Modeling of genetic networks

from microarray data

Junior Barrera

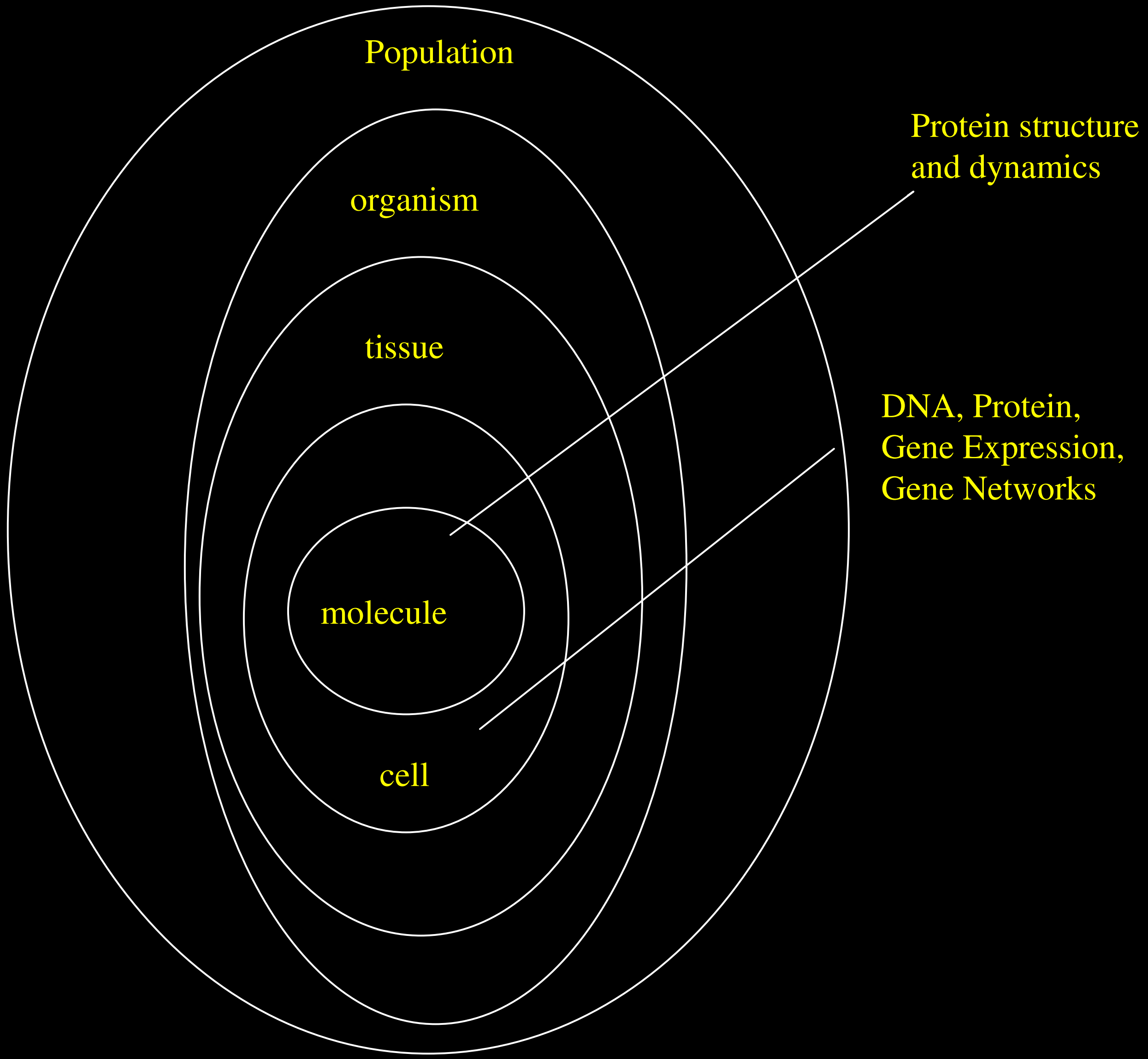
BIOINFO-USP

Layout

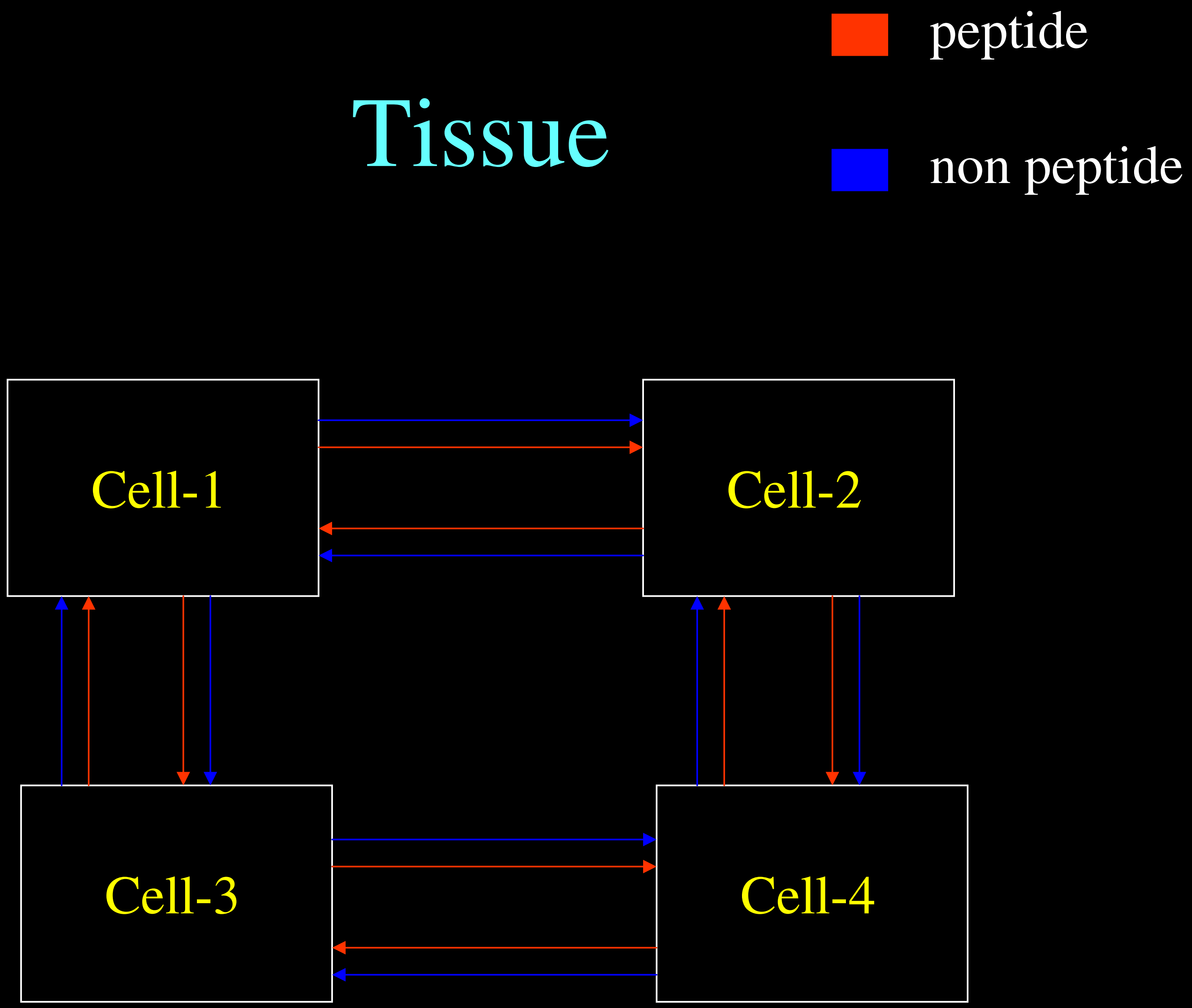
- Introduction
- Lattices and Operators
- Lattice operator design
- Lattice Dynamical System Identification
- Modeling of genetic networks
- Examples
- Conclusion

Introduction

Biological Systems



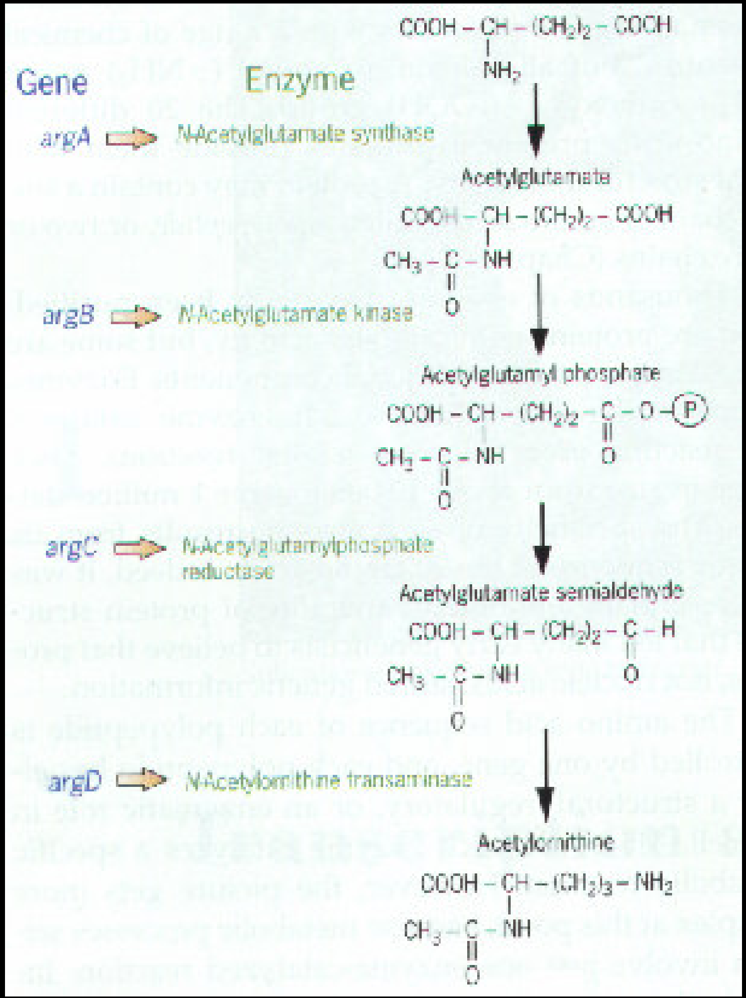
Tissue



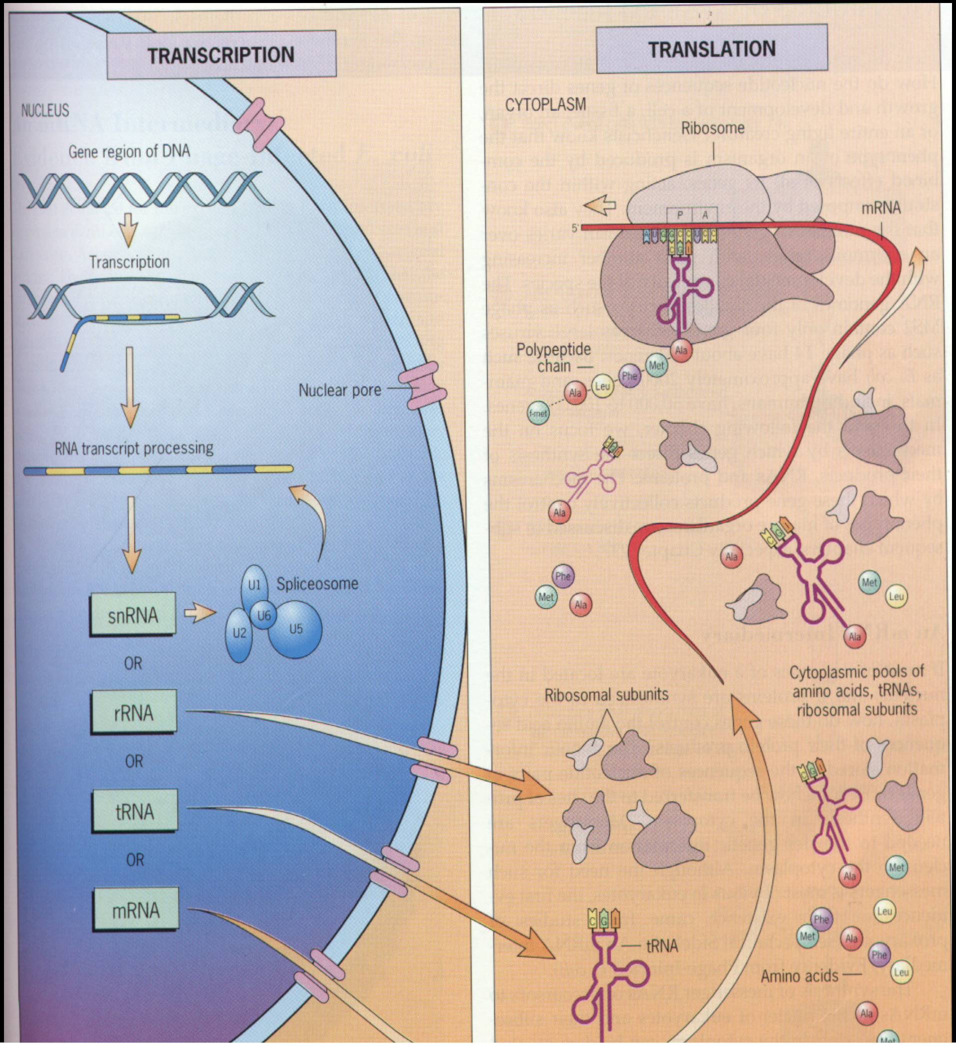
Cell

- Pathways
- Gene networks
- Protein Signals
- Protein Interactions

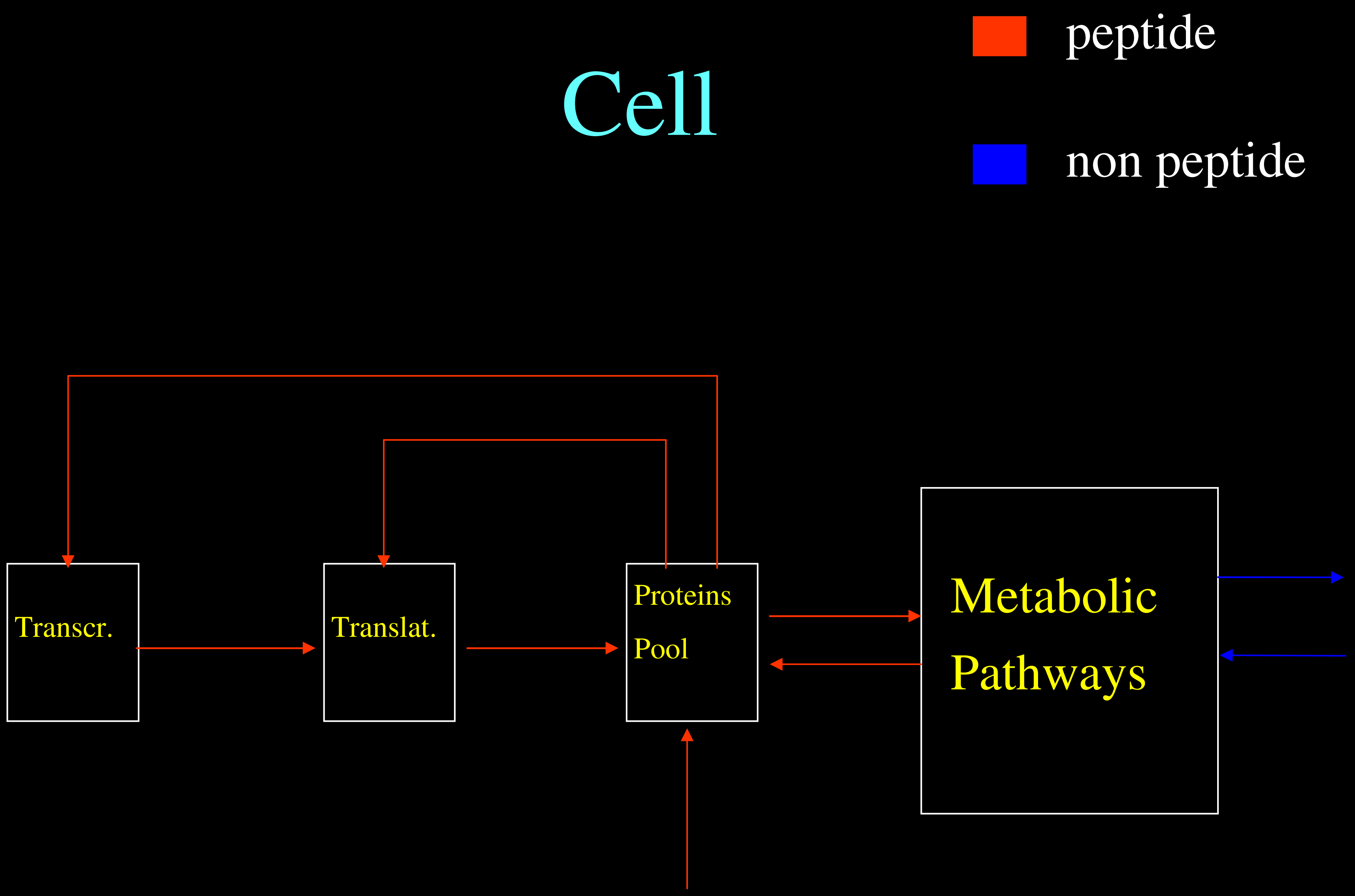
Pathway


COOH – CH – (CH2)2 – COOH
NH2
Gene
Enzyme
argA
N-Acetylglutamate synthase
Acetylglutamate
COOH – CH – (CH2)2 – COOH
CH3 – C – NH
O
argB
N-Acetylglutamate kinase
Acetylglutamyl phosphate
COOH – CH – (CH2)2 – C – O – P
CH3 – C – NH
O
O
argC
N-Acetylglutamylphosphate reductase
Acetylglutamate semialdehyde
COOH – CH – (CH2)2 – C – H
CH3 – C – NH
O
O
argD
N-Acetylornithine transaminase
Acetylornithine
COOH – CH – (CH2)3 – NH2
CH3 – C – NH
O

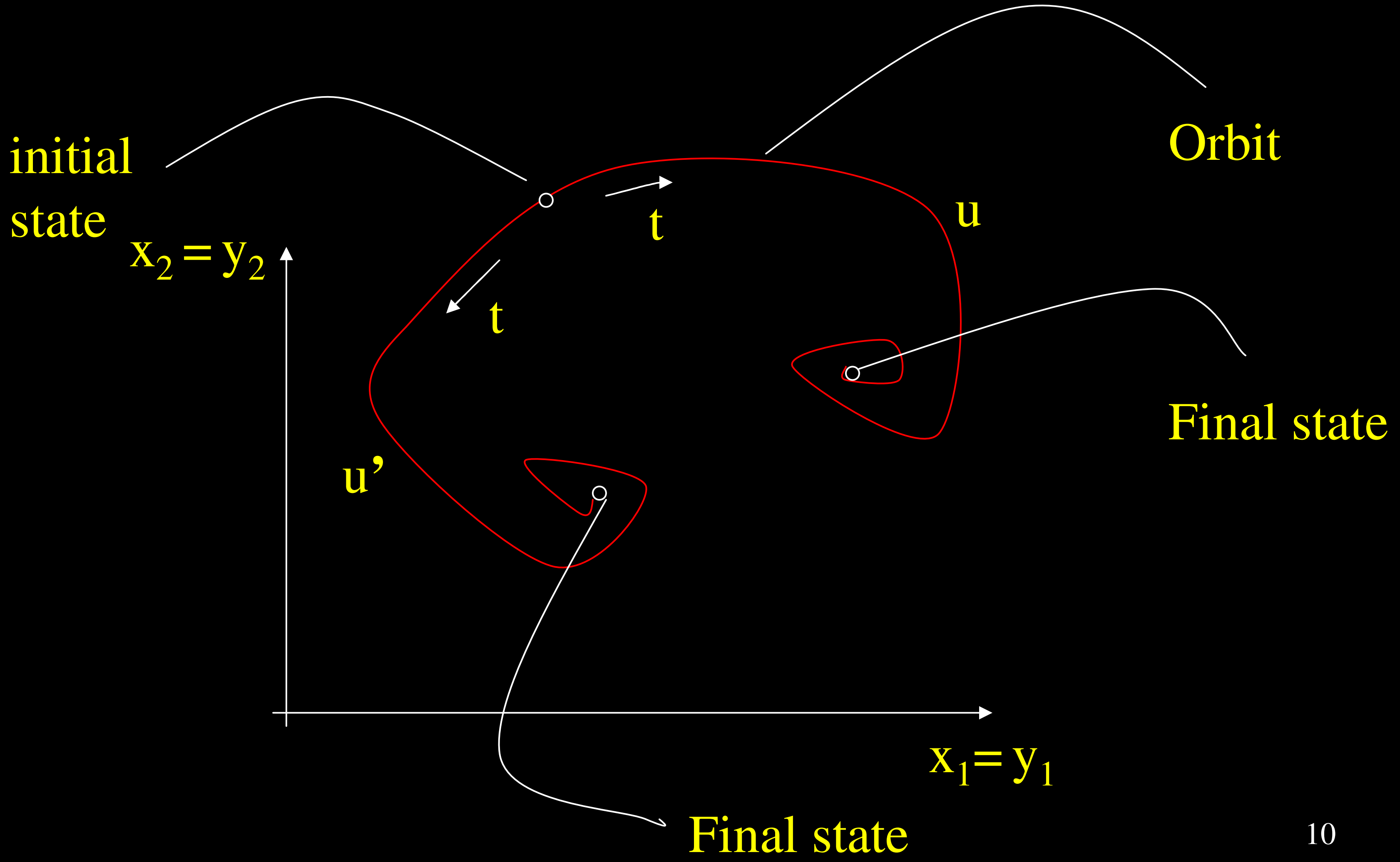
Transcription and translation



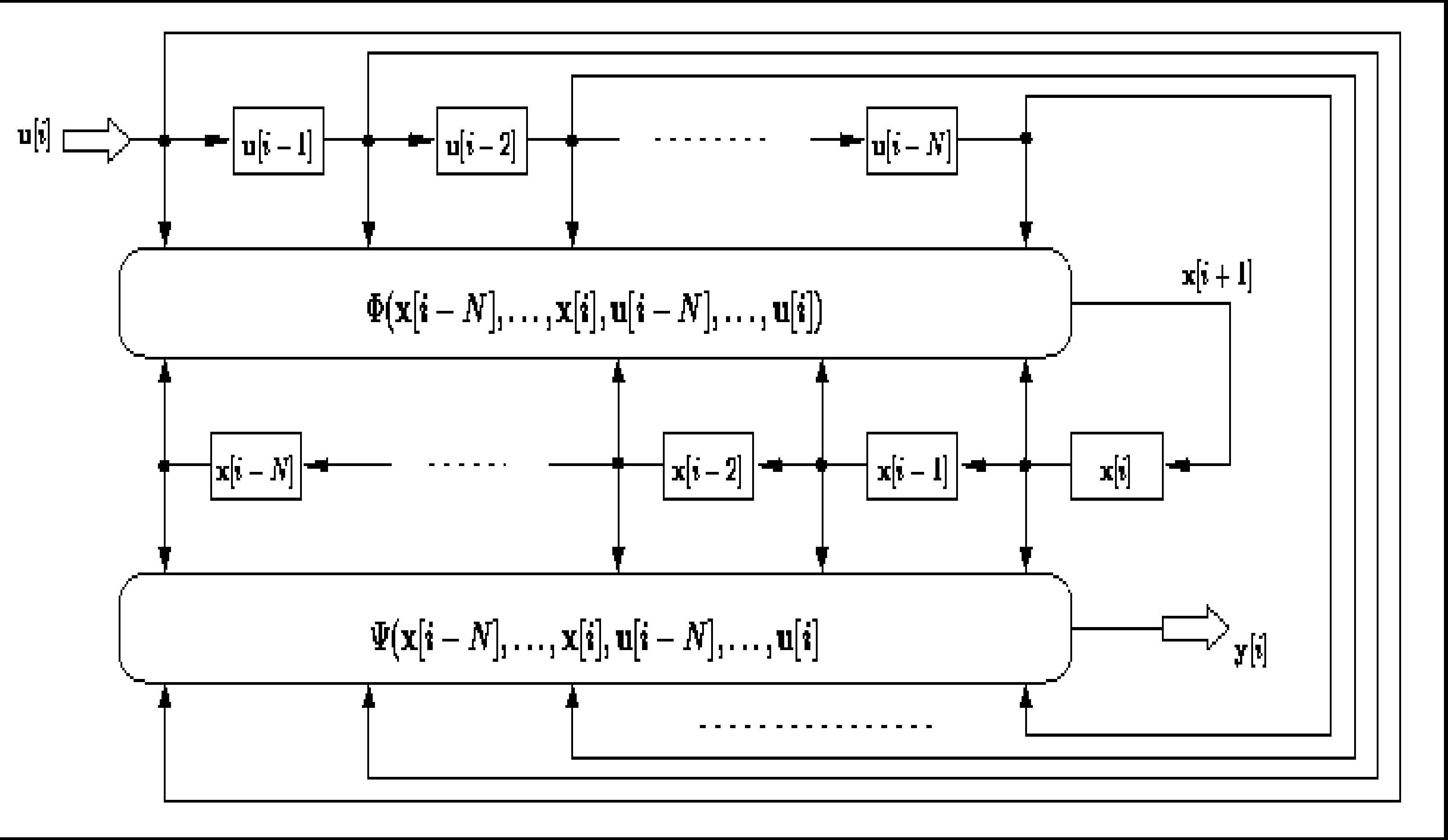
Cell

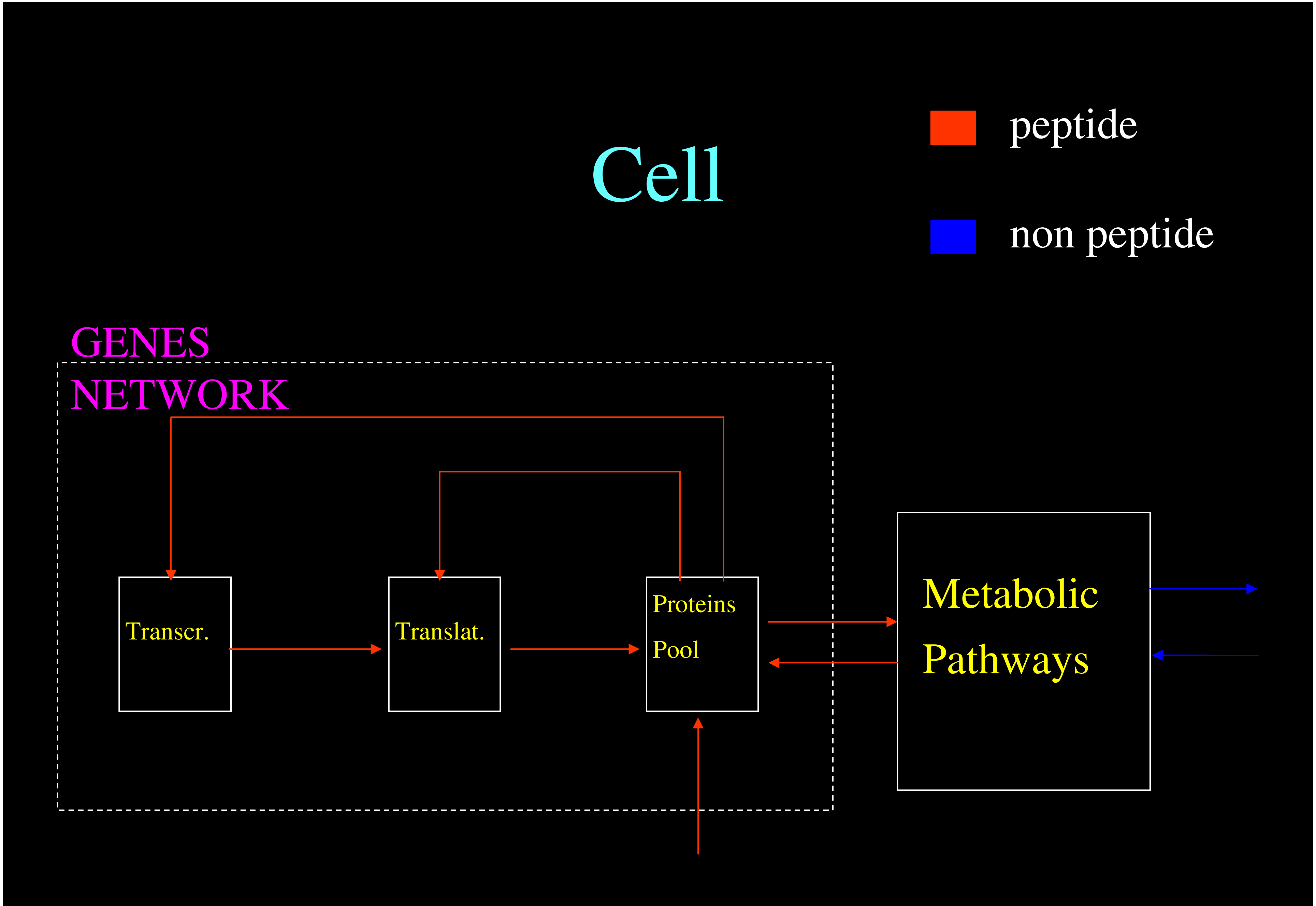

peptide
non peptide
Transcr.
Translat.
Proteins
Pool
Metabolic
Pathways

Modeling Dynamical Systems

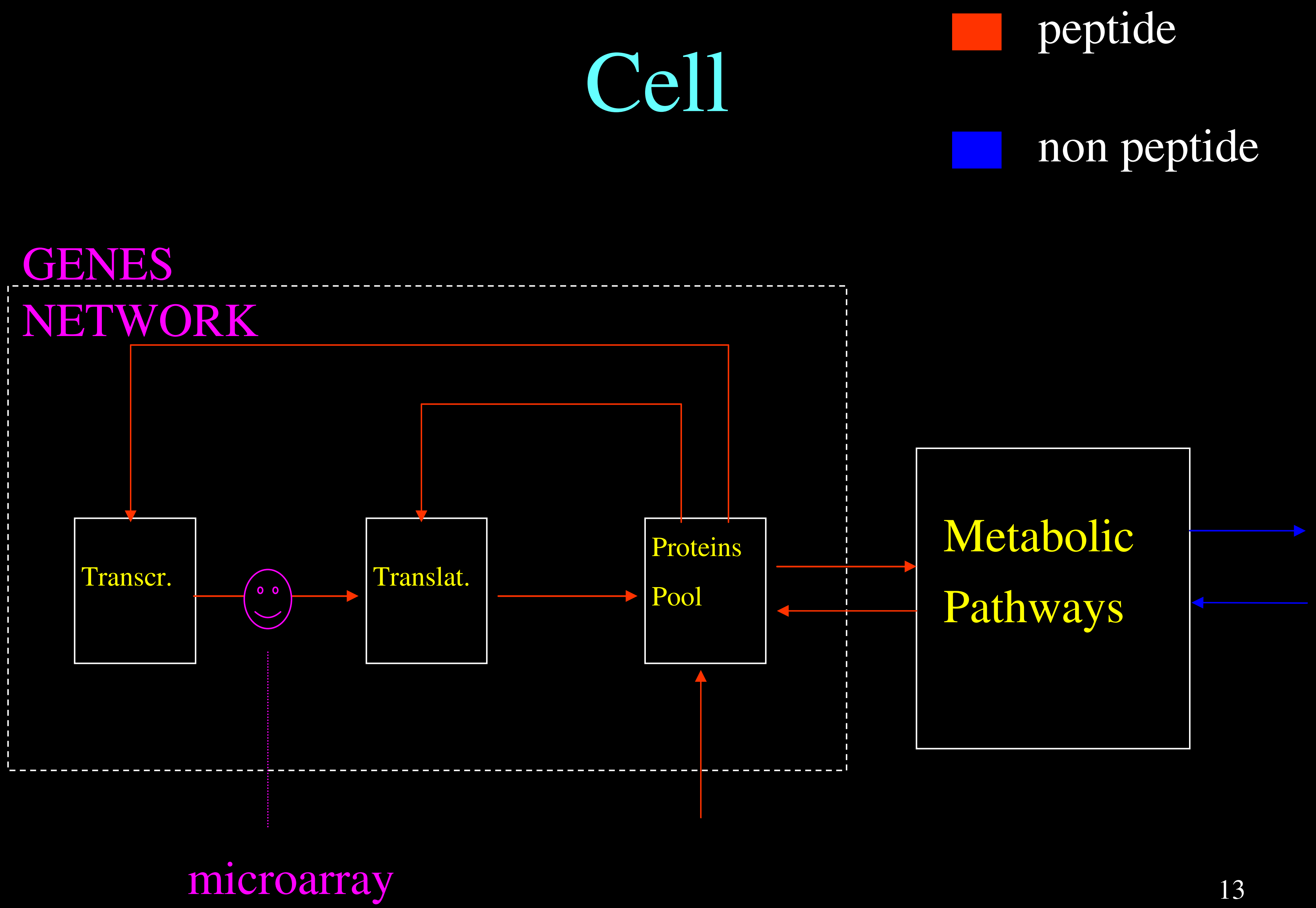


Modeling Dynamical Systems

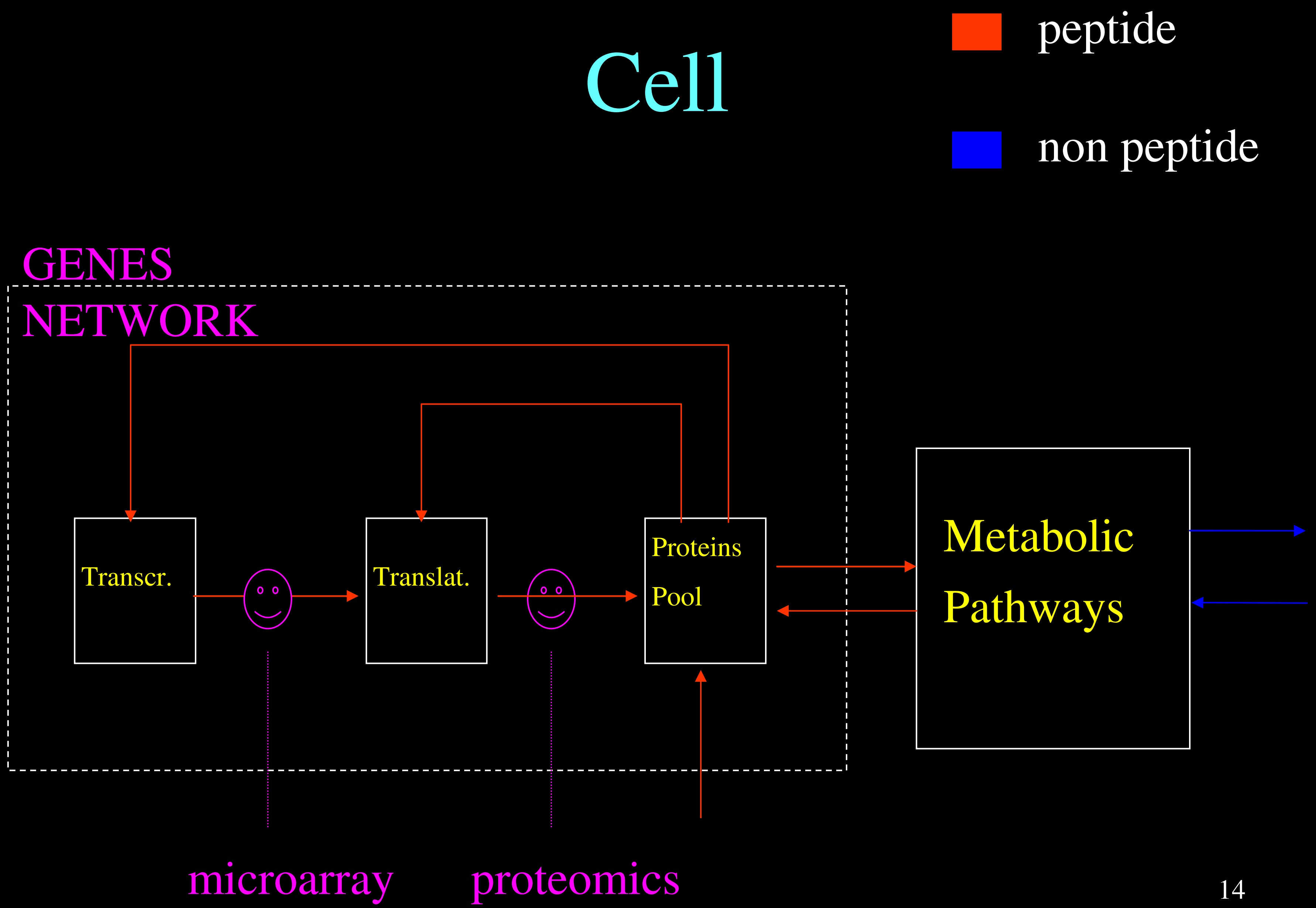



Cell
peptide
non peptide
GENES
NETWORK
Transcr.
Translat.
Proteins
Pool
Metabolic
Pathways

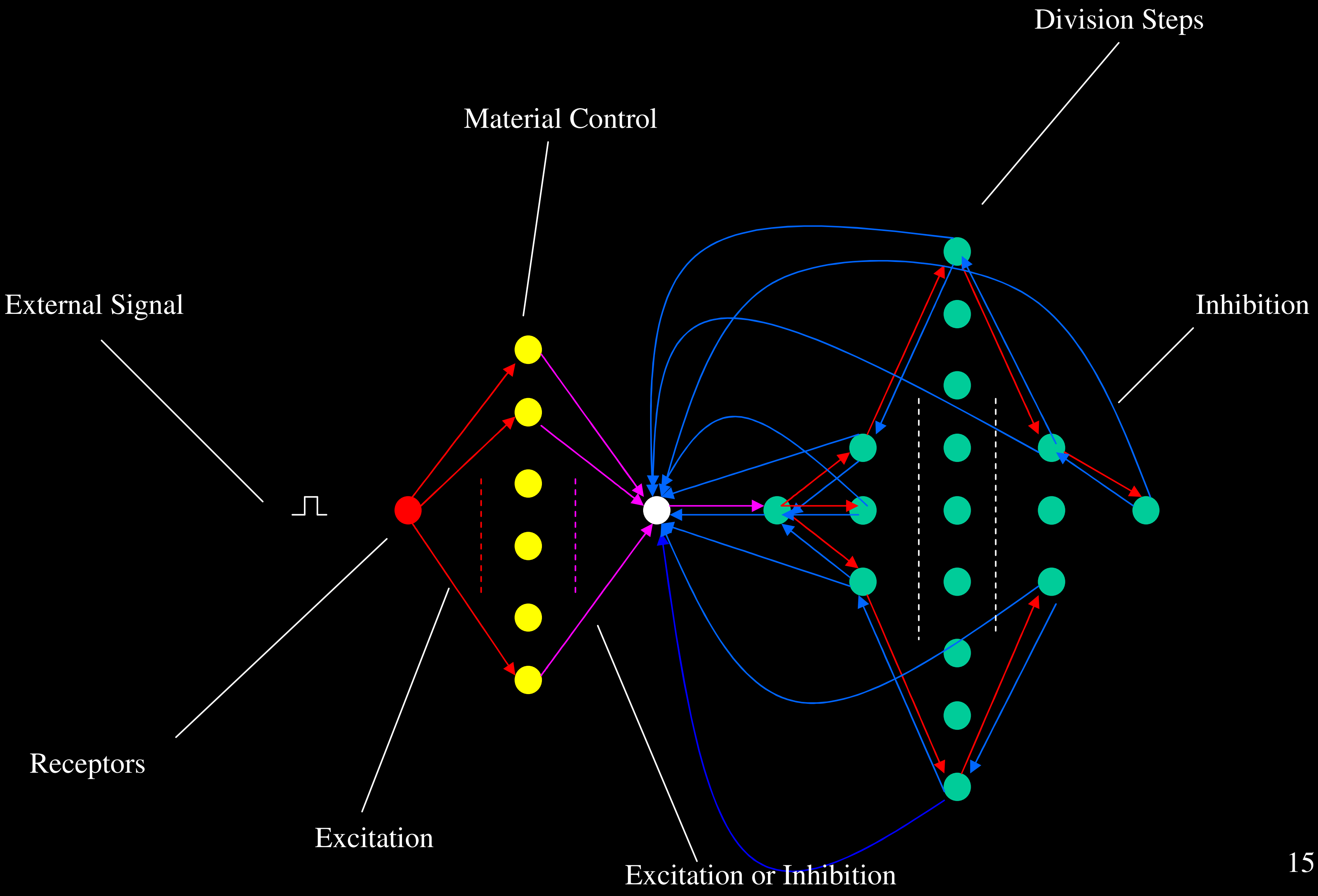
Cell

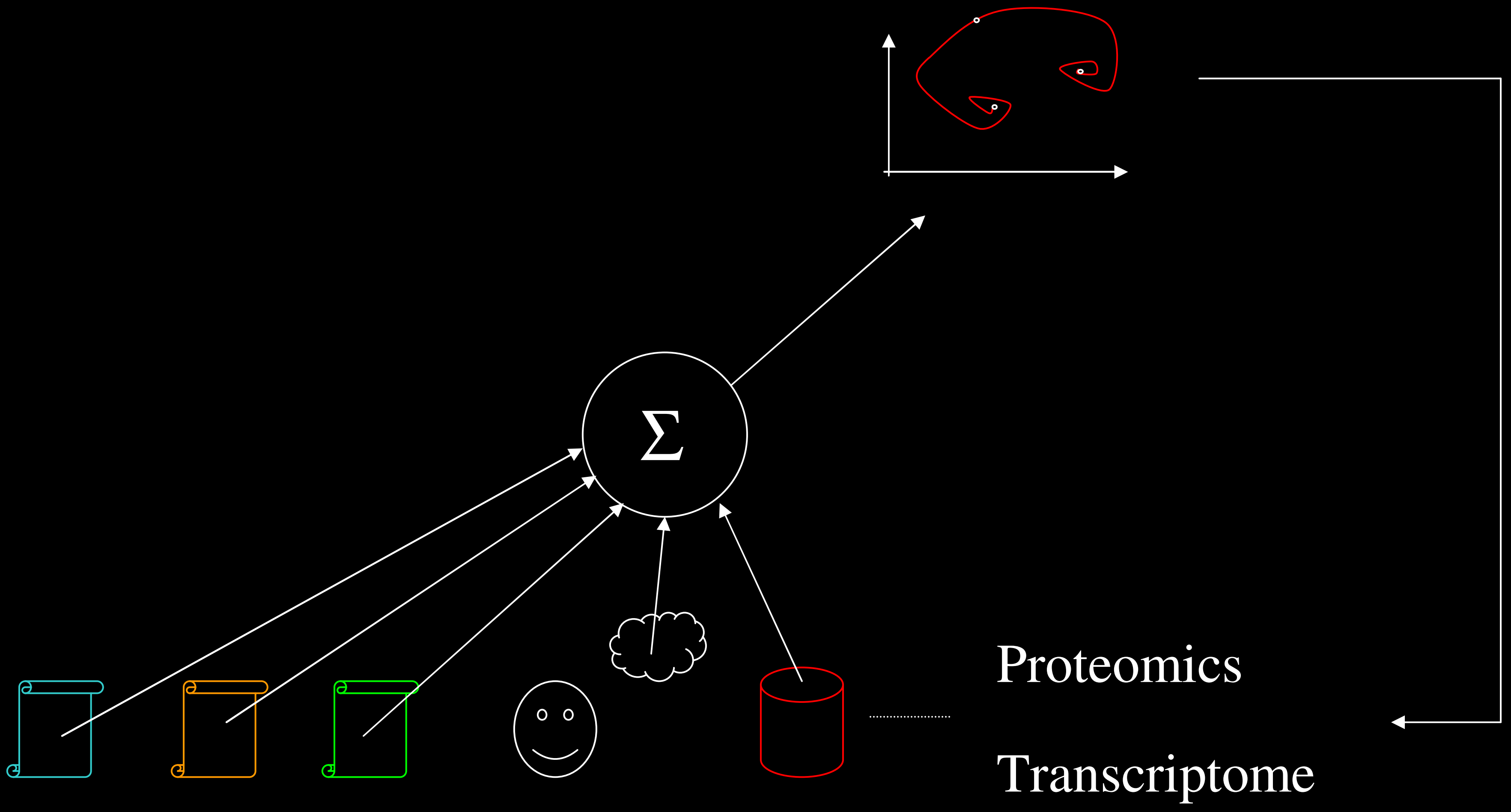

peptide
non peptide
GENES
NETWORK
Transcr.
Translat.
Proteins
Pool
Metabolic
Pathways
microarray

Cell


peptide
non peptide
GENES
NETWORK
Transcr.
Translat.
Proteins
Pool
Metabolic
Pathways
microarray
proteomics

Genetic Network

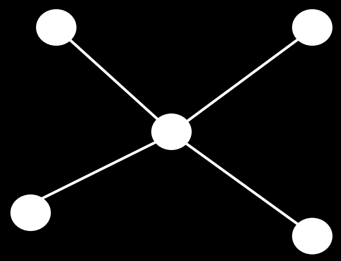



Σ
Proteomics
Transcriptome

Lattices and Operators

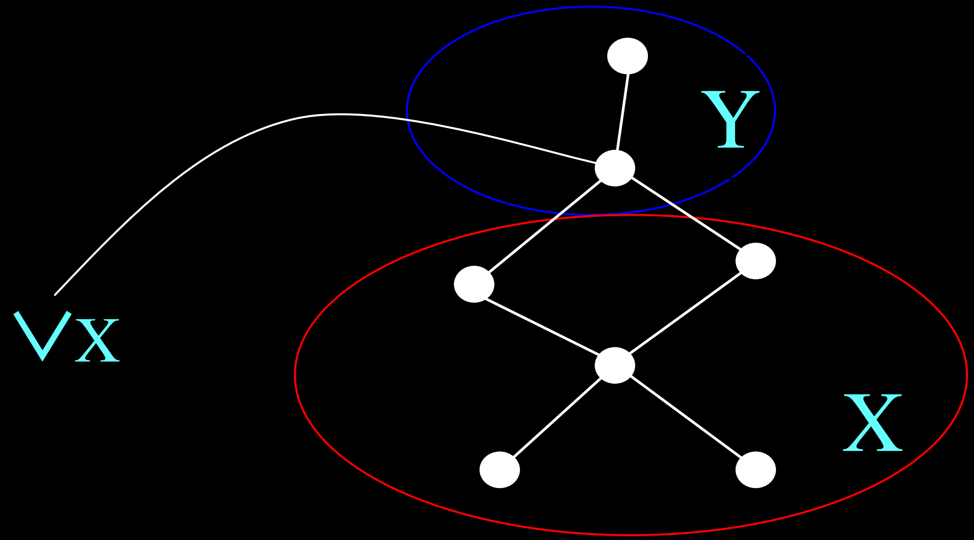
Poset

- A partially ordered set $(L, \leq)$ is a set L with a partial order relation (i.e., reflexive, transitive and anti-symmetric) $\leq$ on L

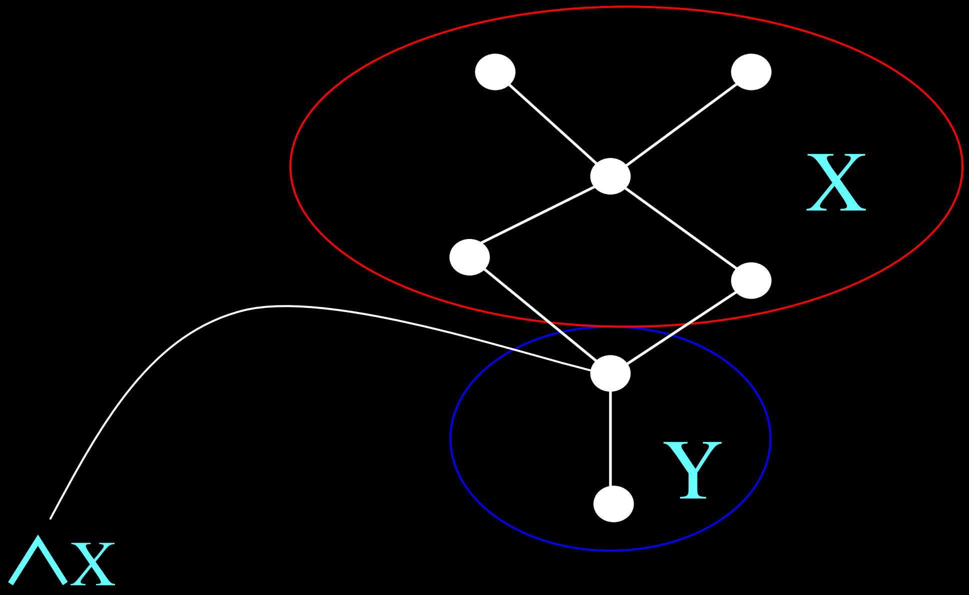


- The upper bound of $X \subseteq L$ is the subset Y of elements $y \in L$ such that $x \leq y$, $\forall x \in X$.

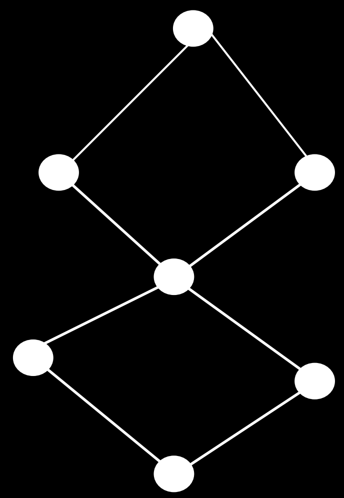
- The union $\bigvee X$ is the least upper bound of X.



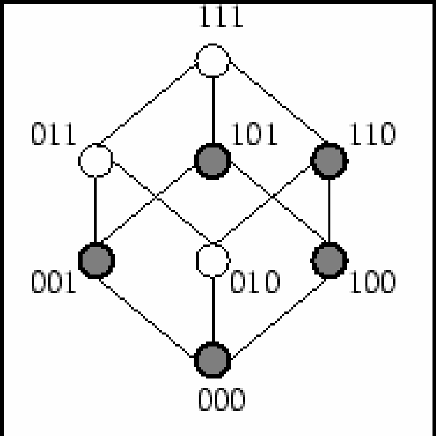
- The lower bound of $X \subseteq L$ is the subset Y of elements $y \in L$ such that $y \leq x, \forall x \in X$.
- The intersection $\wedge X$ is the greatest lower bound of X.



- A lattice L is a poset such that $\forall X \subseteq L$ there exist $\bigwedge X$ and $\bigvee X$.



- Lattice functions $\psi : \mathrm{Fun}[W, L] \rightarrow L$



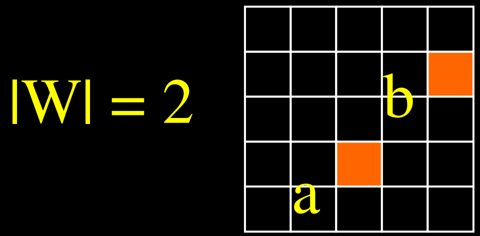
Boolean

	-2	-1	0	1	2
2	0	1	2	2	2
1	0	1	2	2	2
0	-1	1	2	2	2
-1	-1	1	1	1	1
-2	-2	-1	-1	-1	-1

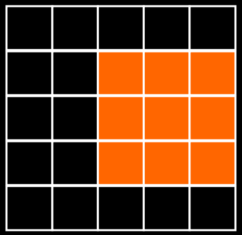
Non Boolean

Intervals

- Let $a, b \in$ Fun$[W,L]$, $a \leq b$ iff $a(x) \leq b(x)$, $x \in W$

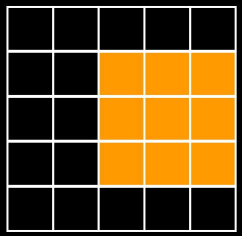


- Interval $[a,b] = \{u \in$ Fun$[W,L]$: $a \leq u \leq b\}$



Sup-generating operator

$$\lambda_{a,b}(u) = 1 \Leftrightarrow u \in [a,b]$$

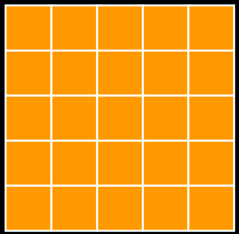


[a,b]

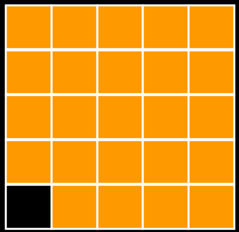
0	0	0	0	0
0	0	1	1	1
0	0	1	1	1
0	0	1	1	1
0	0	0	0	0

Kernel of ψ at y: $K(\psi)(y) = \{u \in \mathrm{Fun}[W,L]: y \leq \psi(u)\}$

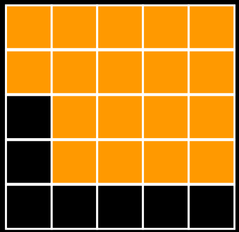
2	0	1	2	2	2
1	0	1	2	2	2
0	-1	1	2	2	2
-1	-1	1	1	1	1
-2	-2	-1	-1	-1	-1
	-2	-1	0	1	2



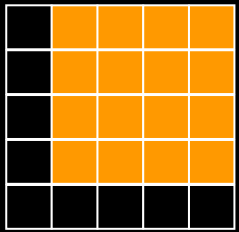
$K(\psi)(-2)$



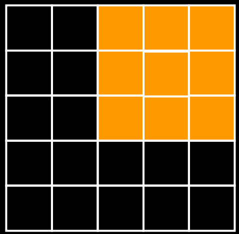
$K(\psi)(-1)$



$K(\psi)(0)$

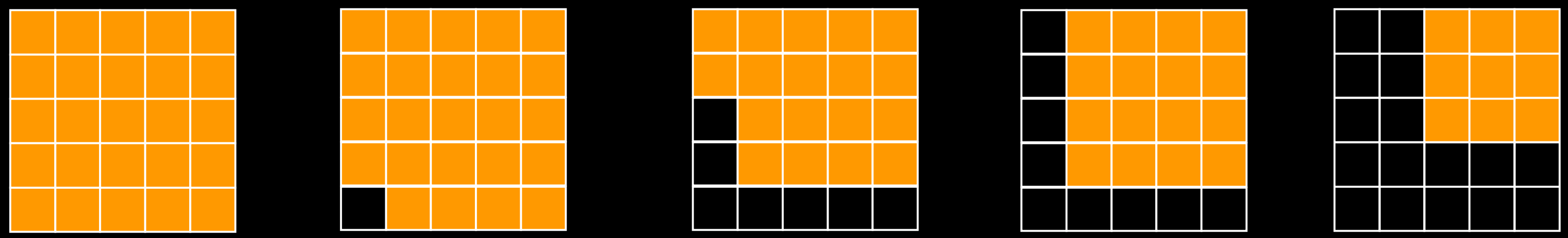


$K(\psi)(1)$

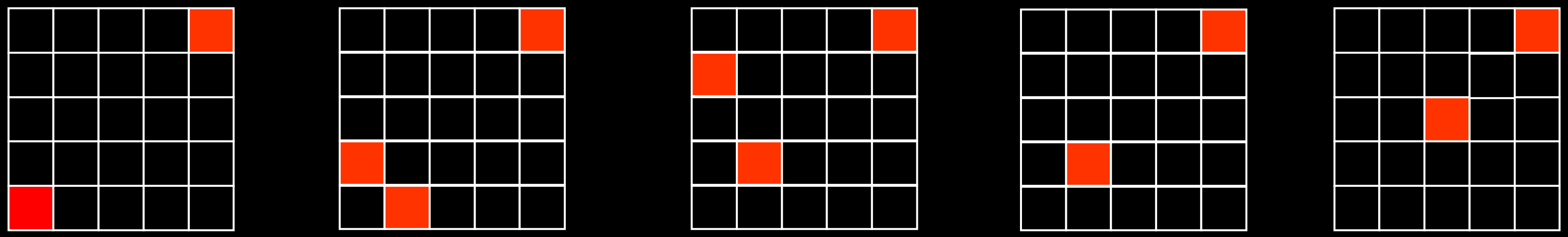


$K(\psi)(2)$

Basis of ψ at *y*: B(ψ) is the set of maximal intervals contained in K(ψ)



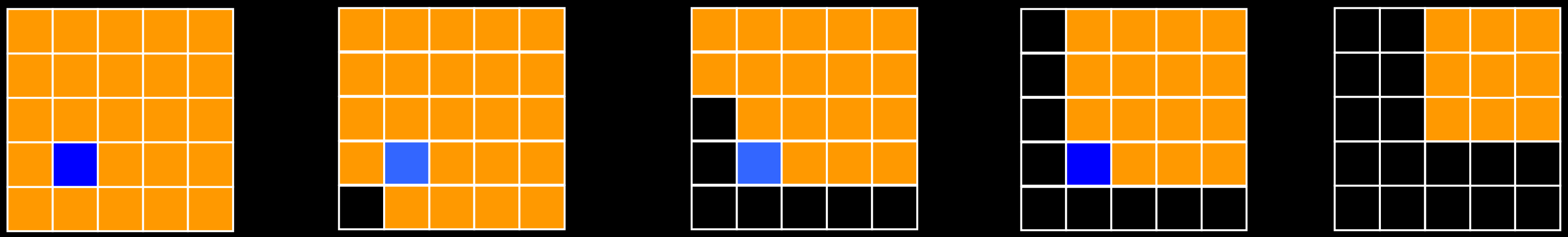
K(ψ)(-2) K(ψ)(-1) K(ψ)(0) K(ψ)(1) K(ψ)(2)



B(ψ)(-2) B(ψ)(-1) B(ψ)(0) B(ψ)(1) B(ψ)(2)

Operator Representation

$$\psi(u) = \bigcup\left\{ y \in L : \bigcup\left\{ \lambda_{a,b}(u) : [a,b] \in B(\psi)(y) \right\} = 1 \right\}$$



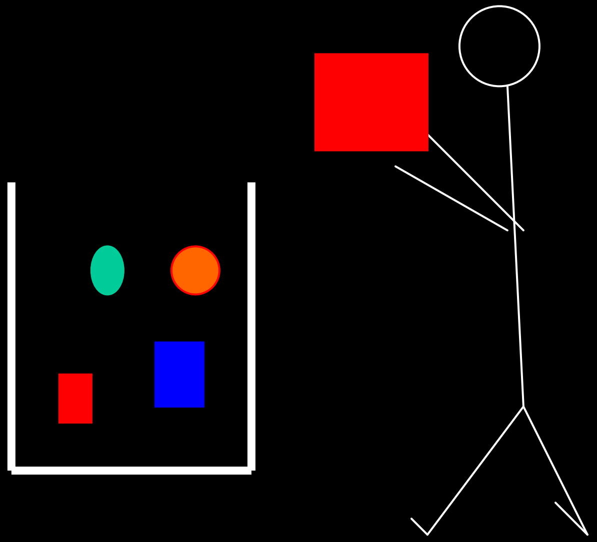
K(ψ)(-2) K(ψ)(-1) K(ψ)(0) K(ψ)(1) K(ψ)(2)

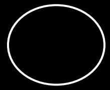
ψ(-1,-1) = 1

Lattice Operator Design

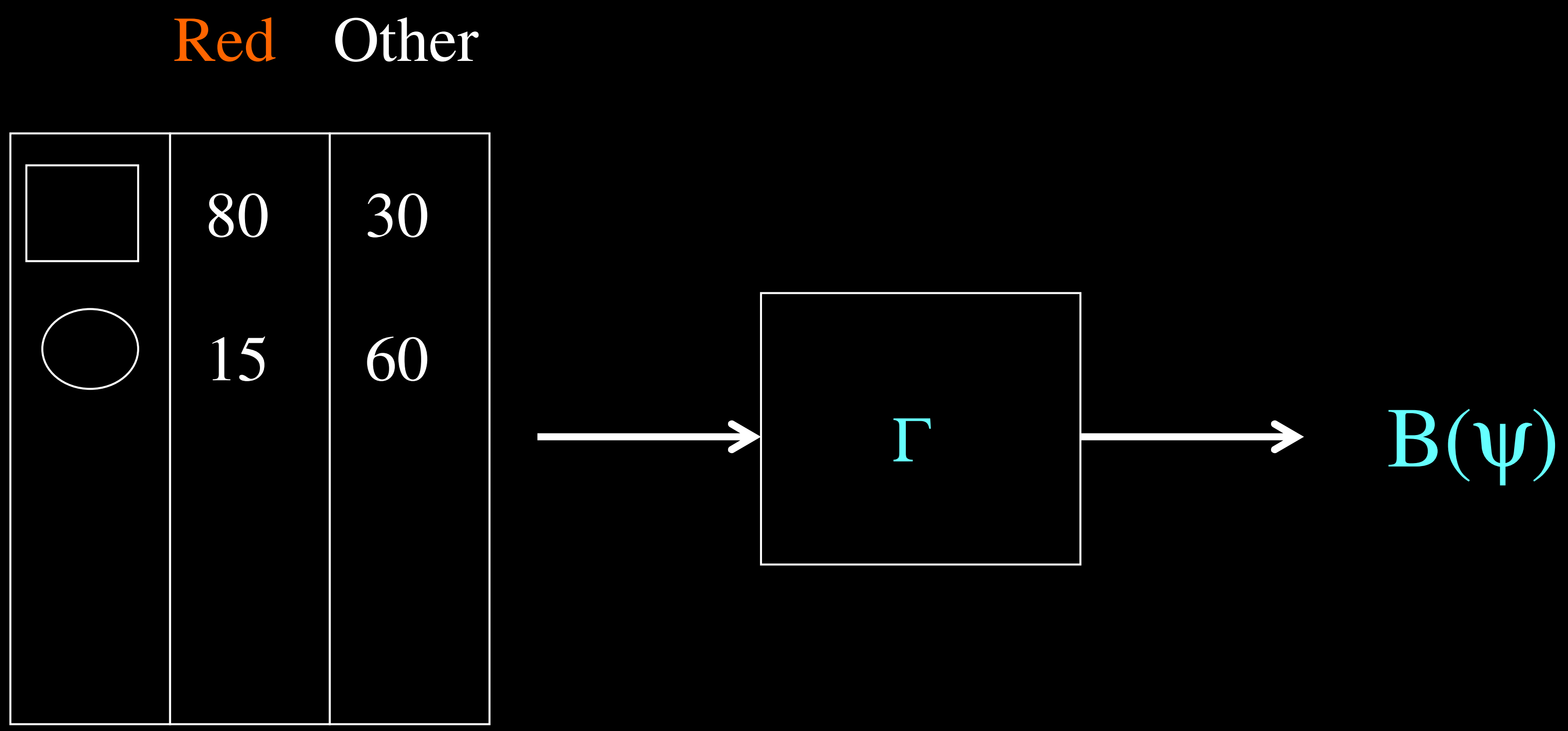
Training

- Domain: planar shapes
- concept: color RED



	Red	Other
	80	30
	15	60

Operator Estimation



Optimization problem

Design goal is to find a function $\psi_{opt} : \text{Fun}[W, L] \rightarrow L$ with minimum error.

Error (expected loss) of a function :

$$Er(\psi) = \mathrm{E}[l(\psi(X), Y)]$$

X is a random function
Y is a random variable

Loss function

$$l : L \times L \rightarrow \Re^{+}$$

Estimation problem

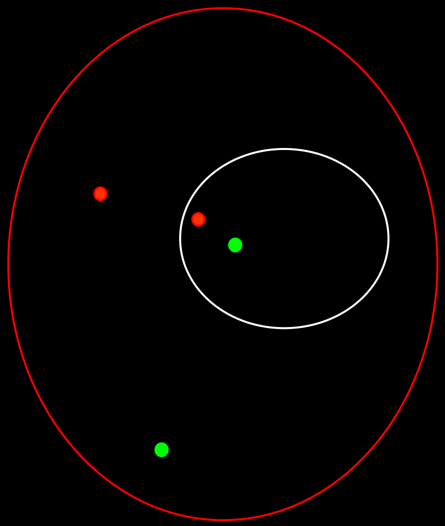
The distribution $P(X,Y)$ is unknown

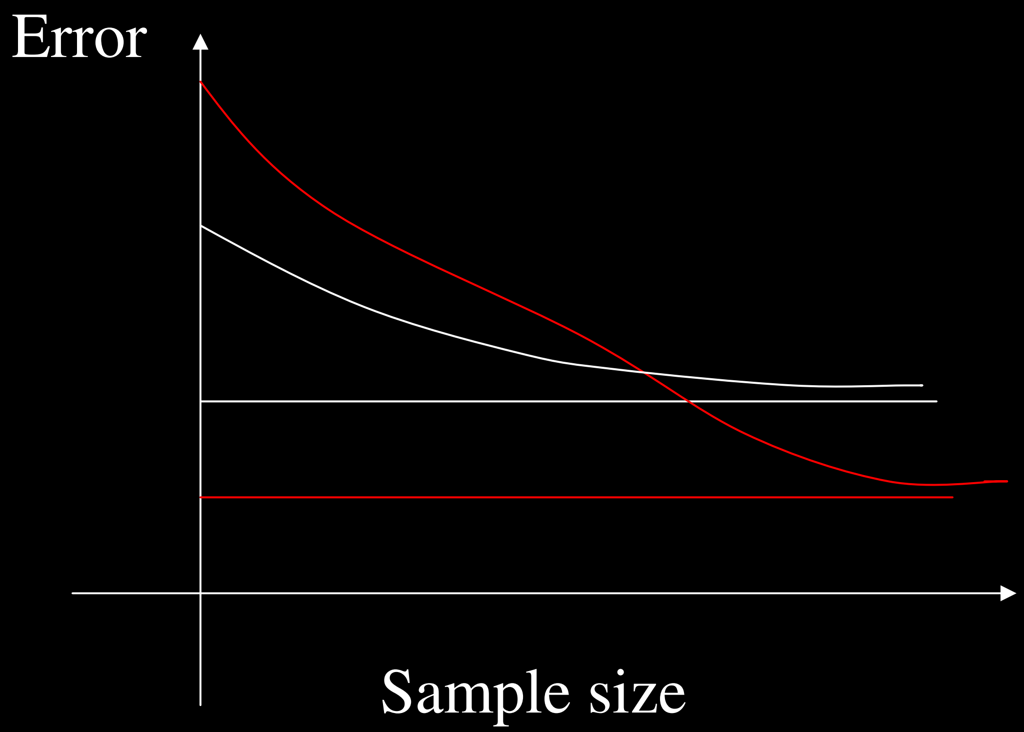
$P(X,Y)$, $Er(\psi)$ and ψ_{opt} should be estimated from realizations of X and Y.

For $m > m(\varepsilon, \delta)$

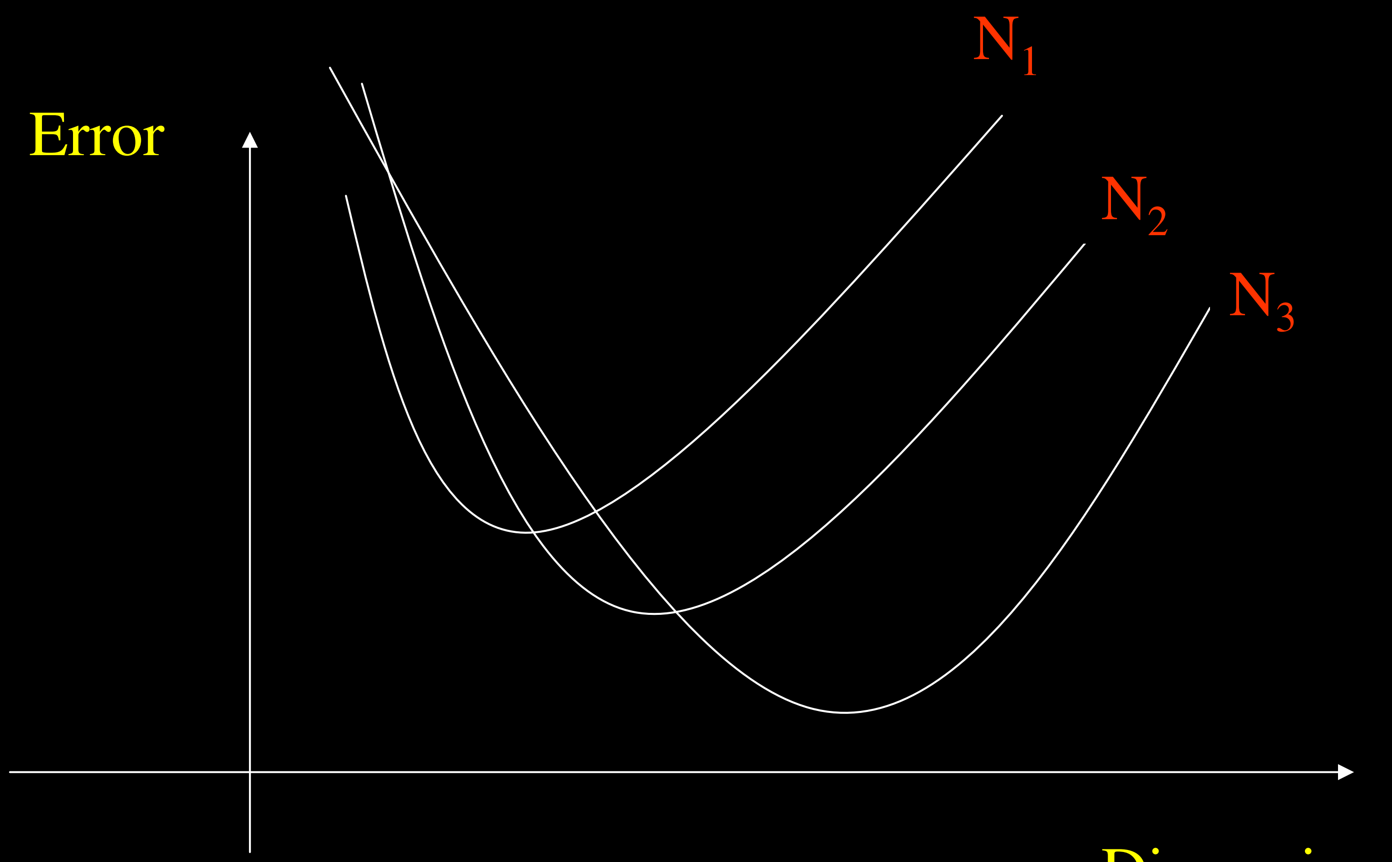
$$Pr(|Er(\psi) - Er(\psi_{opt})| < \varepsilon) > 1 - \delta$$

The constrained estimation problem





Variables for a sample of size N



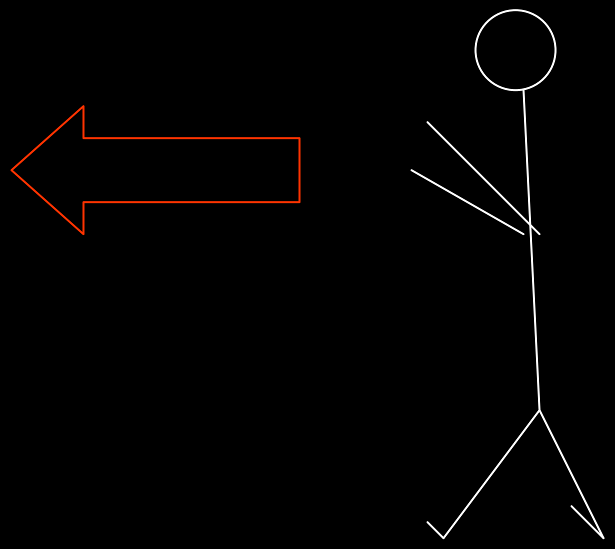
$N_1 < N_2 < N_3$

Formalization of prior information

$\psi(\psi(x)) = \psi(x), \forall x \in L$

$\psi(x) \leq x, \forall x \in L$

$x \leq y \Rightarrow \psi(x) \leq \psi(y), \forall x, y \in L$

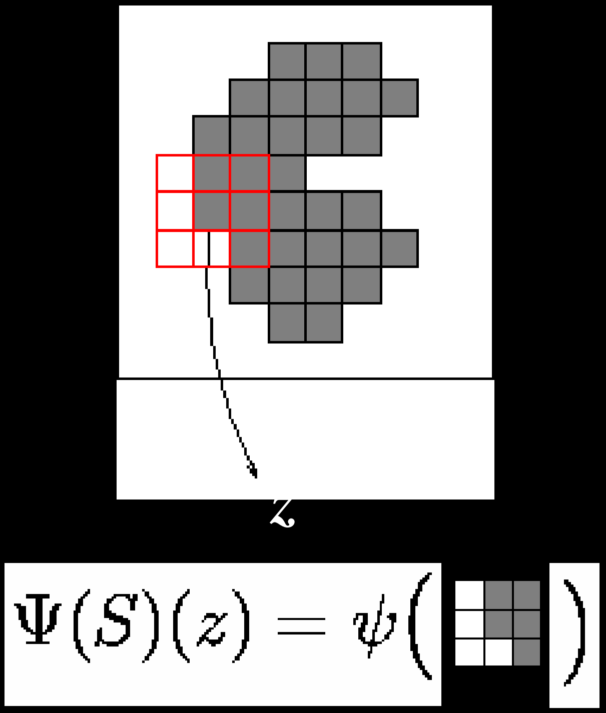


Stack filter

- Spatially translation invariant
- Spatially locally defined
- Increasing
- Commutes with threshold

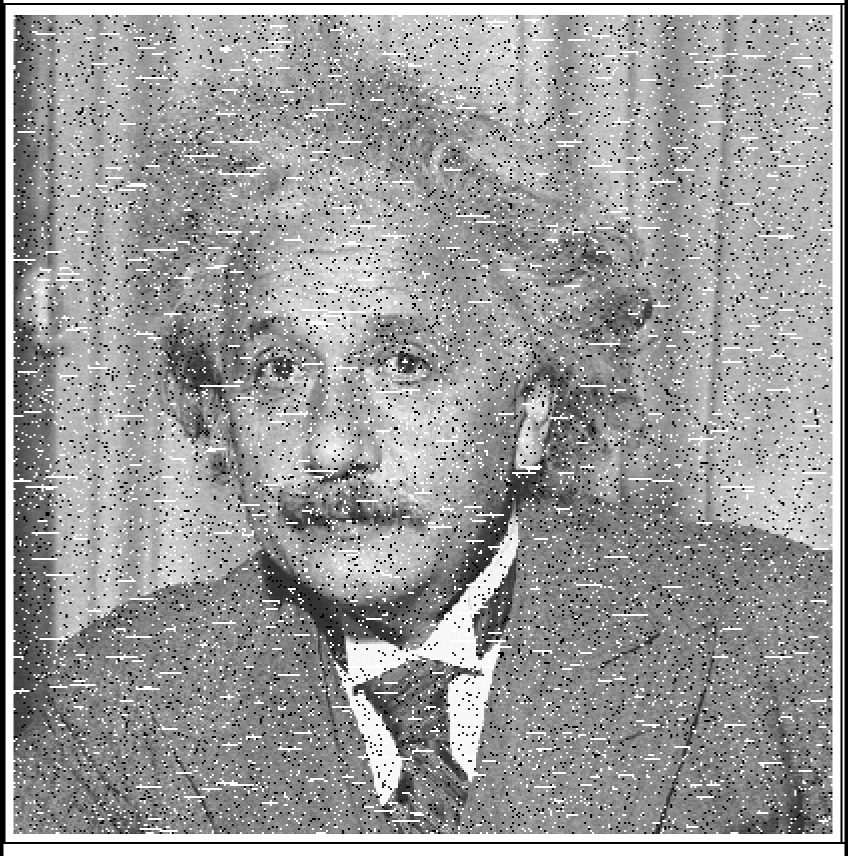
W-operators

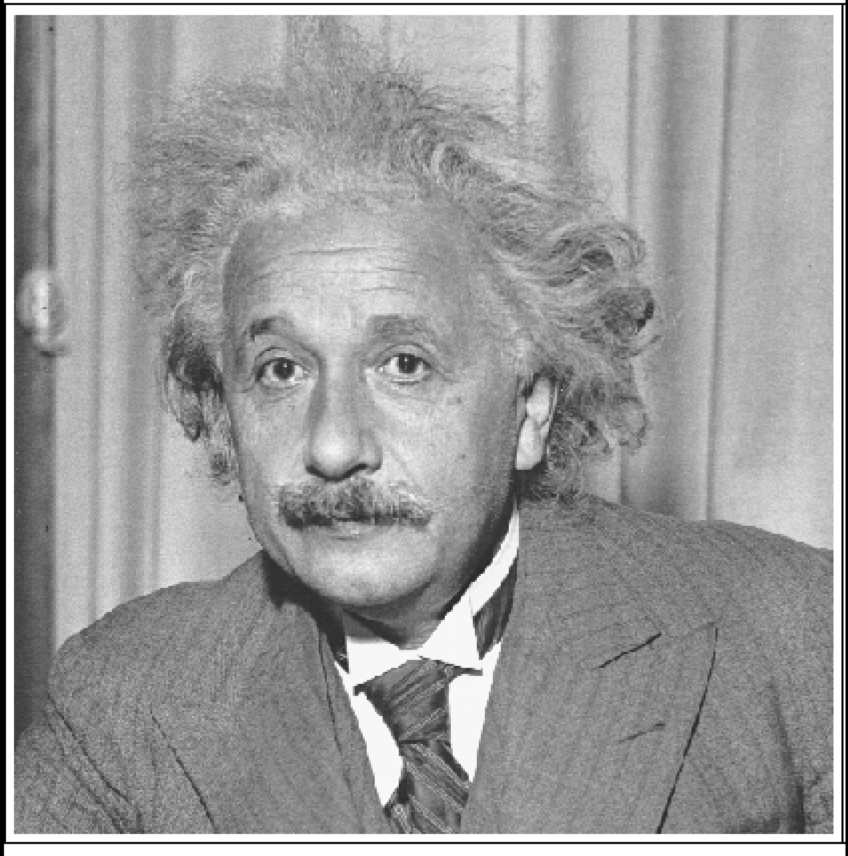
Translation invariance
+
local definition within W
=
W-operators



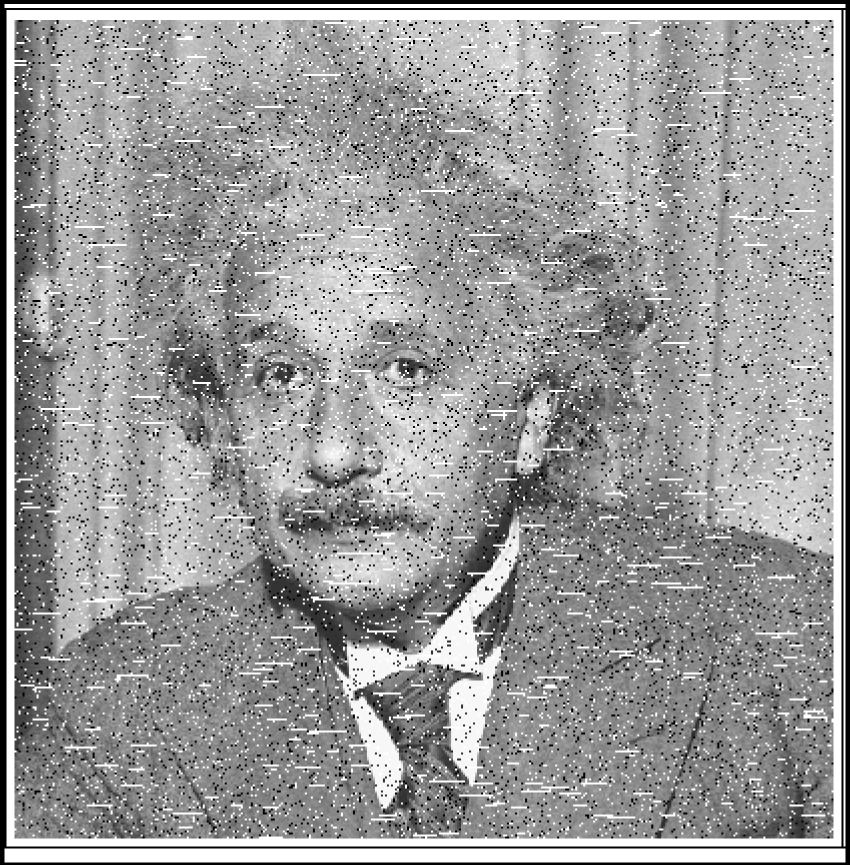
W-operators are characterized by Boolean functions

Noise elimination

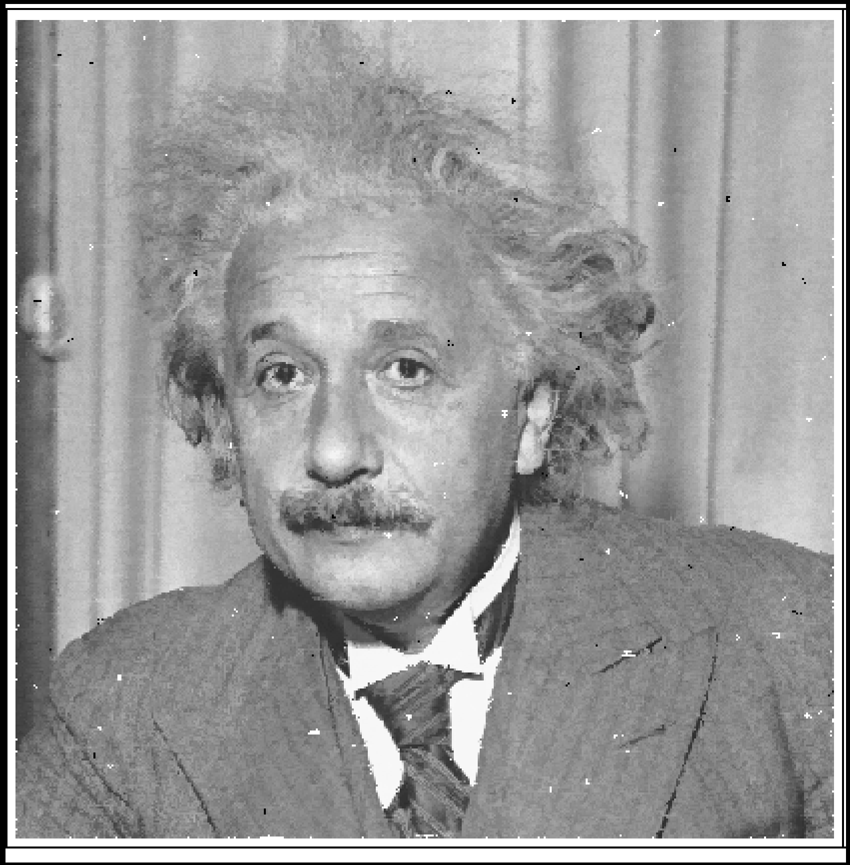




training images



test image



restaured

test image

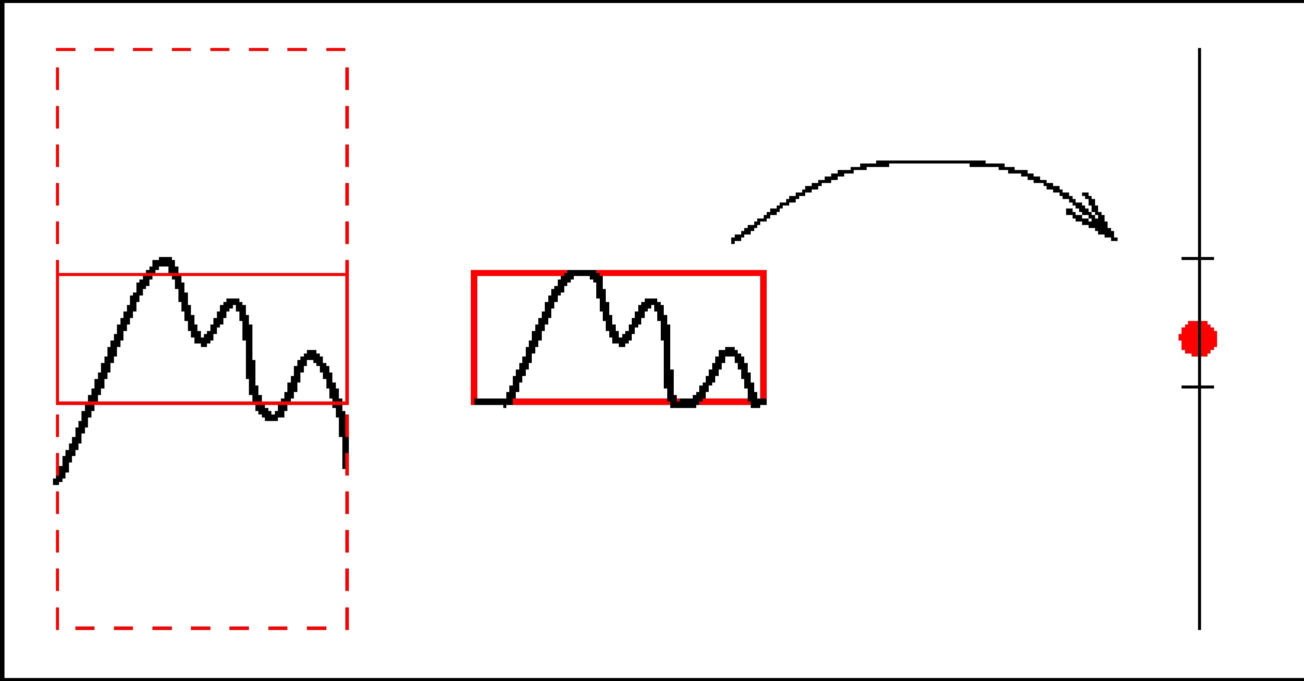
restaured

Apertures

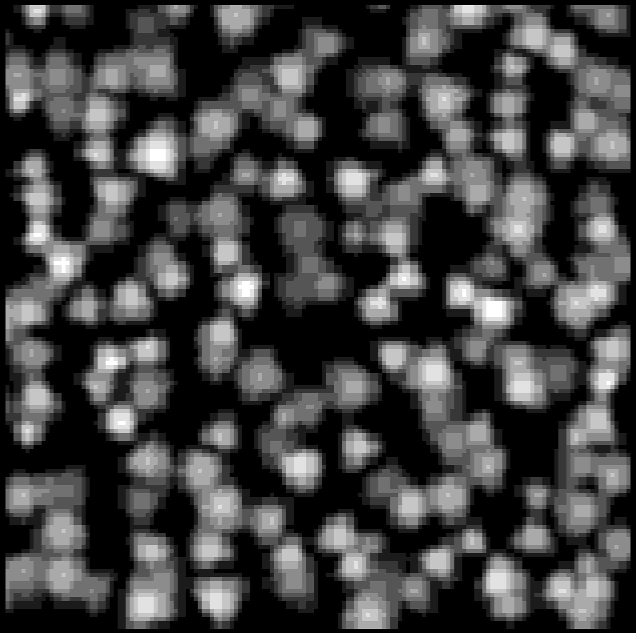
- Spatially translation invariant
- Spatially locally defined
- Range translation invariant
- Range locally defined

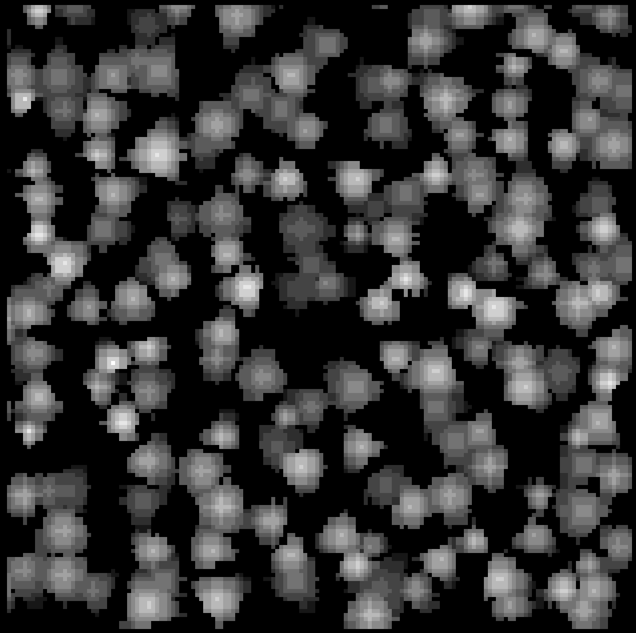
Design of Aperture Filters

Windowing in the space and range

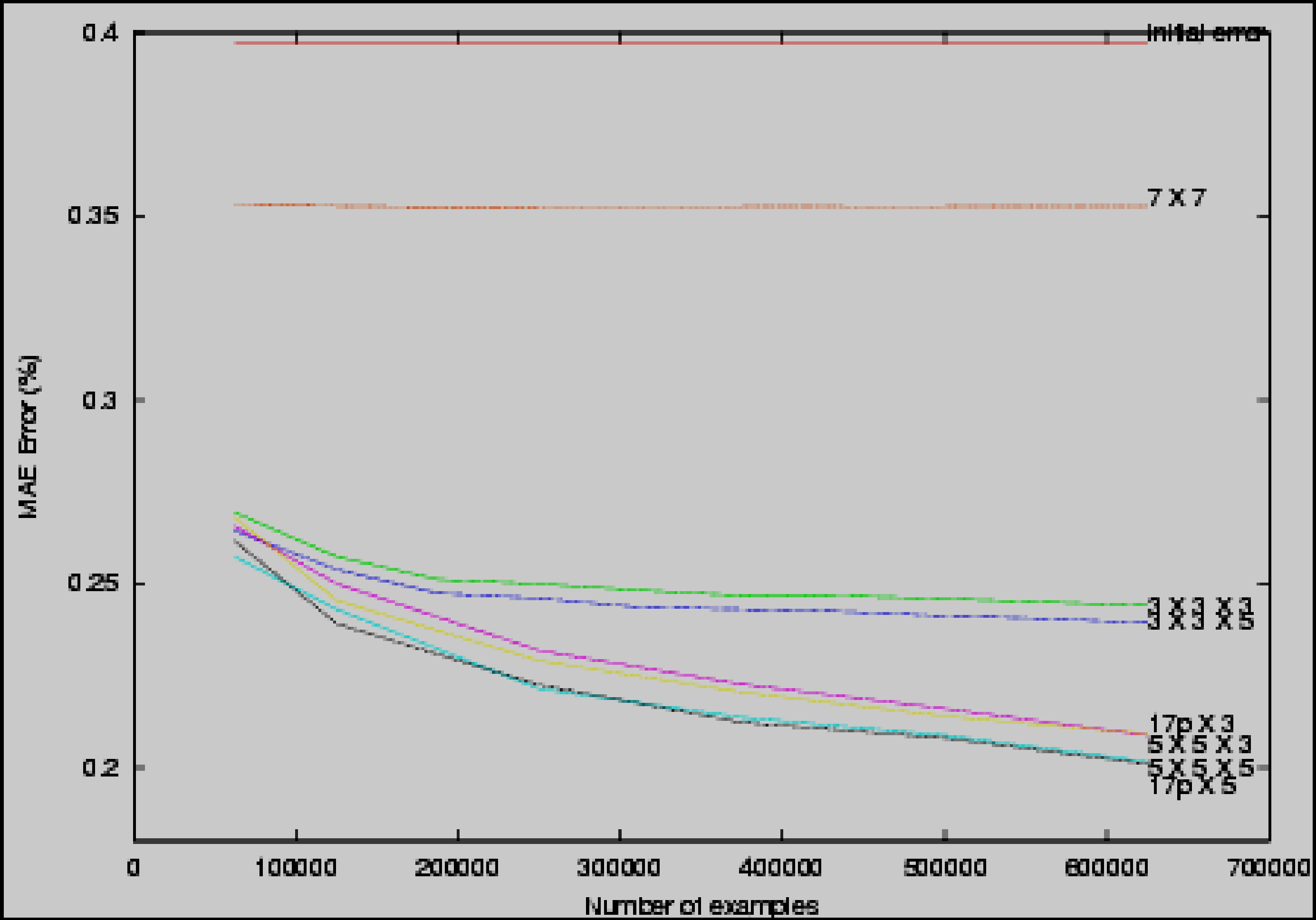


Deblurring





training images



MAE Error (%)
0.4
0.35
0.3
0.25
0.2
0
100000
200000
300000
400000
500000
600000
700000
Number of examples
initial error
7 X 7
3 X 3 X 3
3 X 3 X 5
17p X 3
5 X 5 X 3
5 X 5 X 5
17p X 5

Lattice Dynamical System (LDS) Identification

Lattice Dynamical System

$\mathbf{x}, \mathbf{u}: T \rightarrow L^n$ $\mathbf{y}: T \rightarrow L^m$ $\mathbf{x}[t] \in L^n$ $\mathbf{y}[t] \in L^m$

$$\mathbf{x}[t] = \phi_t(\mathbf{x}[t - N], \ldots, \mathbf{x}[t], \mathbf{u}[t - N], \ldots, \mathbf{u}[t])$$

$$\mathbf{y}[t] = \psi_t(\mathbf{x}[t - N], \ldots, \mathbf{x}[t], \mathbf{u}[t - N], \ldots, \mathbf{u}[t])$$

$$S(\phi_t, \psi_t)$$

Representation

$\mathbf{x}[t][j], \mathbf{y}[t][j] \in L$

$$\mathbf{x}[t][j] = \phi_{t,j}(\mathbf{x}[t - N], \ldots, \mathbf{x}[t], \mathbf{u}[t - N], \ldots, \mathbf{u}[t])$$

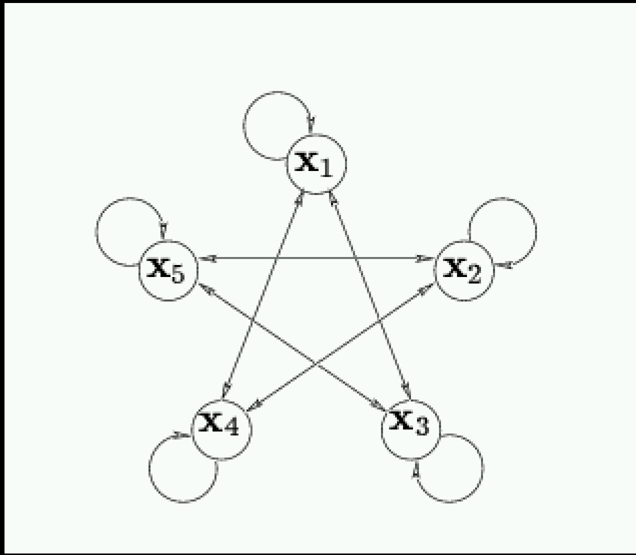
$$\mathbf{y}[t][j] = \psi_{t,j}(\mathbf{x}[t - N], \ldots, \mathbf{x}[t], \mathbf{u}[t - N], \ldots, \mathbf{u}[t])$$

The component functions $\phi_{t,j}$ and $\psi_{t,j}$ have canonical morphological representations

Architecture

Graph representing the transition function components ($\phi_{t,j}$)

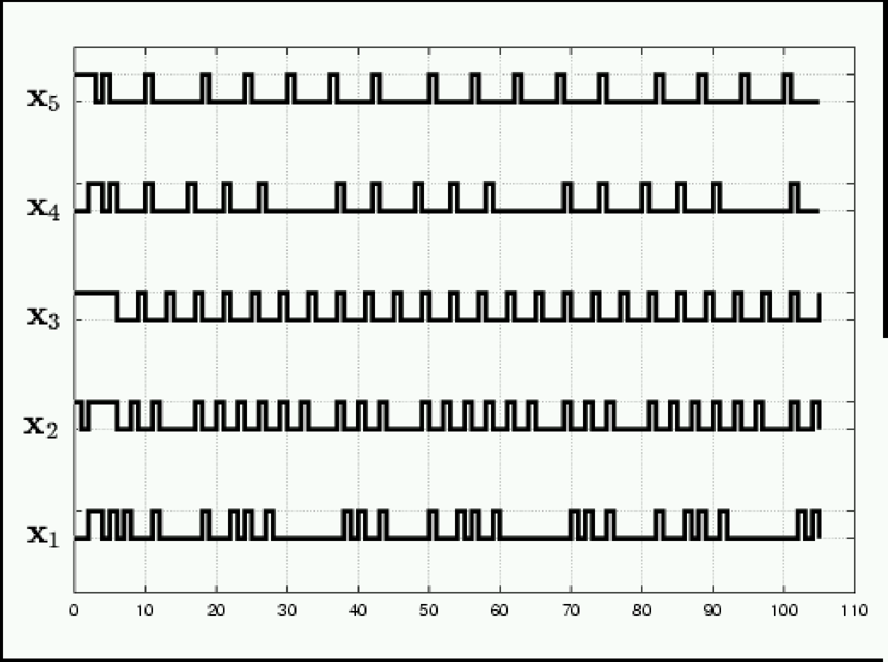
states association



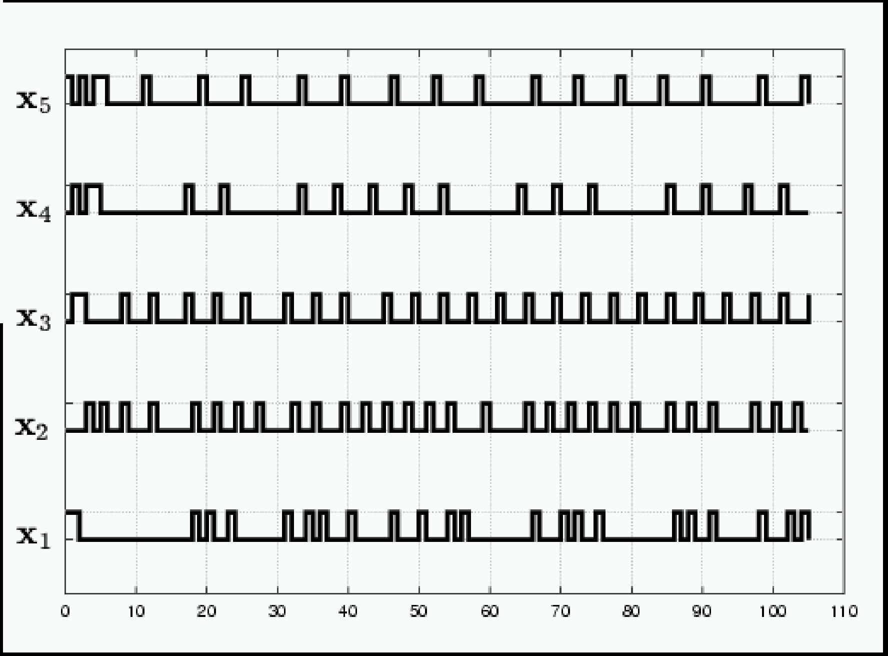
Dynamics

$$\mathbf{x}_1[t+1] = 1 \iff \begin{cases} \mathbf{x}_1[t] = 0 \\ \text{and} \\ \Big[\Big((\mathbf{x}_3[t] = 1 \text{ or } \mathbf{x}_3[t-1] = 1 \text{ or } \mathbf{x}_3[t-2] = 1) \text{ and} \\ \quad (\mathbf{x}_4[t] = 1 \text{ or } \mathbf{x}_4[t-1] = 1 \text{ or } \mathbf{x}_4[t-2] = 1) \Big) \\ \quad \text{or} \\ \quad \Big(\mathbf{x}_3[t] = \mathbf{x}_3[t-1] = \mathbf{x}_3[t-2] = \mathbf{x}_3[t-3] = \mathbf{x}_3[t-4] = 0 \text{ and} \\ \quad \mathbf{x}_4[t] = \mathbf{x}_4[t-1] = \mathbf{x}_4[t-2] = \mathbf{x}_4[t-3] = \mathbf{x}_4[t-4] = 0 \Big) \Big] \end{cases}$$

Simulation



x5
x4
x3
x2
x1
0
10
20
30
40
50
60
70
80
90
100
110



x5
x4
x3
x2
x1
0
10
20
30
40
50
60
70
80
90
100
110

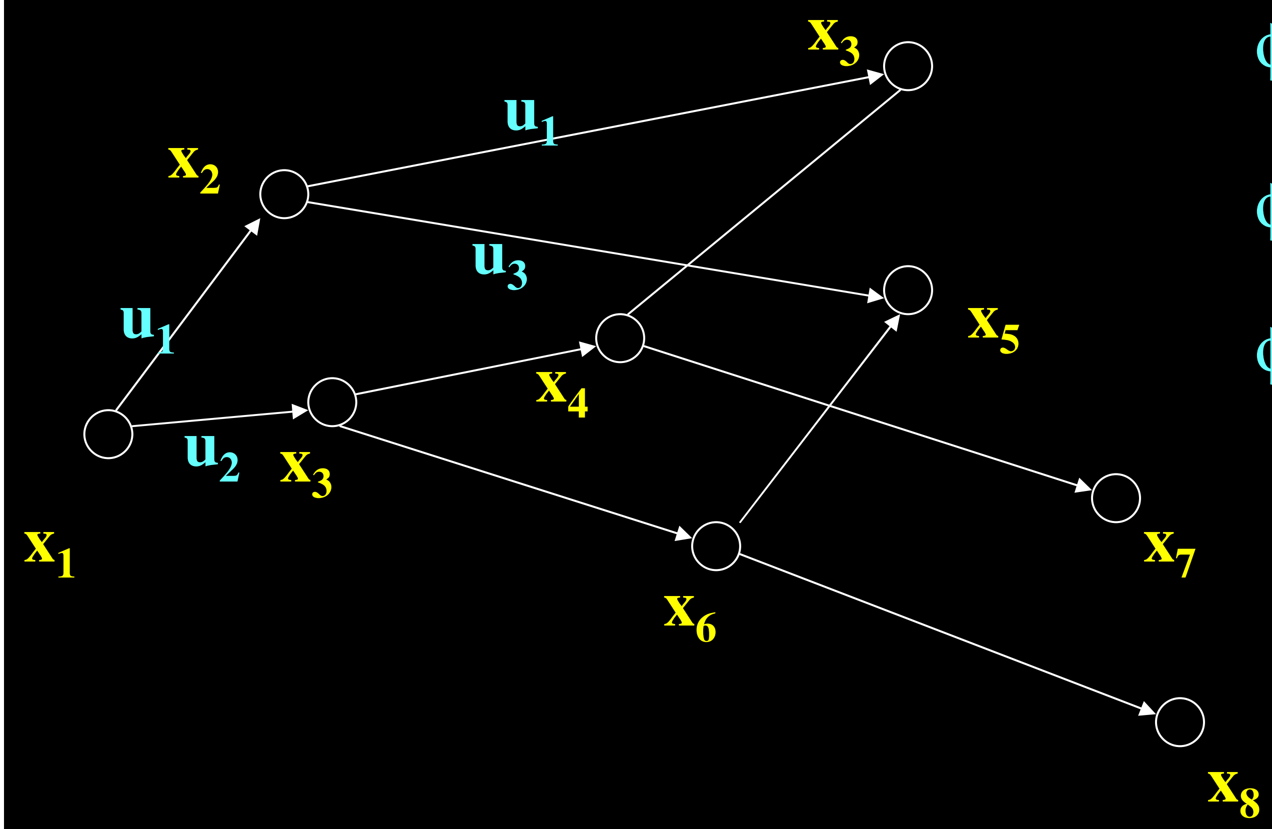
Equivalent System

$\mathbf{x}, \mathbf{u}: T \rightarrow L^{nN}$ $\mathbf{y}: T \rightarrow L^{mN}$ $\mathbf{x}[t] \in L^{nN}$ $\mathbf{y}[t] \in L^{mN}$

$$\mathbf{x}[t+1] = \phi_t(\mathbf{x}[t], \mathbf{u}[t])$$

$$\mathbf{y}[t] = \psi_t(\mathbf{x}[t], \mathbf{u}[t])$$

State Transition Graph



$\phi_t \Leftrightarrow \{\phi_{xi} : L^{mN} \rightarrow L^{nN}\}$

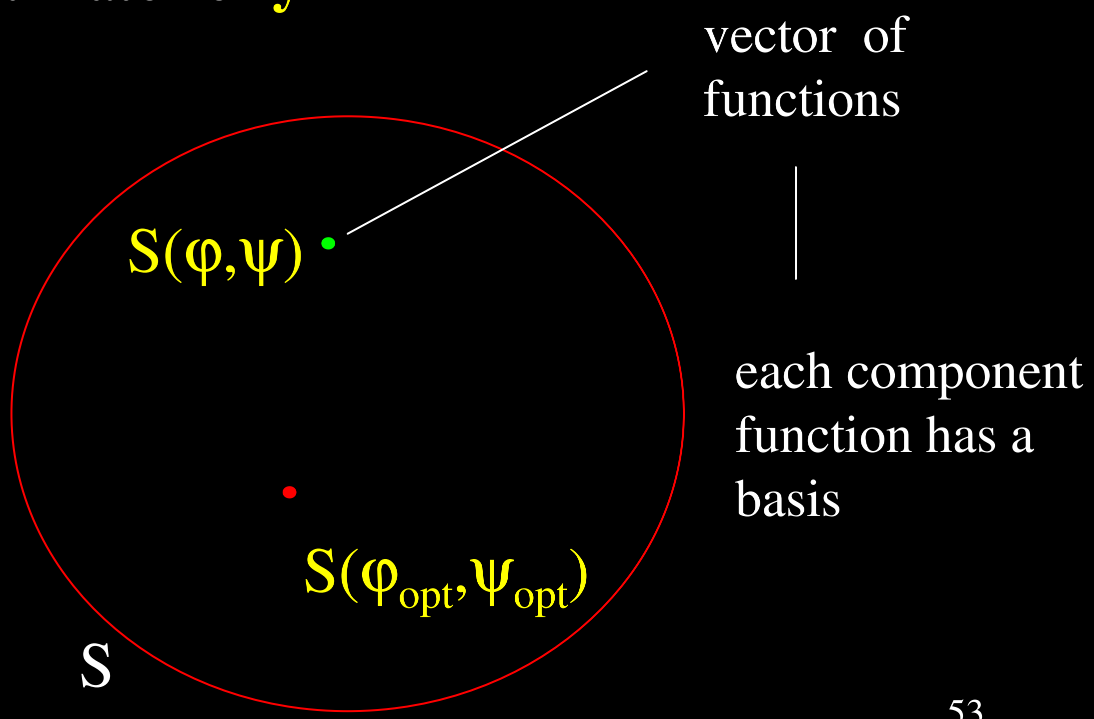
$\phi_{x1}(\mathbf{u}_1) = \mathbf{x}_2$

$\phi_{t,j} \Leftrightarrow \{\phi_{xi,j} : L^{mN} \rightarrow L\}$

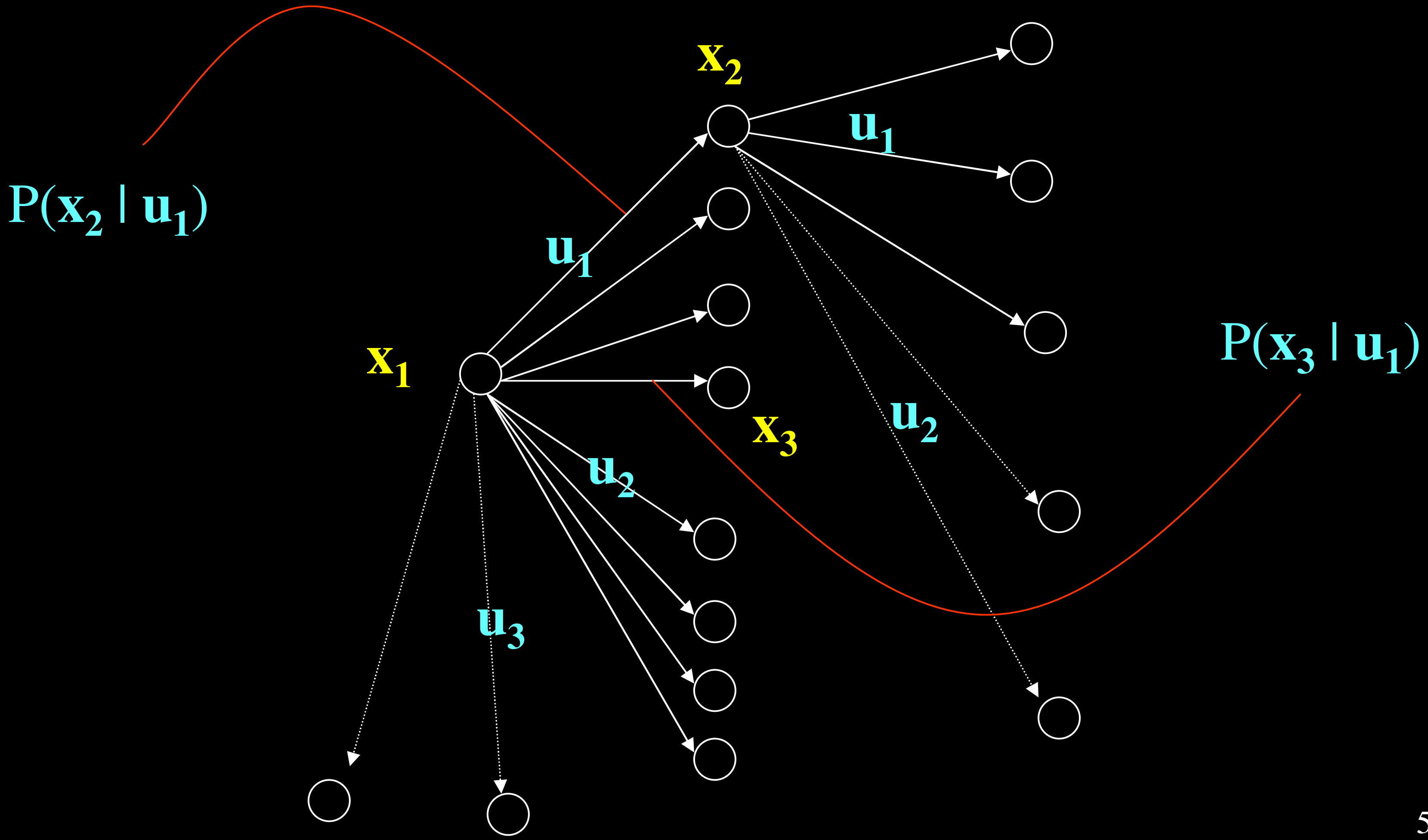
System identification

u and **y are random processes**

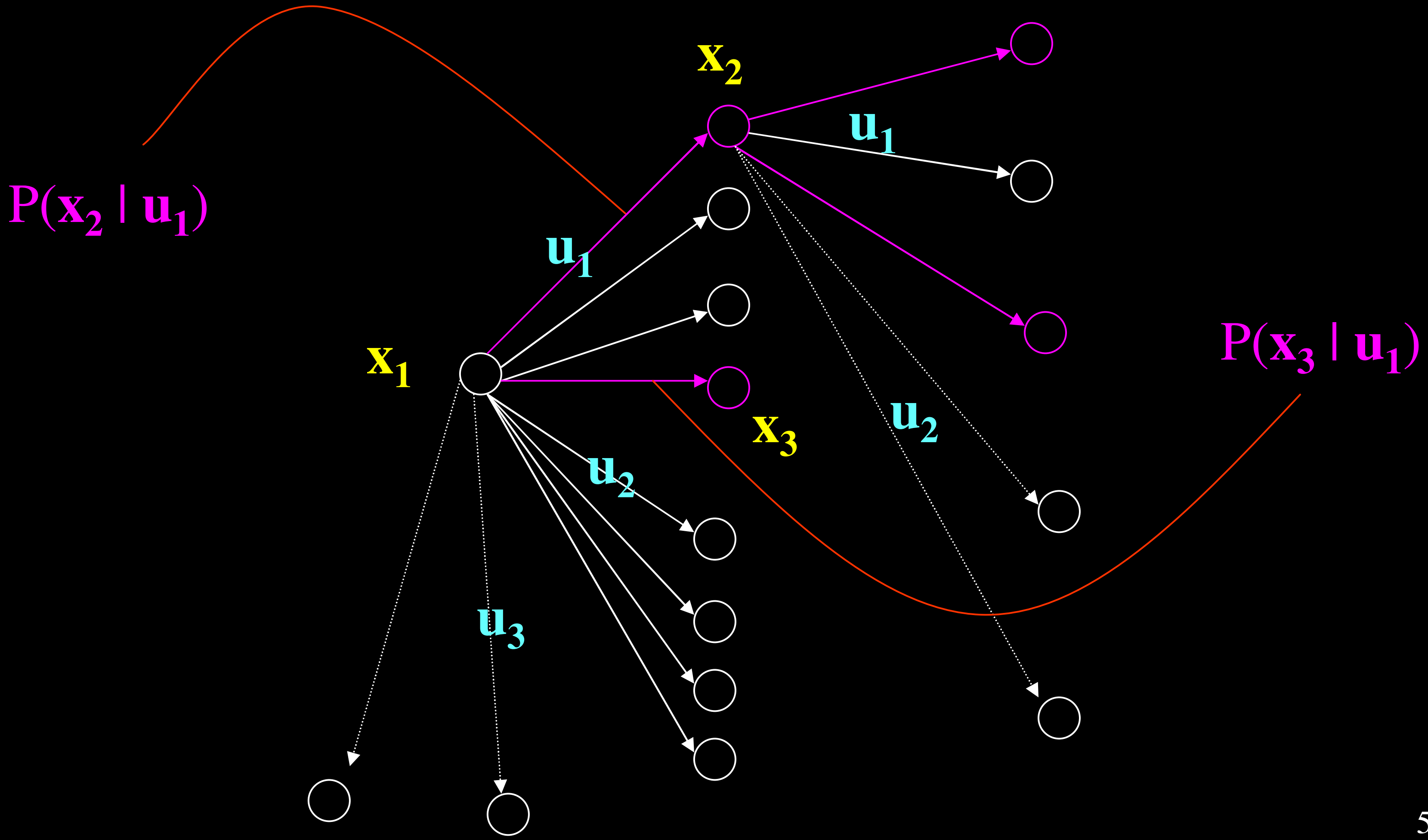
$S_{xo}(\phi_t, \psi_t)(\mathbf{u})$ is an estimator of **y**



Cumulative Error in T steps



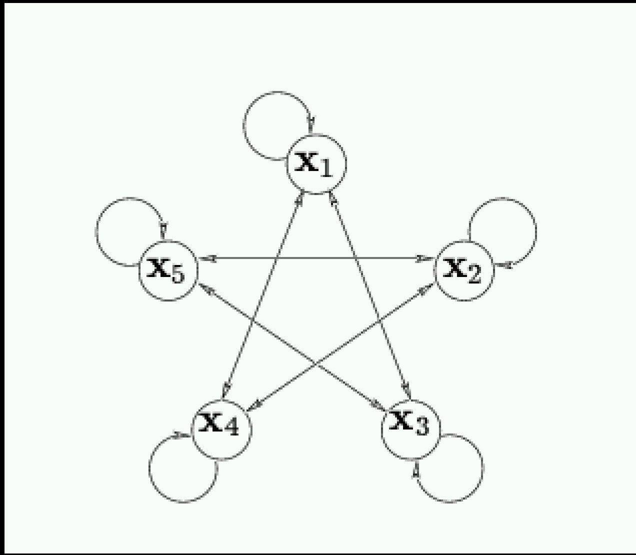
Estimation of the Error

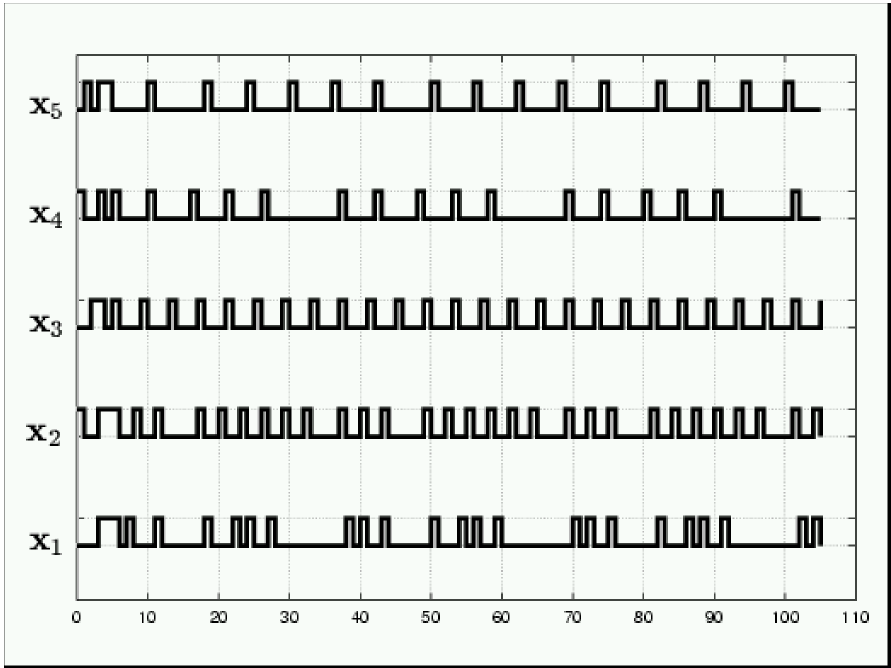


Generalization

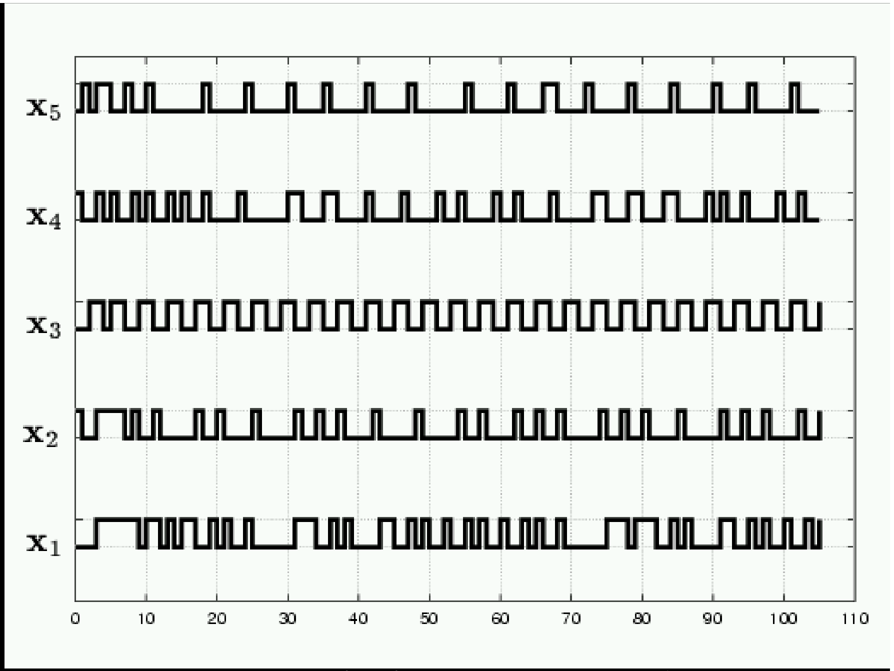
- The best paths estimated give a sample $\phi_{xi}(\mathbf{u}) = \mathbf{x} \Leftrightarrow \phi(\mathbf{x}i, \mathbf{u}) = \mathbf{x} \Leftrightarrow \phi_j(\mathbf{x}i, \mathbf{u}) = \mathbf{x}[j]$, $j \in [1,nN]$
- ϕ_j should be generalized and represented
- System constraints imply in generalization rules
- Learning algorithms build the basis from the sample and constraints

Example of System Identification

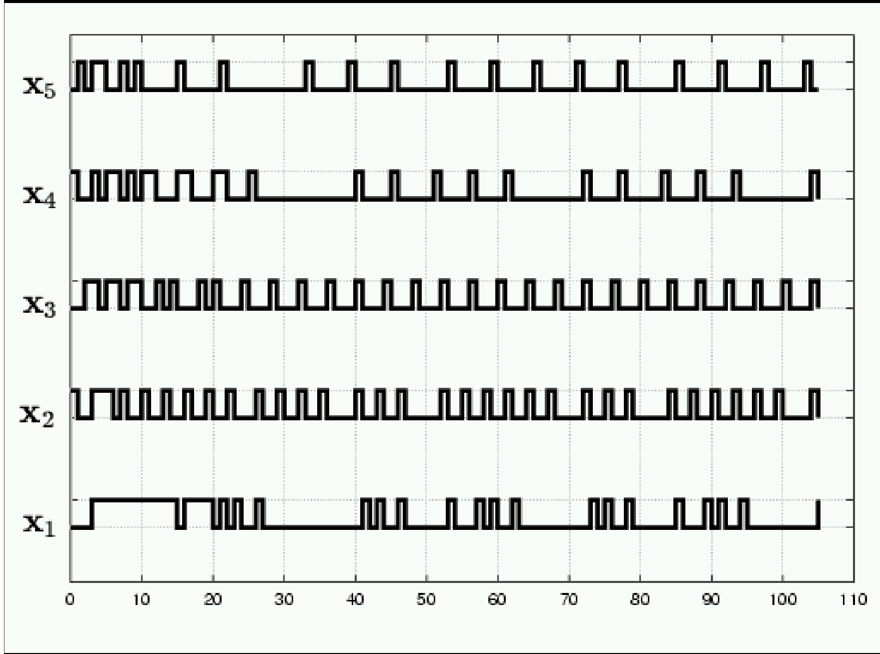




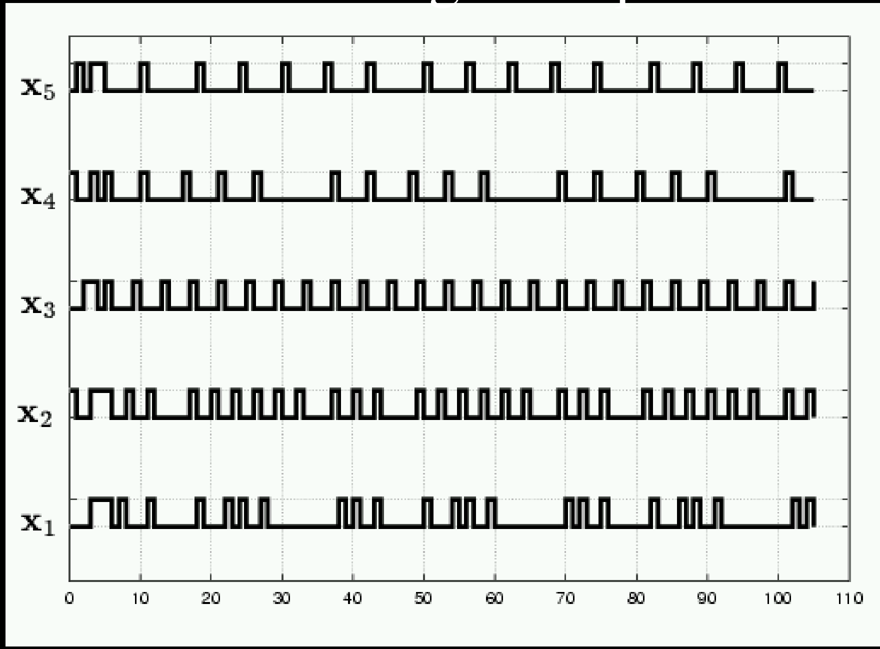
Ideal



Est.: 100 training examples

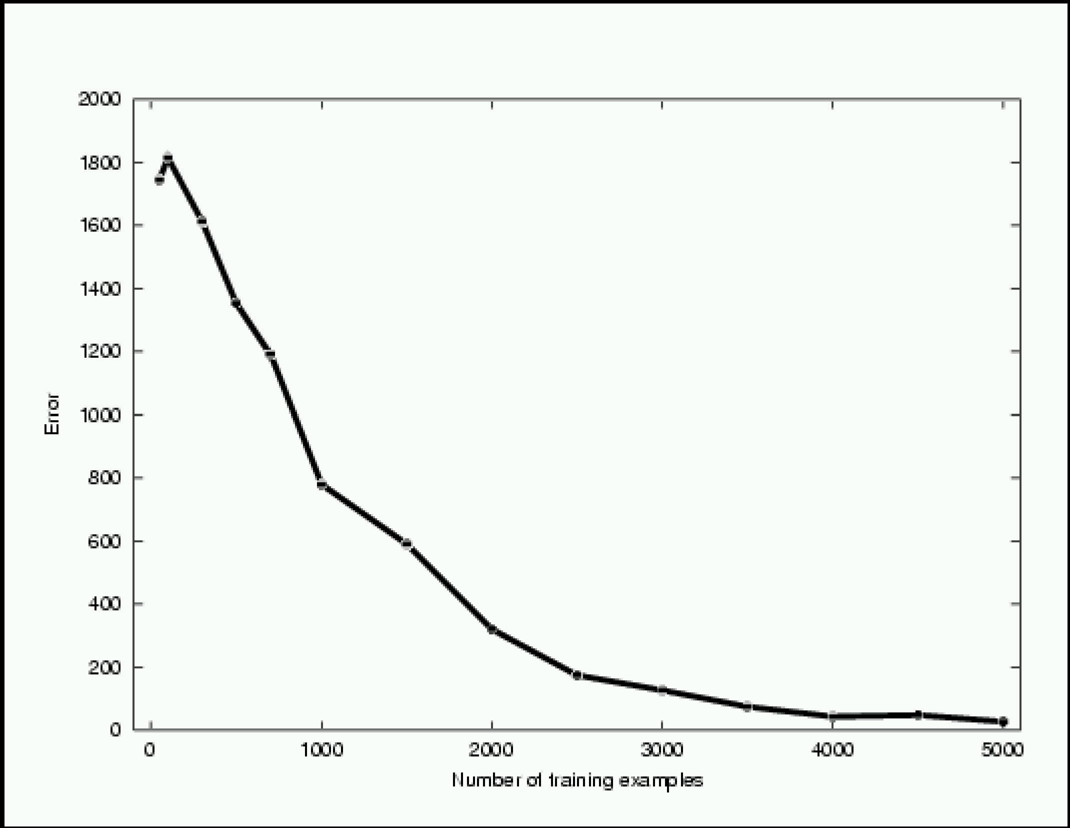


Est.: 500 training examples

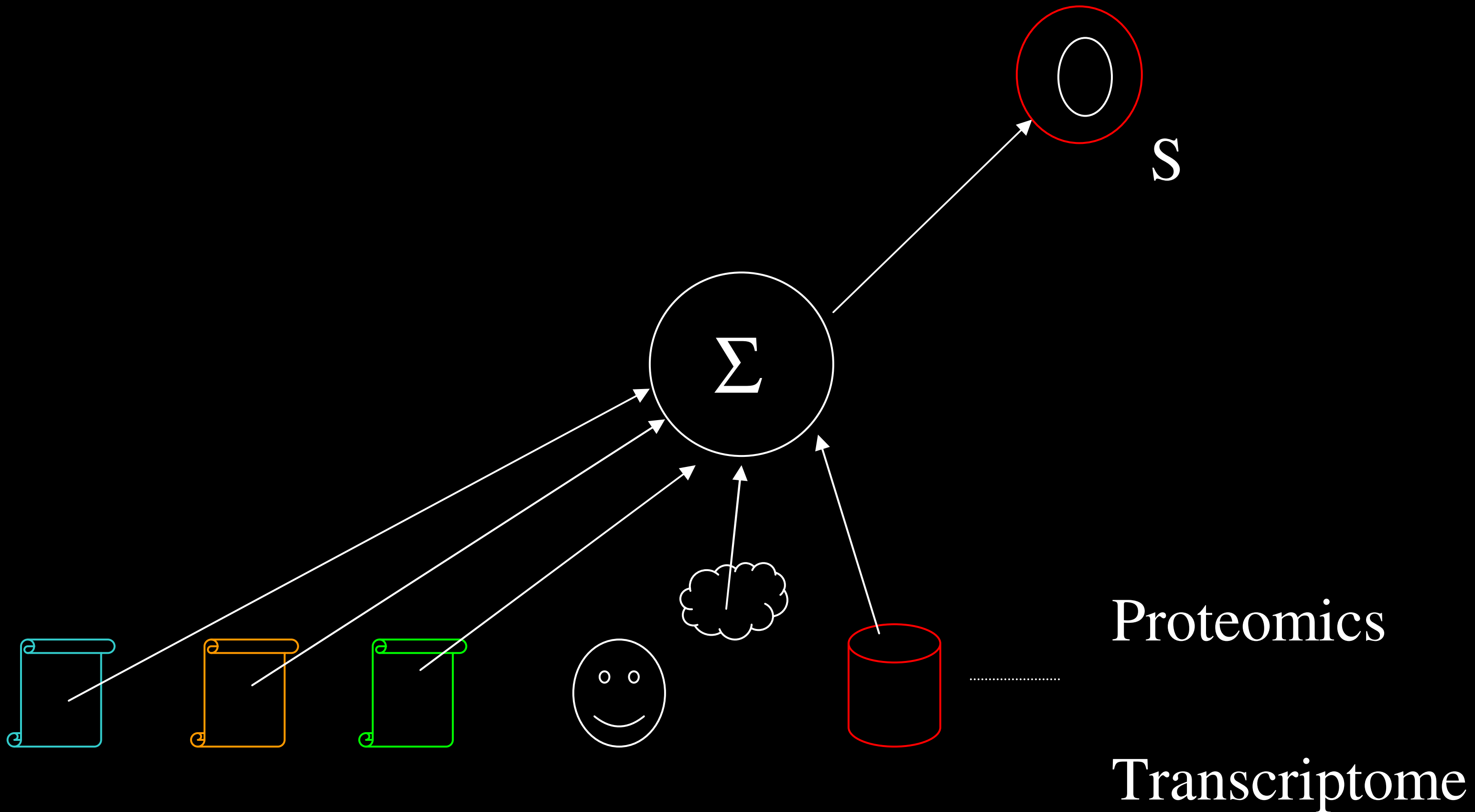


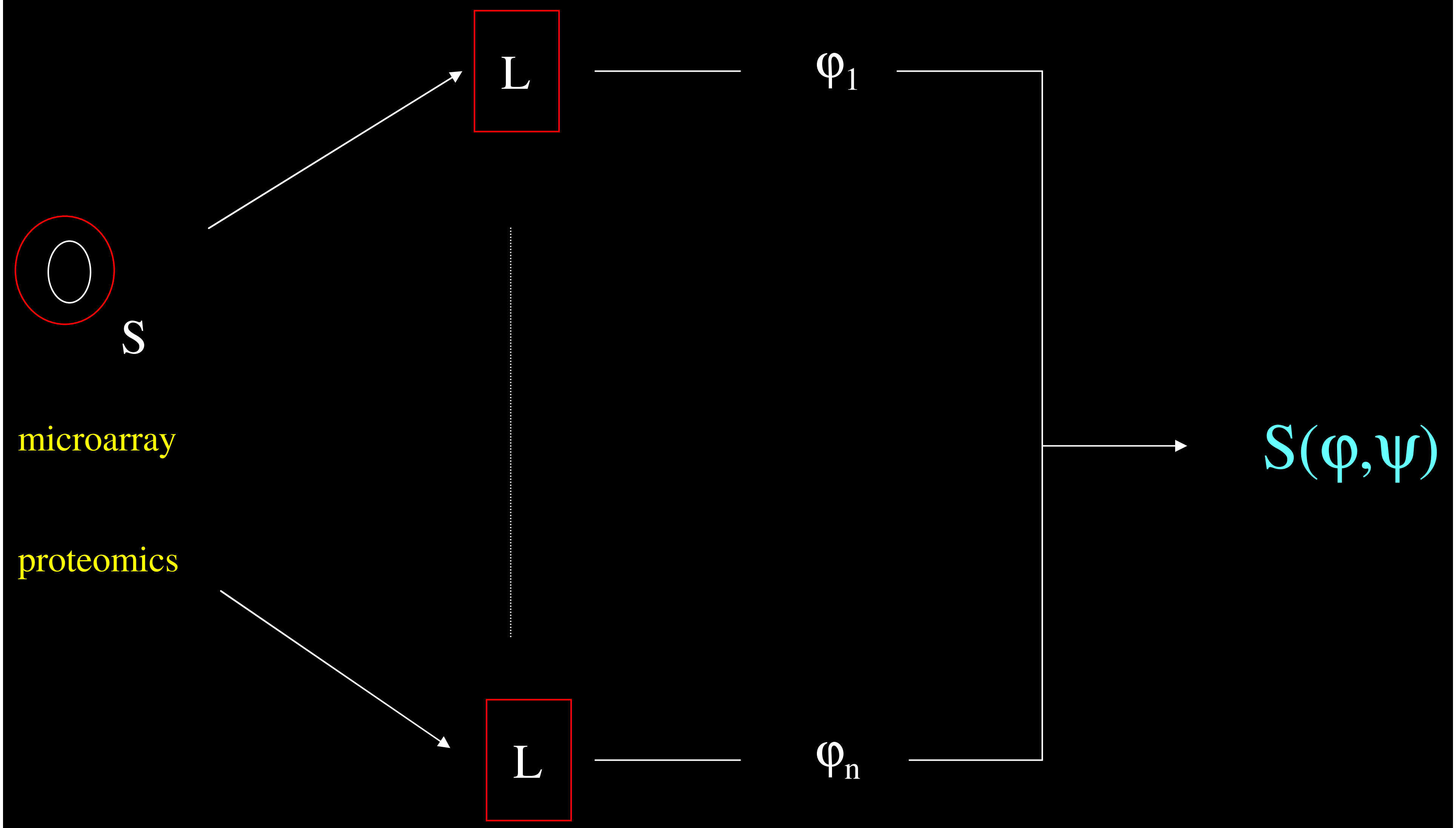
Est.: 1500 training examples

Empirical Cumulative Error



Modeling of Genetic Networks


S
Σ
Proteomics
Transcriptome

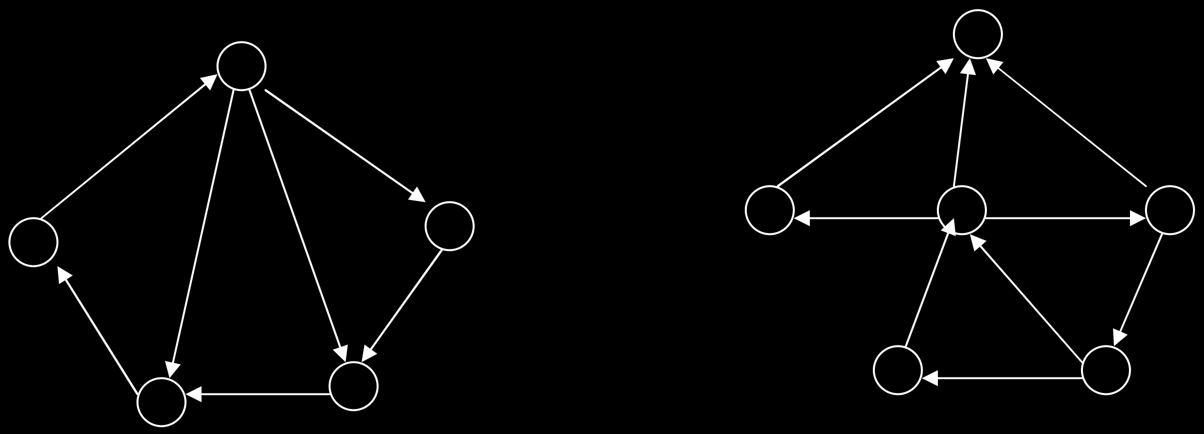

L
φ1
S
microarray
proteomics
L
φn
S(φ,ψ)

System Complexity

- Complexity of a LDS is the number of possible orbits
- For systems of dimension n, the complexity increase with the increase of |L|
- The size of the space of LDS also increase with the increase of n and |L|
- Hence, n and |L| are parameters for adjusting the model complexity

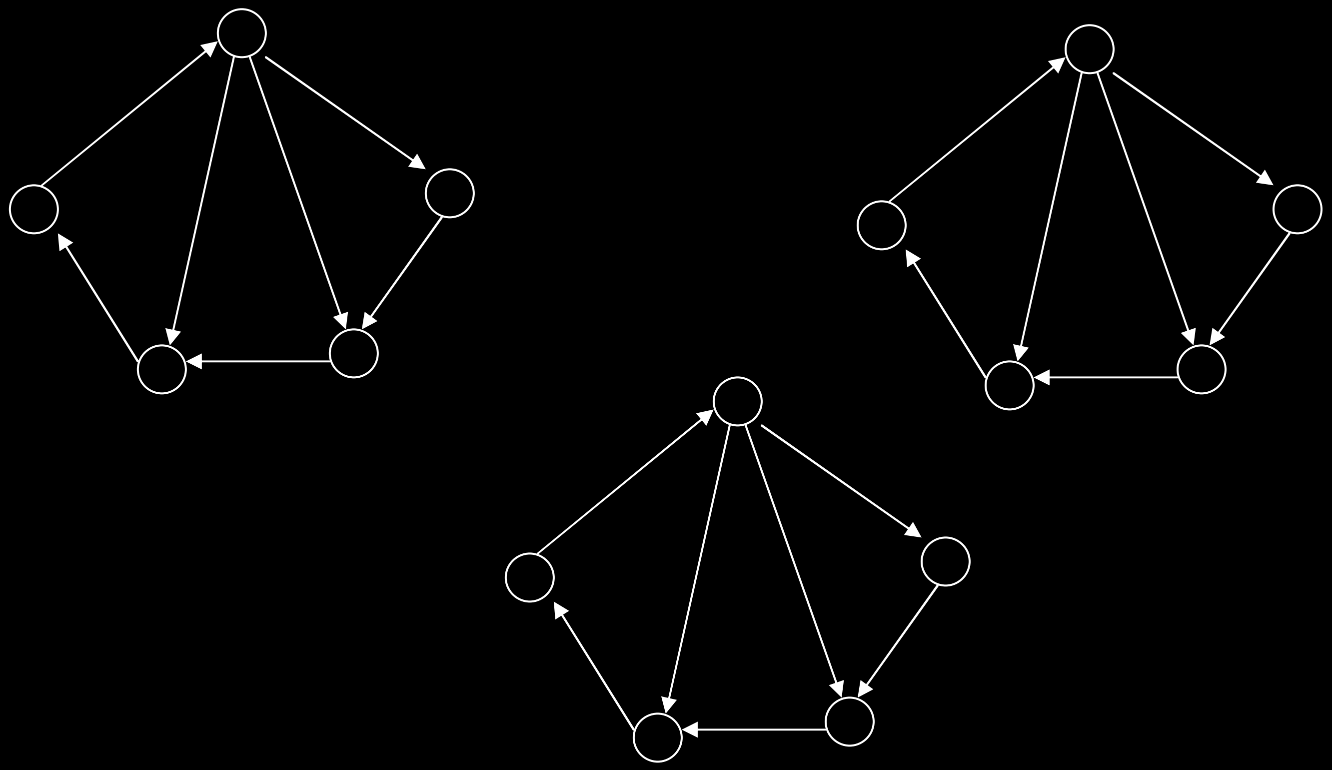
Independent Subsystems

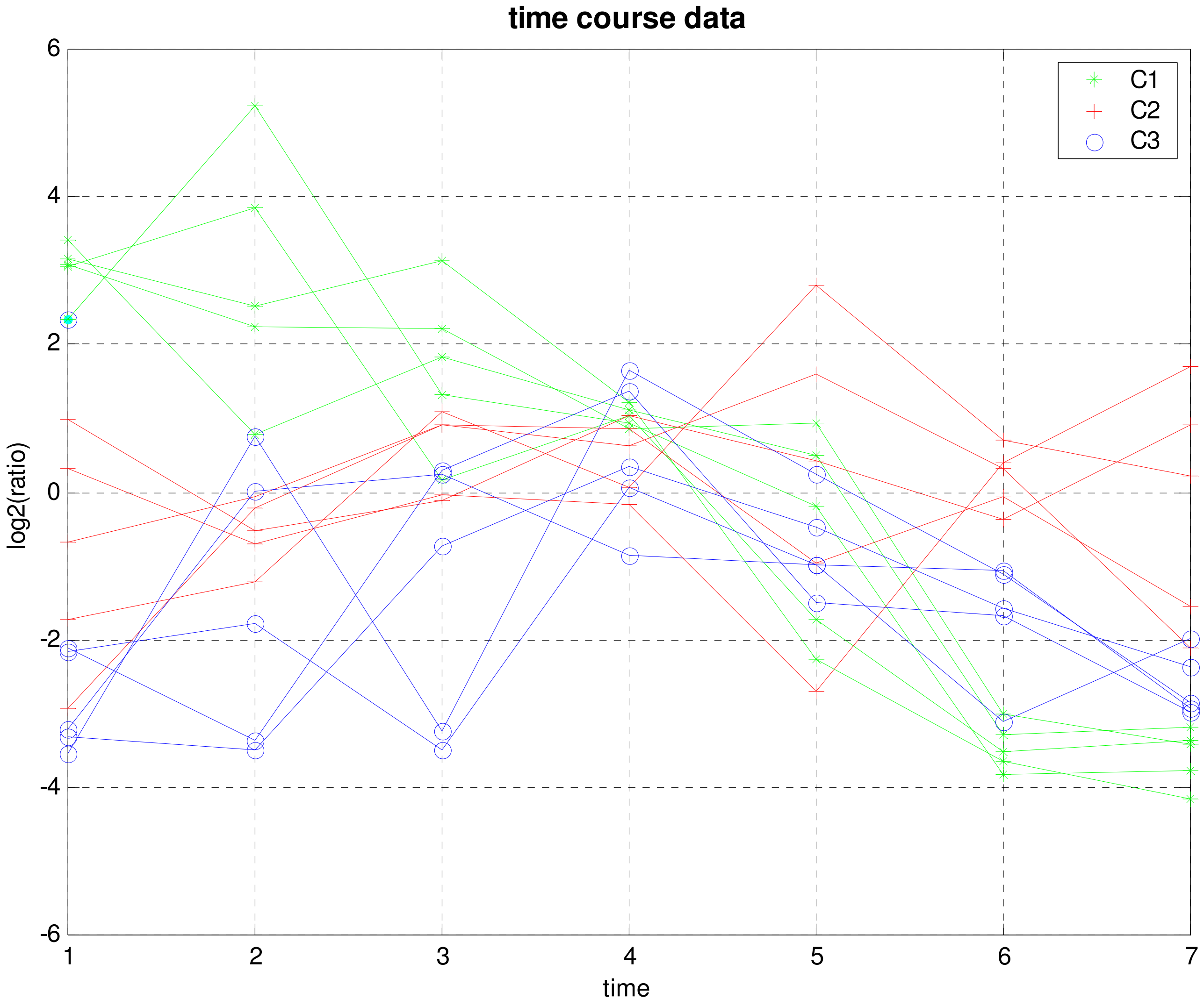
- If the system architecture has more than one connected component, it is composed of independent subsystems



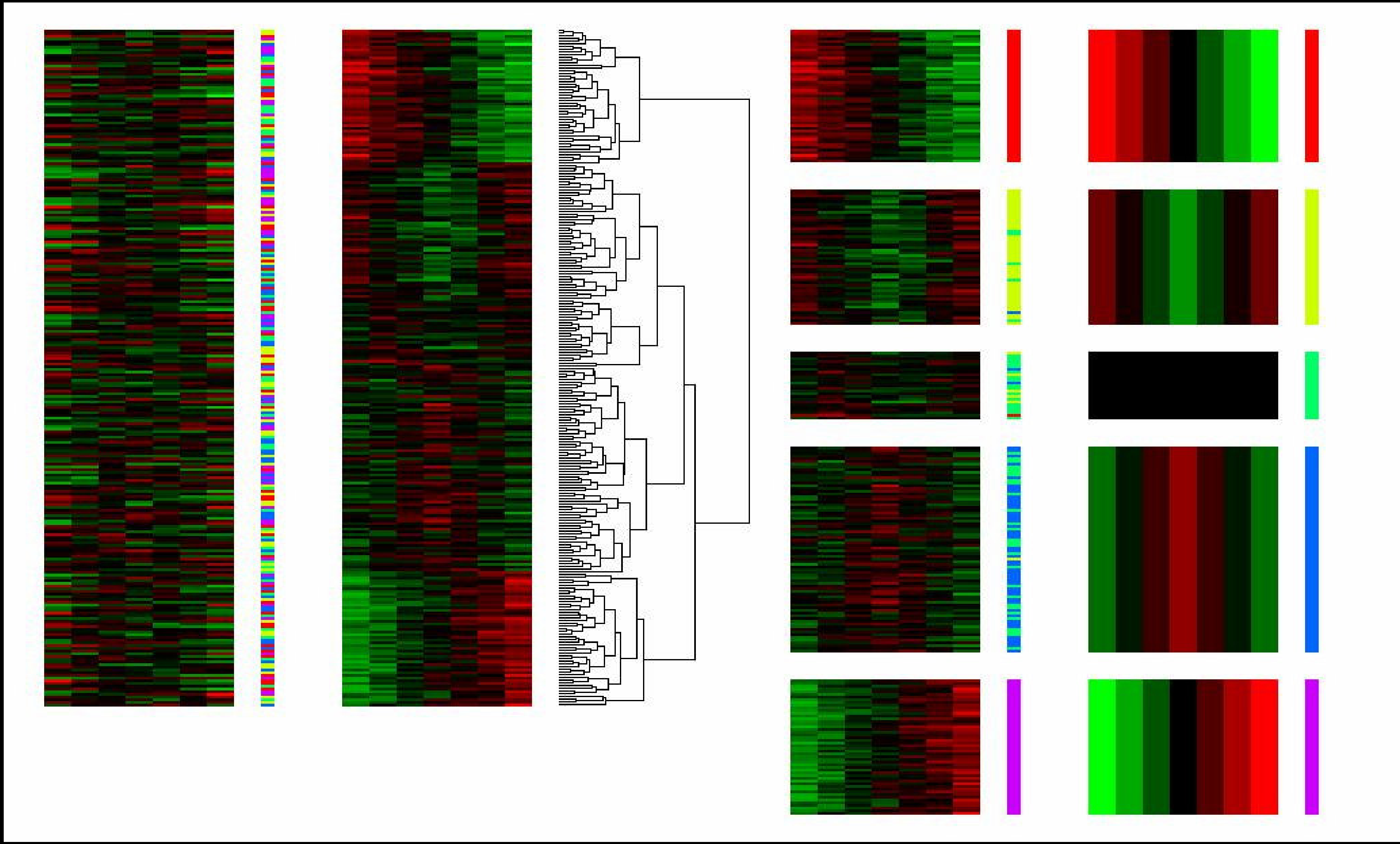
Replication

- Crucial systems may be replicated for safety



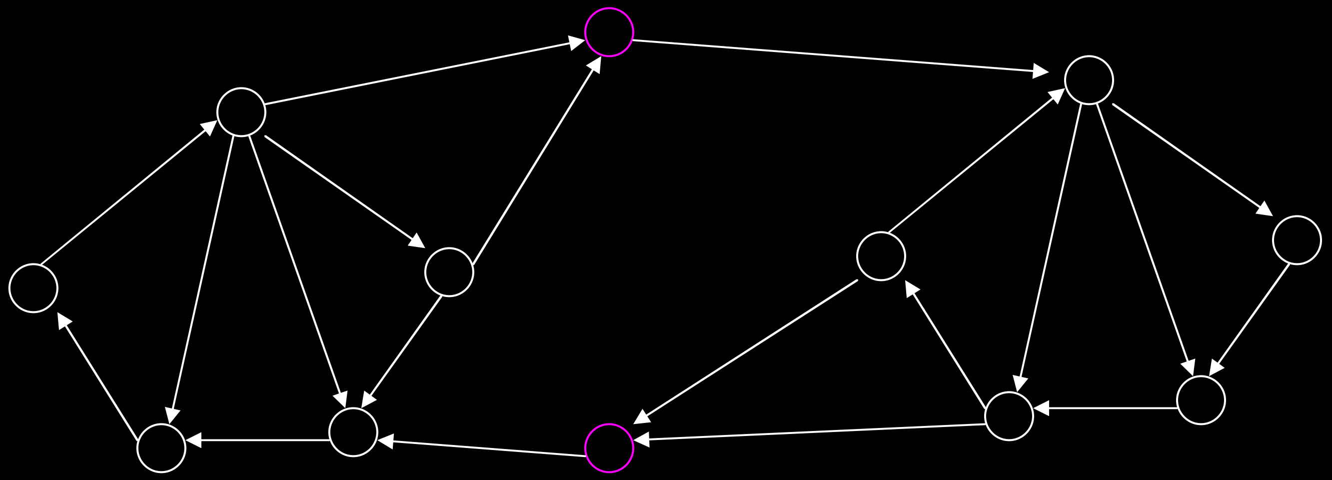


time course data
C1
C2
C3
log2(ratio)
time
6
4
2
0
-2
-4
-6
1
2
3
4
5
6
7

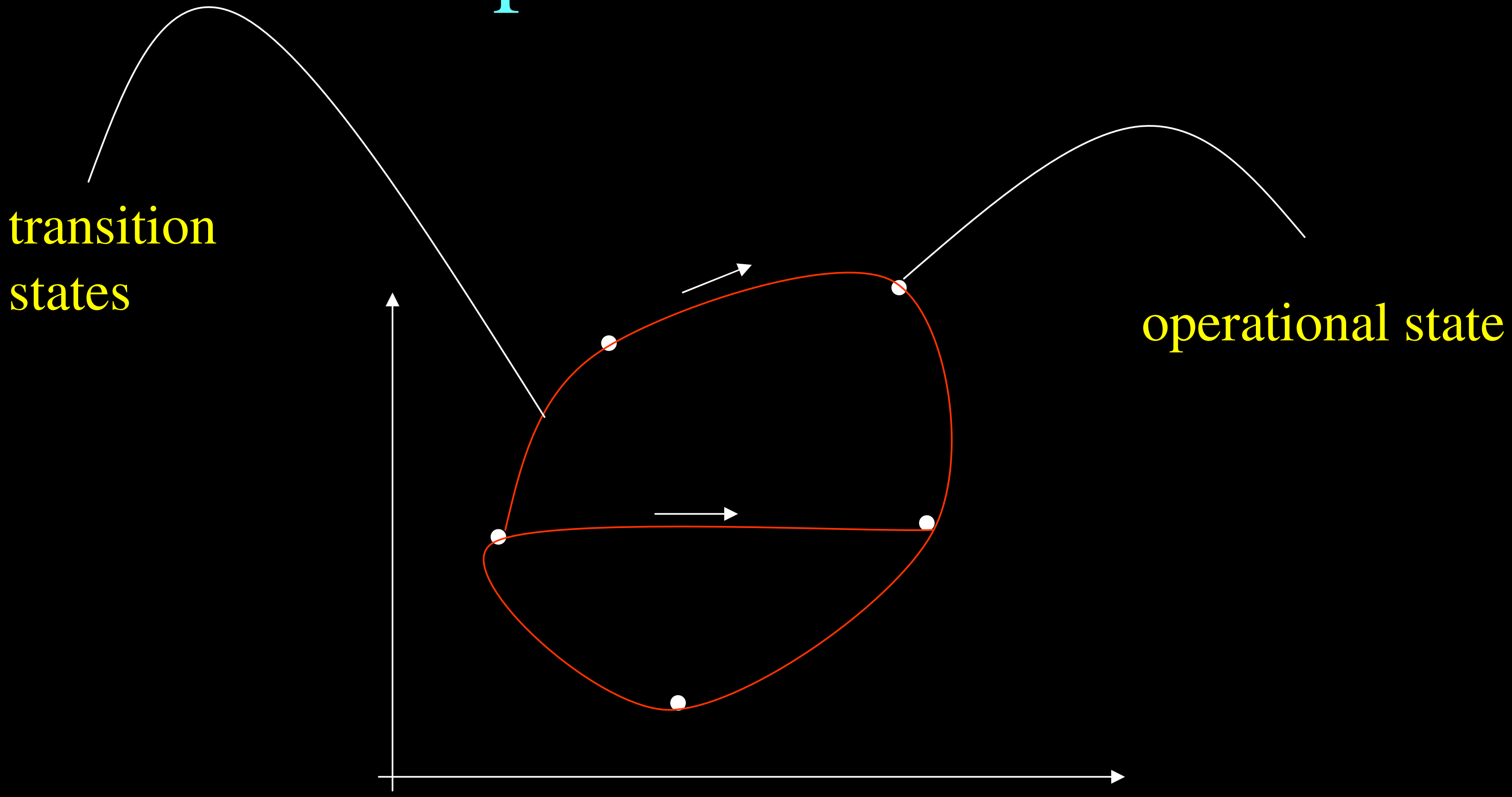


System Connections

- If the system is reduced to independent subsystems by fixing some arguments of $\phi_{t,j}$ it is said weekly connected



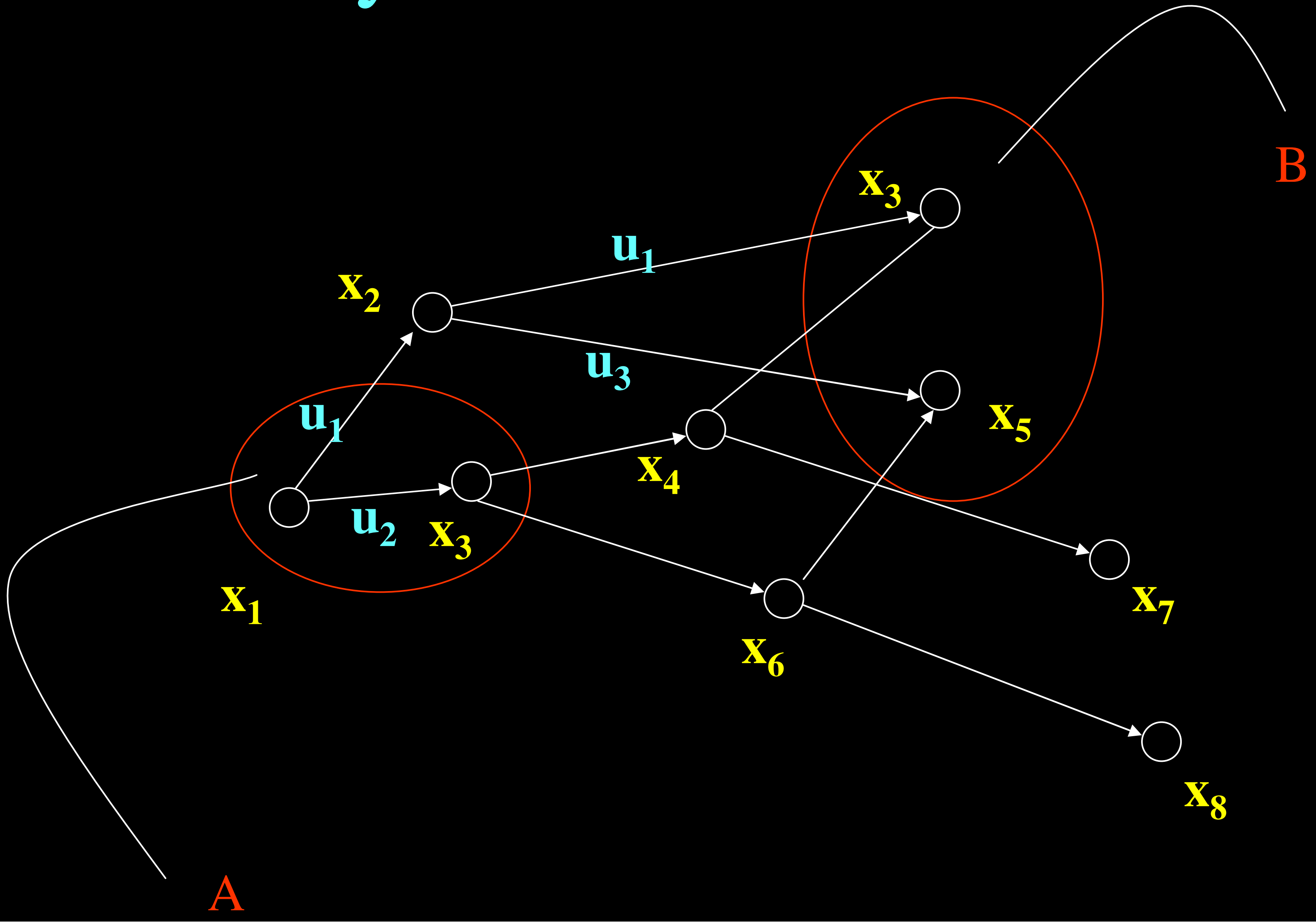
Operational States



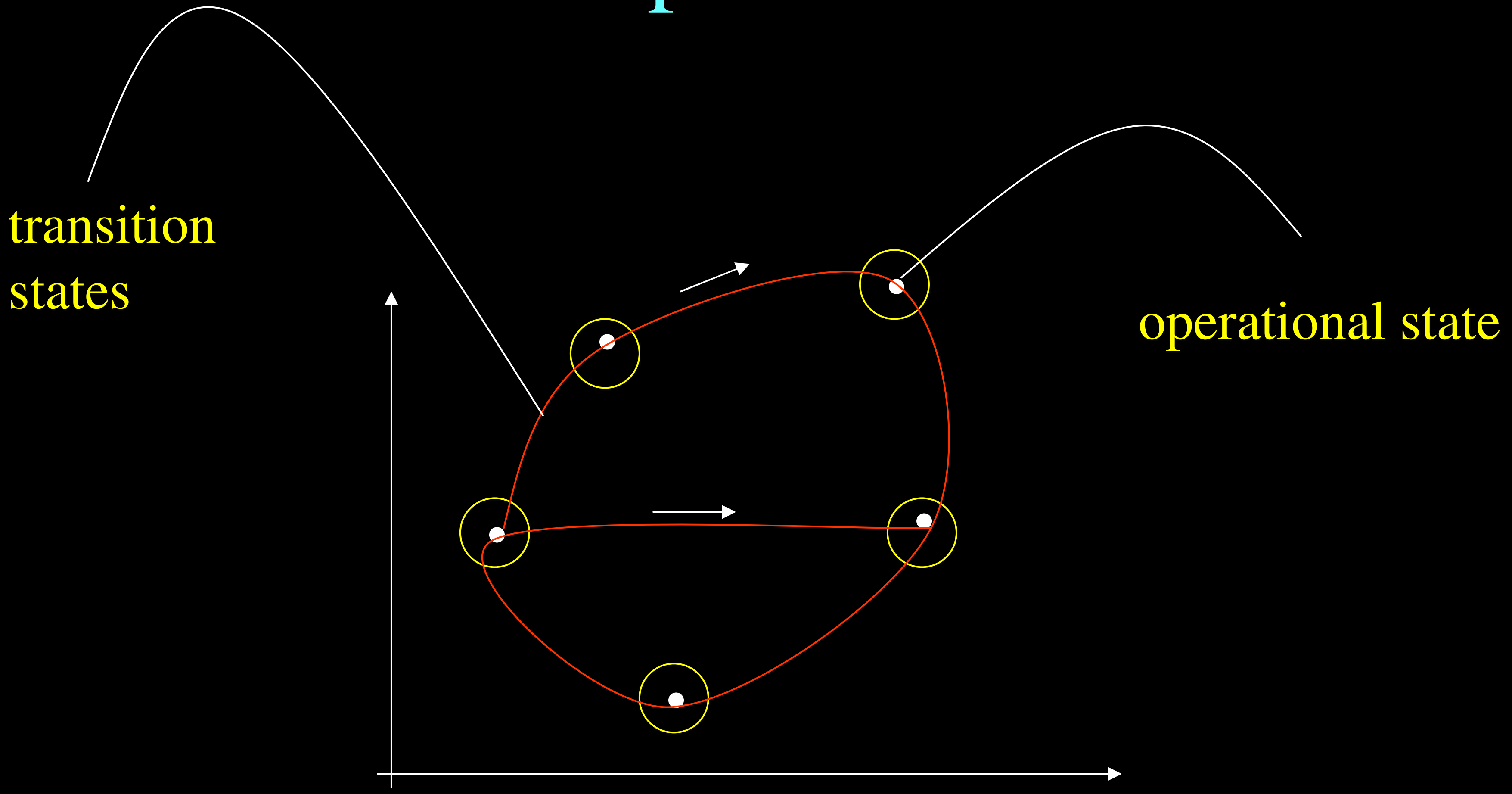
AB-Controlability

- Let A and B be subsets of possible states
- A LDS is AB-controlable if, for every $a \in A$ and $b \in B$, there is a path in the transition graph from a to b.

System AB-Controlable



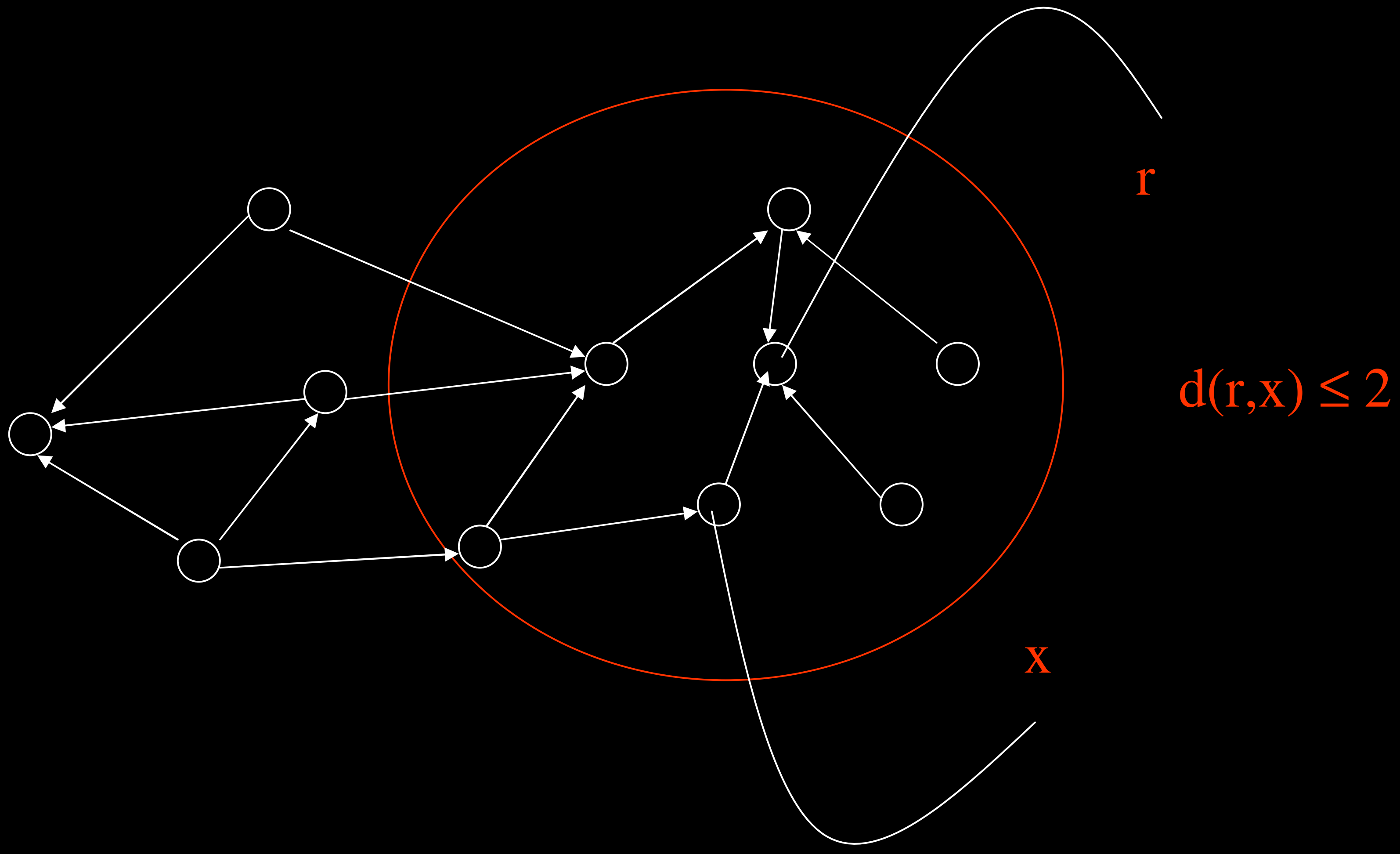
Robust Operational States



N-Robustness

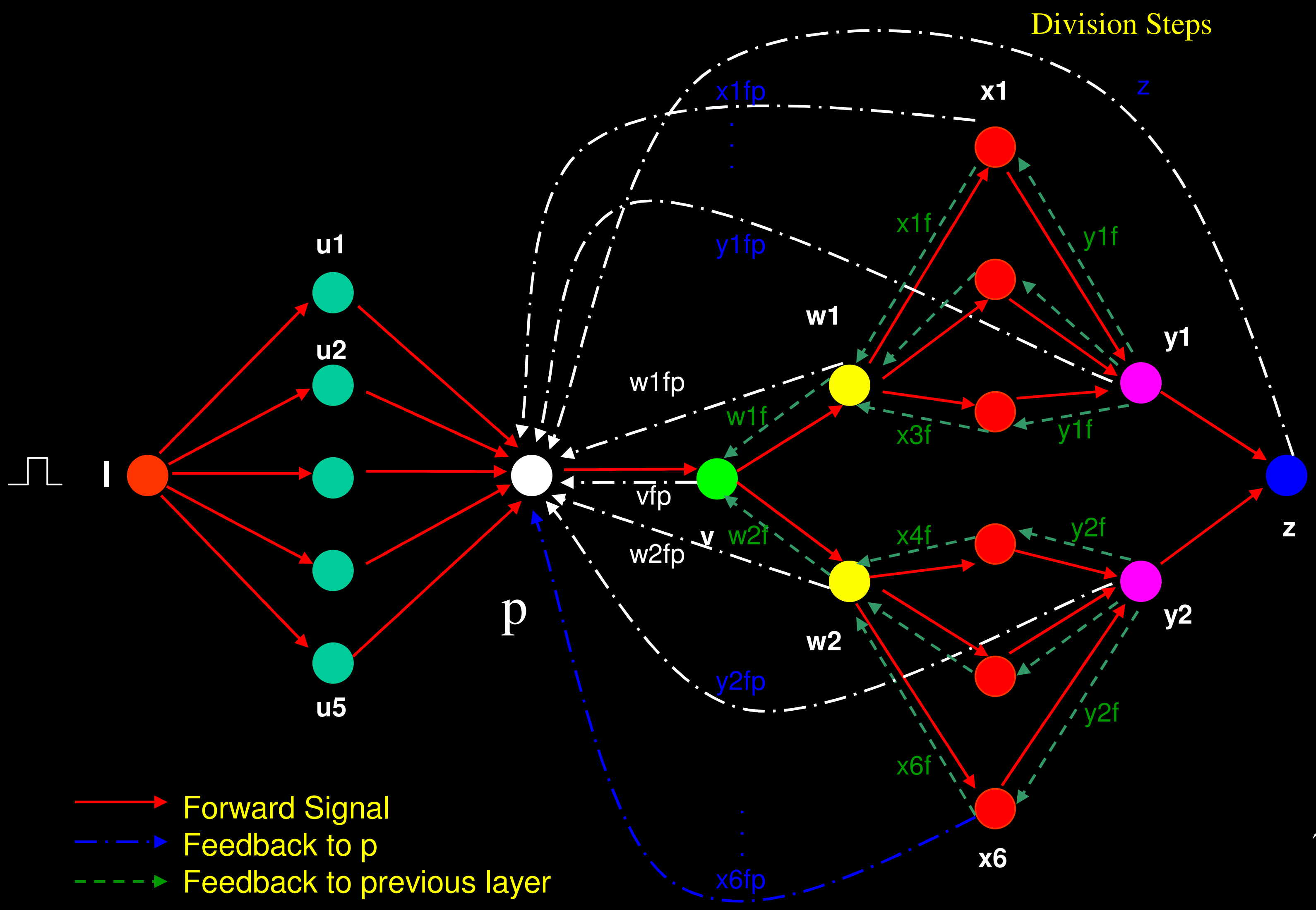
- An operational state r is N-robust if there is an unconditional path in the transition graph from every x such that $d(r,x) \leq N$.

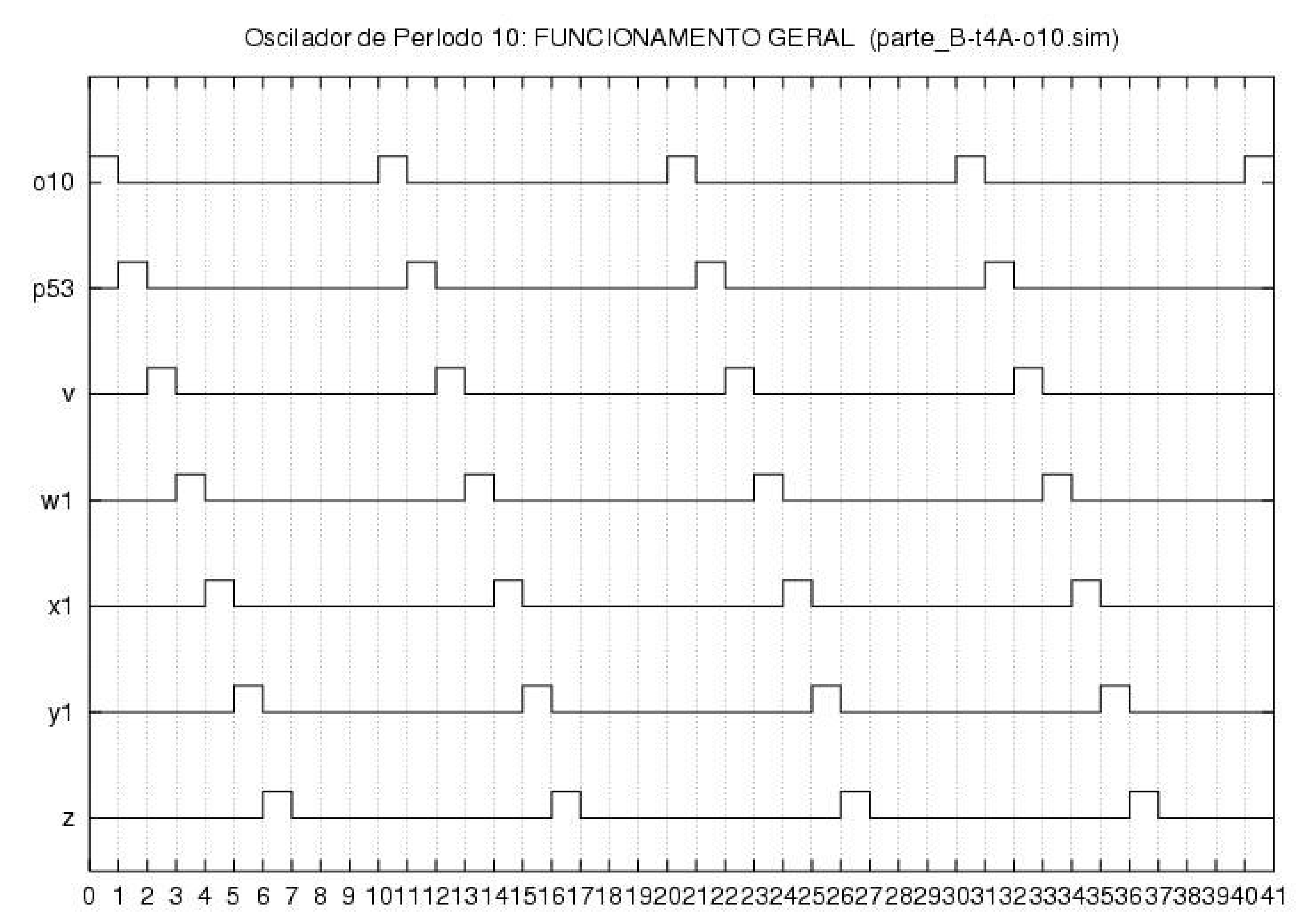
Robust and no Robust States

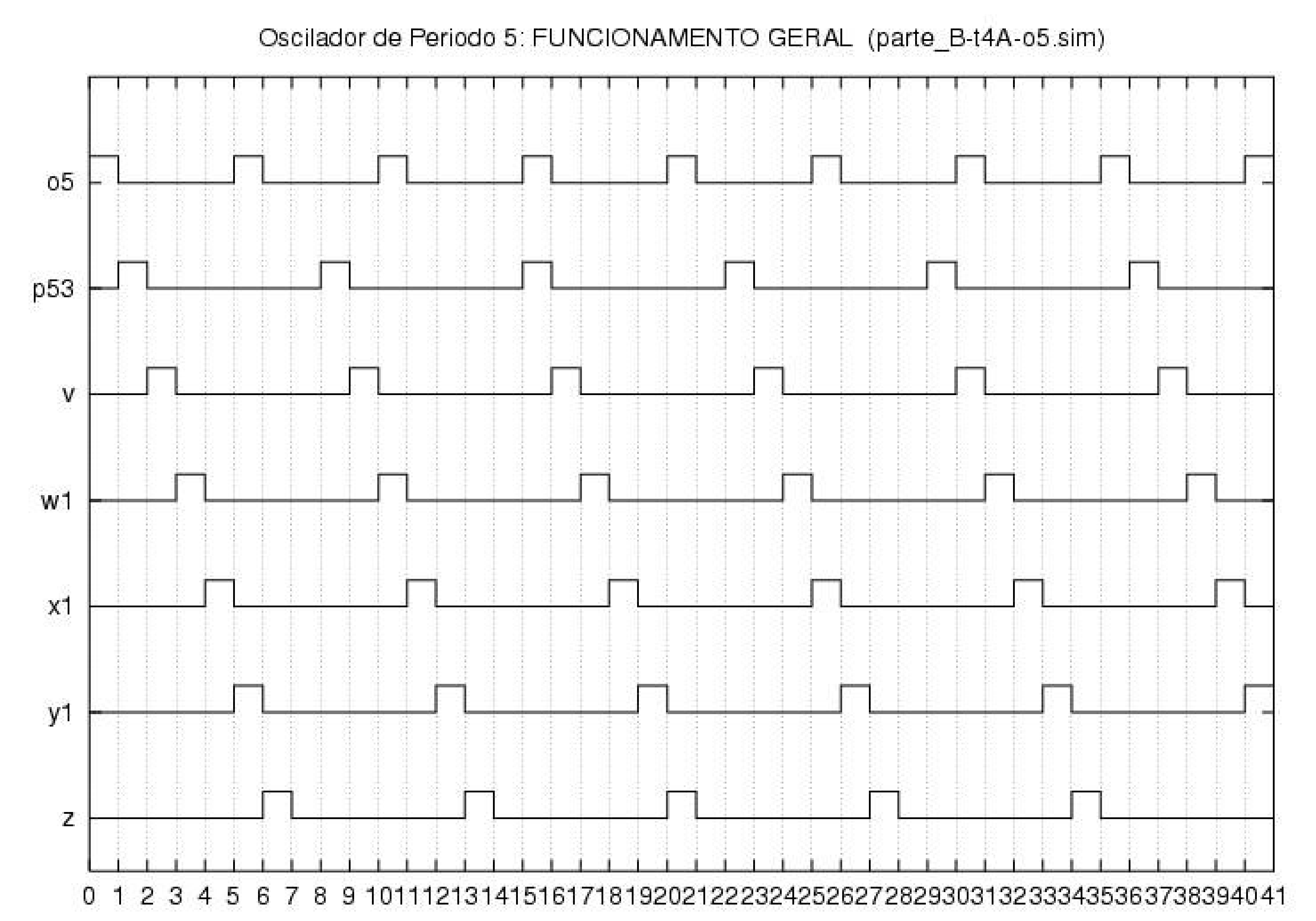


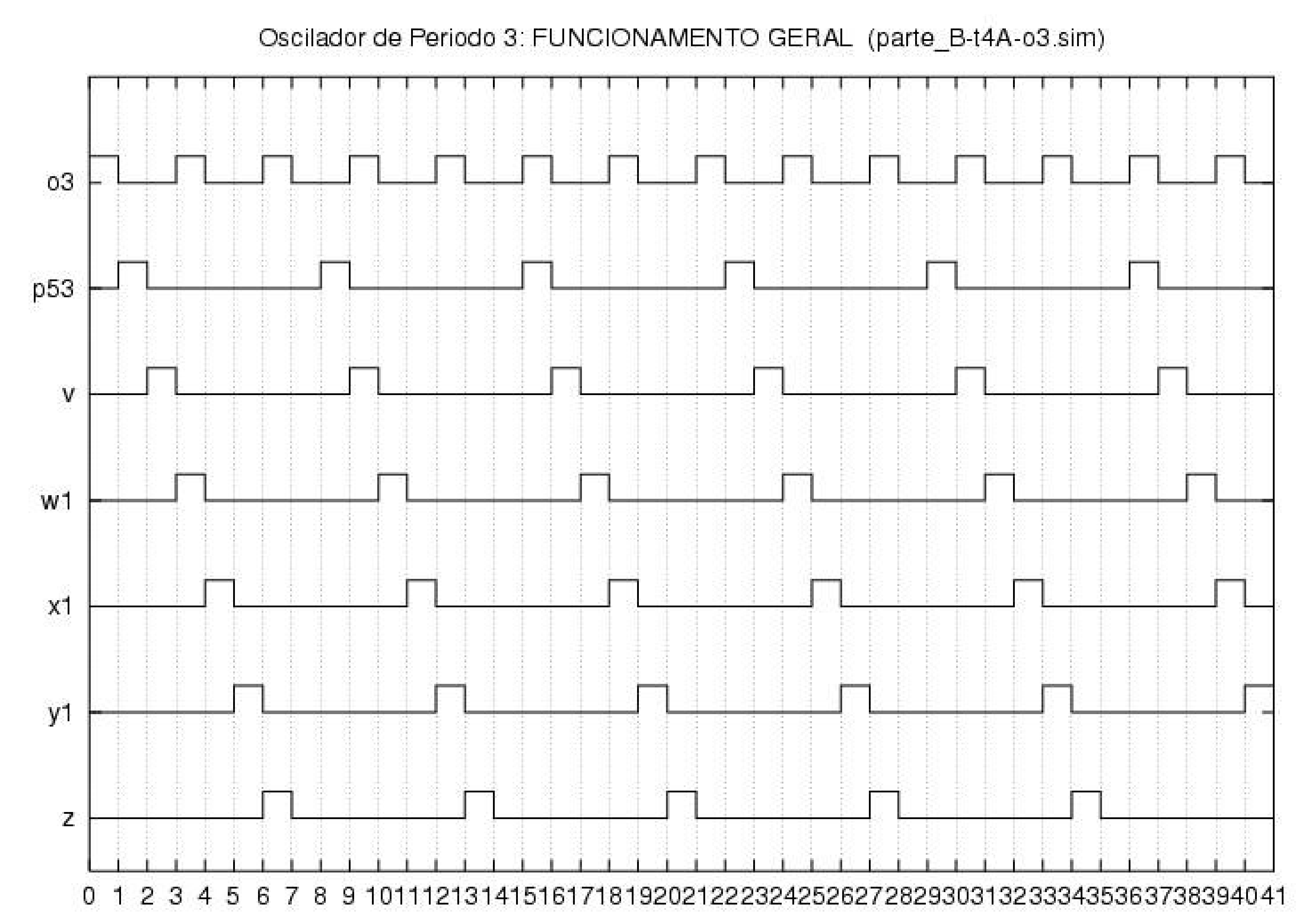
Examples

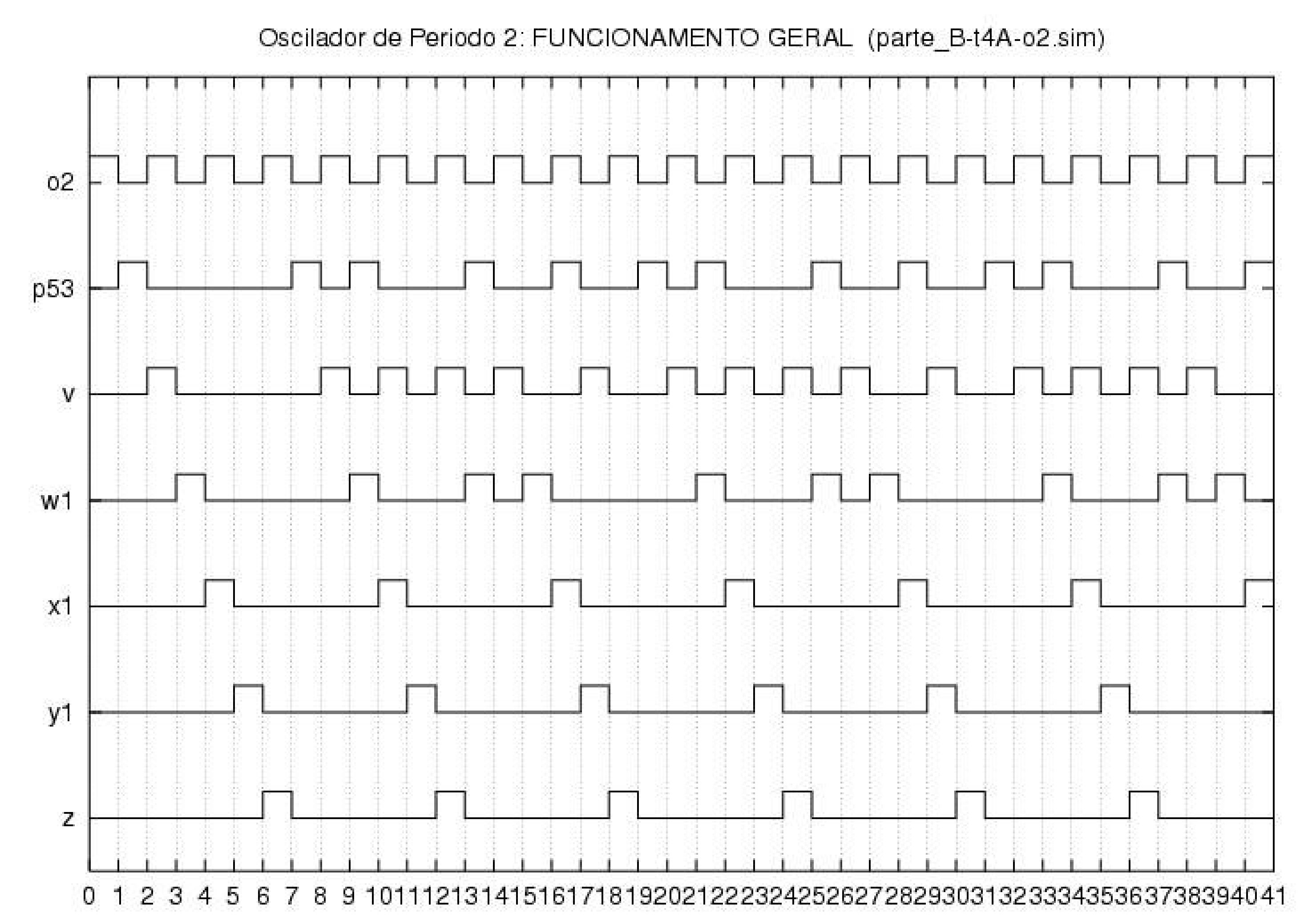
Cell Cycle Modeling



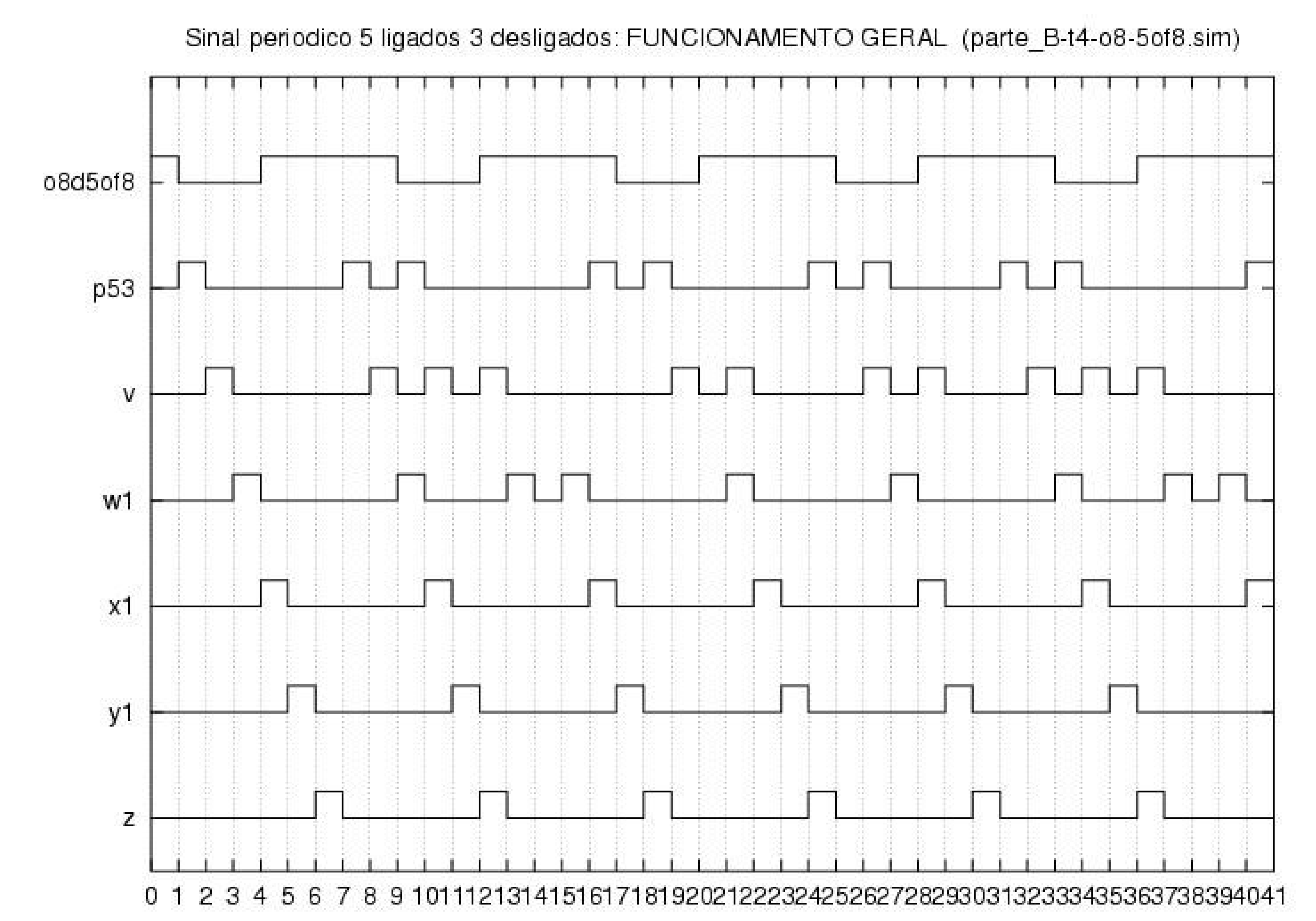

Oscilador de Período 10: FUNCIONAMENTO GERAL (parte_B-t4A-o10.sim)
o10
p53
v
w1
x1
y1
z
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41


Oscilador de Periodo 5: FUNCIONAMENTO GERAL (parte_B-t4A-o5.sim)
o5
p53
v
w1
x1
y1
z
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41

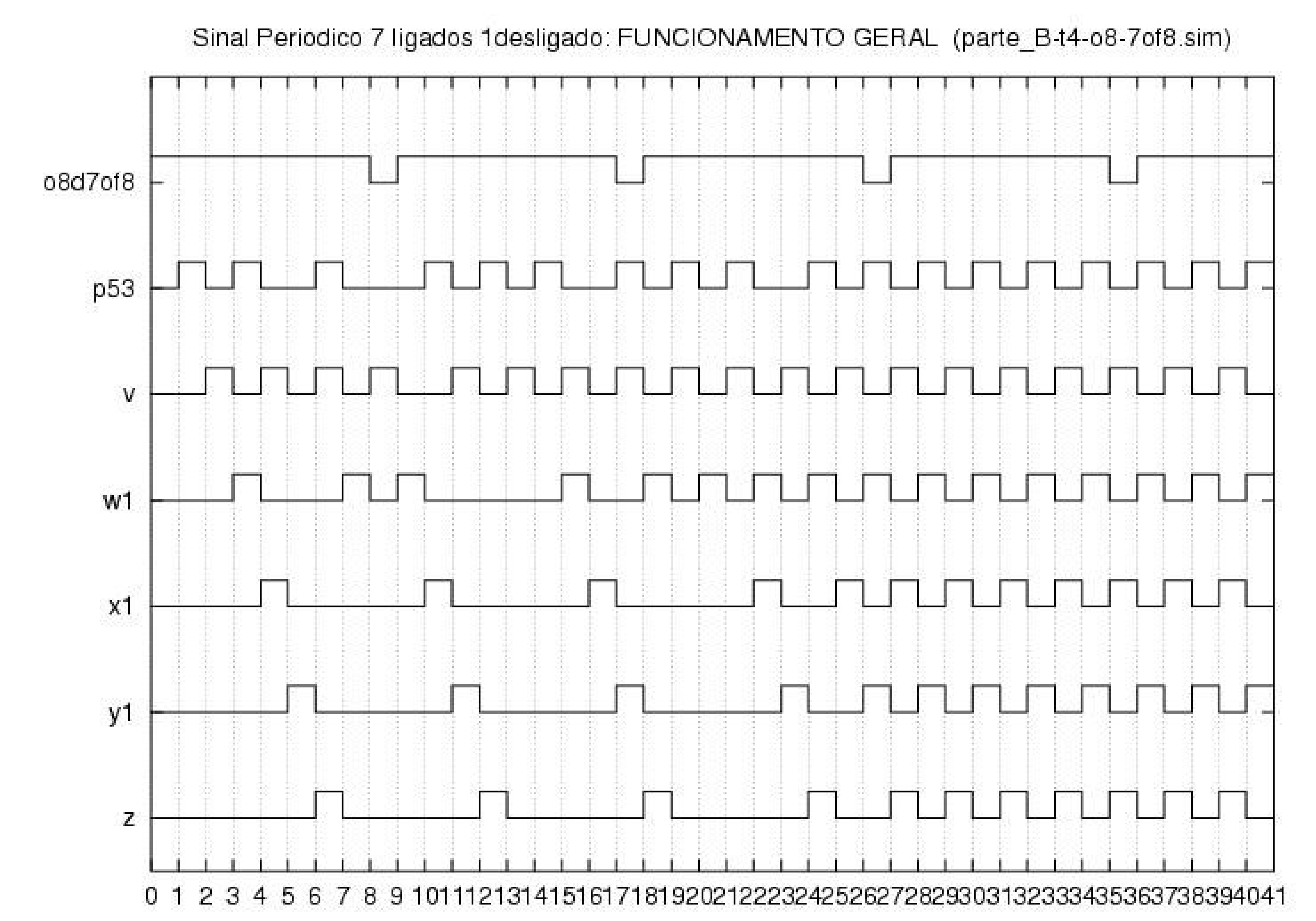

Oscilador de Periodo 3: FUNCIONAMENTO GERAL (parte_B-t4A-o3.sim)
o3
p53
v
w1
x1
y1
z
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41



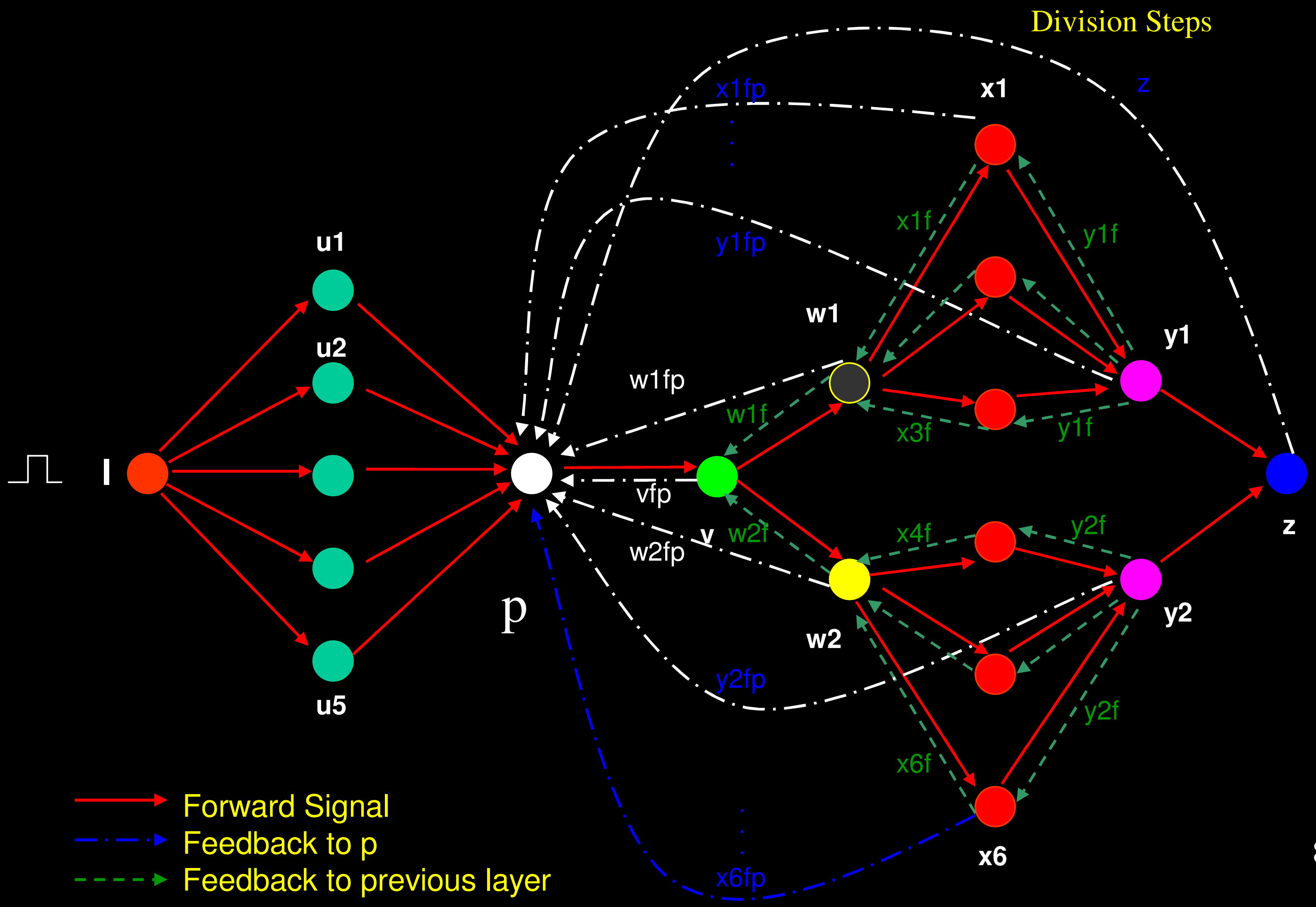
Oscilador de Periodo 2: FUNCIONAMENTO GERAL (parte_B-t4A-o2.sim)
o2
p53
v
w1
x1
y1
z
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41



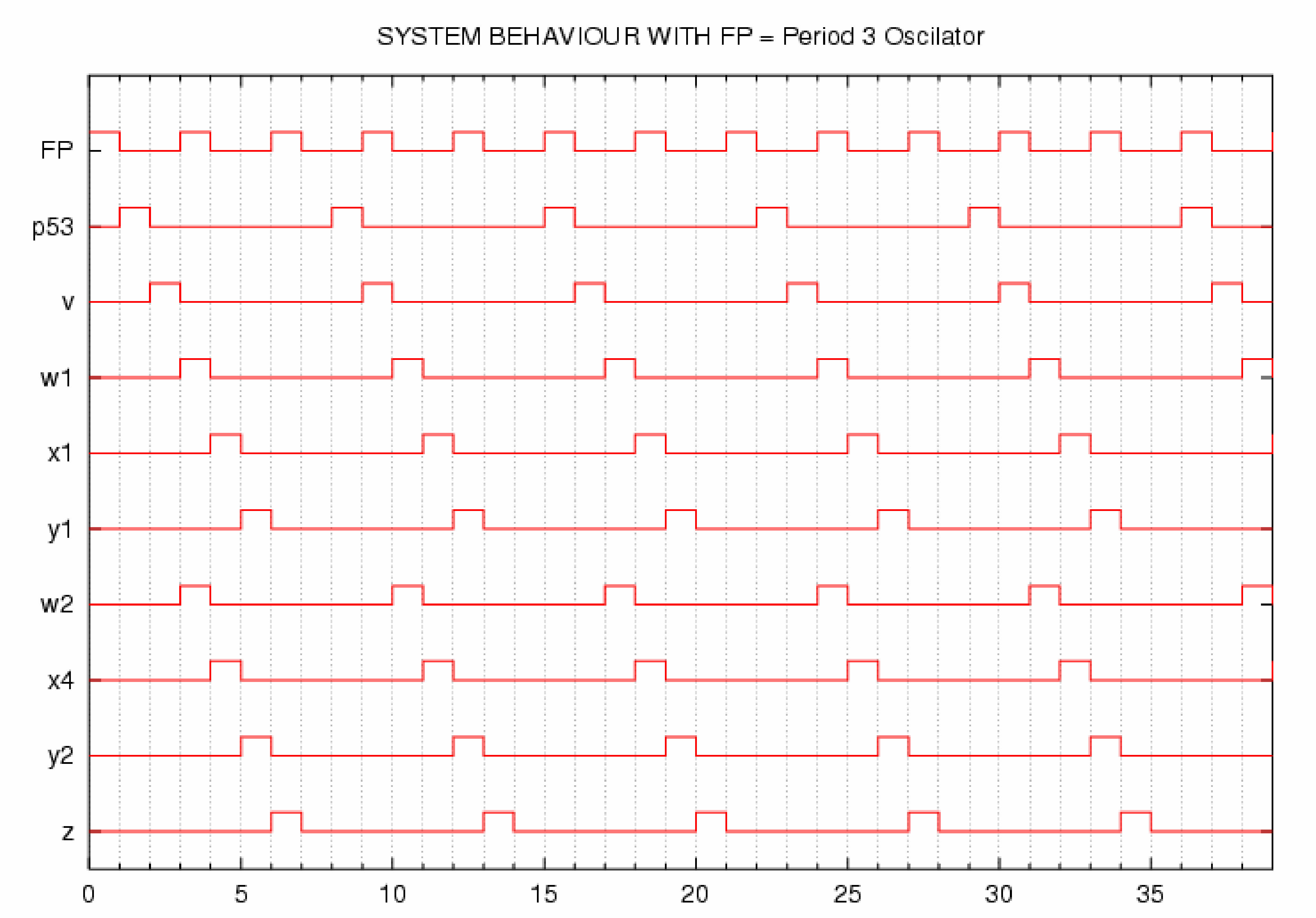
Sinal periodico 5 ligados 3 desligados: FUNCIONAMENTO GERAL (parte_B-t4-o8-5of8.sim)
o8d5of8
p53
v
w1
x1
y1
z
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41

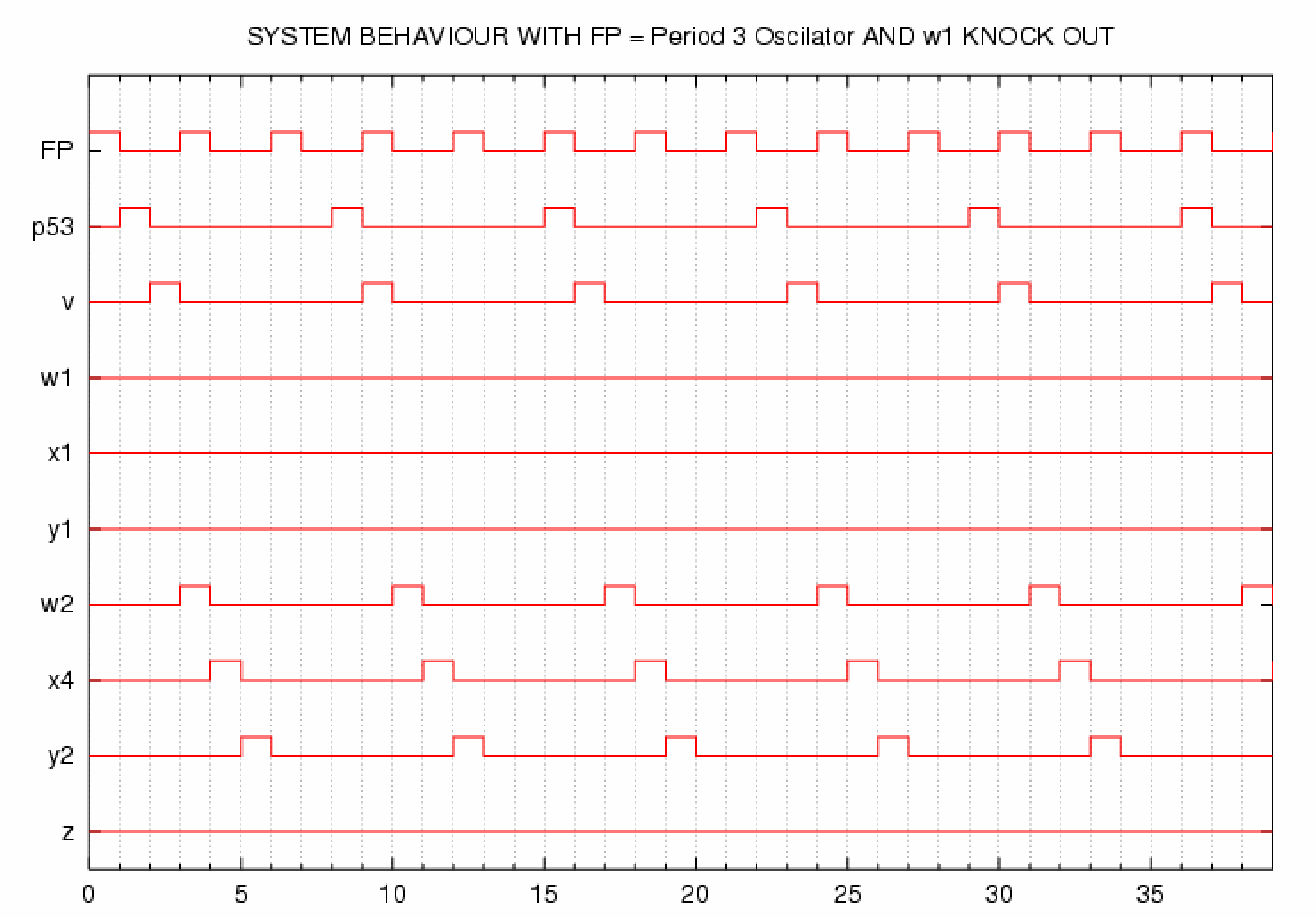

Sinal Periodico 7 ligados 1desligado: FUNCIONAMENTO GERAL (parte_B-t4-o8-7of8.sim)
o8d7of8
p53
v
w1
x1
y1
z
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41

Nockout

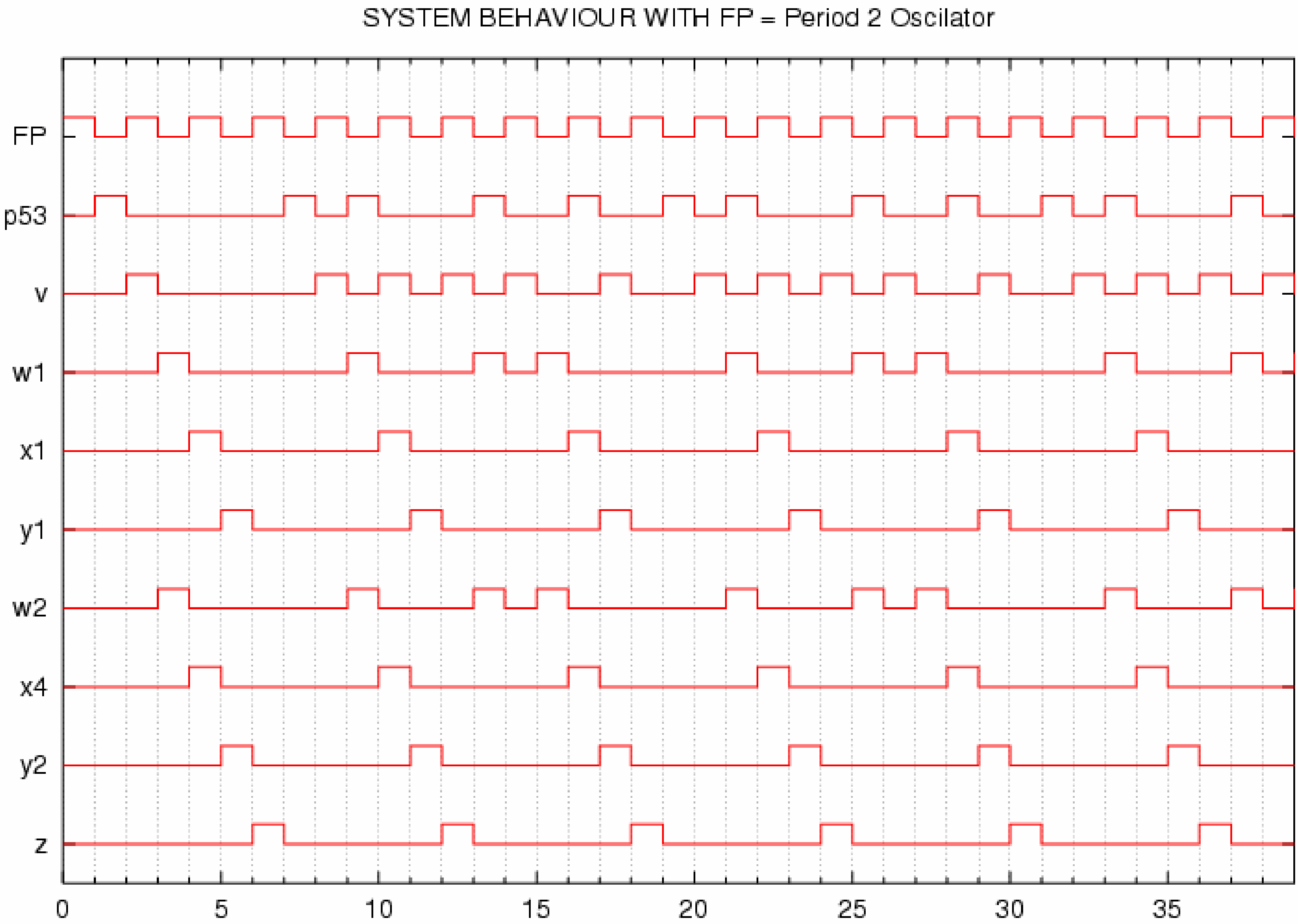


Division Steps
x1fp
z
x1
u1
y1fp
x1f
y1f
w1
y1
u2
w1fp
w1f
x3f
y1f
I
vfp
z
v
w2f
x4f
y2f
w2fp
p
w2
y2
u5
y2fp
y2f
x6f
x6
x6fp
Forward Signal
Feedback to p
Feedback to previous layer

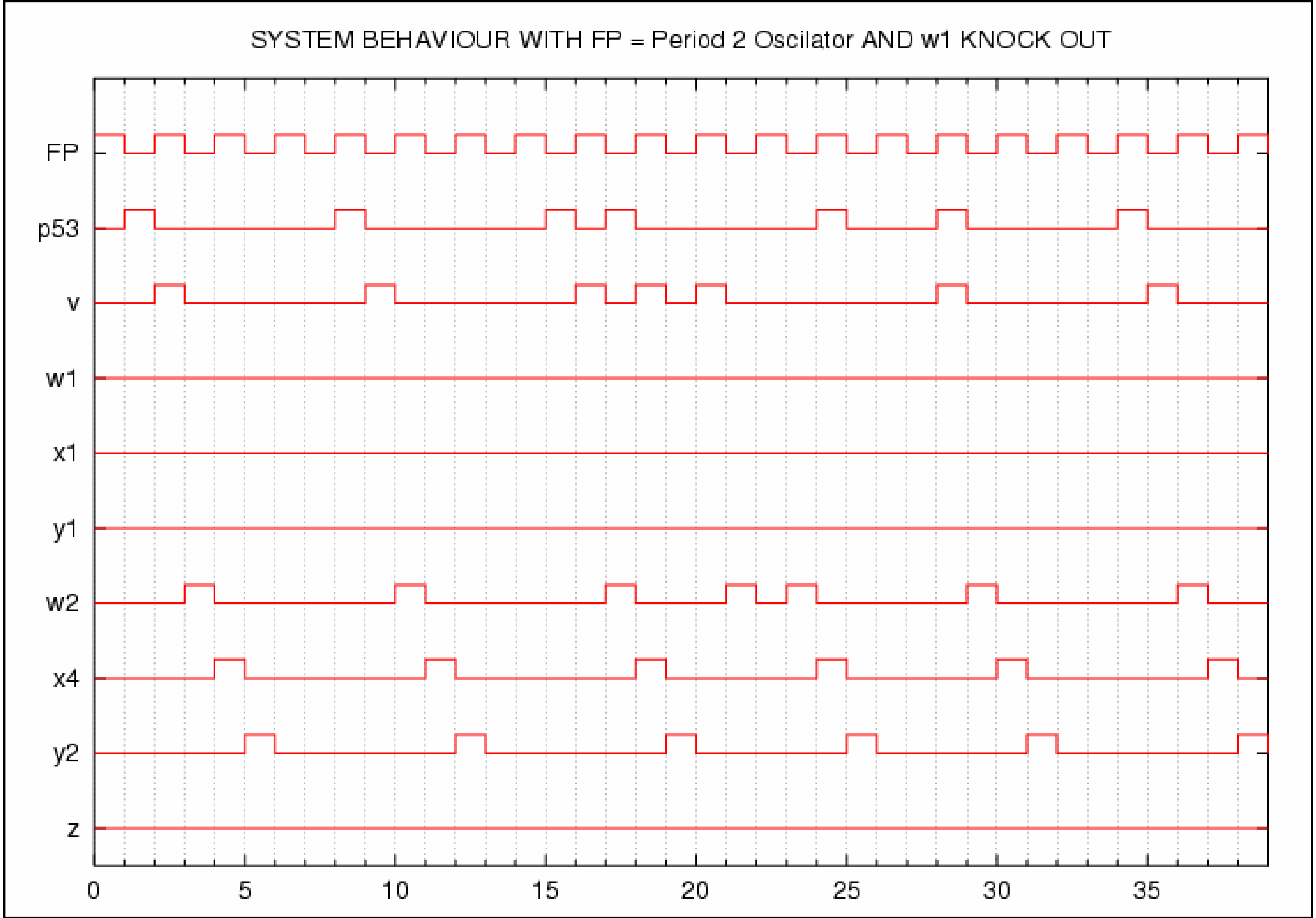

SYSTEM BEHAVIOUR WITH FP = Period 3 Oscilator
FP
p53
v
w1
x1
y1
w2
x4
y2
z
0
5
10
15
20
25
30
35



SYSTEM BEHAVIOUR WITH FP = Period 3 Oscilator AND w1 KNOCK OUT
FP
p53
v
w1
x1
y1
w2
x4
y2
z
0
5
10
15
20
25
30
35



SYSTEM BEHAVIOUR WITH FP = Period 2 Oscilator
FP
p53
v
w1
x1
y1
w2
x4
y2
z
0
5
10
15
20
25
30
35


SYSTEM BEHAVIOUR WITH FP = Period 2 Oscilator AND w1 KNOCK OUT
FP
p53
v
w1
x1
y1
w2
x4
y2
z
0
5
10
15
20
25
30
35

Biological Model

- Cell cycle control by Fibroblast Growth Factor 2 (FGF2) and Adrenocorticotropic Hormone (ACTH) in the Y1 adrenocortical cell line
- FGF2 has long been considered a candidate for participating in cell cycle control, but its molecular mechanisms remain obscure

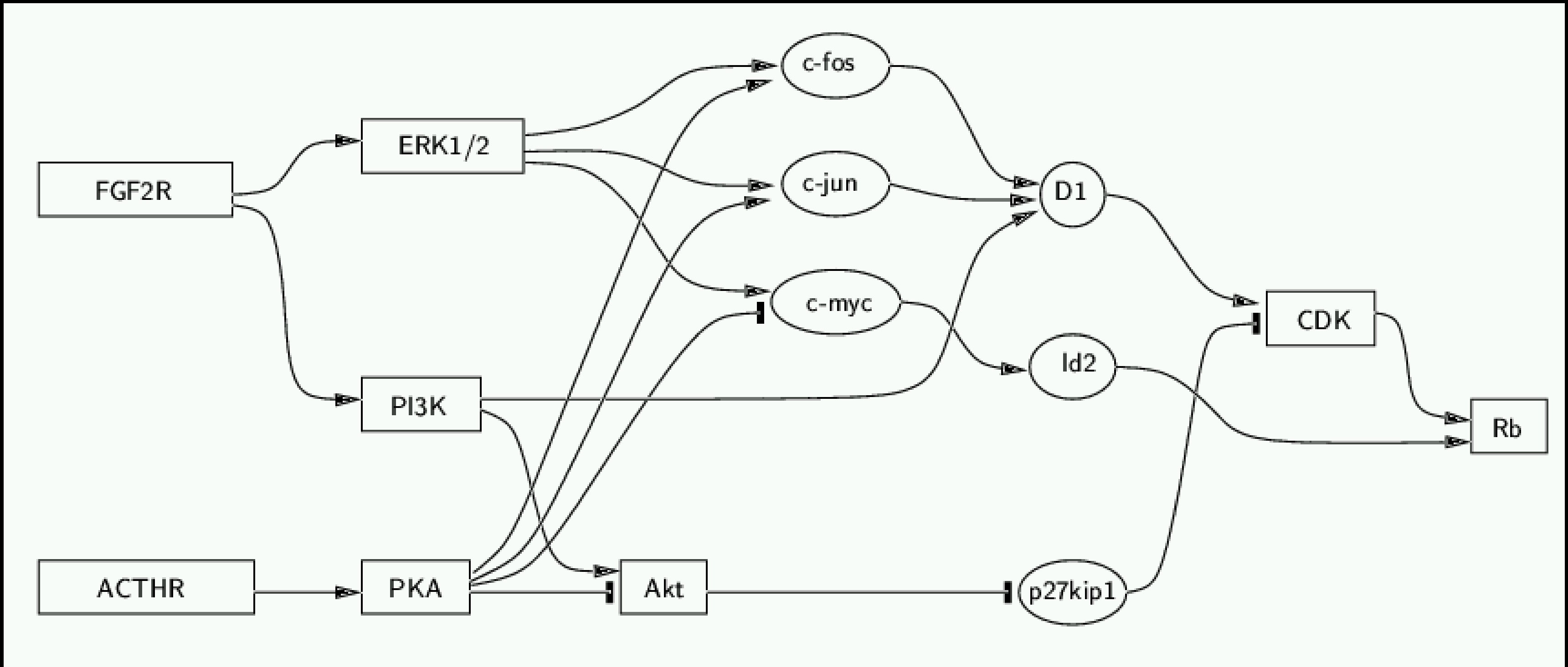
Mitogenic response in G0/G1 cell cycle to FGF2:

- Rapid and transient activation of extra cellular signal-regulated kinases
- Transcription activation of c-fos, c-jun and c-myc genes
- Induction of c-Fos and c-Myc proteins and cyclin D1 protein
- DNA synthesis stimulation

Anti-mitogenic response in G0/G1 cell cycle to ACTH:

- Blocks FGF2 mitogenic response
- Keeps ERK activation and c-Fos and cyclin D1 induction on
- Down regulates the levels of the c-Myc protein
- Down regulates the active form of Akt/PKB enzyme

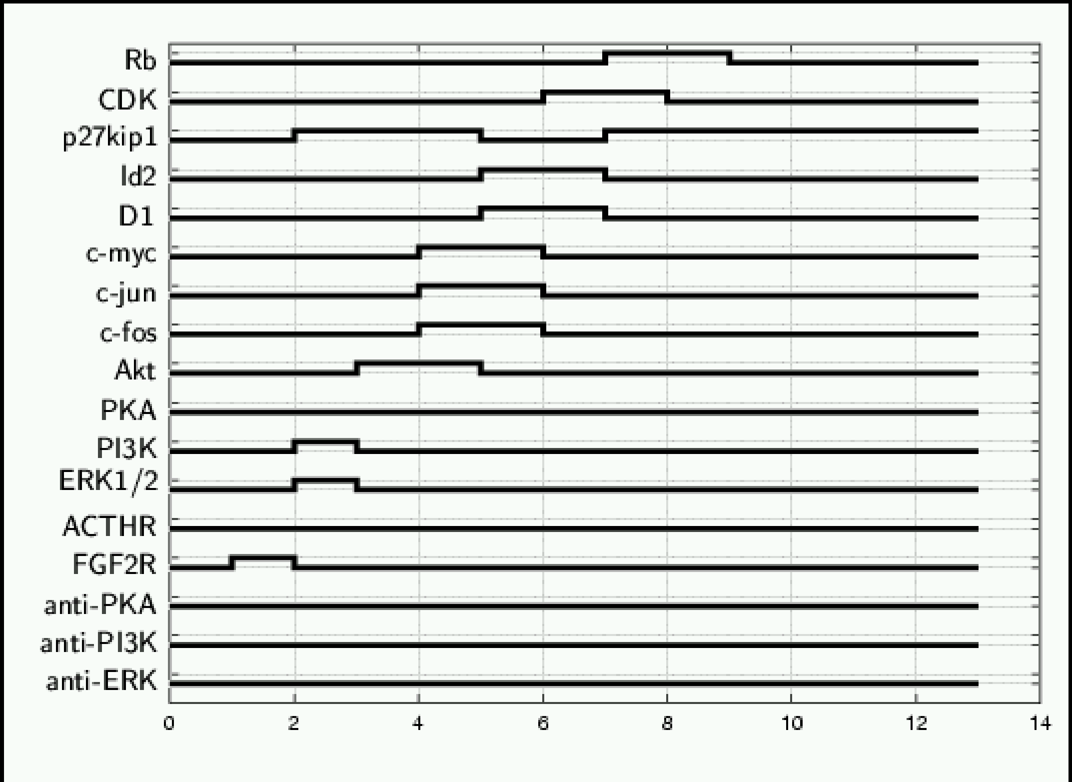
A model for the FGF2/ACTH influence on cell cycle



Model Formalization

Element	Rule
$\mathrm{FGF2R}[i]$	receives external signal
$\mathrm{ACTHR}[i]$	receives external signal
$\mathrm{ERK1/2}[i]$	$\mathrm{FGF2R}[i-1] \cdot \overline{\text{anti-ERK}[i-1]}$
$\mathrm{PI3K}[i]$	$\mathrm{FGF2R}[i-1] \cdot \overline{\text{anti-PI3K}[i-1]}$
$\mathrm{PKA}[i]$	$\mathrm{ACTHR}[i-1] \cdot \overline{\text{anti-PKA}[i-1]}$
$\mathrm{Akt}[i]$	$(\mathrm{PI3K}[i-2] + \mathrm{PI3K}[i-1]) \cdot \overline{\mathrm{PKA}[i-2]} \cdot \overline{\mathrm{PKA}[i-1]}$
$\text{c-fos}[i]$	$\mathrm{ERK1/2}[i-3] + \mathrm{ERK1/2}[i-2] + \mathrm{PKA}[i-3] + \mathrm{PKA}[i-2]$
$\text{c-jun}[i]$	$\mathrm{ERK1/2}[i-3] + \mathrm{ERK1/2}[i-2] + \mathrm{PKA}[i-3] + \mathrm{PKA}[i-2]$
$\text{c-myc}[i]$	$\overline{\mathrm{PKA}[i-2]} \cdot \overline{\mathrm{PKA}[i-1]} \cdot (\mathrm{ERK1/2}[i-3] + \mathrm{ERK1/2}[i-2])$
$\mathrm{D1}[i]$	$\text{c-fos}[i-1] \cdot \text{c-jun}[i-1] \cdot (\mathrm{PI3K}[i-4] - \mathrm{PI3K}[i-3])$
$\mathrm{Id2}[i]$	$\text{c-myc}[i-1]$
$\mathrm{p27kip1}[i]$	$\overline{\mathrm{Akt}[i-2]}$
$\mathrm{CDK}[i]$	$\mathrm{D1}[i-1] \cdot \overline{\mathrm{p27kip1}[i-1]}$
$\mathrm{Rb}[i]$	$\mathrm{CDK}[i-1] \cdot \mathrm{Id2}[i-2]$

Effect of one pulse of FGF2



Effect of one pulse of ACTH