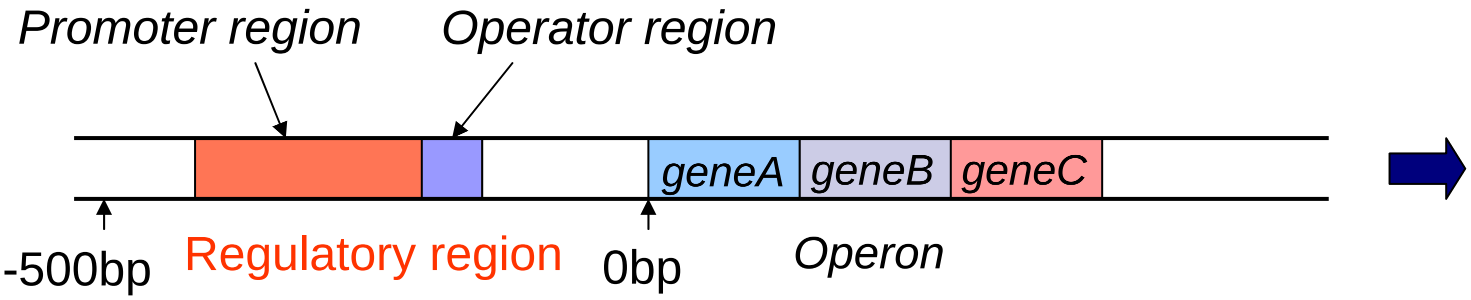
Estimation of Gene Interaction Models

- from a system-theoretic point of view -

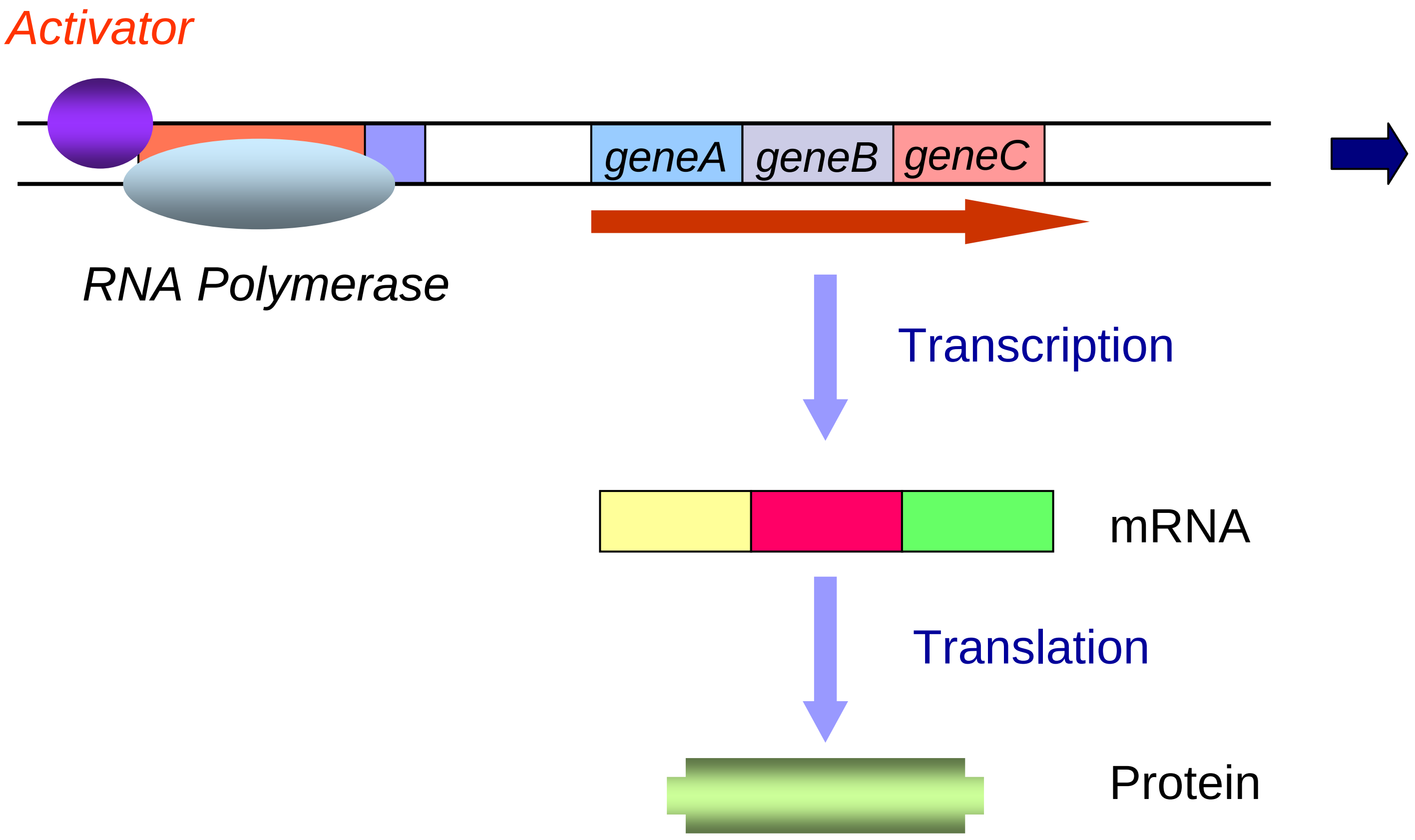
Kunihiko HIRAISHI
School of Information Science,
JAIST

AB2005 Tutorial 2005/11/28

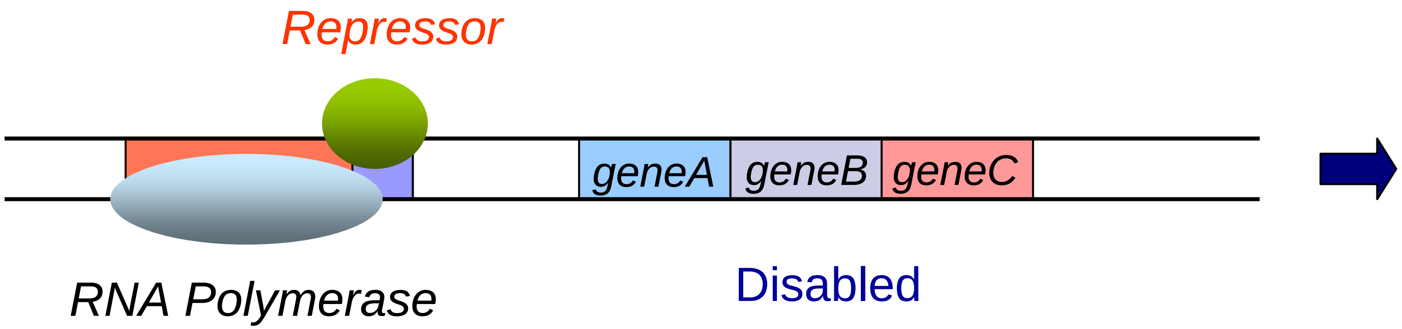
Gene expression mechanism



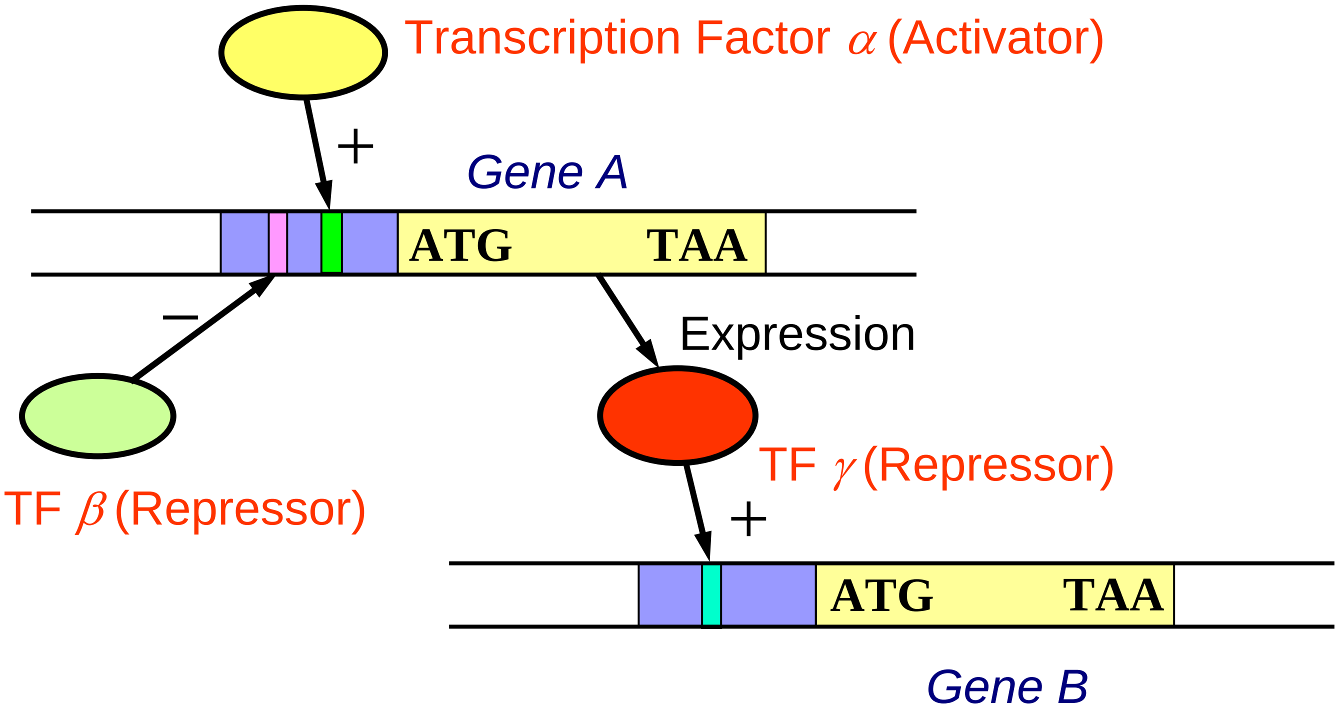
Gene expression mechanism



Gene expression mechanism



Gene interaction network



Transcription Factor α (Activator)

+

Gene A

−

Expression

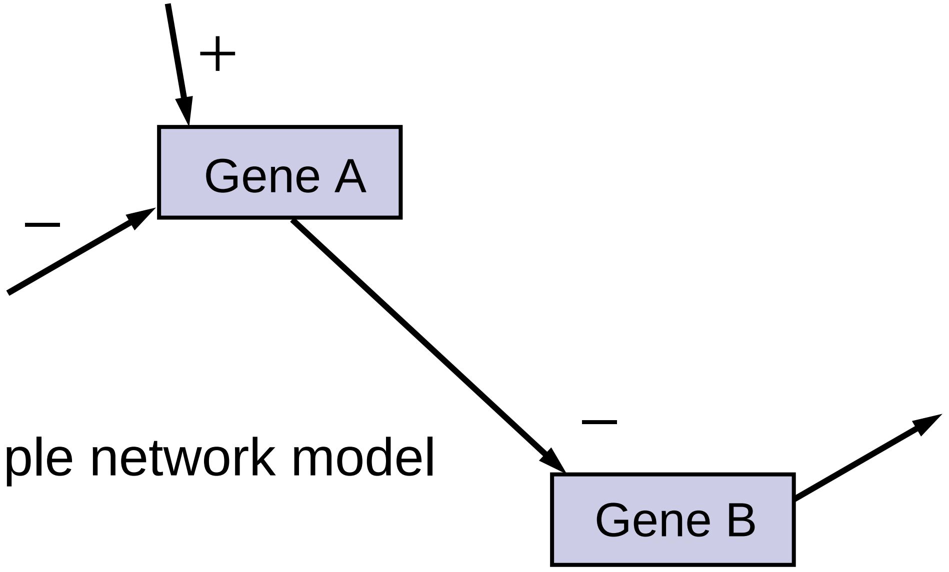
TF β (Repressor)

TF γ (Repressor)

+

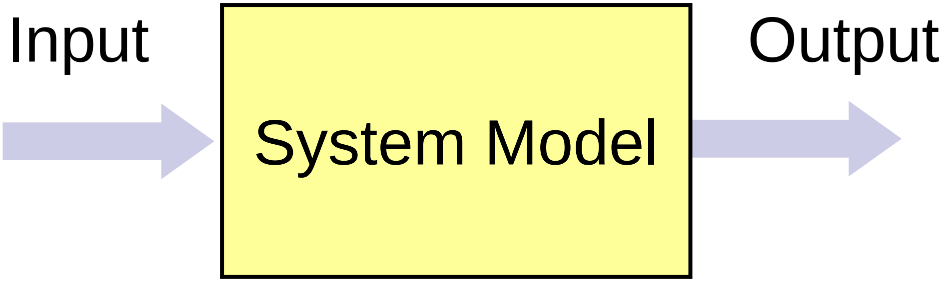
Gene B

Gene regulatory network



a simple network model

System identification problem

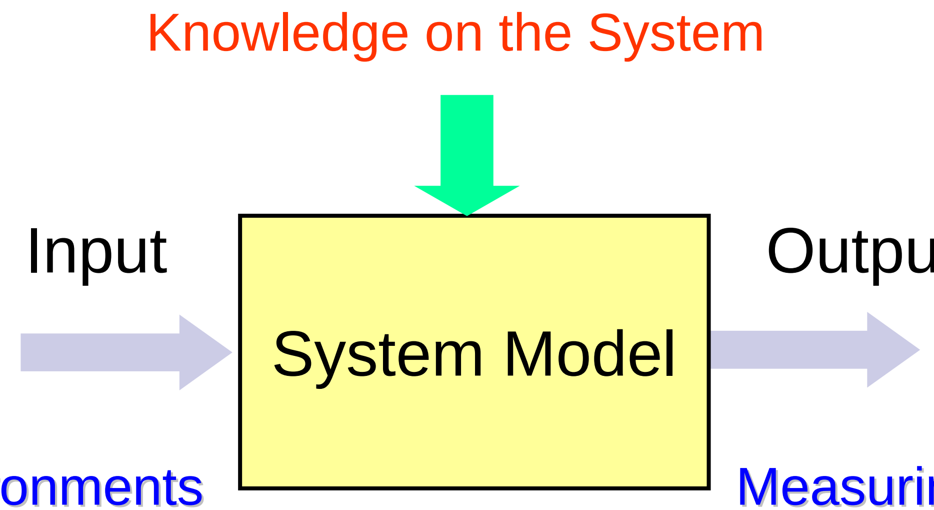


Input

System Model

Output

System identification problem



Knowledge on the System

Input

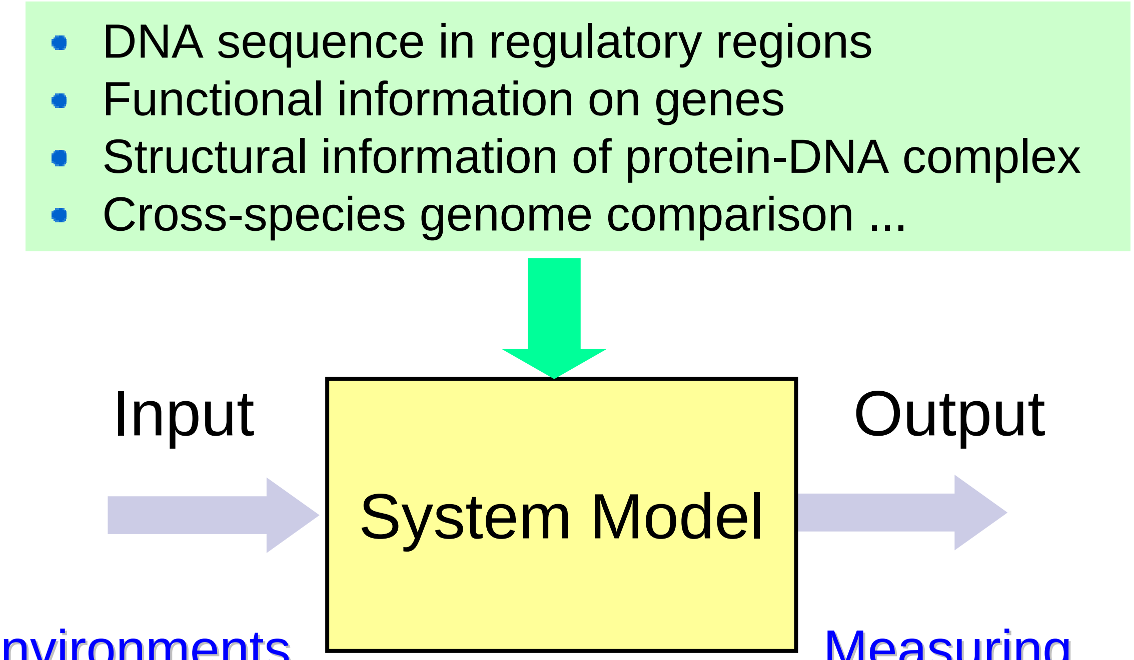
System Model

Output

- Different environments
- Experimental operations
  (e.g., disruption of particular genes)

Measuring
Gene expression level
(DNA Microarray)

System identification problem



- DNA sequence in regulatory regions
- Functional information on genes
- Structural information of protein-DNA complex
- Cross-species genome comparison ...

Input

System Model

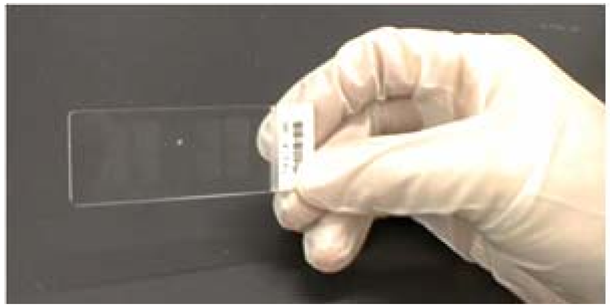
Output

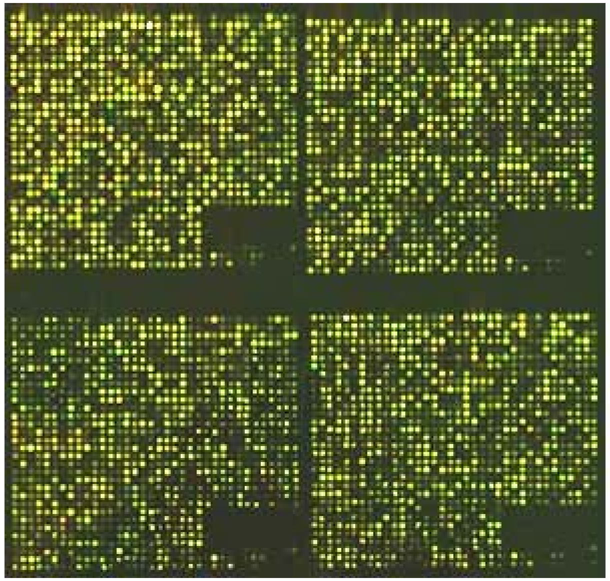
- Different environments
- Experimental operations (e.g., disruption of particular genes)

Measuring
Gene expression level
(DNA Microarray)

DNA microarrays

DNA microarrays are used for measuring the expression levels of large numbers of genes simultaneously.





DNA microarrays

Data#	1	2	3	4
$gene_1$	3.18	0.45	7.63	1.2
$gene_2$	0.25	5.53	0.87	4.61
$gene_3$	1.26	0.01	6.17	0.87

different environments

time	0	1	2	3
$gene_1$	1.43	0.56	3.78	0,03
$gene_2$	0.84	4.11	0.36	1.56
$gene_2$	2.39	0.43	2.28	0.56

time-series

Sequence analysis

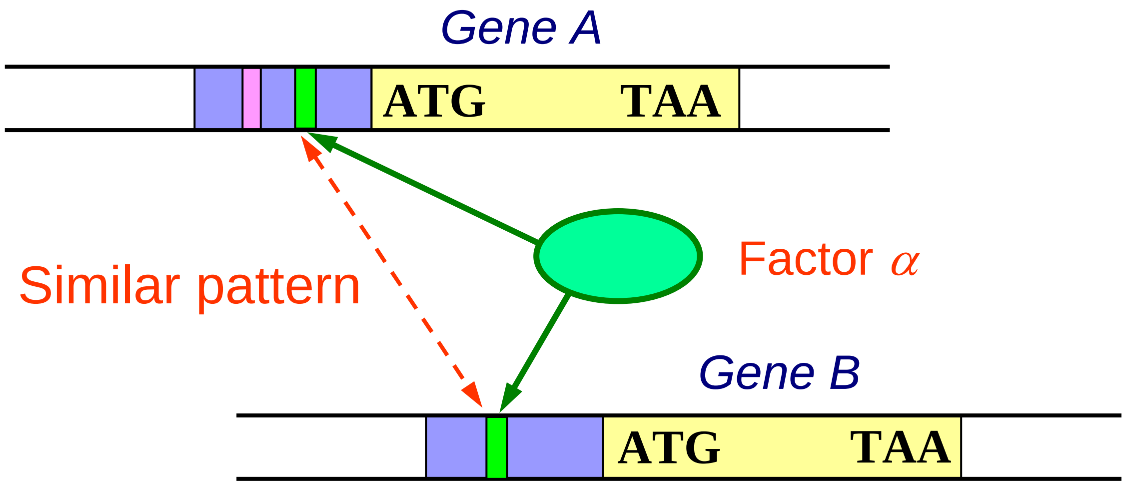
Transcription factors recognize patterns.

Transcription factor: *PerR*

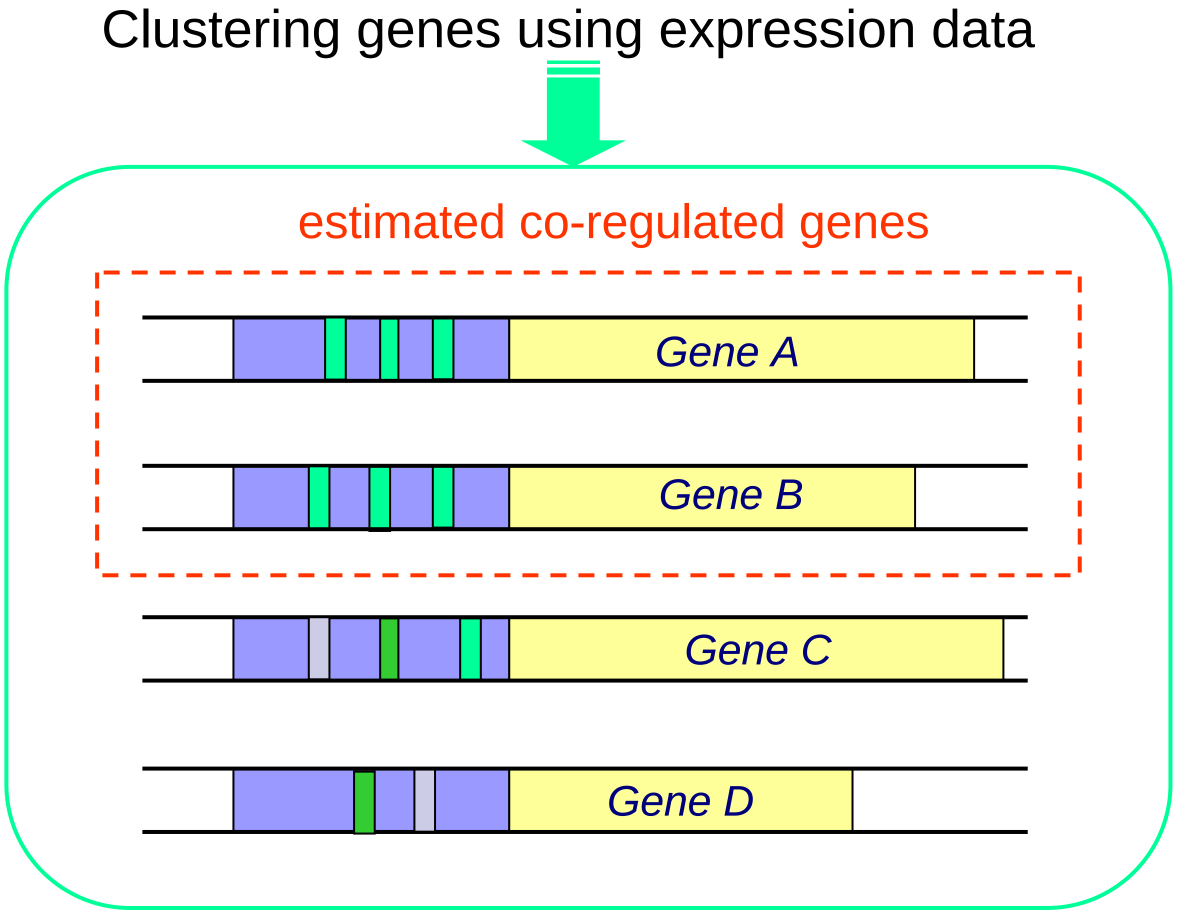
Regulated gene	Operon	Sigma	Regulation	Absolute position	Location	Binding seq.(cis-element)	Exp.	Reference	Year
ahpC	ahpCF	ND	Negative	4118058..4118119	ND	CTTGACAAAAAATATATATTAATTAATAATTCATATATAATTAGAATTATTATTGAAAGCGA	FT	Herbig, A. F., et al.	2001
fur	ND	ND	Negative	2449580..2449594	-49:-35	TTATAATAATTATAG	FT	Fuangthong, M., et al.	2002
hemA	ND	ND	Negative	2878294..2878322	ND	AGAAACTATGTTATAATTATTATAAATAA	FT	Herbig, A. F., et al.	2001
hemA	ND	ND	Negative	2878248..2878289	ND	TTCTATGTTAGAATGATTATAAATTAAGATTGGGTGTTGGGG	FT	Herbig, A. F., et al.	2001
katA	ND	ND	Negative	960520..960577	ND	CTATTTTATAATAATTATAAAATAATATTGACTTTTTACTTAGAGATGATATTATGTT	FT	Herbig, A. F., et al.	2001
mrgA	ND	ND	Negative	3382535..3382589	ND	TCTAAATTATAATTATTATAATTTAGTATTGATTTTTATTTAGTATATGATATAA	FT	Herbig, A. F., et al.	2001
perR	ND	ND	Negative	943933..943958	-13:+13	TTACACTAATTATAAACATTACAATG	FT	Fuangthong, M., et al.	2002
perR	ND	ND	Negative	943942..943964	-4:+18	TTATAAACATTACAATGTAAGAA	FT	Fuangthong, M., et al.	2002
ykvW	ND	ND	Negative	1450655..1450705	-75:-25	TAATGATAATTATTATCAAAAAGAAATTAAAATAATTATAATTGAAATTCT	FT	Gaballa, A., et al.	2002

http://dbtbs.hgc.jp/

Sequence analysis



Sequence analysis

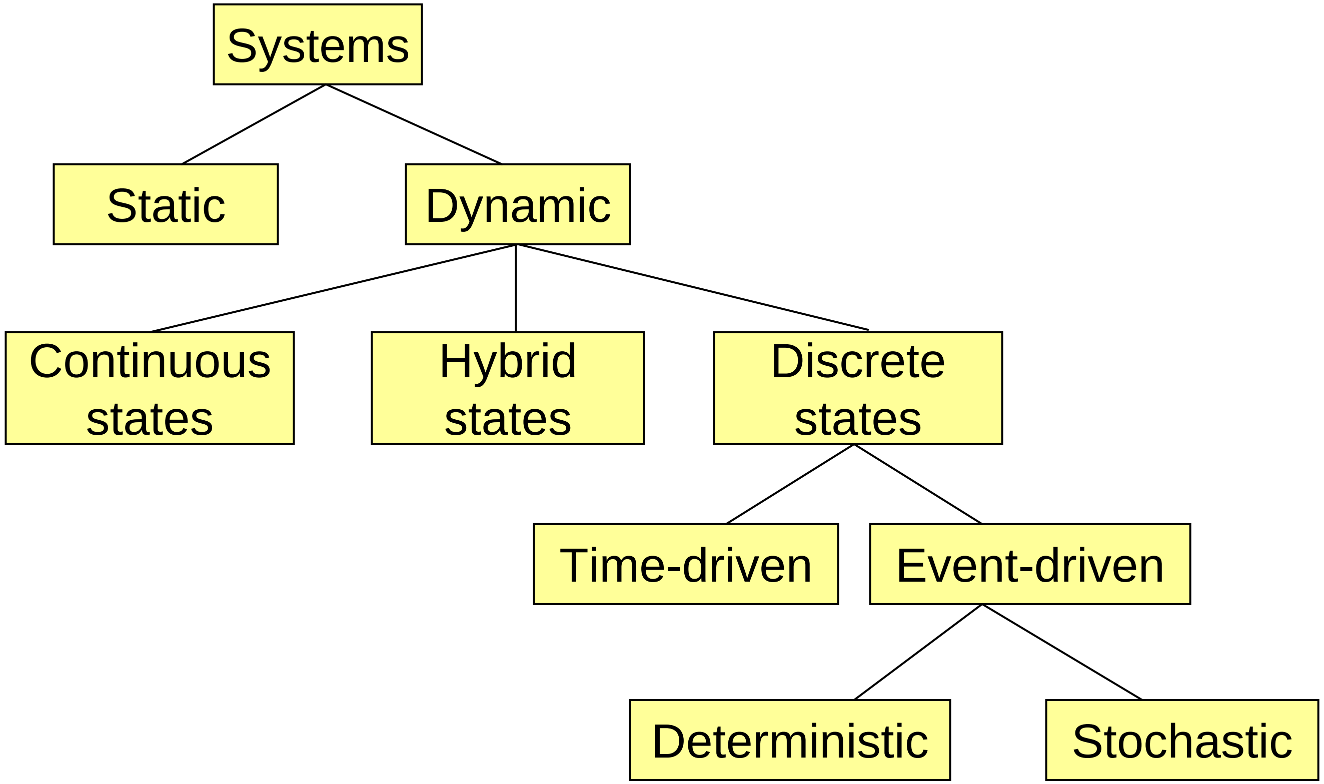

Clustering genes using expression data

estimated co-regulated genes

Question

- Which formalisms can be used for modeling gene regulatory networks?
- Each formalism should deal with
  - *vector states,*
  - *dependence on genes, and*
  - *dynamics on gene expression.*

Classification of systems



Systems

Static

Dynamic

Continuous states

Hybrid states

Discrete states

Time-driven

Event-driven

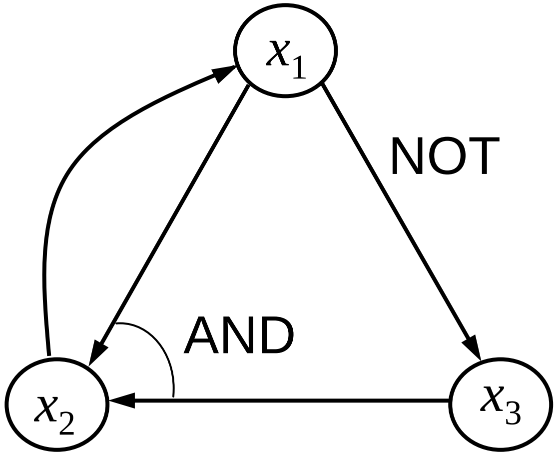
Deterministic

Stochastic

Classification of models

- *Deterministic / Stochastic* : Is the model deterministic (e.g. Boolean networks) or stochastic (e.g. Bayesian networks)?
- *Discrete / Continuous / Hybrid* : Is the model defined on discrete state changes (e.g. automata) or continuous state changes (e.g. differential equations) or a mixture of discrete and continuous state change (e.g. hybrid automata)?
- *Qualitative / Quantitative* : Does the model capture the qualitative nature of the system (e.g. qualitative reasoning and simulation models) or run on numerical values?
- *Detailed or not*: Does the model include the information on which TFs are bound to the regulatory region of each gene or include the information only on the positive/negative effect of genes to other genes?

Boolean networks



next-state function

$x'_1 := x_2$

$x'_2 := x_1 \wedge x_3$

$x'_3 := \neg x_1$

Deterministic, Discrete-value, Discrete-time, Synchronous

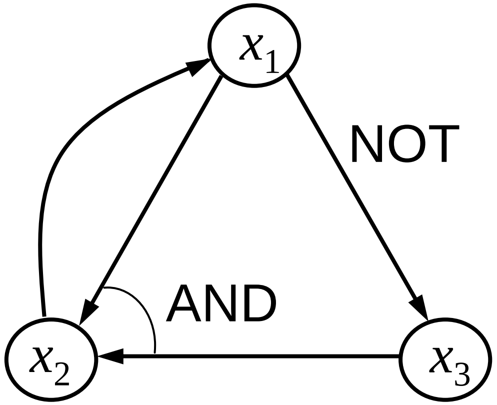
Boolean networks - learning

[Akutsu 1999]

Given a time-series data $D = \{ X_1, ..., X_k \}$ of instances of $[x_1, ..., x_n]$, find a network that best matches D.

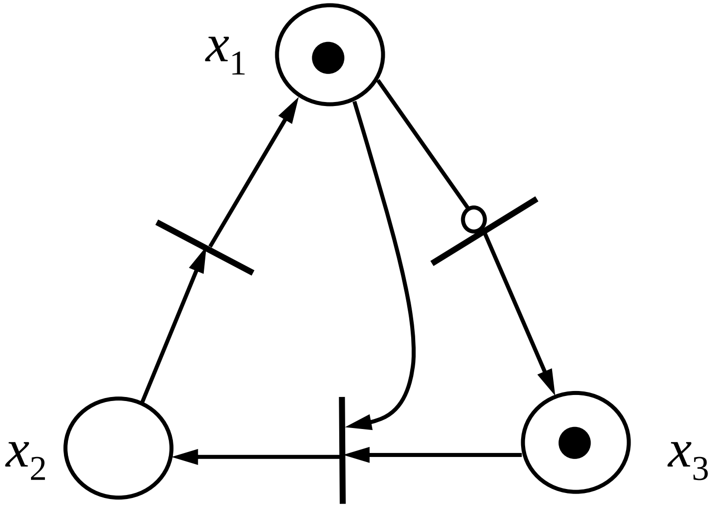
time	0	1	2	3
x_1	1	0	1	0
x_2	0	1	0	1
x_3	1	0	1	0

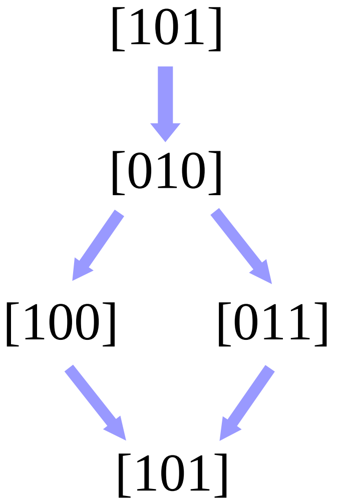
time-series expression data



Each value of $x_i(t)$ is a logical value indicating whether gene i is expressed or not at time step t.

Petri nets




[101]

[010]

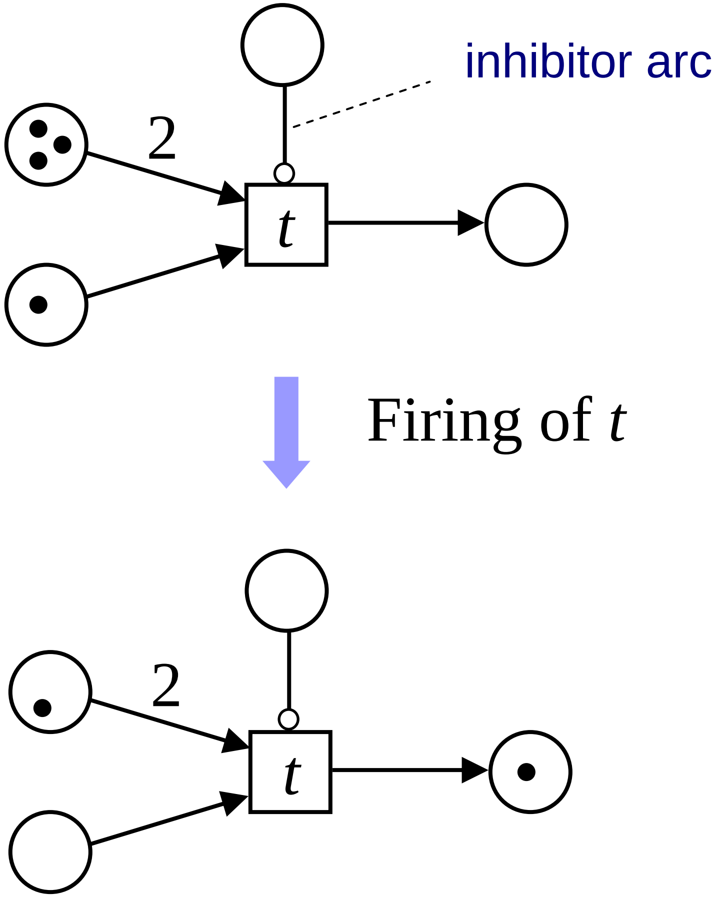
[100] [011]

[101]

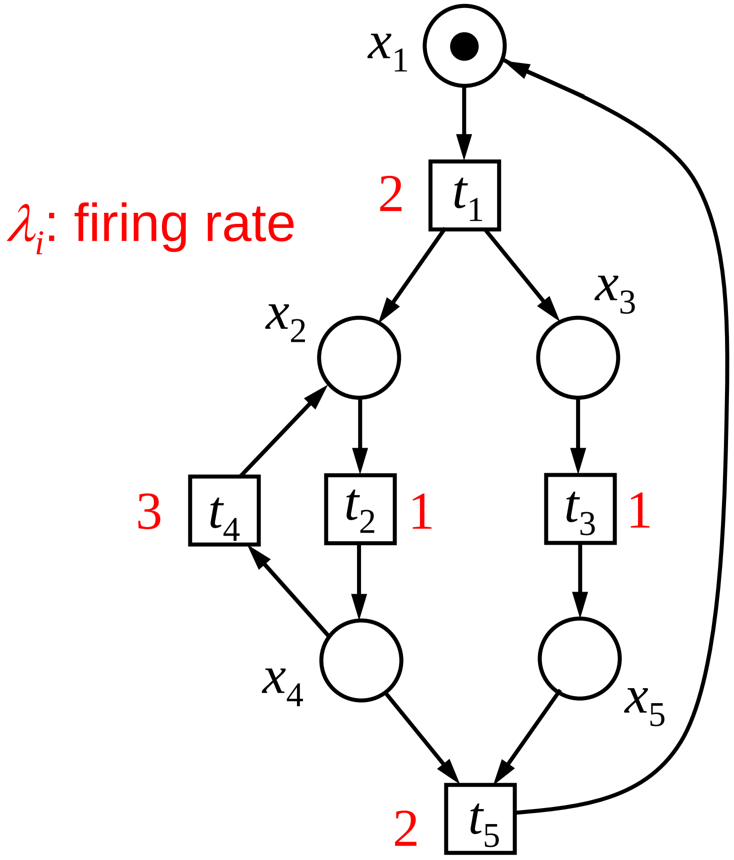
Reachability tree

Deterministic, Discrete-value, Discrete-time, Asynchronous

Dynamics of Petri Nets



Stochastic Petri nets



λ_i: firing rate

Each transition has an exponentially distributed random variable that represents firing delay.

Stochastic, Discrete-value, Continuous-time, Event-driven

Differential equations

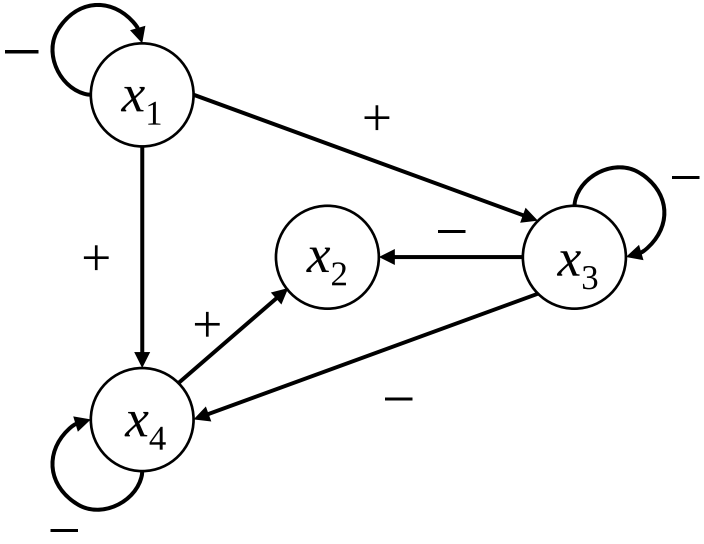
General form:

$$\frac{dx_i}{dt} = F_i^+(x_1, \cdots, x_n) - F_i^-(x_1, \cdots, x_n) \quad (i = 1, \cdots, n)$$

An approximation form (S-system) [Okamoto 2001]:

$$\frac{dx_i}{dt} = \alpha_i \prod_{j=1}^{n} x_j^{g_{ij}} - \beta_i \prod_{j=1}^{n} x_j^{h_{ij}} \quad (i = 1, \cdots, n)$$

Differential equations

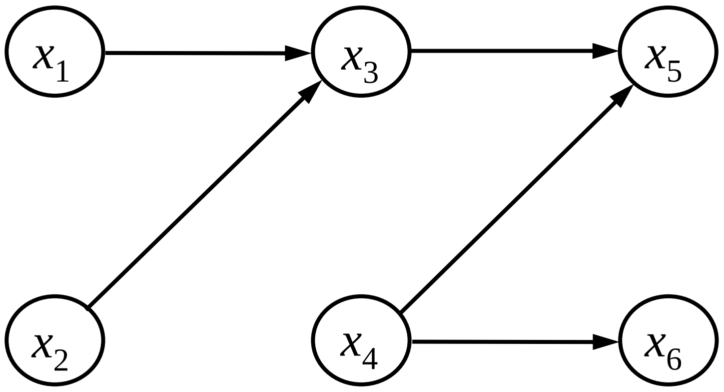


linear differential equations

$$dx_1(t) / dt = c_1 - a_{11}\, x_1(t)$$
$$dx_2(t) / dt = a_{31}\, x_3(t) + a_{42}\, x_4(t)$$
$$dx_3(t) / dt = c_3 + a_{13}\, x_1(t) - a_{33}\, x_3(t)$$
$$dx_4(t) / dt = a_{14}\, x_1(\mathrm{t}) - a_{34}\, x_3(\mathrm{t}) - a_{44}\, x_4(\mathrm{t})$$

Deterministic, Continuous-value, Continuous-time

Bayesian networks



conditional probabilities

$$P(x_3 \mid \pi(x_3)),\ \pi(x_3) = \{\ x_1, x_2\ \},$$
$$P(x_5 \mid \pi(x_5)),\ \pi(x_5) = \{\ x_3, x_4\ \},$$
$$P(x_6 \mid \pi(x_6)),\ \pi(x_6) = \{\ x_4\},$$

$$\begin{aligned} P(x_1, \ldots, x_6) &= \Pi_{i=1,n}\, P(xi \mid p(xi)) \\ &= P(x_1)P(x_2)\ P(x_3 \mid \pi(x_3))P(x_4)\ P(x_5 \mid \pi(x_5))\ P(x_6 \mid \pi(x_6)). \end{aligned}$$

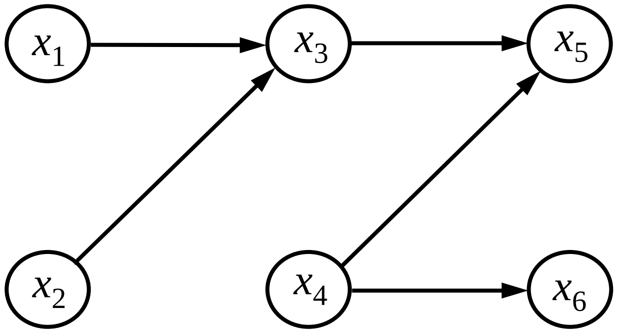
Stochastic, Continuous/Discrete-value, Static

Bayesian networks - learning

[Friedman 1999]

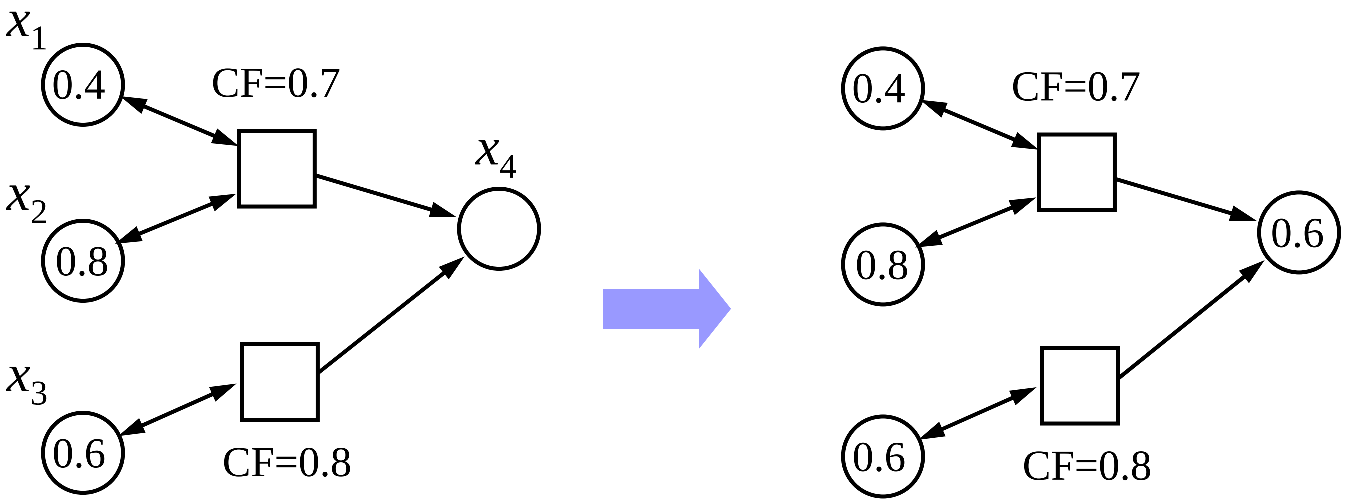
Given a training set $D = \{ X_1, ..., X_k \}$ of instances of $[x_1, ..., x_n]$, find a network that best matches D.

Data	1	2	3	4
x_1	3.18	0.45	7.63	1.2
x_2	0.25	5.53	0.87	4.61
x_3	1.26	0.01	6.17	0.87



Each value of $x_i(j)$ represents the expression level of gene i under environment j.

Fuzzy Petri nets

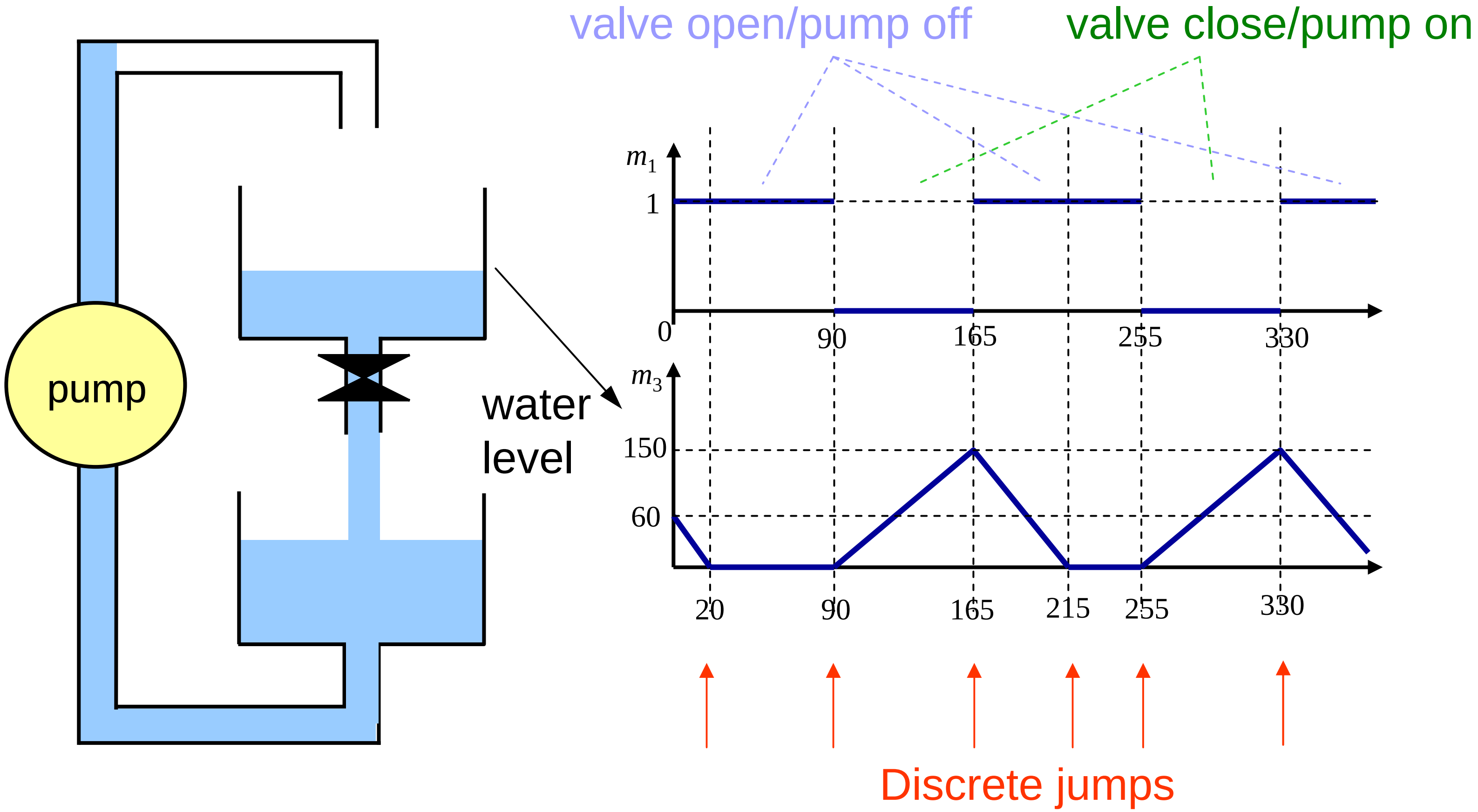


Each place has a fuzzy value in interval [0, 1].

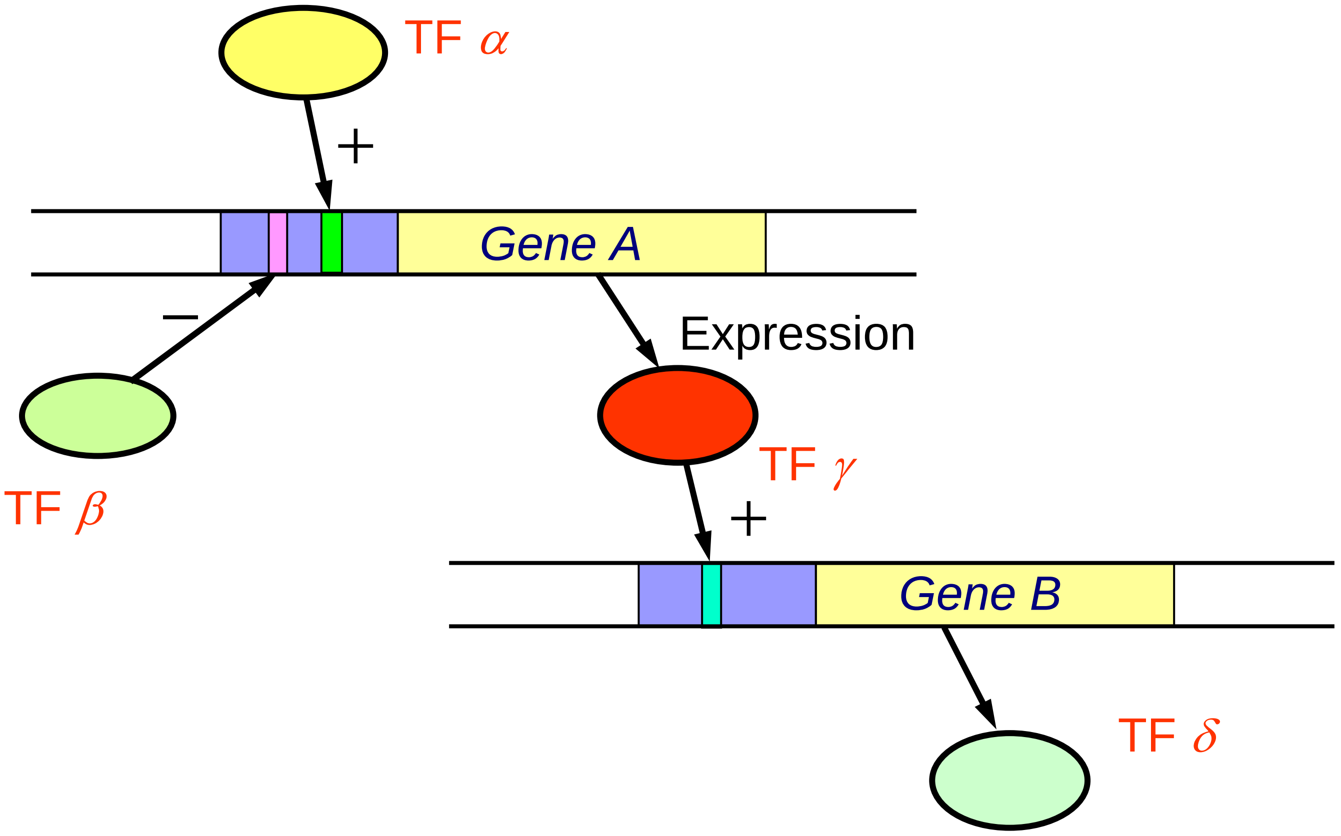
0.6 = max(0, min(0.7, min(0.4, 0.8)), min(0.8, 0.6)).

Fuzzy, Continuous-value, Static

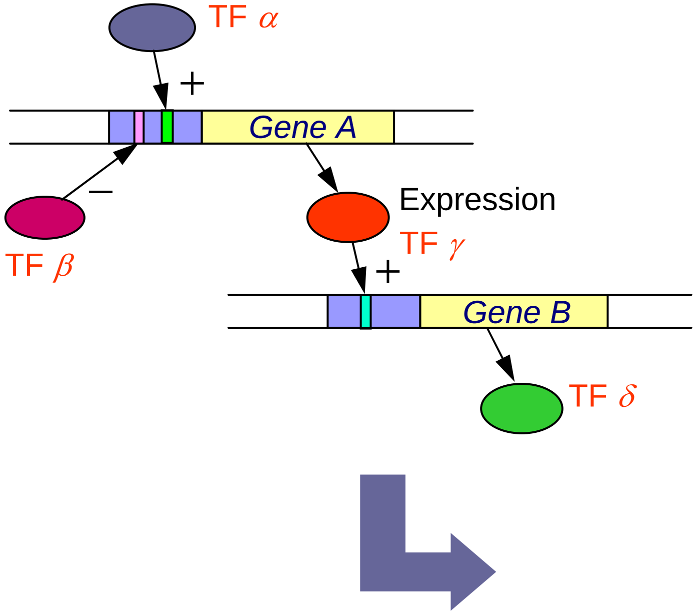
Hybrid systems

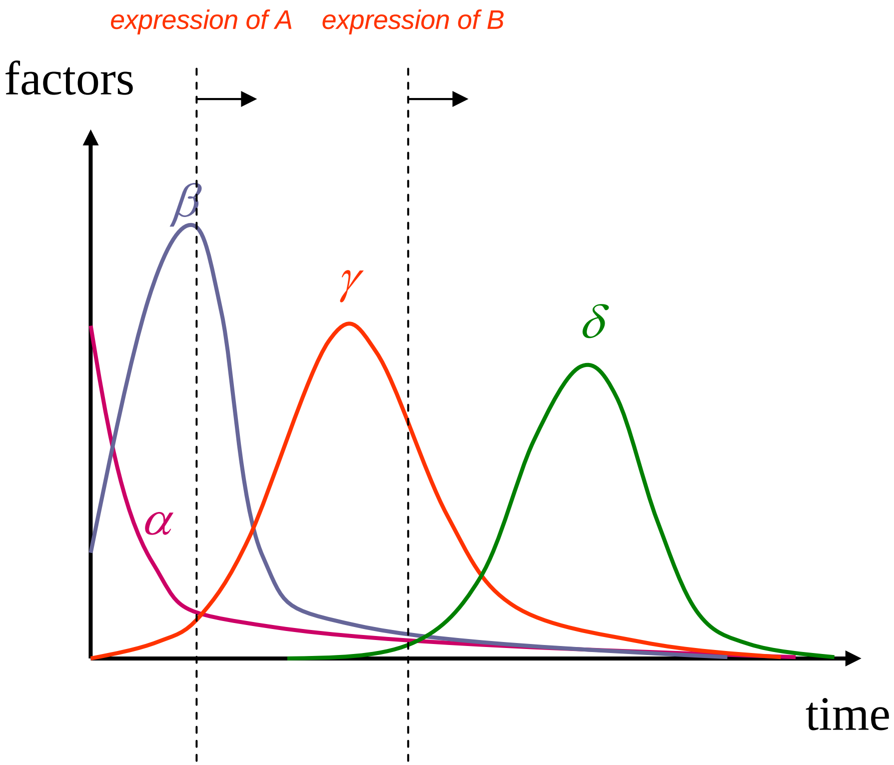


Gene interaction as a hybrid system

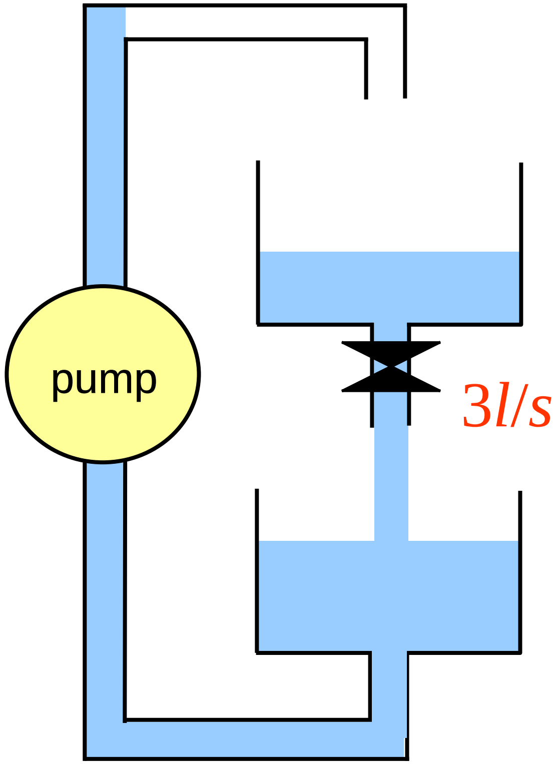


Gene interaction as a hybrid system

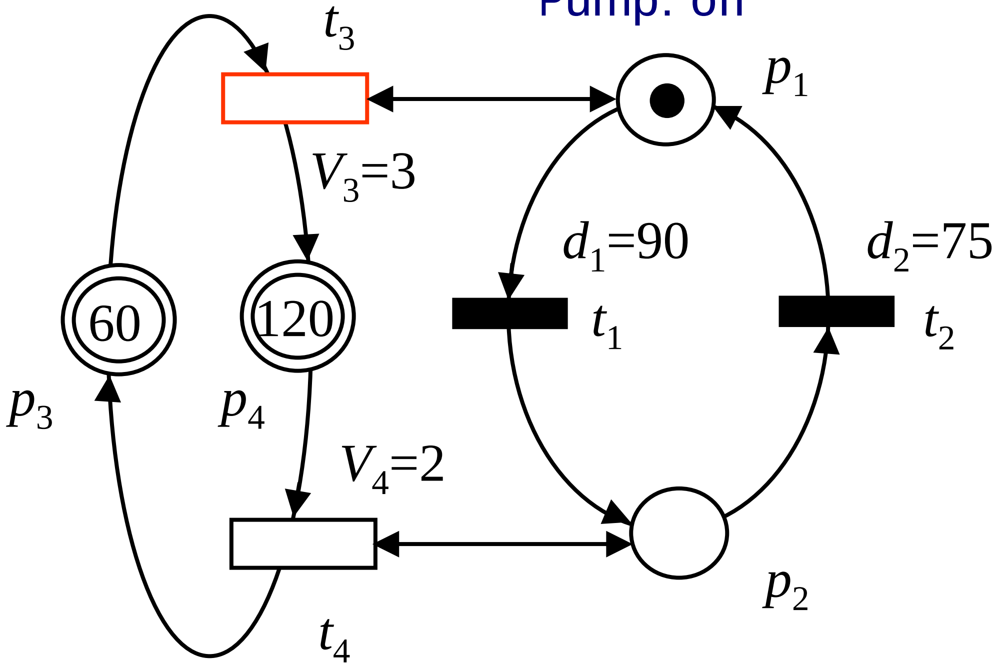




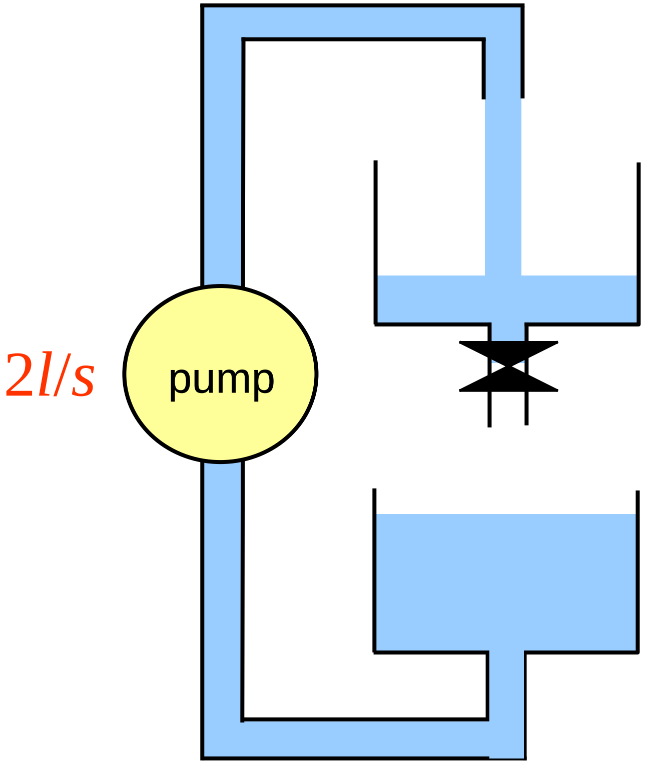
Hybrid Petri nets

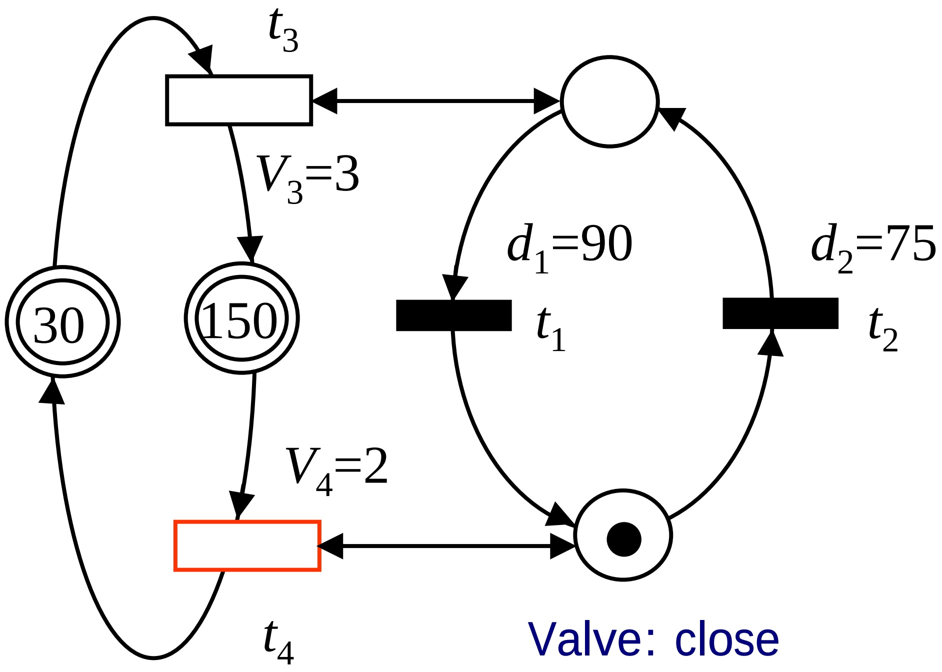


Valve: open
Pump: off



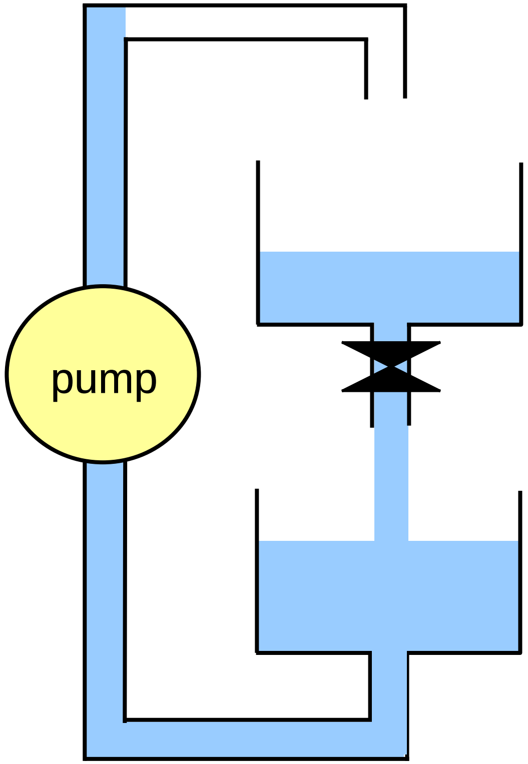
Hybrid Petri nets

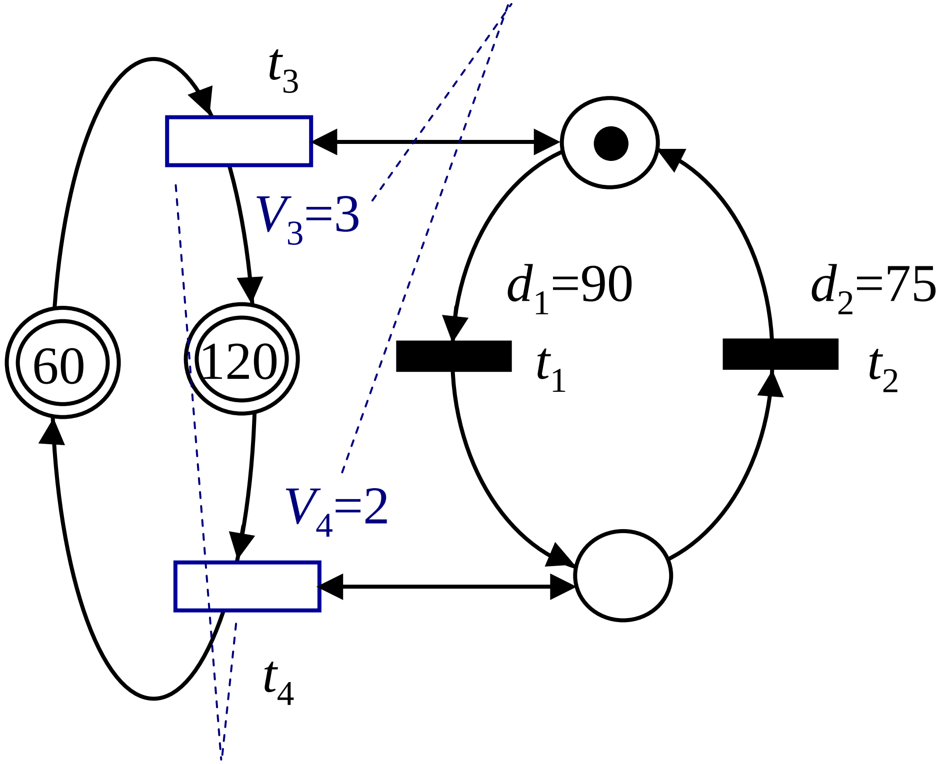




Valve: close
Pump: on

Hybrid Petri nets

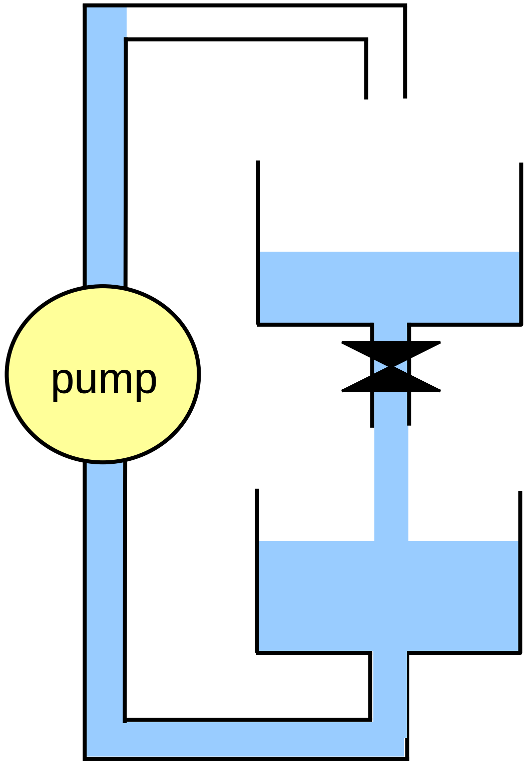


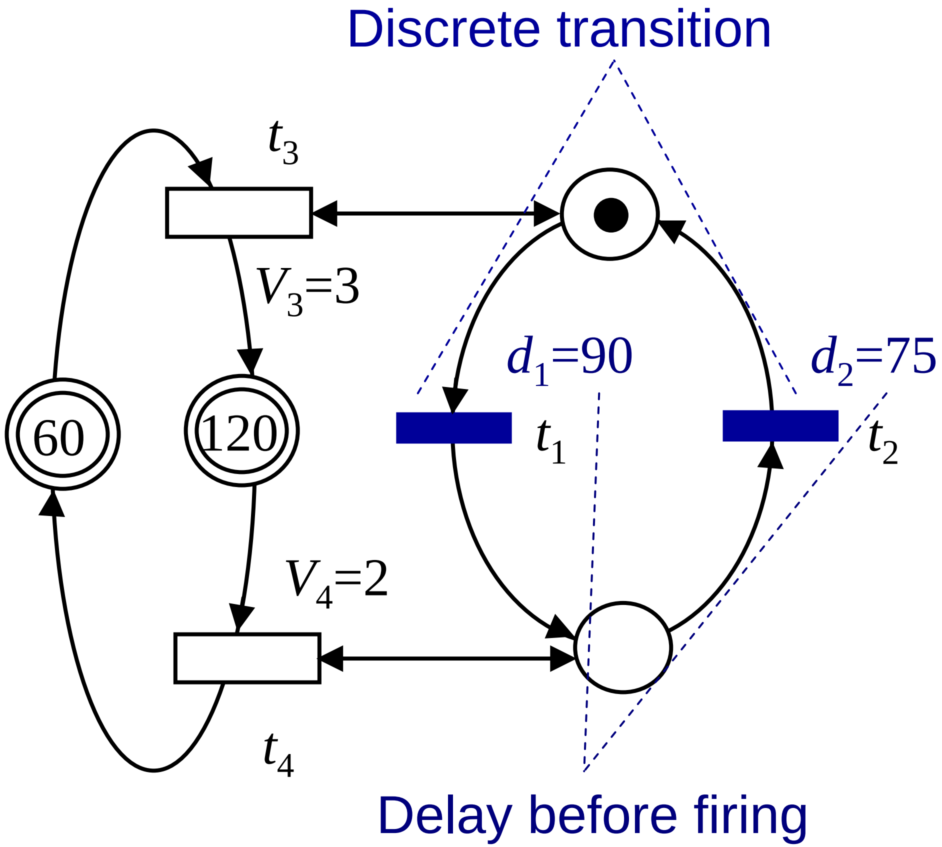


Maximum firing rate

Continuous transition

Hybrid Petri nets





Symbolic execution by CLP

conditions on event occurrence

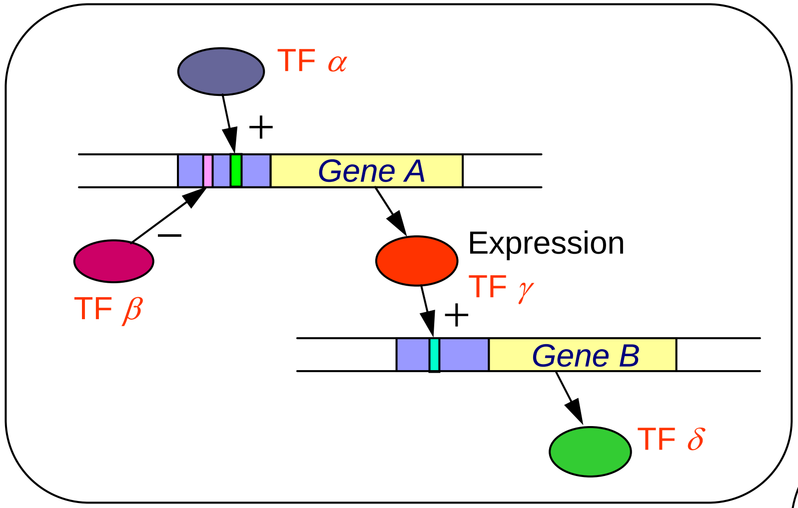
```
nextevent([1, 0, 0, P4, D1, D2], T, [0, 1, 0, P4, D1, 0], T + D):- !, D = 90 - D1.
nextevent([1, 0, P3, P4, D1, D2], T, [0, 1, P3 - 3 * D, P4 + 3 * D, D1, 0], T + D):- D = 90 - D1.
nextevent([1, 0, P3, P4, D1, D2], T, [1, 0, 0, P4 + 3 * D, D1 + D, D2], T + D):- D = P3 / 3.
nextevent([0, 1, P3, 0, D1, D2], T, [1, 0, P3, 0, 0, D2], T + D):- !, D = 75 - D2.
nextevent([0, 1, P3, P4, D1, D2], T, [1, 0, P3 + 2 * D, P4 - 2 * D, 0, D2], T + D):- D = 75 - D2.
nextevent([0, 1, P3, P4, D1, D2], T, [0, 1, P3 + 2 * D, 0, D1, D2 + D], T + D):- D = P4 / 2.

exec(S, T):- min(Tnext, nextevent(S, T, S1, Tnext)), exec(S1, Tnext)).
```

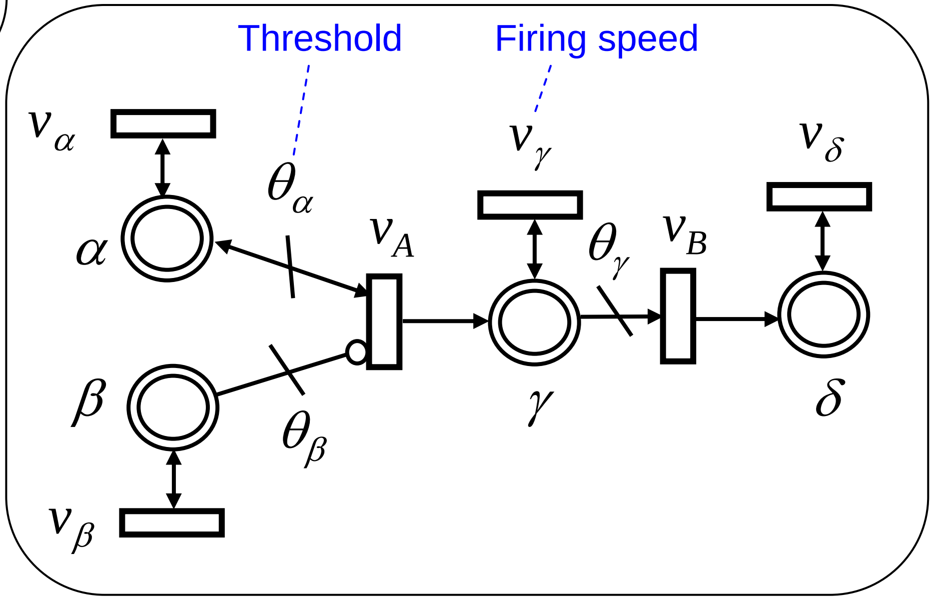
current state　　finding the earliest event　　next state

CLP (Constraint Logic Programming)

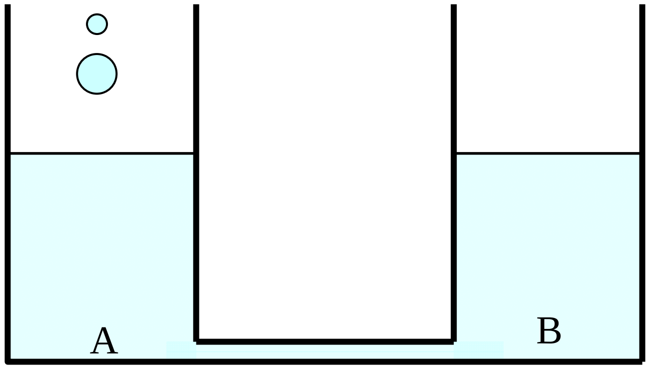
Gene interaction as a hybrid system



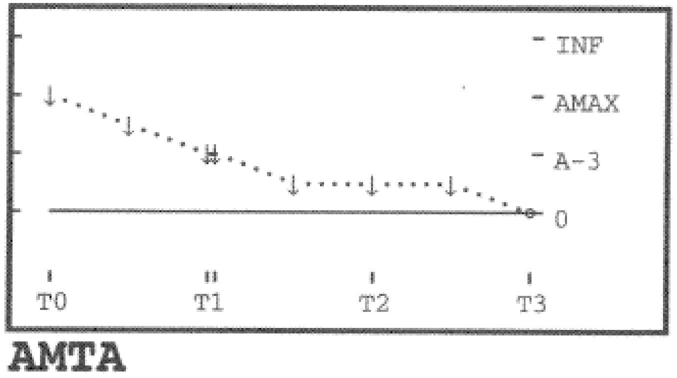
Hybrid Petri net



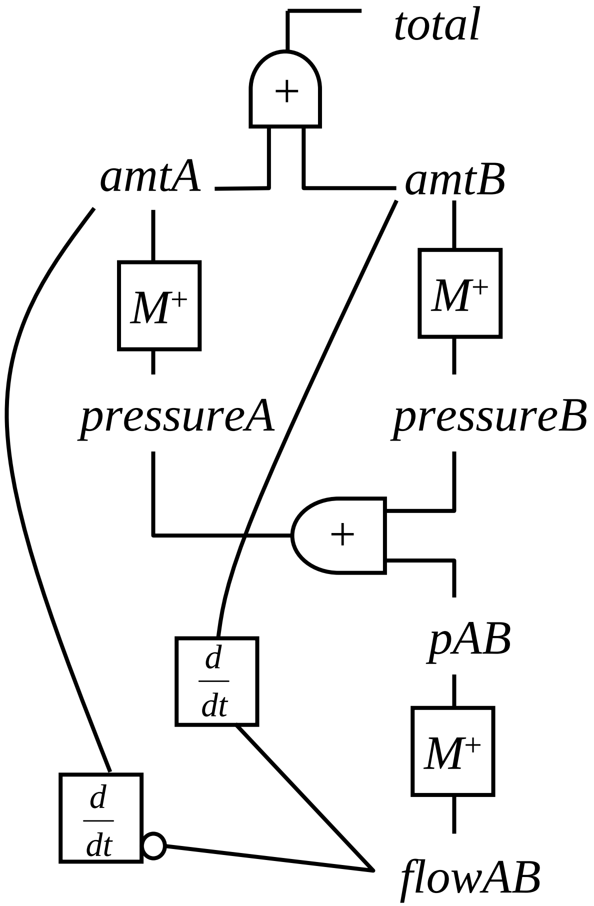
Models using incomplete knowledge



physical system

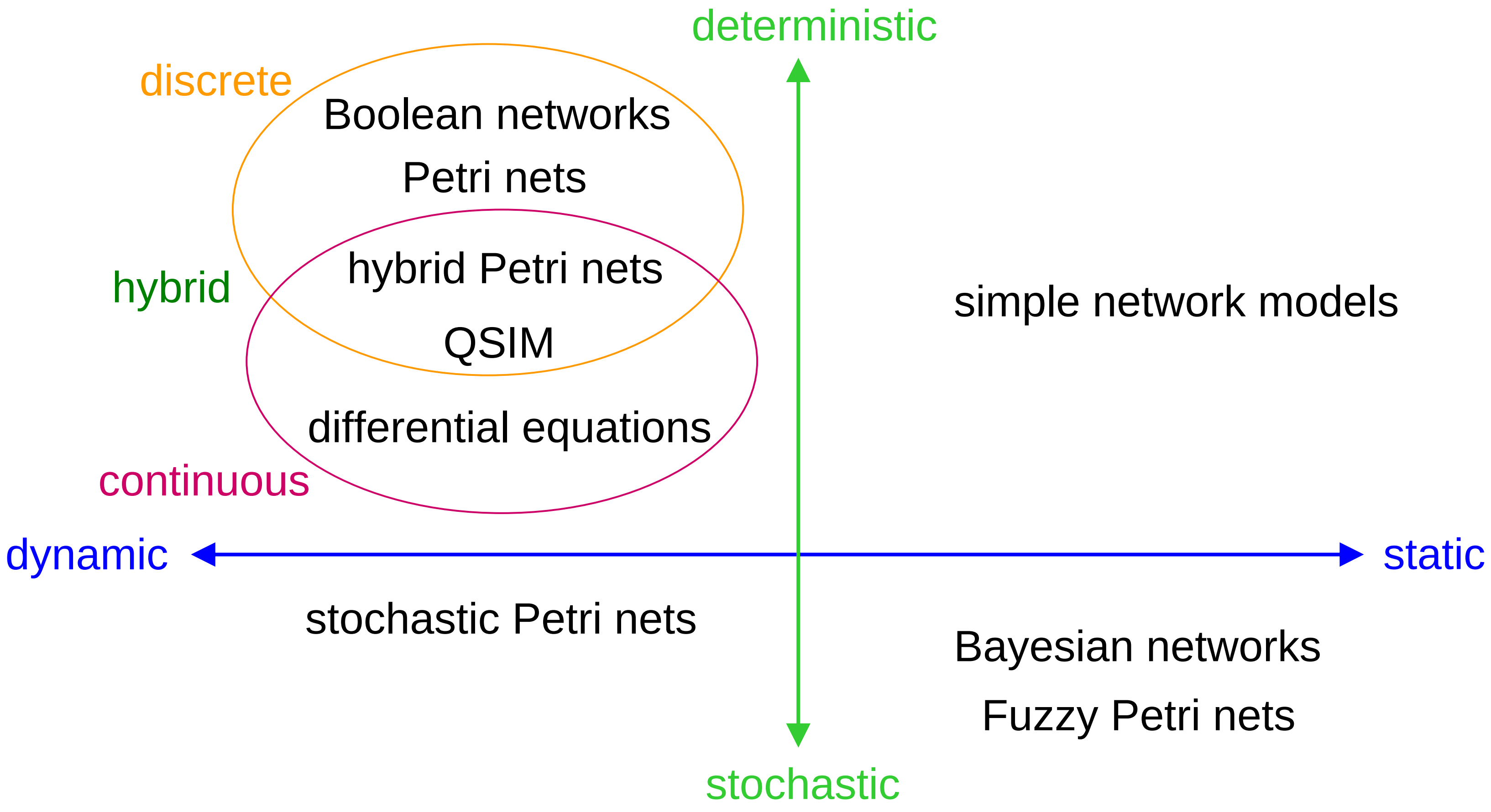


qualitative behavior



QSIM Model [Kuipers]

Summary



Models for representing gene interaction