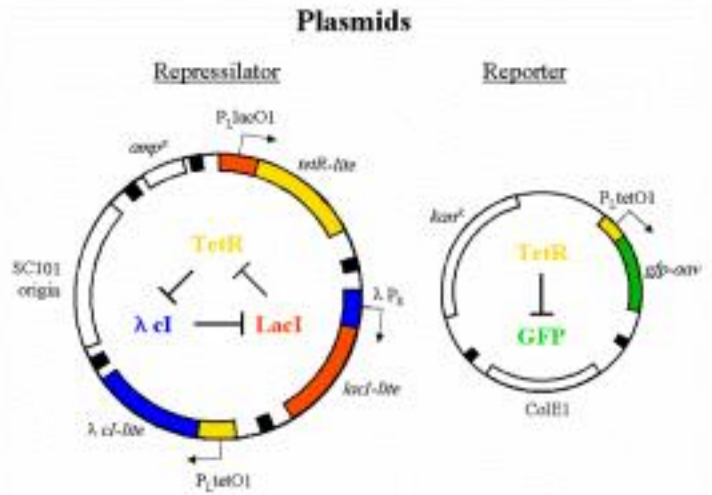
λ cI→ ¬ *lacI; lacI*→ LacI

uet et al., Antoniotti et al., Wigler & Mishra

Cycles of Repression

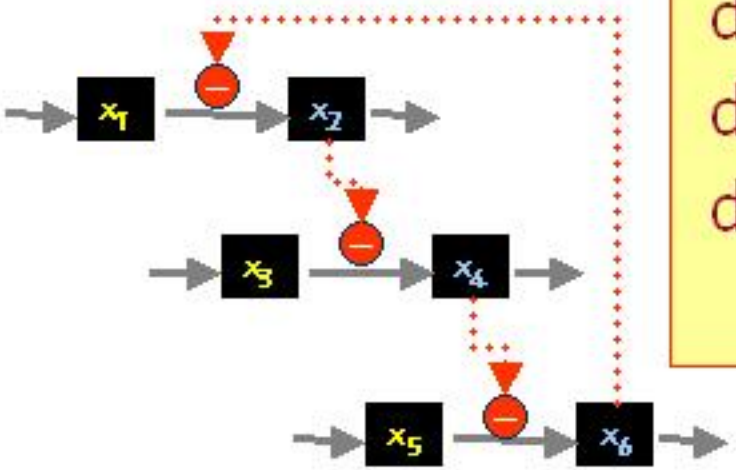
- The first repressor protein, LacI from E. coli inhibits the transcription of the second repressor gene, tetR from the tetracycline-resistance transposon Tn10, whose protein product in turn inhibits the expression of a third gene, cI from l phage.
- Finally, CI inhibits lacI expression,
- completing the cycle.

Biological Model



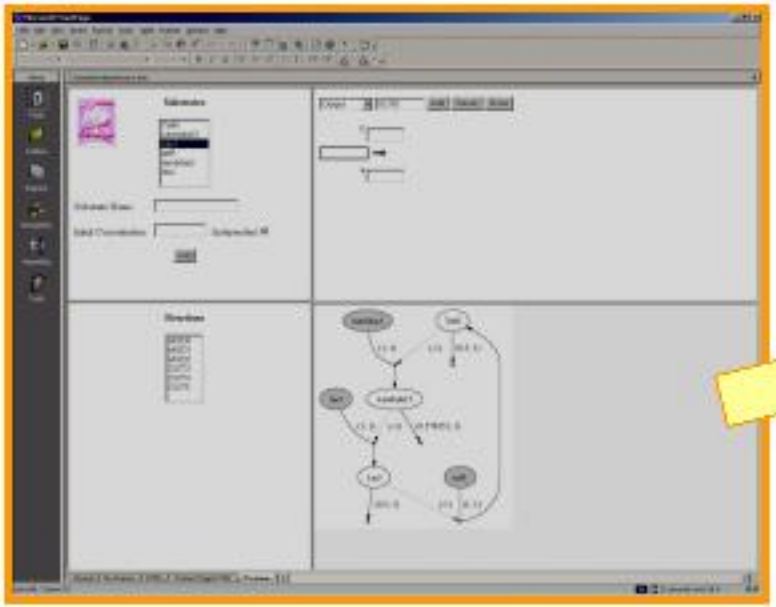
- Standard molecular biology: Construct
  - A low-copy plasmid encoding the repressilator and
  - A compatible higher-copy reporter plasmid containing the tet-repressible promoter PLtet01 fused to an intermediate stability variant of gfp.

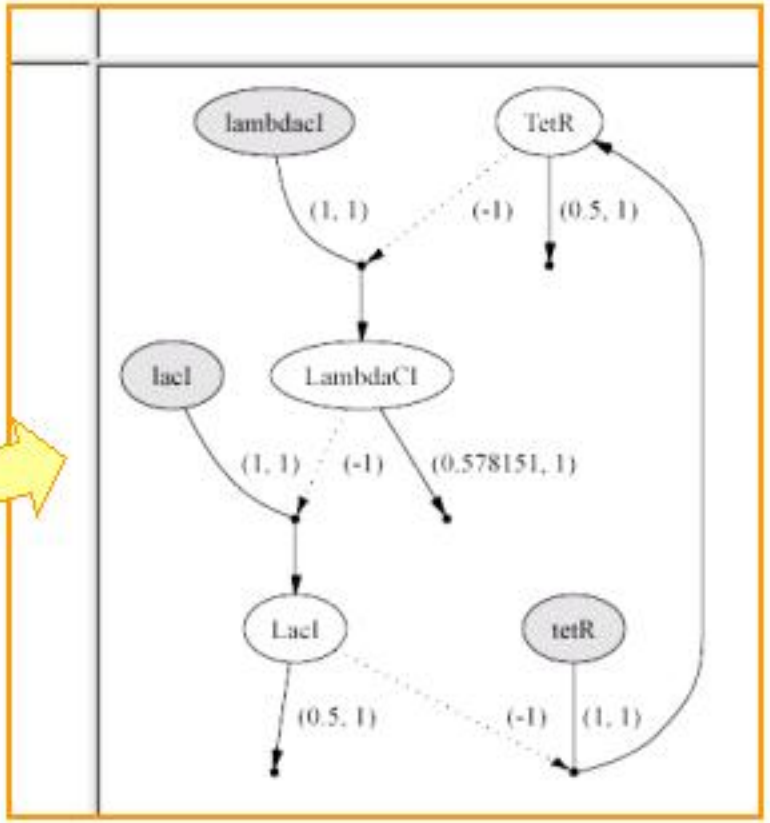
Cascade Model: Repressilator?



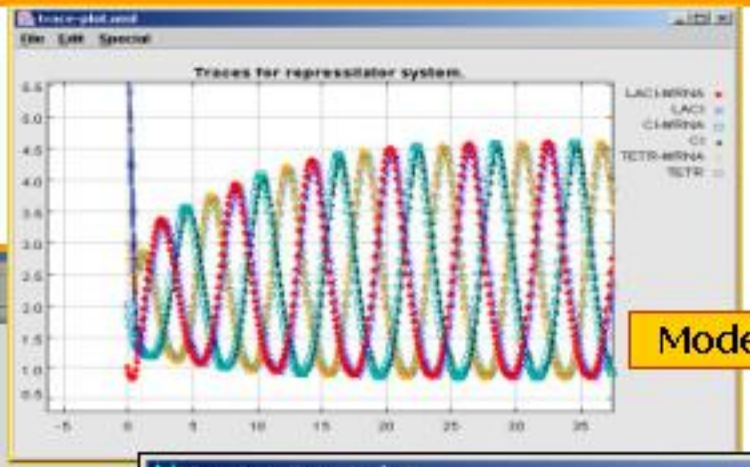
$$dx_2/dt = \alpha_2 X_6^{g26} X_1^{g21} - \beta_2 X_2^{h22}$$
$$dx_4/dt = \alpha_4 X_2^{g42} X_3^{g43} - \beta_4 X_4^{h44}$$
$$dx_6/dt = \alpha_6 X_4^{g64} X_5^{g65} - \beta_6 X_6^{h66}$$
$$X_1, X_3, X_5 = \text{const}$$

SimPathica System

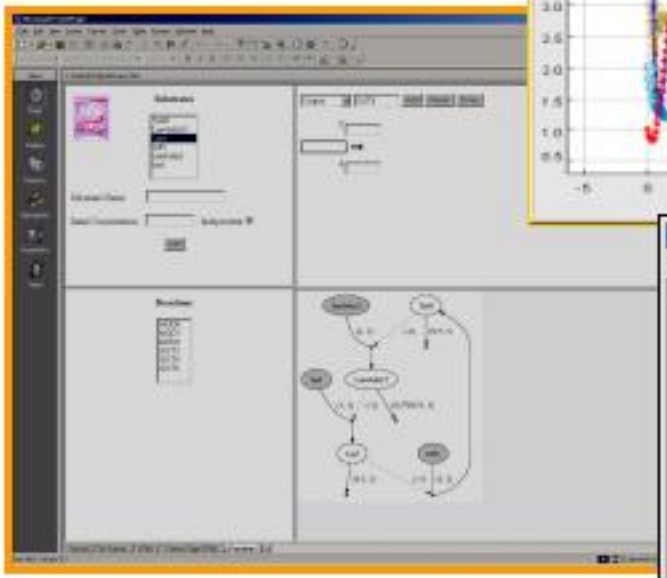




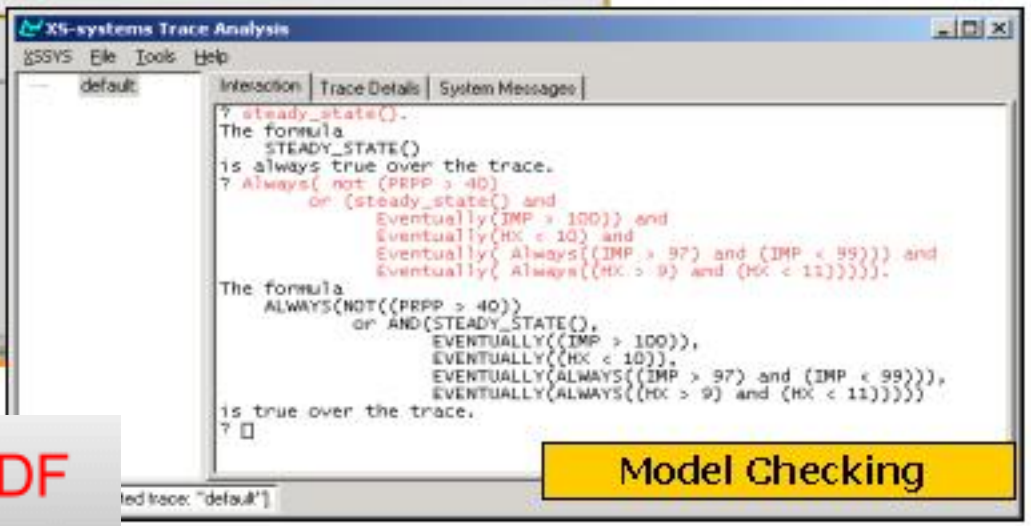
Simpathica System



Model Simulation



Model Building



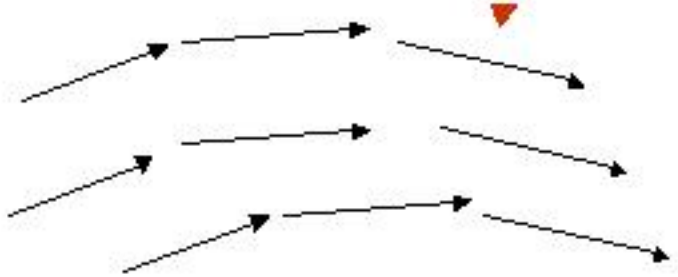
Model Checking

Symbolic Analysis

Invariant: $\Phi(s(t))$

$f(s(t), s(t+\Delta t), \Delta t)$

Invariant: $\Phi(s(t+\Delta t))$



$$\Phi(s) = \mu X.\ [X(s(t)) \wedge f(s(t), s(t+\Delta t), \Delta t) \Rightarrow X(s(t+\Delta t))]$$

Algebraic Approaches

- **Ritt-Kolchin:** Ideal Theoretic Approach
- **Kolchin-Singer:** Galois-Theoretic Approach
- **Lie:** Group-Theoretic Approach

- Understanding their interrelationship
- Effectiveness of various approaches

Differential Algebra

Assume that the system (SISO) is described as shown below:

$$\begin{aligned}
\dot{x}_1 &= p_1(X, u, \dot{u}, \ldots, u^{(k)}) \\
&\vdots \\
\dot{x}_r &= p_r(X, u, \dot{u}, \ldots, u^{(k)}) \\
0 &= q_1(X, u) \\
&\vdots \\
0 &= q_s(X, u) \\
y &= h(X, u)
\end{aligned}$$

Consider the following differential ideal I in the differential ring $\mathbb{R}\{X, u, y\}$:

$$I = [\dot{x}_1 - p_1, \ldots, \dot{x}_r - p_r, q_1, \ldots, q_s, y - h].$$

The input-output relation is then obtained by finding the contraction I^c of the ideal I to the ring $\mathbb{R}\{u, y\}$. The generators of $I^c = I \cap \mathbb{R}\{u, y\}$ give the differential polynomials involving u and y. However, the underlying algorithmic questions for ain largely unsolved.

Example System

Example Consider the following system (adapted from Forsman [Forsman92]):

$$A \to B,$$

with the following kinetic equations:

$$[\dot{B}] = [A]^{0.5} - [B]^{0.5}.$$

The input u controls the concentration $[A]$ as follows:

$$[\dot{A}] = u[A]^{-2} - [A]^{-1.5},$$

and the output y is simply $[B]$:

$$y = [B].$$

We can simplify the above system to a polynomial system by following transformations:

$$x_1^2 = [A] \quad \text{and} \quad x_2^2 = [B].$$

Input-Output Relations

Thus,

$$I = [2x_1^5\dot{x}_1 + x_1 - u, 2x_2\dot{x}_2 + x_2 - x_1, x_2^2 - y].$$

After eliminating x_1 and x_2, we obtain the following input-output relation:

$$\begin{gathered}(20\dot{y}^8y^2 - 4\dot{y}^{10}y - 40\dot{y}^6y^3 + 40\dot{y}^4y^4 - 20\dot{y}^2y^5 + 4y^6)\ddot{y}^2 \\ + (4u\dot{y}^5y - 4\dot{y}^6y - 20\dot{y}^4y^2 + 40u\dot{y}^3y^2 + 20\dot{y}^2y^3 + 20u\dot{y}y^3 + 4y^4)\ddot{y} \\ - \dot{y}^2y^5 + 5\dot{y}^4y^4 - 10\dot{y}^6y^3 + 20u\dot{y}^3y^2 + 10\dot{y}^8y^2 + y^2 - 8\dot{y}^6y + 10u\dot{y}^5y \\ - u^2y + 2u\dot{y}y - \dot{y}^2y - 5\dot{y}^{10}y + \dot{y}^{12} + 8\dot{y}^2y^3 + 2u\dot{y}y^3 = 0.\square\end{gathered}$$

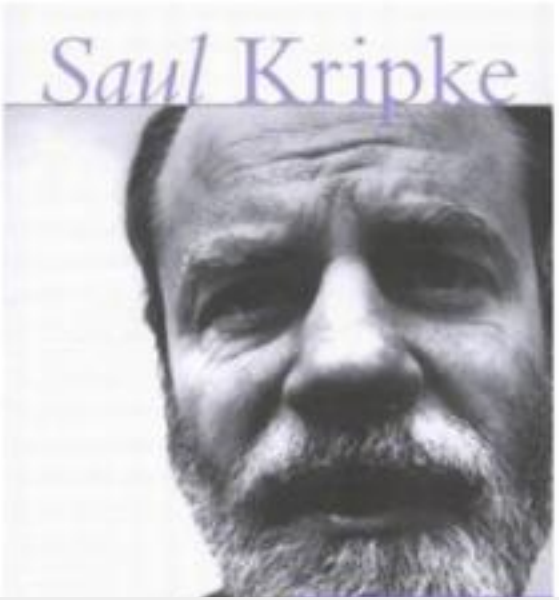
Obstacles

- Various Approaches:
  - Ideas based on the H-bases (*Gröbner Bases*).
  - Ideas based on Ritt's Characteristic Sets.
  - *Obstacles*: Failure of a Hilbert-basis like theorem (only a weaker version, *Ritt-Raudenbusch Basis Theorem*, holds), existence of non-recursive differential ideals, etc.

Simpler Computational Models

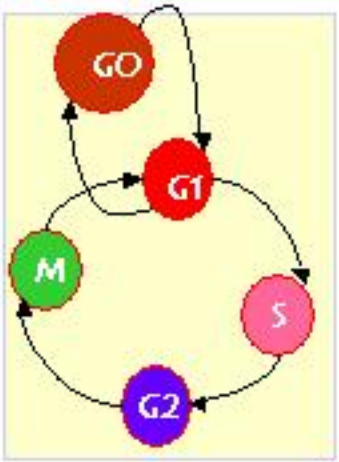
- Kripke Models/Discrete Event Systems
- Hybrid Automata
- Their Connection to
  - Turing Machines
  - "Real" Turing Machines

Kripke Structure

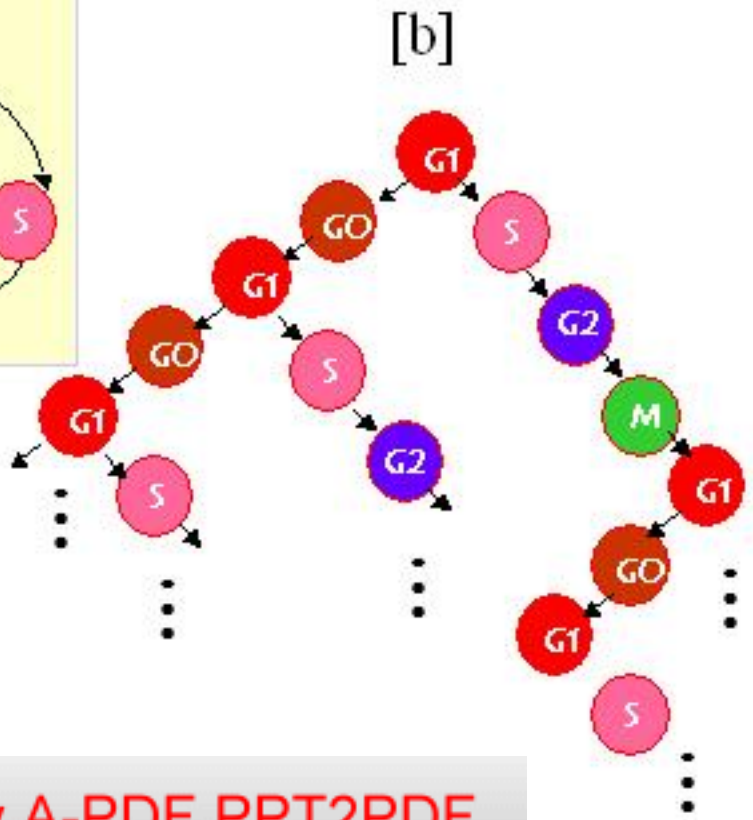


- Formal Encoding of a Dynamical System:
- Simple and intuitive pictorial representation of the behavior of a complex system
  - A **Graph** with **nodes** representing **system states** labeled with information true at that state
  - The **edges** represent **system transitions** as the result of some action

Computation Tree



[a]



[b]

- Finite set of states; Some are initial states
- **Total** transition relation: every state has at least one next state i.e. infinite paths
- There is a set of basic environmental variables or features ("atomic propositions")
- In each state, some atomic propositions are true

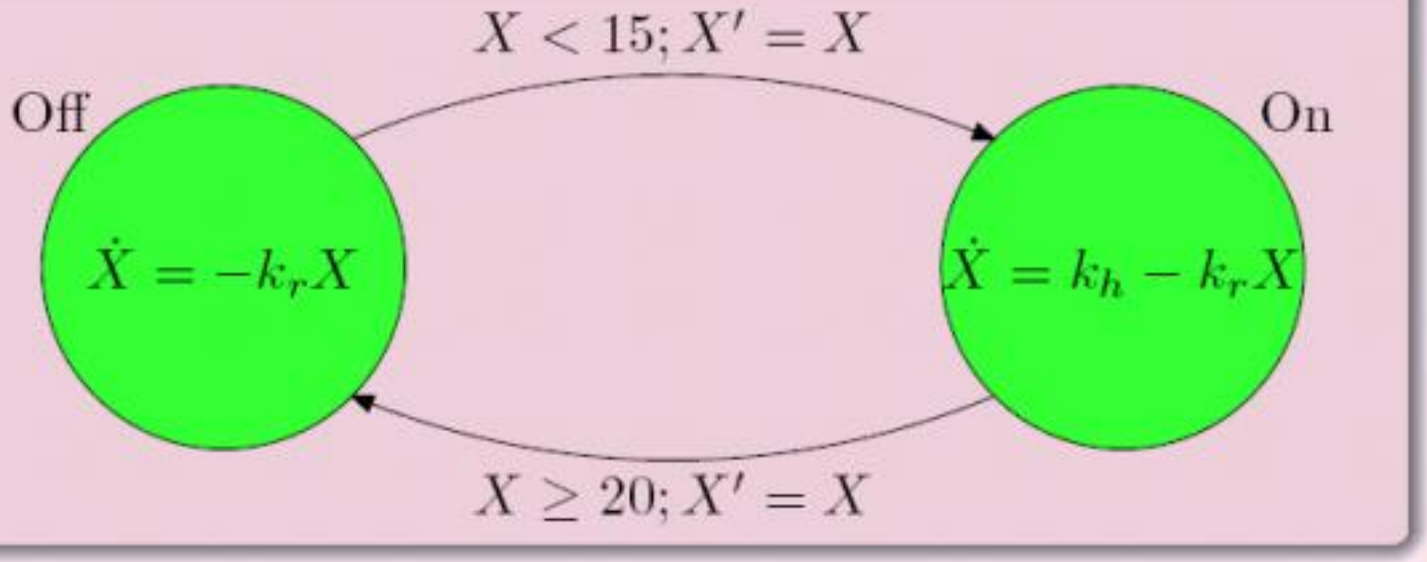
Hybrid Automata

Definition 1 (Hybrid Automata - Syntax). *A* hybrid automaton $H = (Z, Z', \mathcal{V}, \mathcal{E}, Inv, \mathcal{F}, Act, Reset)$ *of dimension* k *consists of the following components:*

1. $Z = \langle Z_1, \ldots, Z_k \rangle$ *and* $Z' = \langle Z'_1, \ldots, Z'_k \rangle$ *are two vectors of variables ranging over the reals* $\mathbb{R}$;
2. $\langle \mathcal{V}, \mathcal{E} \rangle$ *is a graph; the objects,* $v \in \mathcal{V}$, *are called* locations;
3. *Each vertex* $v \in \mathcal{V}$ *is labeled by the formula* $Inv(v)[Z]$;
4. $\mathcal{F}$ *is a function assigning to each vertex* $v \in \mathcal{V}$ *a continuous vector field over* $\mathbb{R}^k$; *we will use* $f_v : \mathbb{R}^k \times \mathbb{R}^+ \longrightarrow \mathbb{R}^k$ *to indicate the solution of the vector field* $\mathcal{F}(v)$ *and* $Dyn(v)[Z, Z', T]$ *to identify the corresponding formula, i.e.,* $Dyn(v)[Z, Z', T] \equiv Z' = f_v(Z, T)$;
5. *Each edge* $e \in \mathcal{E}$ *is labeled by the two formulæ* $Act(e)[Z]$ *and* $Reset(e)[Z, Z']$; $\overline{Reset}(e)[Z'] \stackrel{\text{def}}{=} \exists Z\, Reset(e)[Z, Z']$.

Thermostat

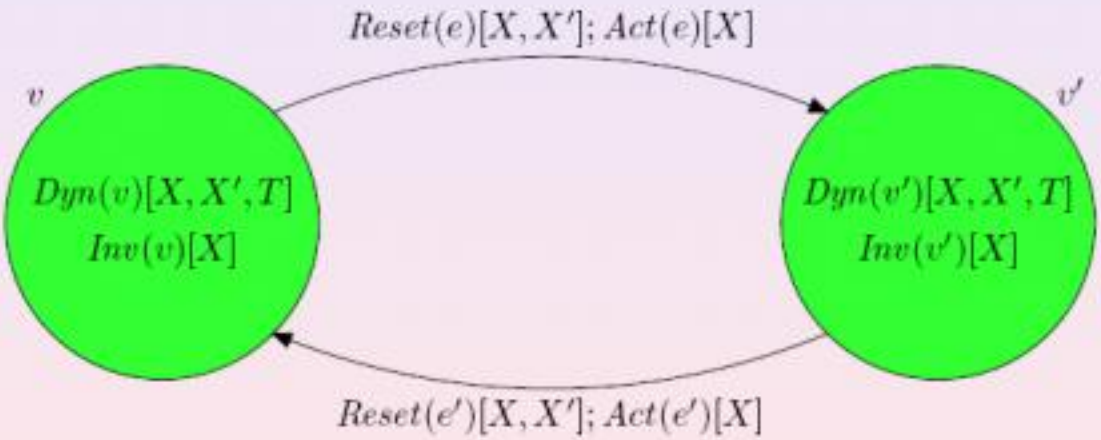
A thermostat model



Intuition

Hybrid Automata - Intuitively

Intuitively, a hybrid automaton is a finite state automaton H with continuous variables X



A state is a pair (v, r) where r is an evaluation for X

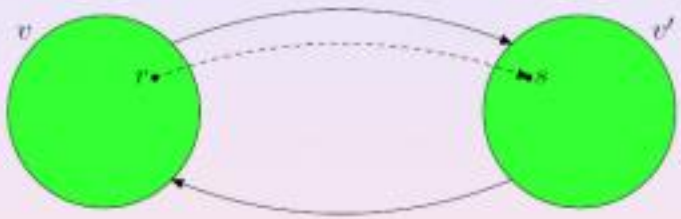
Semantics

Hybrid Automata - Sematics

Definition (Continuous Transition)

$\langle v, r\rangle \xrightarrow{t}_C \langle v, s\rangle \iff$ there exists a continuous $f : \mathbb{R}^+ \mapsto \mathbb{R}^k$ such that $r = f(0)$, $s = f(t)$, and for each $t' \in [0, t]$ the formulæ $Inv(v)[f(t')]$ and $Dyn(v)[r, f(t'), t']$ hold

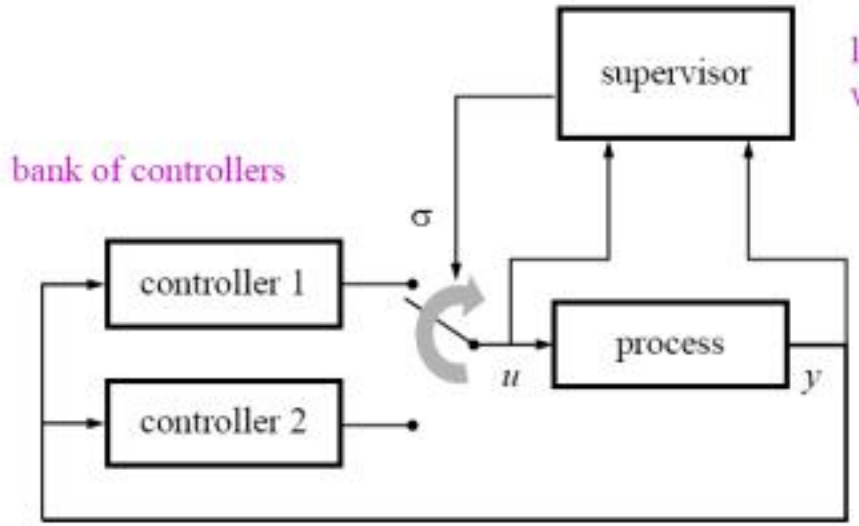
Hybrid Automata - Sematics



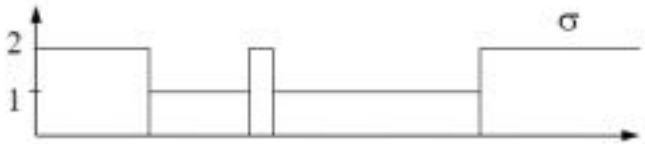
Definition (Discrete Transition)

$\langle v, r\rangle \xrightarrow{\langle v,v'\rangle}_D \langle v', s\rangle \iff \langle v, v'\rangle \in \mathcal{E}$ and $Inv(v)[r]$, $Act(\langle v, v'\rangle)[r]$, $Reset(\langle v, v'\rangle)[r, s]$, and $Inv(v')[s]$ hold

Engineered Systems



$\sigma \equiv$ switching signal taking values in the set $\{1,2\}$



Chemotaxis

- Escherichia coli has evolved a strategy for responding to a chemical gradient in its environment
  - It detects the concentration of ligands through a number of receptors
  - It reacts by driving its flagella motors to alter its path of motion.
  - Either it "runs" – moves in a straight line by moving its flagella counterclockwise (CCW), or it "tumbles" – randomly change its heading by moving its flagella clockwise (CW).
- The response is mediated through the molecular concentration of CheY in a phosphorylated form, which in turn is determined by the bound ligands at the receptors that appear in several forms.
- The more detailed pathway involves other
  - CheB (either with phosphorylation or without, B_p and B_0),
  - CheZ (Z),
  - bound receptors (LT) and
  - unbound receptors (T)
- Their continuous evolution is determined by a set of differential
  rough kinetic mass action formulation.

Non-Stochastic Chemotaxis

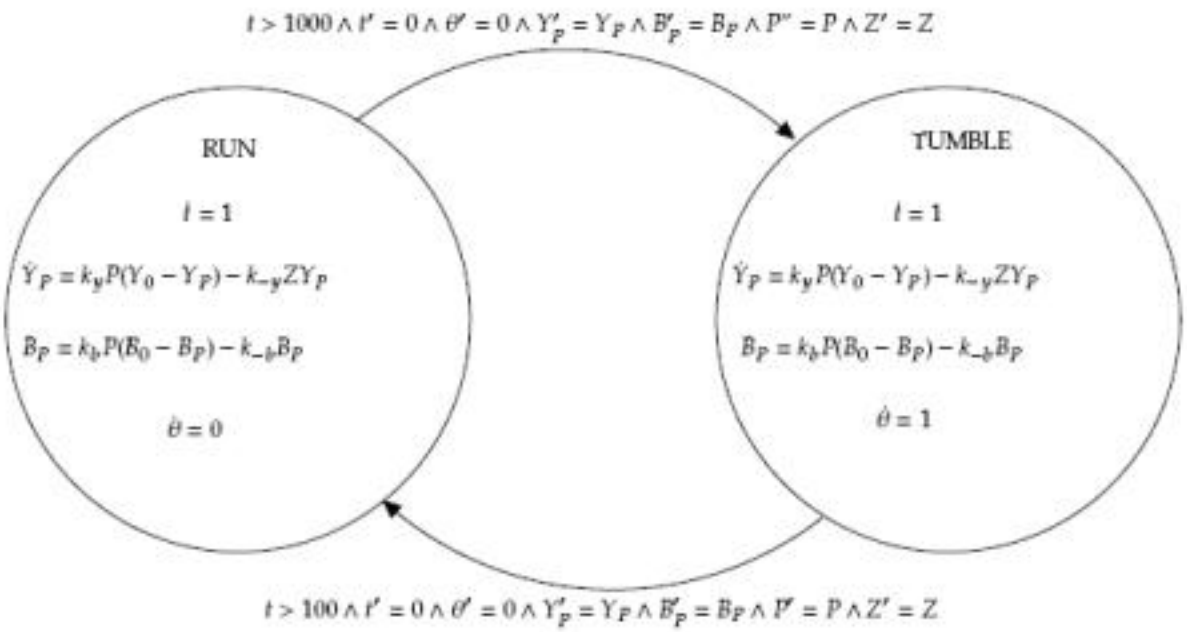
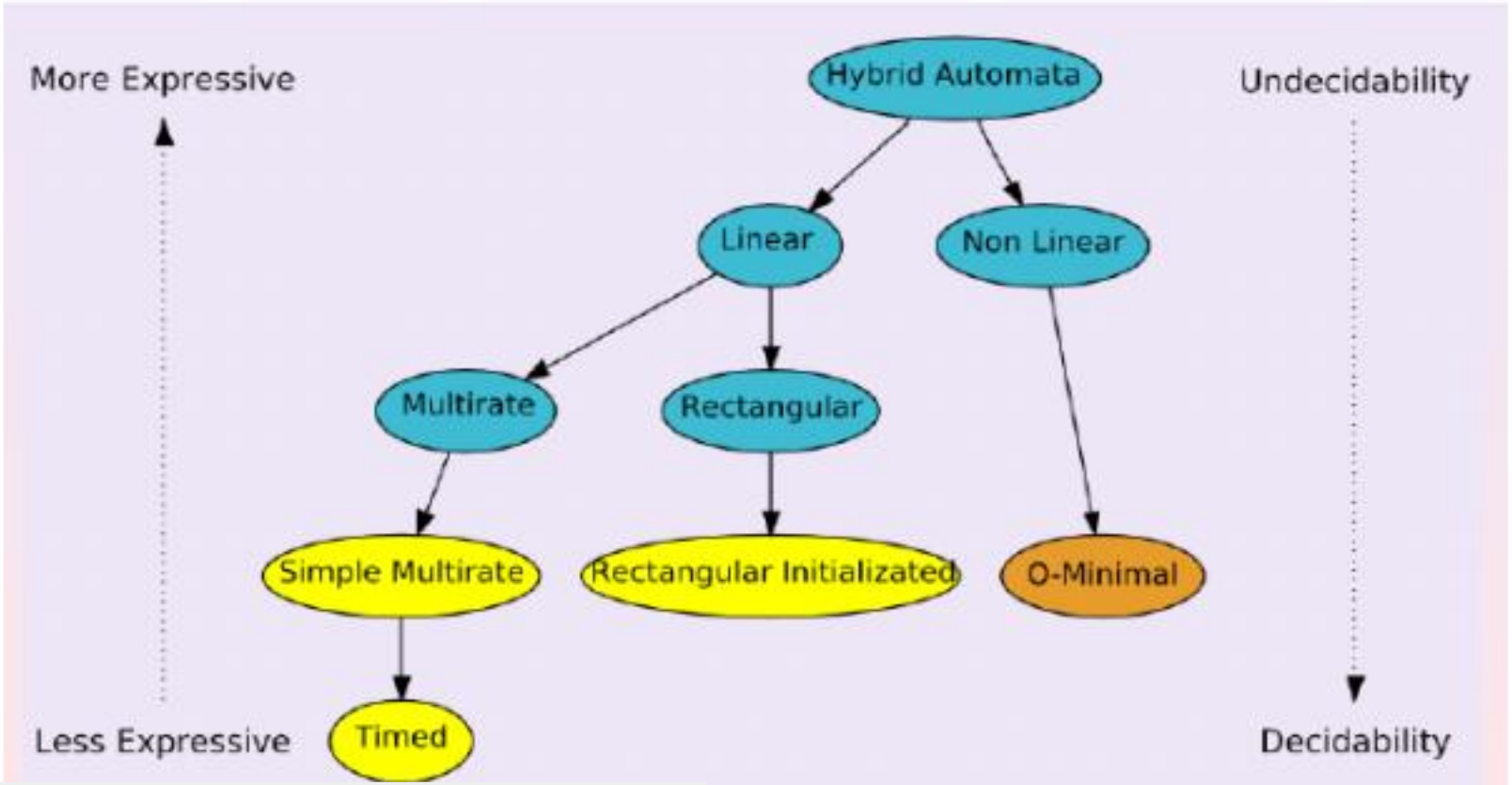


Fig. 2. An IDA capturing the run-tumble mechanism of *E. coli*.

Questions of Interest

- Controllability:
  - Assume that the system is at the "origin" initially. Can we find a control signal so that the state reaches a given position at a fixed time?
- Observability:
  - Can the state x be determined from observations of the output y over some time interval.
- Reachability: A computationally simpler problem:
  - Can we determine what states are reached as the system evolves autonomously or under a class of control signal.
- "HALTING" Problem:
  - Can the system reach a designated state at some time and then stay there?

Decision problems



Dynamics

- Replacing differential equations by "equivalent" dynamics:

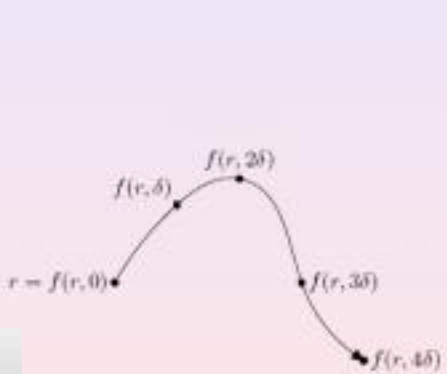
If $f(X,T)$ is the solution of $\dot{X} = \mathcal{F}(X,T)$, then

$$\dot{X} = \mathcal{F}(X,T) \quad \text{and} \quad Dyn[X,X',T] \equiv X' = f(X,T),$$

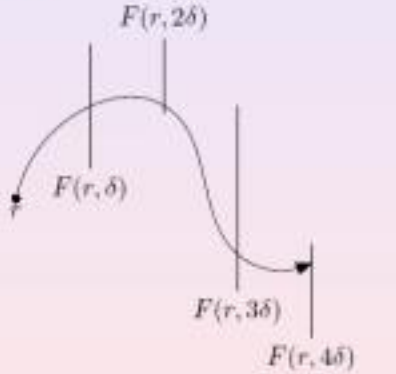
are equivalent

Inclusion Dynamics

We are interested in inclusion dynamics defined by formulæ



$Dyn[X,X',T] \equiv X' = f(X,T)$



$Dyn[X,X',T] \equiv X' \in F(X,T)$

Michael's Form

- Let $F_x^v(T) \equiv \{X' \mid Dyn(v)[X, X', T] \wedge Inv(v)[X']\}$

- A Hybrid automaton is in Michael's form if
  - F_x^v is lower semi-continuous
  - For each $t \in I_X^v$ the set $F_x^v(t)$ is closed and convex
  - where I_X^v is the largest $[0, t')$ such that $F_x^v(t) \neq \emptyset$, $\forall\, t \in [0, t')$.

Theorem

If H is in Michael's form, then $s \in F_r^v(t)$ iff $\langle v, r\rangle \xrightarrow{t}_C \langle v, s\rangle$.

Reachability

Michael's Form and Reachability

For each automaton in Michael's form, we can write a formula $Reach(H, ph)[X, X', T]$, where $ph = v_0, \ldots, v_n$ is a path on $\langle \mathcal{V}, \mathcal{E} \rangle$, such that

$Reach(H, ph)[X, X', T]$ holds $\iff$ H reaches $\langle v_n, X' \rangle$ from $\langle v_0, X \rangle$ with a trajectory corresponding to ph

The path must not be infinite!!

Two New Models

FOCoRe and IDA

FOCoRe (First Order Constant Reset hybrid automata) are first-order hybrid automata:

- in Michael's form
- with constant resets (i.e., $Reset(e)[X, X']$ does not depend on X)

IDA (Independent Dynamics hybrid Automata) allows identity resets between locations whose dynamic does not change

We can reduce reachability problem for either FOCoRe or IDA $\mathcal{T}$-automata to a satisfiability problem for formulæ of $\mathcal{T}$

First Order Theory of Reals

- Tarski's theorem says that the first-order theory of reals with $+$, $\times$, $=$, and $>$ allows quantifier elimination. Algorithmic quantifier elimination implies decidability.
- Every quantifier-free formula composed of polynomial equations and inequalities, and Boolean connectives defines a semialgebraic set. Thus a set S is semi-algebraic if:

$$S = \{\langle \xi_1, \cdots, \xi_n \rangle \in \mathcal{R}^n | \psi(\xi_1, \cdots, \xi_n) = True\}, \text{ or}$$

$$S = \bigcup_{i=1}^{I} \bigcap_{j=1}^{J_i} \{\langle \xi_1, \cdots, \xi_n \rangle \in \mathcal{R}^n | sign(f_{i,j}(\xi_1, \cdots, \xi_n)) = s_{i,j}\}$$

where $\psi(\xi_1, \cdots, \xi_n)$ *is a quantifier-free formula involving* n *algebraic variables,* $f_{i,j}s$ *are multivariate polynomials over* R *and the* $s_{i,j}s$ *are* $\{-1, 0, +1\}$.

SaCoRe

- Hybrid Automata's inclusion dynamics, approximated by semi-algebraic formula. Dyn[X,X', T] $\equiv$ Semialgebraic Set
- A more realistic approximation, for time invariant systems:
  - Dyn[X, X', h]

    $\approx \{X' \mid X' = X + \mathcal{F}(X,0)\, h + \delta, |\delta| < \varepsilon\}$,

    for a suitably chosen

    $$\varepsilon \equiv |\mathcal{F}(X,0)\, h^2/2! + \mathcal{F}(X,0)\, h^3/3! + \cdots|$$

Another Example: Biological Pattern Formation

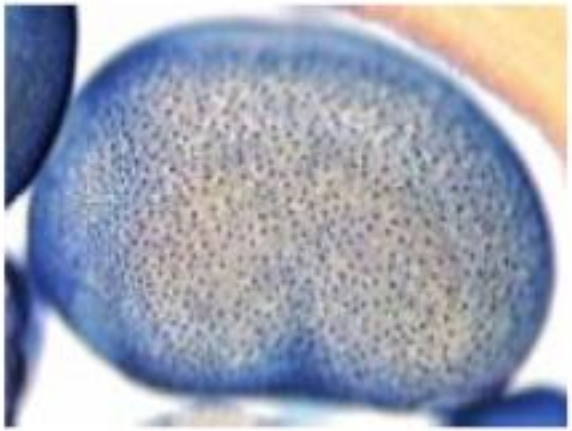
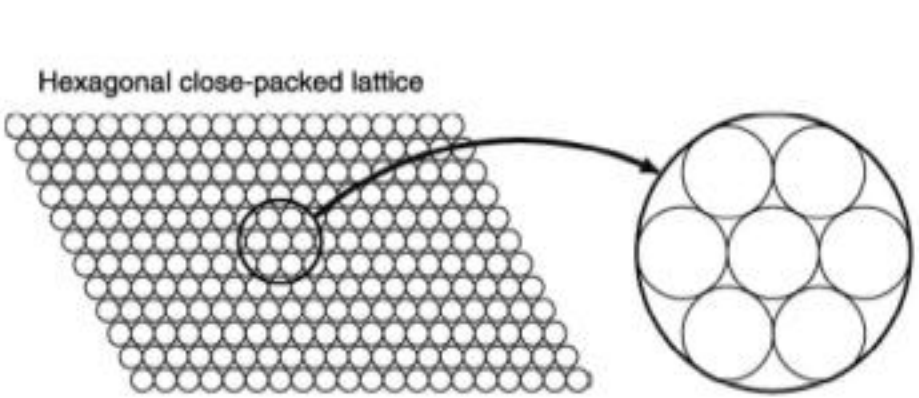
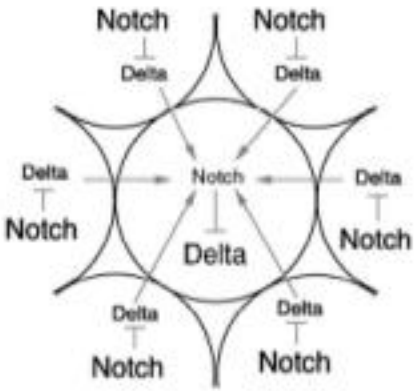


Figure 3: *Xenopus* embryo labeled by a marker for ciliated cell precursors seen as black dots.[1]

- Embryonic Skin Of The South African Claw-Toed Frog
- "Salt-and-Pepper" pattern formed due to lateral inhibition in the Xenopus epidermal layer

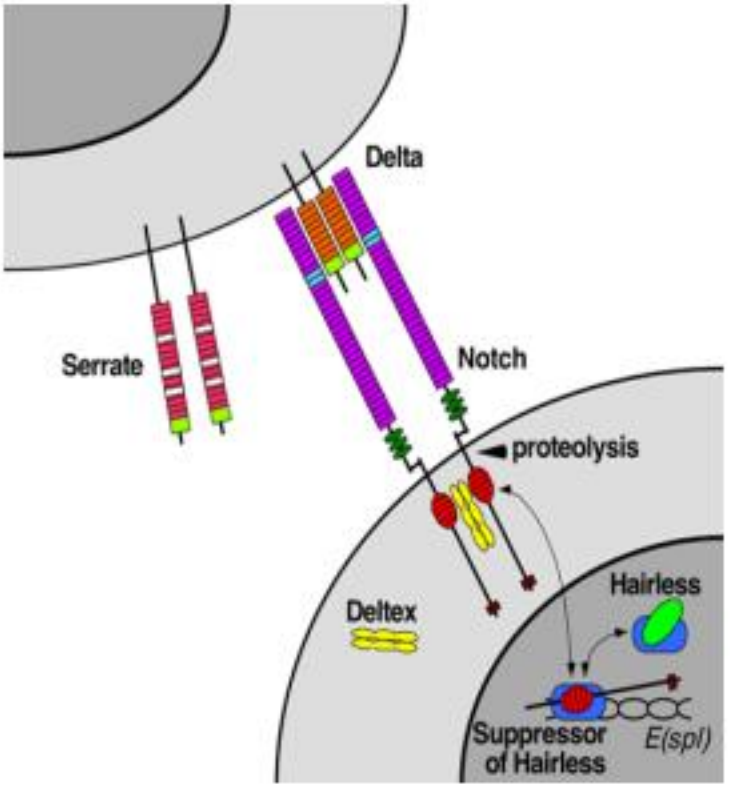
Delta-Notch Signalling





Physically **adjacent** cells **laterally inhibit** each other's ciliation (Delta production)

Delta-Notch Pathway



- Delta binds and activates its receptor Notch in neighboring cells (proteolytic release and nuclear translocation of the intracellular domain of Notch)
- Activated Notch suppresses ligand (Delta) production in the cell
- A cell producing more ligands forces its neighboring cells to produce less

Pattern formation by lateral inhibition with feedback: a mathematical model of Delta-Notch intercellular signalling

Collier et al.(1996)

$$\frac{\mathrm{d}(N_P/N_0)}{\mathrm{d}\tau} = F(\bar{D}_P/D_0) - \mu N_P/N_0,$$

$$\frac{\mathrm{d}(D_P/D_0)}{\mathrm{d}\tau} = G(N_P/N_0) - \rho D_P/D_0.$$

Rewriting...

$$\dot{n}_P = f(\bar{d}_P) - n_P,$$

$$\dot{d}_P = v\{g(n_P) - d_P\}.$$

Where:

$$f(x) = \frac{x^k}{a + x^k}, \quad g(x) = \frac{1}{1 + bx^h},$$

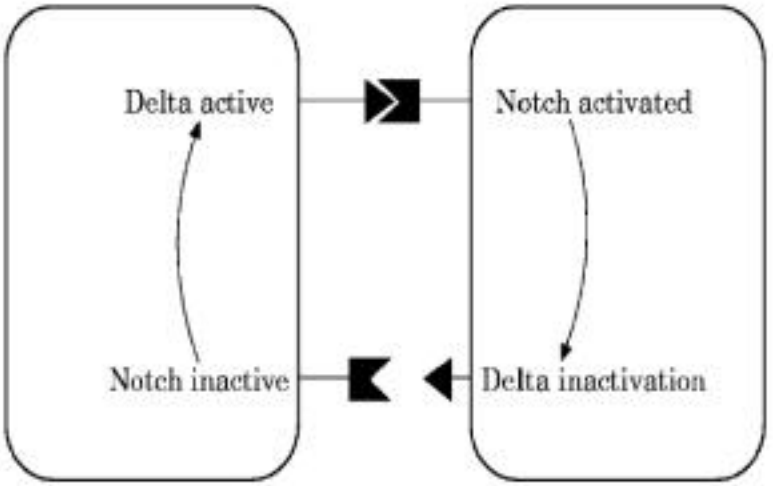
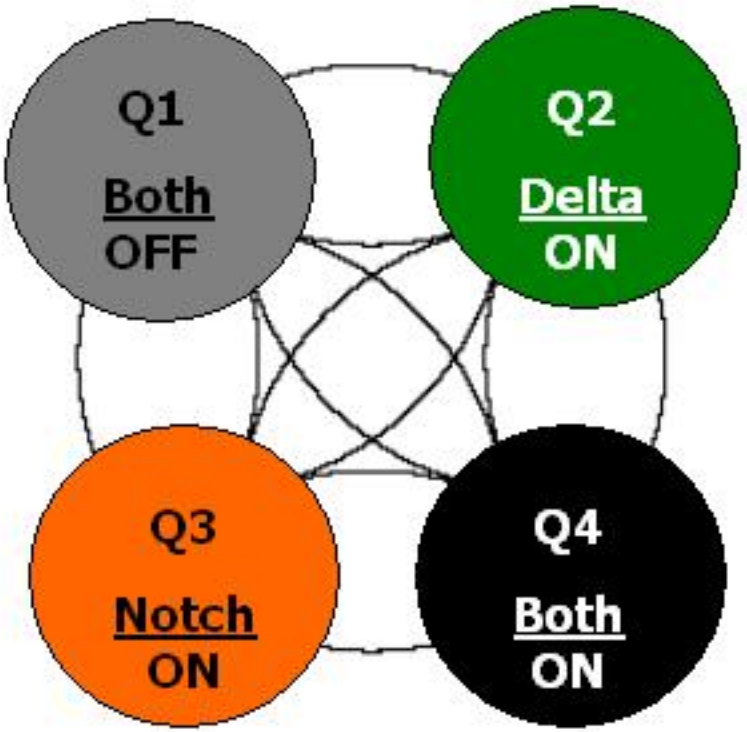


FIG. 1. Diagrammatic representation of the effective feedback loop between Notch and Delta in neighbouring cells. Details of the Notch signalling pathway are omitted for clarity. *Key*: ▶ Delta; ◀ Notch.

Collier et al.

One-Cell Delta-Notch Hybrid Automaton



(a) Transition diagram for a single cell

$$H_{one\_cell} = (Q, X, \Sigma, Init, f, Inv, R)$$

$$Q = q_1, q_2, q_3, q_4$$

$$X = (x_1, x_2)^T \in \Re^2$$

$$\Sigma = \left\{ u_N = \sum_{i=1}^{6} x_{Delta,i} \right\}$$

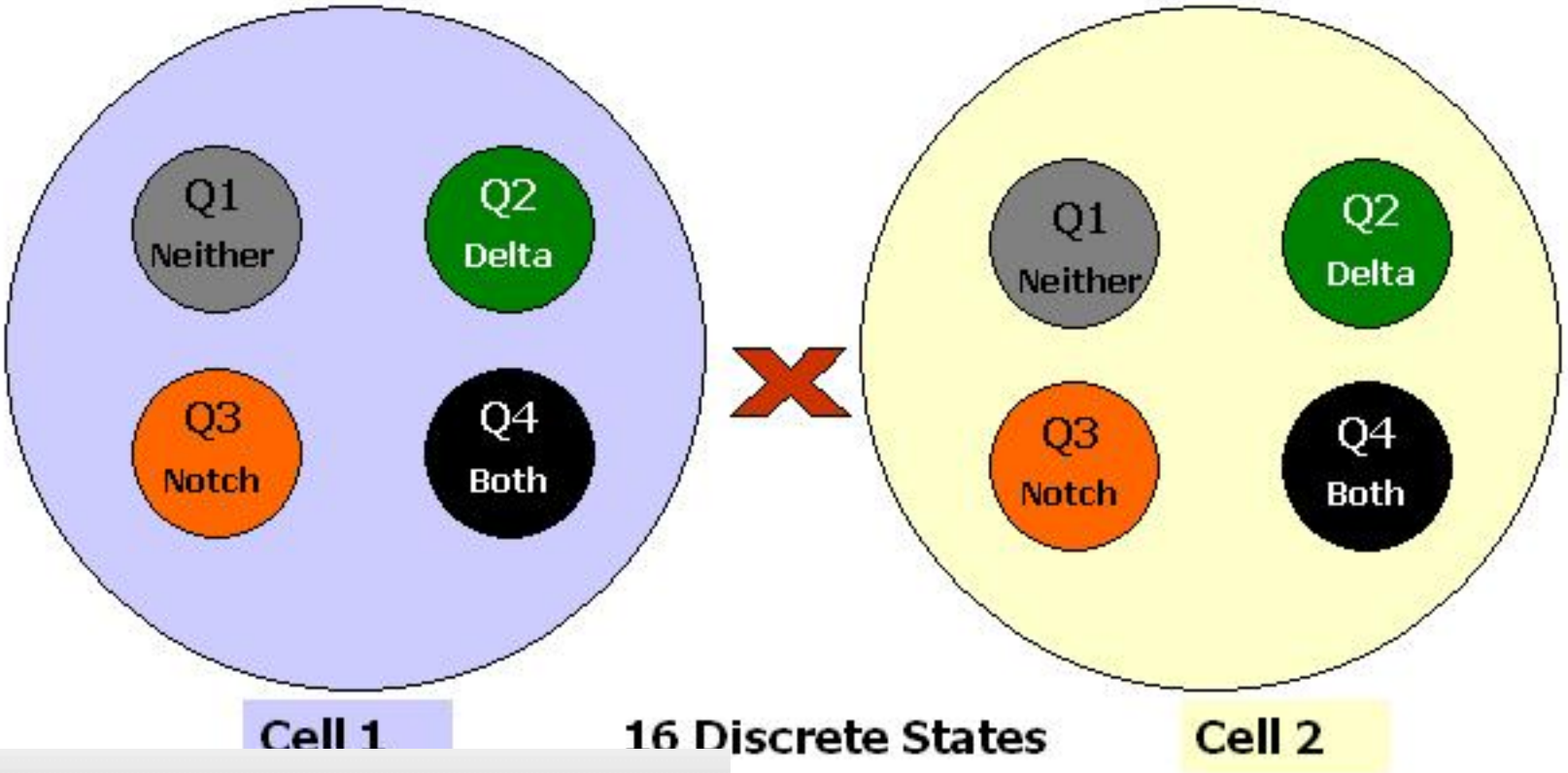
$$Init = Q \times \{X \subset \Re^2 : x_1, x_2 > 0\}$$

$$f(q,x) = \begin{cases} [-\lambda_D x_1; -\lambda_N x_2]^T & \text{if } q = q_1 \\ [R_D - \lambda_D x_1; -\lambda_N x_2]^T & \text{if } q = q_2 \\ [-\lambda_D x_1; R_N - \lambda_N x_2]^T & \text{if } q = q_3 \\ [R_D - \lambda_D x_1; R_N - \lambda_N x_2]^T & \text{if } q = q_4 \end{cases}$$

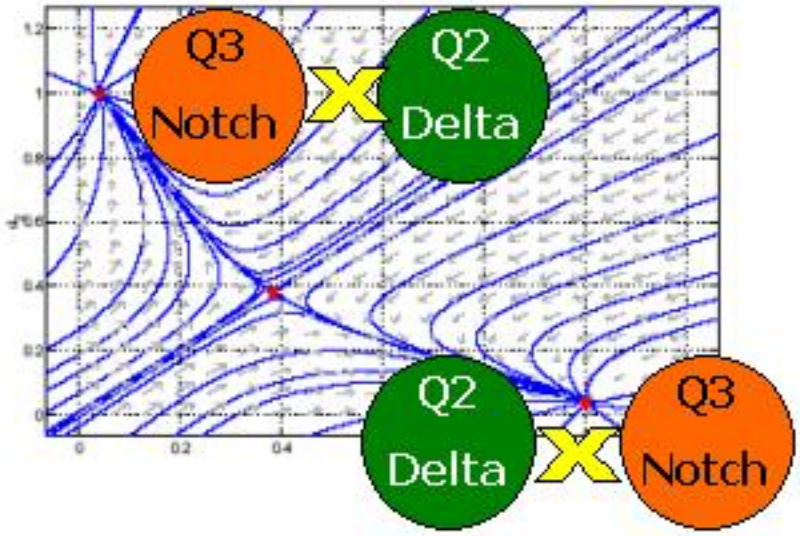
$$Inv = \{q_1, \{-x_2 < h_D, u_N < h_N\}\} \cup \{q_2, \{-x_2 \geq h_D, u_N < h_N\}\} \cup \{q_3, \{-x_2 < h_D, u_N \geq h_N\}\} \cup \{q_4, \{-x_2 \geq h_D, u_N \geq h_N\}\}$$

Ghosh et al.

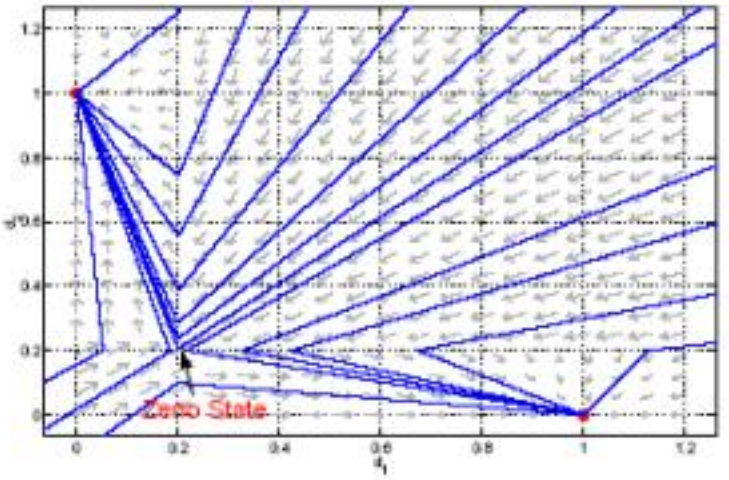
Two-Cell Delta-Notch System



System Properties: True & Approximate



(a) Nonlinear model



(b) Hybrid systems model

Fig. 7. Phase plane projections for two cell system showing equilibria. Labels d_1 and d_2 are the Delta protein concentrations in cell 1 and 2 respectively.

State Reachability

Reaching State q_7 (2,3) When we ask $True\ \exists \mathcal{U}\ [-2n_1 > -1 \wedge 5d_2 < 1 \wedge -2n_2 < -1 \wedge 5d_1 > 1]$, we get:

Iteration 1: $5d_1 - 1 \geq 0 \wedge 2n_1 - 1 \leq 0 \wedge 5d_2 - 1 \leq 0 \wedge 2n_2 - 1 \geq 0$

Iteration 2: $n_1 - 1 \leq 0 \wedge [[2n_1 - 5d_1 \leq 0 \wedge 5d_2 - 1 \leq 0 \wedge 8n_2 - 5d_2 - 3 \geq 0 \wedge n_2 + n_1 - 1 = 0] \vee [8n_1 - 5d_1 - 3 \leq 0 \wedge 4d_2 + d_1 - 1 = 0 \wedge 2n_2 - 1 \geq 0 \wedge 8n_2 + 5d_1 - 5 \geq 0] \vee [5d_1 - 1 \geq 0 \wedge 2n_1 - 5d_1 \leq 0 \wedge 5d_2 + 2n_1 - 2 \leq 0 \wedge 2n_2 - 1 \geq 0] \vee [5d_1 - 1 \geq 0 \wedge 2n_1 - 1 \leq 0 \wedge 5d_2 - 1 \leq 0 \wedge 8n_2 - 5d_2 - 3 \geq 0] \vee [2n_1 - 1 \leq 0 \wedge 5d_2 - 1 \leq 0 \wedge 8n_2 - 5d_2 - 3 \geq 0 \wedge 8n_2 + 5d_1 - 5 \geq 0] \vee [2n_1 - 5d_1 \leq 0 \wedge 5d_2 - 1 \leq 0 \wedge 2n_2 - 1 \geq 0 \wedge 8n_2 + 5d_1 - 5 \geq 0]]$

$\equiv f_7$ (say).

State Reachability

Reaching State q_{10} (3,2) When we ask $True\ \exists\mathcal{U}\ [-2n_1 < -1 \wedge 5d_2 > 1 \wedge -2n_2 > -1 \wedge 5d_1 < 1]$, we get:

Iteration 1: $5d_1 - 1 \leq 0 \wedge 2n_1 - 1 \geq 0 \wedge 5d_2 - 1 \geq 0 \wedge 2n_2 - 1 \leq 0$

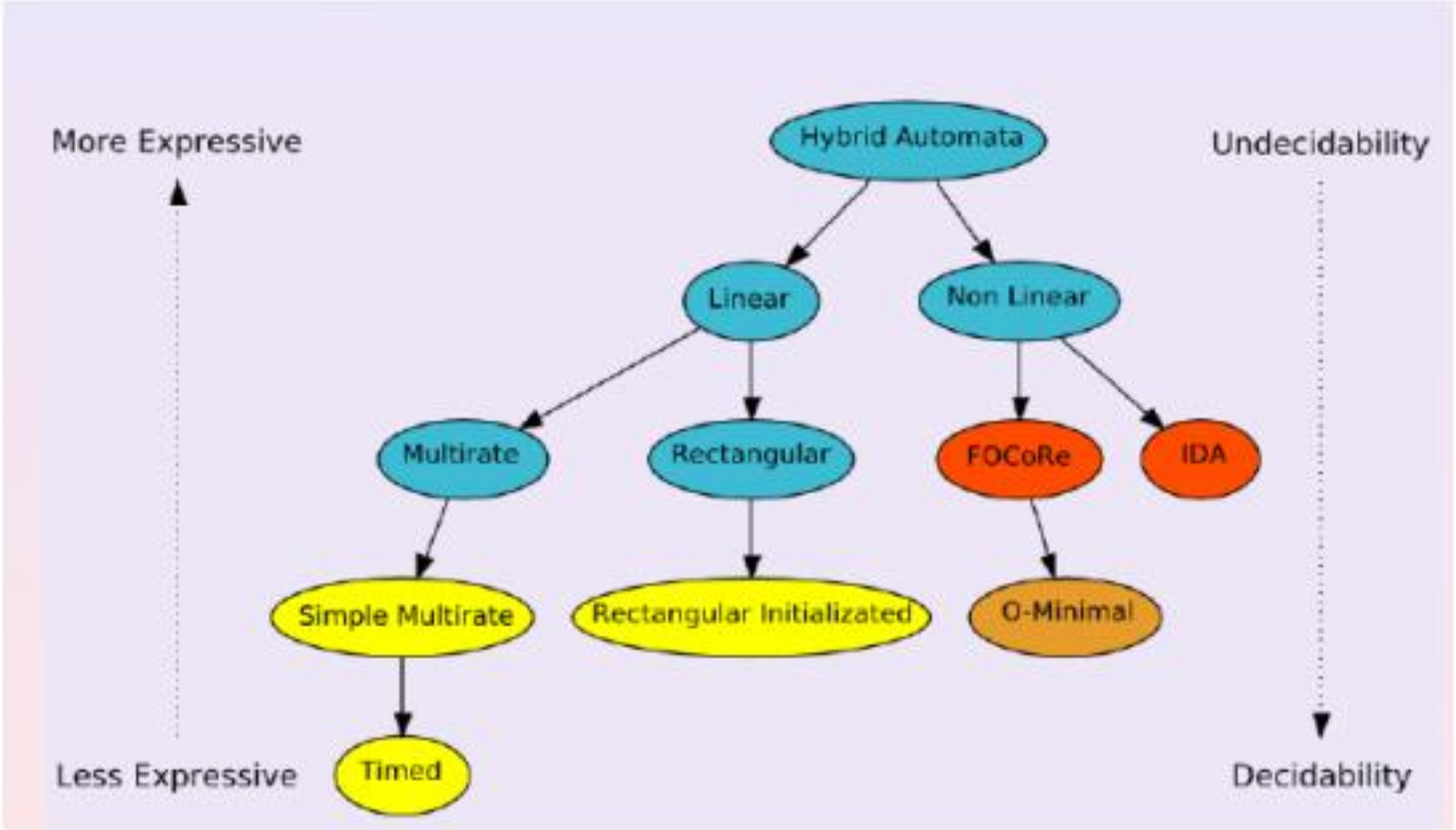
Iteration 2: $n_2 - 1 \leq 0 \wedge [[2n_1 - 1 \geq 0 \wedge 5d_2 + 8n_1 - 5 \geq 0 \wedge d_2 + 4d_1 - 1 = 0 \wedge 2n_2 + 5d_1 - 2 \leq 0] \vee [2n_1 - 1 < 0 \wedge 8n_1 - 5d_1 - 3 \geq 0 \wedge 5d_2 + 8n_1 - 5 \geq 0 \wedge n_2 + n_1 - 1 = 0] \vee [8n_1 - 5d_1 - 3 \geq 0 \wedge 5d_2 + 8n_1 - 5 < 0 \wedge 5d_2 + 2n_1 - 2 \geq 0 \wedge n_2 + n_1 - 1 = 0] \vee [2n_1 - 1 \geq 0 \wedge 5d_2 - 1 \geq 0 \wedge 2n_2 + 5d_1 - 2 \leq 0 \wedge n_2 + n_1 - 1 < 0] \vee [5d_1 - 1 \leq 0 \wedge 2n_1 - 1 \geq 0 \wedge 5d_2 + 8n_1 - 5 \geq 0 \wedge 2n_2 - 5d_2 \leq 0] \vee [5d_1 - 1 \leq 0 \wedge 2n_1 - 1 \geq 0 \wedge 5d_2 + 8n_1 - 5 \geq 0 \wedge 2n_2 - 1 \leq 0] \vee [8n_1 - 5d_1 - 3 \geq 0 \wedge 5d_2 - 1 \geq 0 \wedge 2n_2 + 5d_1 - 2 \leq 0 \wedge 2n_2 - 1 \leq 0]]$

Impossibility Of Reaching Wrong Equilibrium:

$$f_7 \wedge \neg f_{10} = n_1 - 1 \leq 0 \wedge [[2n_1 - 5d_1 \leq 0 \wedge 5d_2 - 1 < 0 \wedge 8n_2 - 5d_2 - 3 \geq 0 \wedge n_2 + n_1 - 1 = 0] \vee [2n_1 - 1 \leq 0 \wedge 5d_2 - 1 \leq 0 \wedge 8n_2 - 5d_2 - 3 \geq 0 \wedge 2n_2 + 5d_1 - 2 > 0] \vee [2n_1 - 1 \leq 0 \wedge 5d_2 + 2n_1 - 2 \leq 0 \wedge 4d_2 + d_1 - 1 = 0 \wedge n_2 + n_1 - 1 > 0] \vee [2n_1 - 5d_1 \leq 0 \wedge 5d_2 - 1 \leq 0 \wedge n_2 + n_1 - 1 > 0 \wedge 2n_2 - 1 \geq 0] \vee [2n_1 - 1 \leq 0 \wedge 5d_2 - 1 < 0 \wedge 8n_2 - 5d_2 - 3 \geq 0 \wedge 8n_2 + 5d_1 - 5 \geq 0] \vee [8n_1 - 5d_1 - 3 < 0 \wedge 4d_2 + d_1 - 1 = 0 \wedge 2n_2 - 1 \geq 0 \wedge 8n_2 + 5d_1 - 5 \geq 0] \vee [5d_1 - 1 \geq 0 \wedge 2n_1 - 5d_1 < 0 \wedge 5d_2 + 2n_1 - 2 \leq 0 \wedge 2n_2 - 1 \geq 0] \vee [5d_1 - 1 \geq 0 \wedge 2n_1 - 1 \leq 0 \wedge 5d_2 - 1 < 0 \wedge 8n_2 - 5d_2 - 3 \geq 0] \vee [2n_1 - 1 < 0 \wedge 5d_2 - 1 \leq 0 \wedge 8n_2 - 5d_2 - 3 \geq 0 \wedge 8n_2 + 5d_1 - 5 \geq 0]]$$

Since we have assumed no upper bound on the initial values and since we have been able to compute only two iterations, this formula does *not* evaluate to *True* given $n_1 < n_2 \wedge d_1 > d_2$. However, when *Qepcad* simplifies the above formula assuming that $n_1 > n_2 \wedge d_1 < d_2$, it immediately evaluates

Hybrid Hierarchy



Logic & Model-Checking

Deciphering Design Principles in a Biological Systems

Step 1. Formally encode the behavior of the system as a hybrid automaton

Step 2. Formally encode the properties of interest in a powerful logic

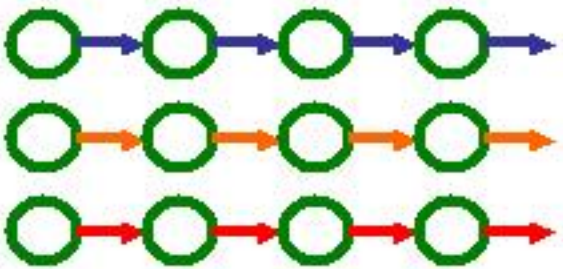
Step 3. Automate the process of checking if the formal model of the system satisfies the formally encoded properties using Model Checking

Temporal Logic

- **First Order Logic**: Time is an explicitly quantified variable
- **Propositional Modal logic**: was invented to formalize modal notions and suppress the quantified variables – with operators "possibly P" and "necessarily P" (similar to "eventually" and "henceforth")

Branching versus Linear Time

- **Temporal Logic**:
  - Short hand for describing the way properties of the system change with time
  - Time is implicit
- **Linear-time**: Only one possible future in a moment
  - Look at individual computations
- **Branching-time**: It may be possible to split to different courses depending on possible futures
  - Look at the tree of computations



Time is Branching

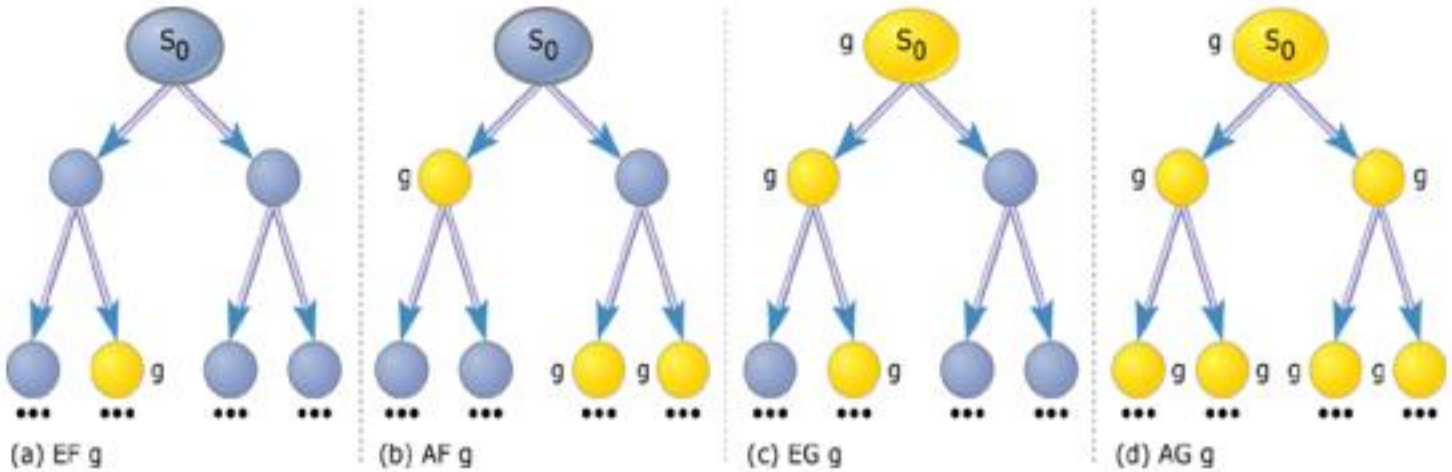
Computation Tree Logic (CTL)

- Branching Time temporal logic: interpreted over an ***execution tree*** where branching denotes non-deterministic actions
- Explicitly quantify over two modes – the path and the time
- Each time we talk about a temporal property, we also specify whether it is true on all possible paths or whether it is true on at least one path - ***Path quantifiers***
  - **A** = "for all future paths"
  - **E** = "for some future path"

Semantics for CTL

- For $p \in AP$:
  $s \models p \Leftrightarrow p \in L(s)$ $\quad$ $s \models \neg p \Leftrightarrow p \notin L(s)$
- $s \models f \wedge g \Leftrightarrow s \models f$ and $s \models g$
- $s \models f \vee g \Leftrightarrow s \models f$ or $s \models g$
- $s \models EX\, f \Leftrightarrow \exists\, \pi = \langle s_0 s_1 \ldots \rangle$ from s $s_1 \models f$
- $s \models E(f\, U\, g) \Leftrightarrow \exists\, \pi = \langle s_0 s_1 \ldots \rangle$ from s
  $\exists j \geq 0 \,[\, s_j \models g$ and $\forall i : 0 \leq i < j \,[s_i \models f\,]\,]$
- $s \models EG\, f \Leftrightarrow \exists\, \pi = \langle s_0 s_1 \ldots \rangle$ from s $\forall i \geq 0$: $s_i \models f$

Some CTL Operators



EF g

AF g

EG g

AG g

CTL Model-Checking

- Straight-forward approach: **Recursive descent on the structure of the query formula**
- Label the states with the terms in the formula:
  - Proceed by marking each point with the set of valid sub-formulas
- **"Global" algorithm:**
  - Iterate on the structure of the property, traversing the whole of the model in each step
  - Use fixed point unfolding to interpret Until:

$$\mathbf{E}(\psi_2 \, \mathbf{U}^+ \, \psi_1) \leftrightarrow \mathbf{E}\,\mathbf{X}(\psi_1 \vee \psi_2 \wedge \mathbf{E}(\psi_2 \, \mathbf{U}^+ \, \psi_1))$$

$$\mathbf{A}(\psi_2 \, \mathbf{U}^+ \, \psi_1) \leftrightarrow \mathbf{A}\,\mathbf{X}(\psi_1 \vee \psi_2 \wedge \mathbf{A}(\psi_2 \, \mathbf{U}^+ \, \psi_1))$$

Naïve CTL Model-Checker

procedure CTL_check (Model $(U, \mathcal{I}, w_0)$, Formula φ) =
  if $w_0 \in$ eval(φ)
  then print("φ is satisfied at w_0 in $(U, \mathcal{I})$")
  else print("φ not satisfied at w_0 in $(U, \mathcal{I})$");

function eval (Formula φ): Pointset =
  case φ **of**
    p : **return** $\mathcal{I}$(p);
    $\bot$: **return** {};
    $(\psi_1 \rightarrow \psi_2)$: **return** $U \setminus$ eval(ψ_1) $\cup$ eval(ψ_2);
    $\mathbf{E}(\psi_2 \, \mathbf{U}^+ \, \psi_1)$: $E1 :=$ eval(ψ_1); $E2 :=$ eval(ψ_2); $E := \{\}$;
      repeat until stabilization
        $E := E \cup \{w \mid (\text{succ}(w) \cap (E1 \cup (E2 \cap E))) \neq \{\}\}$;
      return E;
    $\mathbf{A}(\psi_2 \, \mathbf{U}^+ \, \psi_1)$: $E1 :=$ eval(ψ_1); $E2 :=$ eval(ψ_2); $E := \{\}$;
      repeat until stabilization
        $E := E \cup \{w \mid \{\} \neq \text{succ}(w) \subseteq E1 \cup (E2 \cap E)\}$;
      return E;

intset = **return** $\{w' \mid (w, w') \in \mathcal{I}(\prec)\}$;

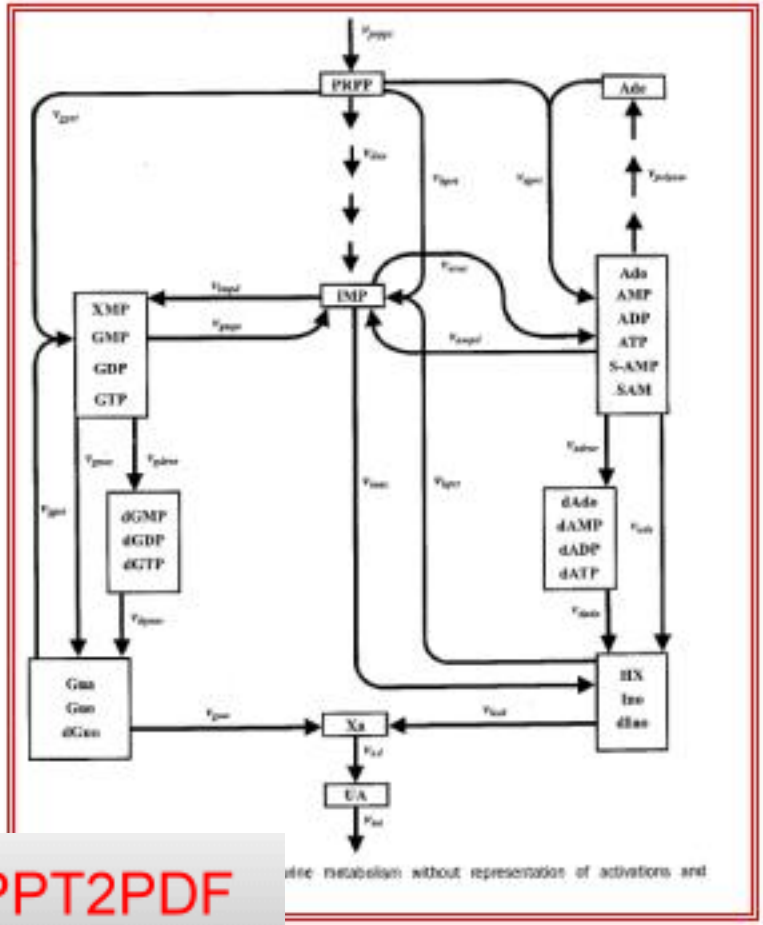
Other Model Checking Algorithms

- LTL Model Checking: Tableu-based…
- CTL* Model Checking: Combine CTL and LTL Model Checkers…
- Symbolic Model Checking
  - Binary Decision Diagram
  - OBDD-based model-checking for CTL
  - Fixed-point Representation
  - Automata-based LTL Model-Checking
- SAT-based Model Checking
- Algorithmic Algebraic Model Checking
- Hierarchical Model Checking

Purine Metabolism

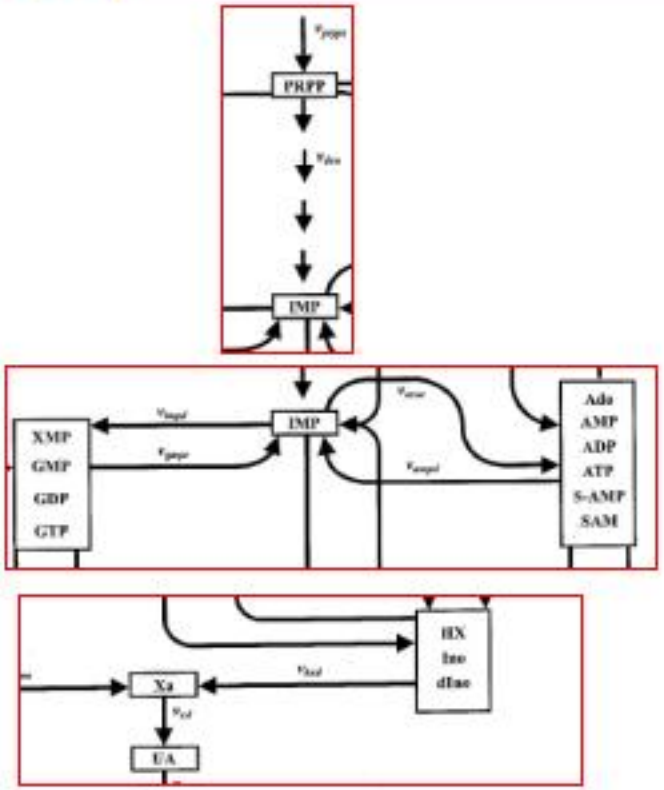
- **Purine Metabolism**
  - Provides the organism with building blocks for the synthesis of DNA and RNA.
  - The consequences of a malfunctioning purine metabolism pathway are severe and can lead to death.
- **The entire pathway is almost closed but also quite complex. It contains**
  - several feedback loops,
  - cross-activations and
  - reversible reactions
- **Thus is an ideal candidate for reasoning with computational tools.**

Simple Model



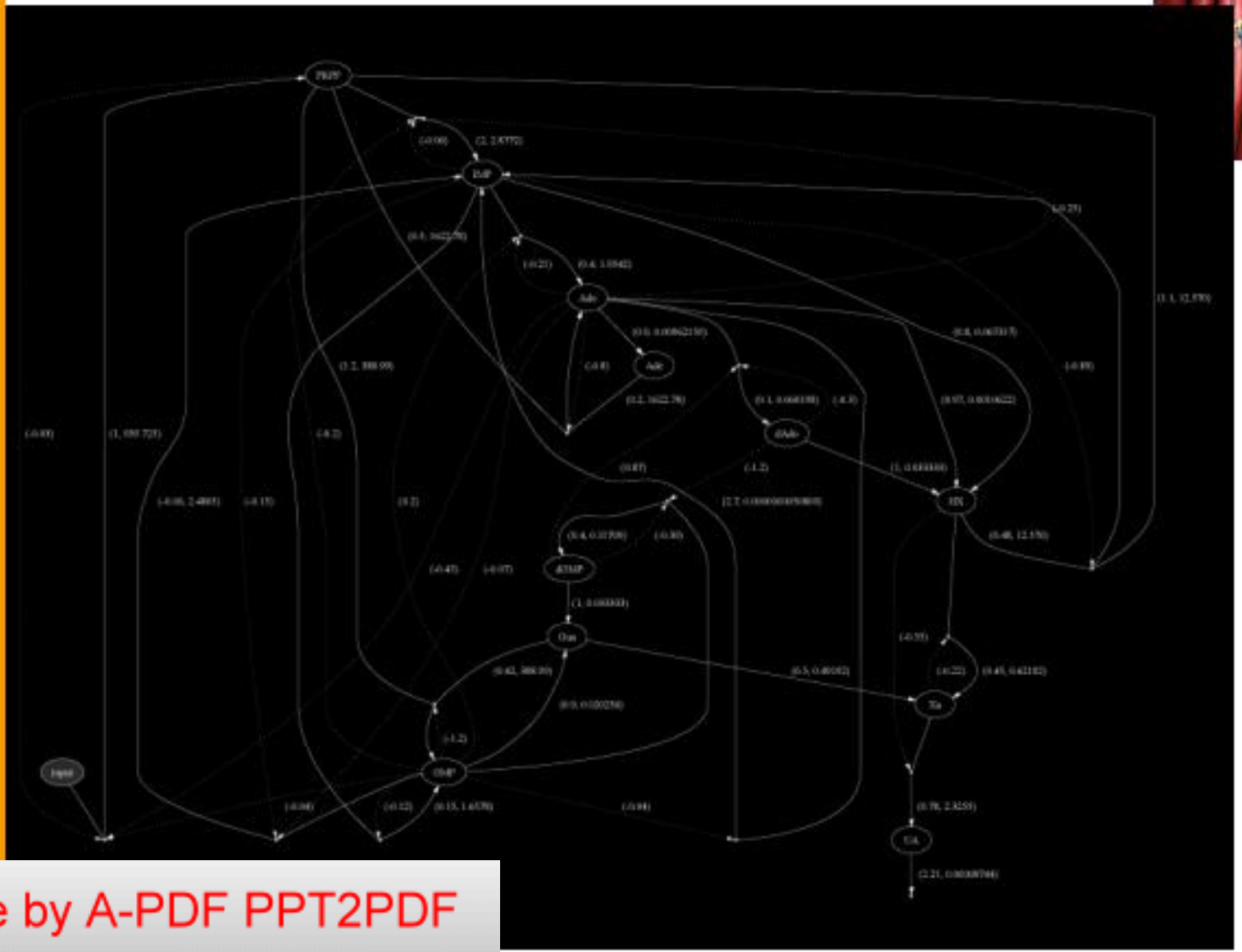
ine metabolism without representation of activations and

Biochemistry of Purine Metabolism



- The main metabolite in purine biosynthesis is *5-phosphoribosyl-a-1-pyrophosphate* (*PRPP*).
  - A linear cascade of reactions converts PRPP into *inosine monophosphate* (*IMP*). IMP is the central branch point of the purine metabolism pathway.
  - IMP is transformed into AMP and GMP.
  - Guanosine, adenosine and their derivatives are recycled (unless used elsewhere) into *hypoxanthine* (*HX*) and *xanthine* (*XA*).
  - XA is finally oxidized into *uric acid* (*UA*).

Purine Metabolism



Queries

- Variation of the initial concentration of PRPP does not change the steady state. **(PRPP = 10 * PRPP1) implies steady_state()**
- This query will be true when evaluated against the modified simulation run (i.e. the one where the initial concentration of PRPP is 10 times the initial concentration in the first run – PRPP1).

TRUE

- Persistent increase in the initial concentration of PRPP does cause unwanted changes in the steady state values of some metabolites.
- If the increase in the level of PRPP is in the order of 70% then the system does reach a steady state, and we expect to see increases in the levels of IMP and of the hypoxanthine pool in a "comparable" order of magnitude.

Always (PRPP = 1.7*PRPP1) implies steady_state()

TRUE

Queries

- Consider the following statement:
- Eventually

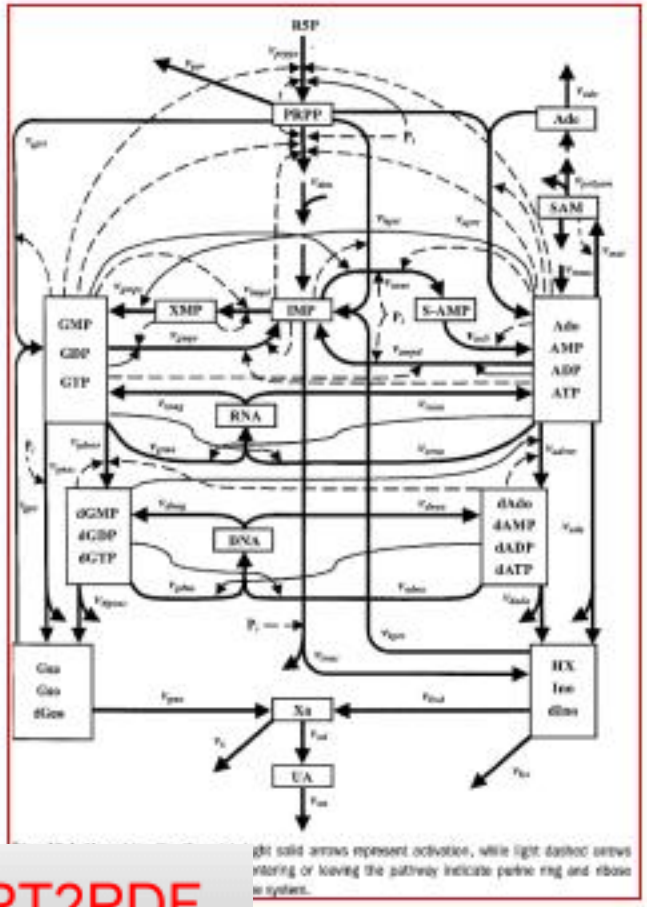
(Always (PRPP = 1.7 * PRPP1)
implies
steady_state()
and Eventually
Always(IMP < 2* IMP1))
and Eventually (Always
(hx_pool < 10*hx_pool1)))

- where IMP1 and hx_pool1 are the values observed in the unmodified trace. The above statement turns out to be false over the modified experiment trace..

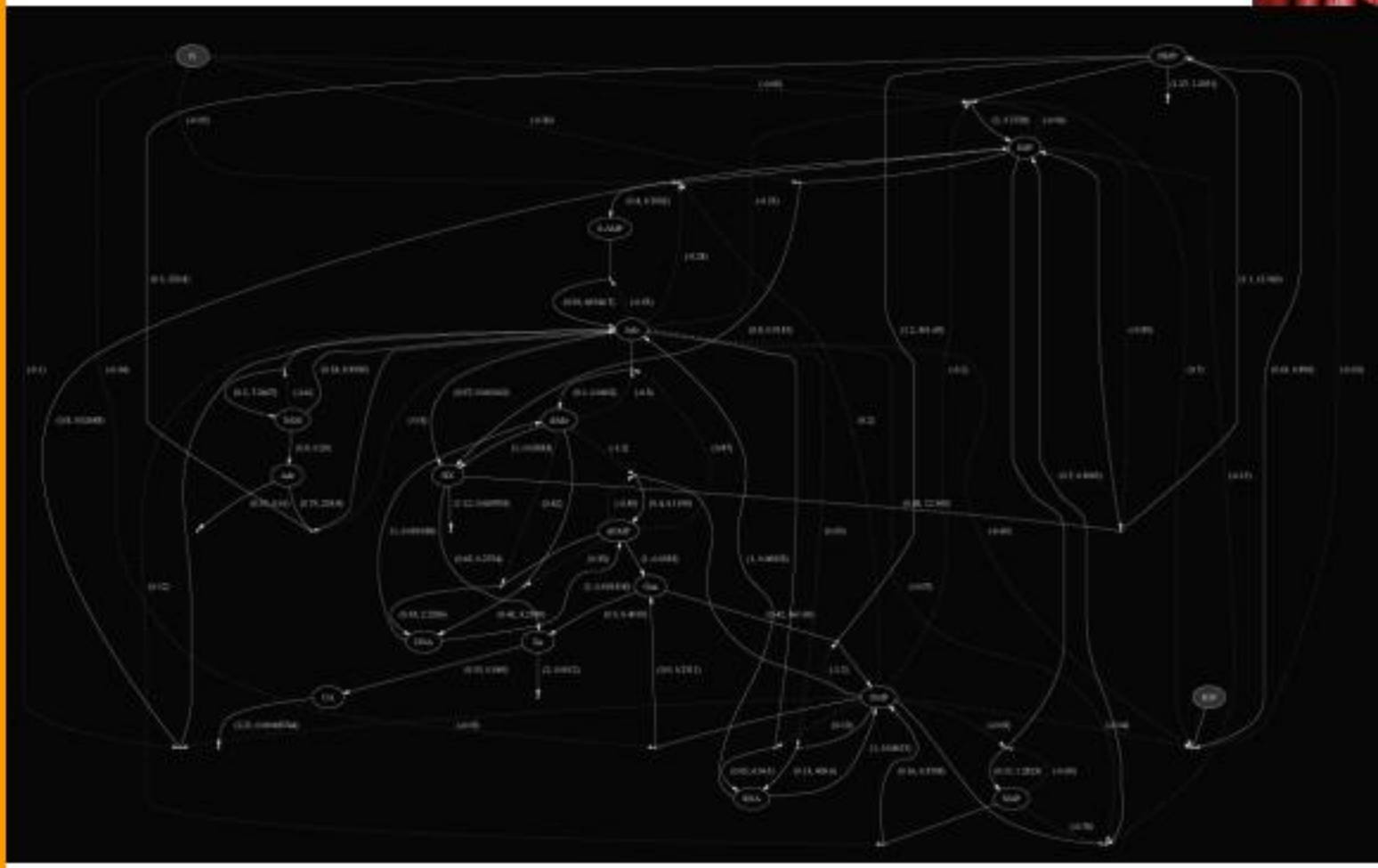
False

- In fact, the increase in IMP is about 6.5 fold while the hypoxanthine pool increase is about 60 fold.
- Since the above queries turn out to be false over the modified trace, we conclude that the model "over-predicts" the increases in some of its products and that it should therefore be amended

Final Model



Purine Metabolism



Continuous-Time Logics

- Linear Time
  - Metric Temporal Logic (MTL)
  - Timed Propositional Temporal Logic (TPTL)
  - Real-Time Temporal Logic (RTTL)
  - Explicit-Clock Temporal Logic (ECTL)
  - Metric Interval Temporal Logic (MITL)
- Branching time
  - Real-Time Computation Tree Logic (RTCTL)
  - Timed Computation Tree Logic (TCTL)

Alur et al,

TCTL: Syntax And Semantics

Basic Syntax And Semantics *The basic syntax of TCTL is:*

$$\phi ::= p \mid \neg\phi \mid \phi_1 \vee \phi_2 \mid \phi_1 \exists\mathcal{U} \phi_2 \mid \phi_1 \forall\mathcal{U} \phi_2 \mid z.\phi$$

- **z.** *The freeze quantification "z." binds the associated variable z to the current time. Thus the formula $z.\phi(z)$ holds at time t iff $\phi(t)$ does.*
- *$\forall\mathcal{U}$ and $\exists\mathcal{U}$: universal and existential "until" operators. It is common notation to subscript the* Until *operator as $p \exists\mathcal{U}_{\leq t} q$ to indicate that q has to be satisfied within t time units. This is just a convenient notation for $p \exists\mathcal{U} (q \wedge (t \leq 5))$.*
  E.g. $(p)\forall\mathcal{U}(q)$ asks whether on any path *leading off the state where the modal formula is being considered, p is true everywhere until the state where q is true. (q is required to be true somewhere, and p is required to be true until the previous instant, but not necessarily at the point where q becomes true)*

T-μ CALCULUS: Syntax

While μ-calculus works for discrete time systems, we have Tμ-calculus for capturing the continuous time properties of hybrid systems:

$$\phi ::= X \mid p \mid \neg\phi \mid \phi_1 \vee \phi_2 \mid \phi_1 \triangleright \phi_2 \mid z.\phi \mid \mu X.\phi$$

Note that though we mention only the least-fixpoint μ, *the* greatest-fixpoint ν *can be expressed as* $\neg\mu X.(\neg\phi[X := \neg X])$.

"Until": T- μ Fixpoint

$$s_1 \; \exists \mathcal{U} \; s_2 \; = \; \mu X.(s_2 \vee (s_1 \triangleright X))$$

- **s2** is true now **or**
- **s1** holds for *one-step on some path* after which **s2** holds **or**
- **s1** holds for *one-step on some path* after which **s1** holds for *one more step on some path* after which **s2** holds **or**
- and so on..

Since the universal unti *can be computed using* $s_1 \forall \mathcal{U} s_2 \; = \; \neg((\neg s_2) \; \exists \mathcal{U} \; (\neg s_1 \wedge \neg s_2))$, *it is sufficient to focus on* $\exists \mathcal{U}$. [3]

TCTL Model Checking

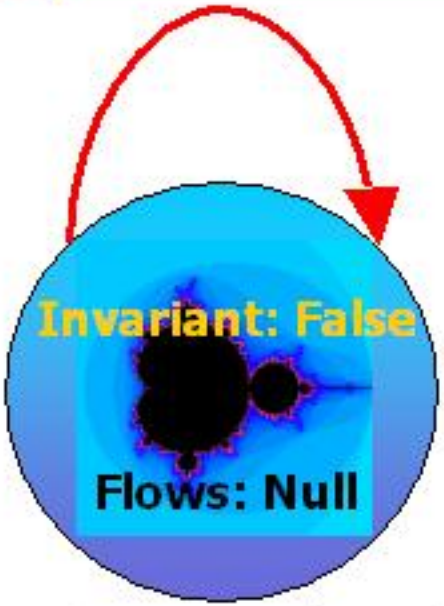
- Only "Until" requires "computation"
- Until: Iterative computation of "one-step" Until
- Least fixpoint computation:

1. $\psi := \text{false}$
2. repeat
   (a) $\phi := \psi$
   (b) $\psi := \phi[X := \phi]$
   (c) until $[\phi] = [\psi]$
3. return ϕ

Semi-Decidability Of TCTL

- Global "time" variable
  - Allows interpretation of the TCTL operators freeze (z.X) and subscripted until (Ua)
- While "one-step until" is decidable, the fixpoint is not guaranteed to converge
- So TCTL is "semi"-decidable

Mandelbrot Hybrid Automaton



$$(x' = x^2 - y^2 + C_r) \wedge (y' = 2xy + C_i)$$

Let:

$$C = C_r + \imath.C_i$$
$$S(t) = x(t) + \imath.y(t)$$

Then:

$$S'(t) = S^2(t) + C$$

$$S'(t) = \{x(t)^2 - y(t)^2 + C_r\} + \imath.\{2x(t)y(t) + C_i\} = \{x(t) + \imath.y(t)\}^2 + \{C_r + \imath.C_i\}$$

Reachability Query: $(x_1^2 + x_2^2 \geq 4)$

Solution

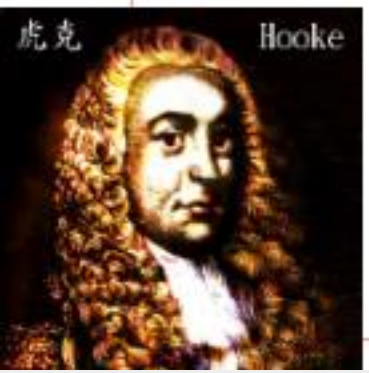
- Bounded Model Checking
  - Fully O-minimal Systems for Dense CTL
- Constrained Systems
  - Linear Systems for Dense CTL
  - O-minimal for Dense CTL
  - SACoRe (Semi algebraic Constrained Reset) for TCTL
  - IDA (Independent Dynamics Automata) for TCTL

Hooke

Thursday 25 May 1676

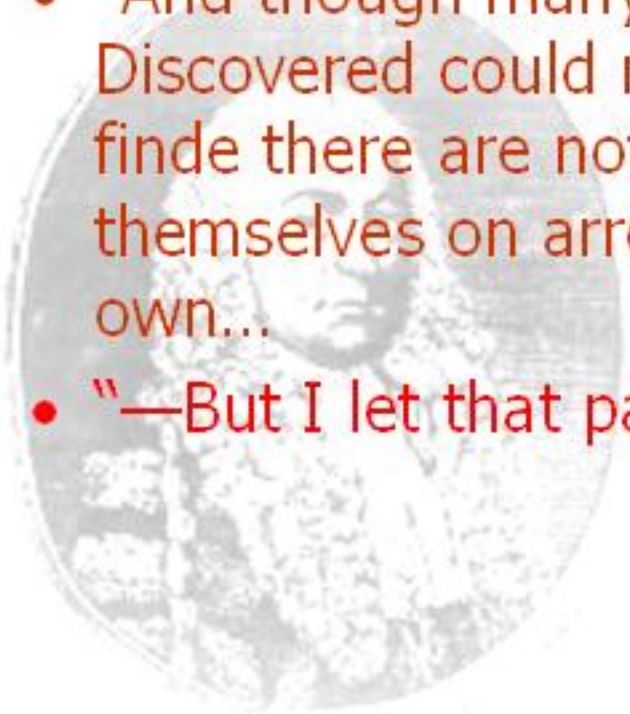
Damned Doggs.

Vindica me deus.



- Commenting on Sir Nicholas Gimcrack character in *The Virtuoso*, a play by Thomas Shadwell.

Hooke
in the Royal Society, 26 June 1689



- "And though many things I have first Discovered could not find acceptance yet I finde there are not wanting some who pride themselves on arrogating of them for their own...
- "—But I let that passe for the present."

Hooke...

- "So many are the links, upon which the true Philosophy depends, of which, if any can be loose, or weak, the whole chain is in danger of being dissolved;
- "it is to begin with the Hands and Eyes, and to proceed on through the Memory, to be continued by the Reason;
- "nor is it to stop there, but to come about to the Hands and Eyes again, and so, by a continuall passage round from one Faculty to another, it is to be maintained in life and strength."

The end...