DNA HOLDS THE KEY TO COGNITION

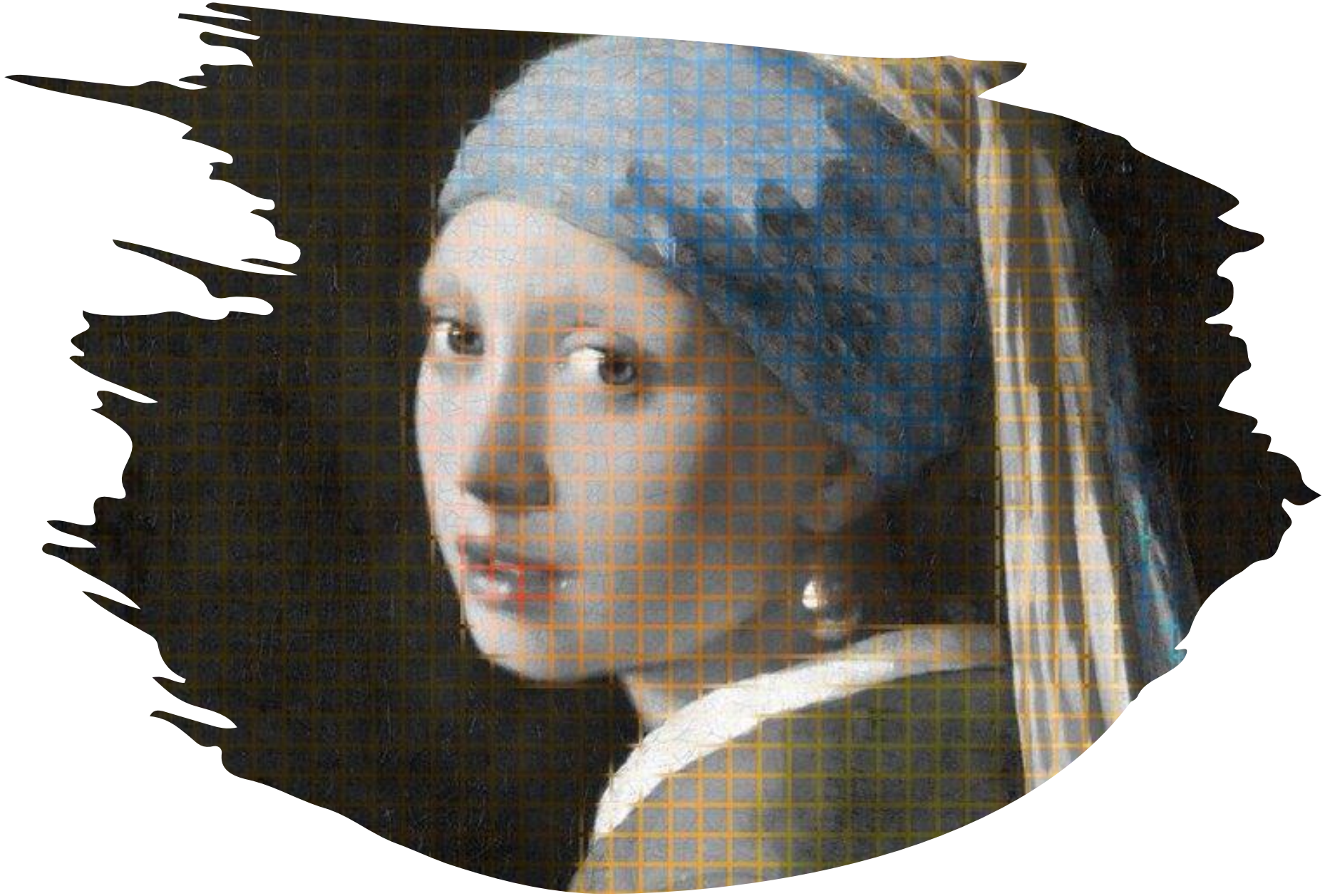
1. INTRODUCTION

GEN EXPRESSION
(BOOLEAN NETWORK)

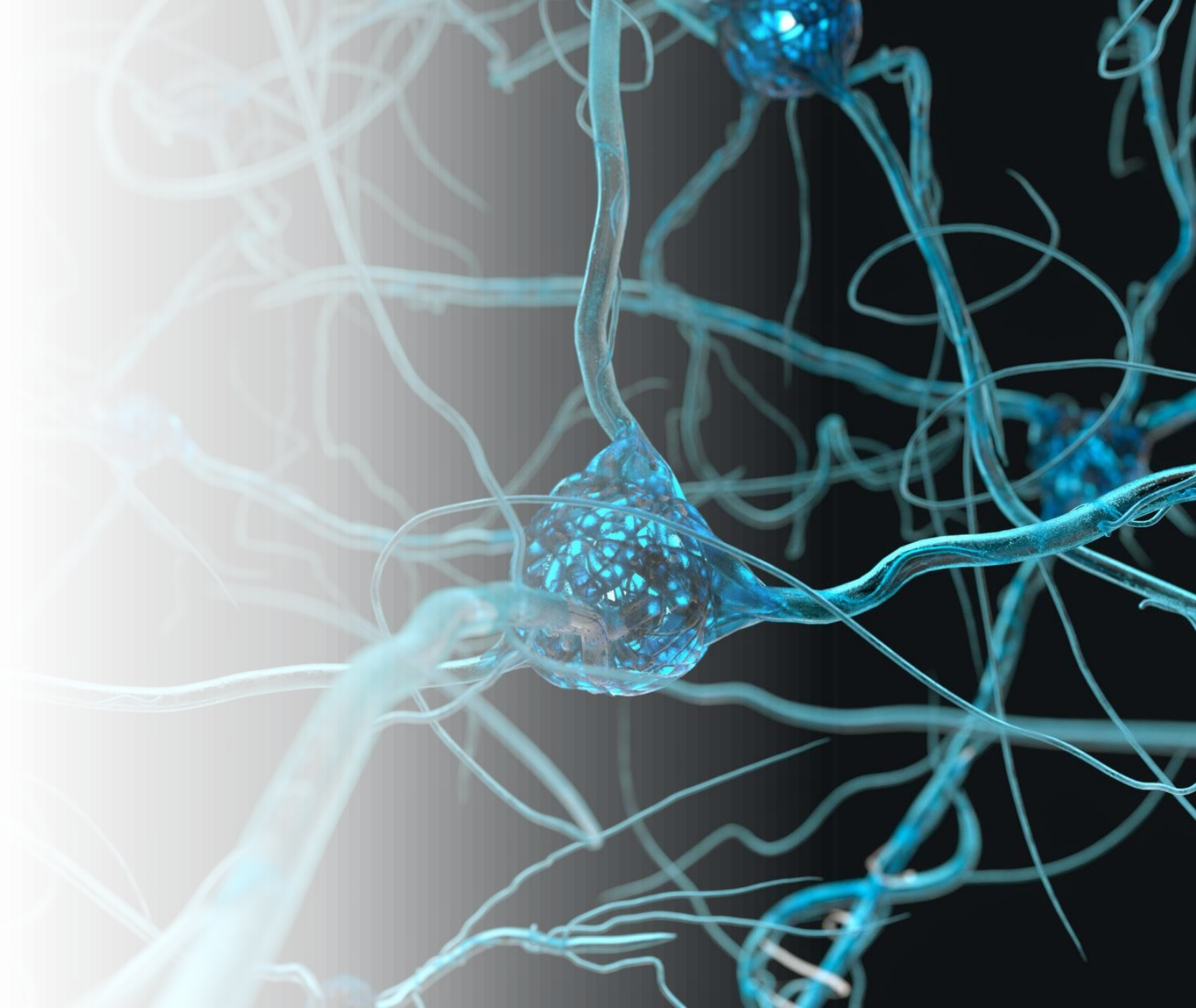
SIGNAL PROCESSING
(EEG & EMG)

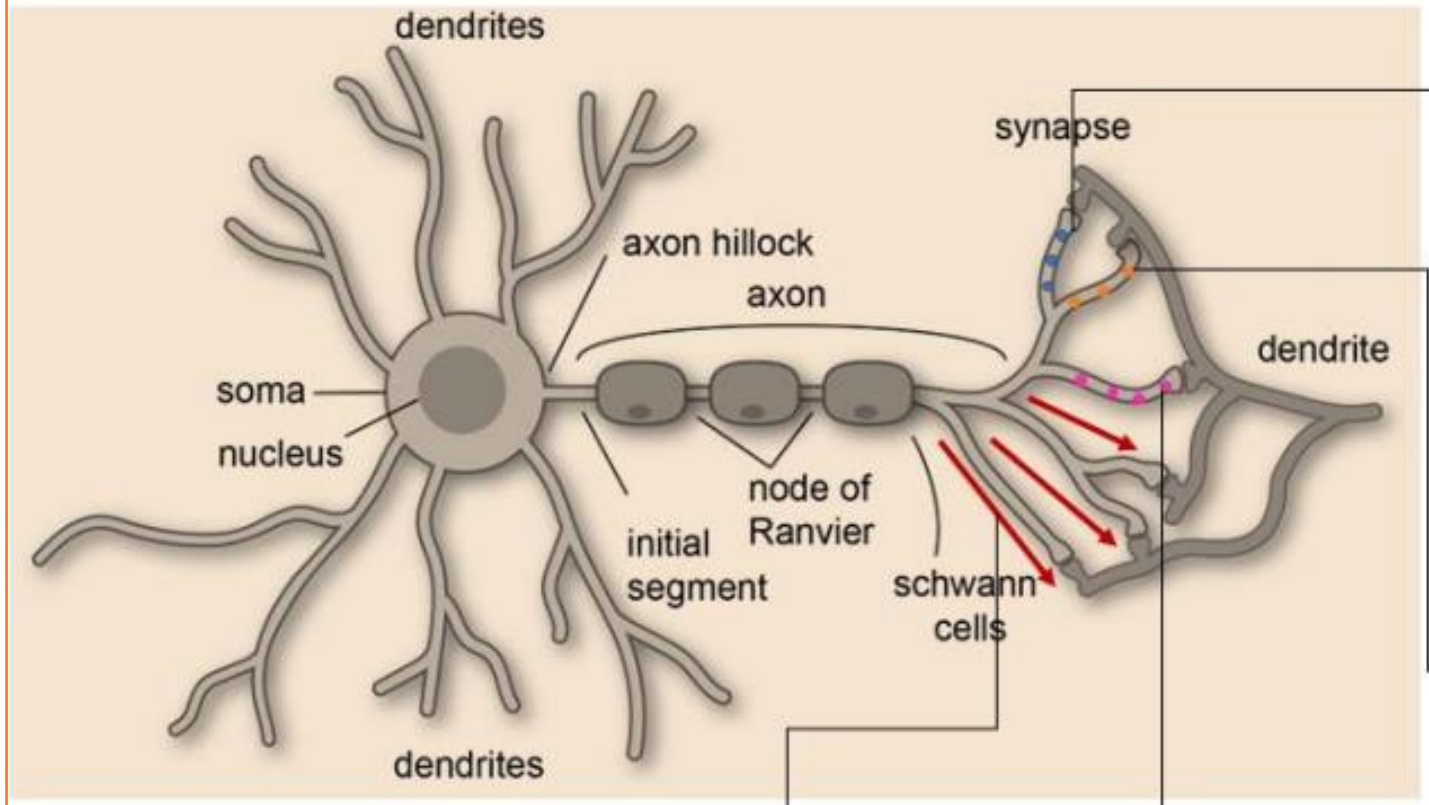
2. Problem identification

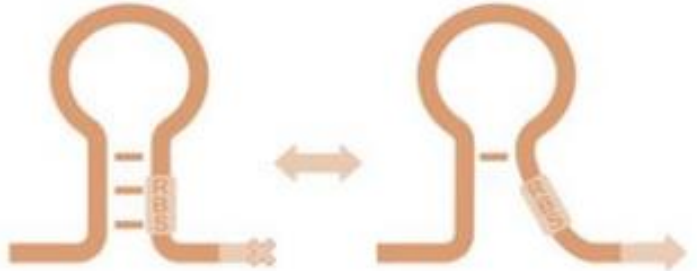
- Unconditional reflex
- Experiment
- Artifical neural network contradiction
- What is algorithsm and its origin?



- Instructions for forming connections in the brain in an organized manner.
- Products derived from DNA directly affect the electrical current in neurons.
- DNA acts as a database and information is extracted by interactions between DNA and proteins – microbial memory.

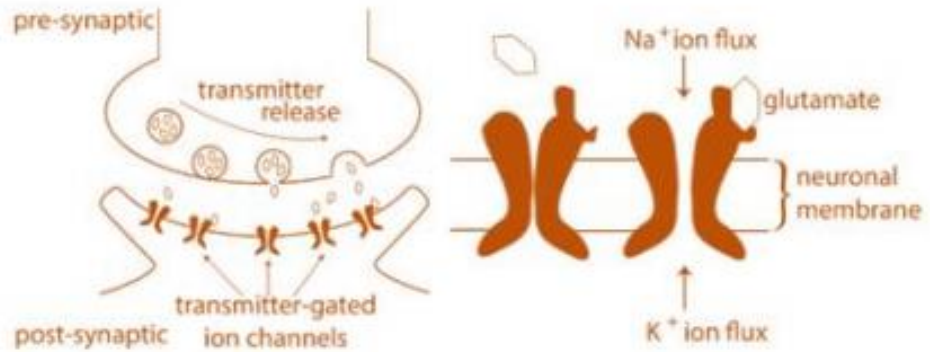


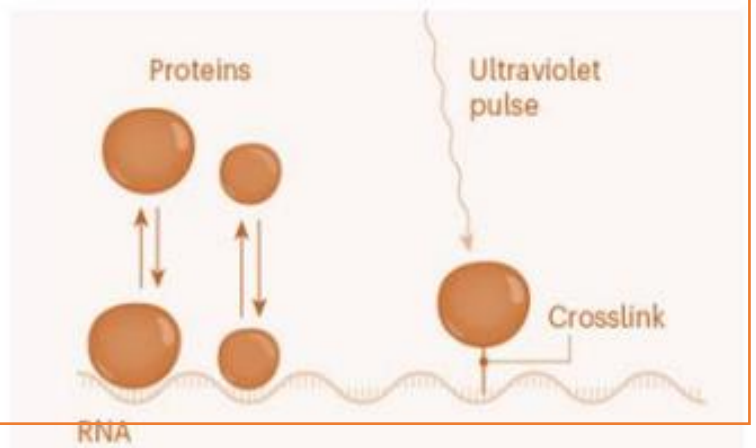

dendrites
synapse
axon hillock
axon
dendrite
soma
nucleus
node of
Ranvier
initial
segment
schwann
cells
dendrites



A

B


pre-synaptic
transmitter
release
transmitter-gated
ion channels
post-synaptic
Na+ ion flux
glutamate
neuronal
membrane
K+ ion flux

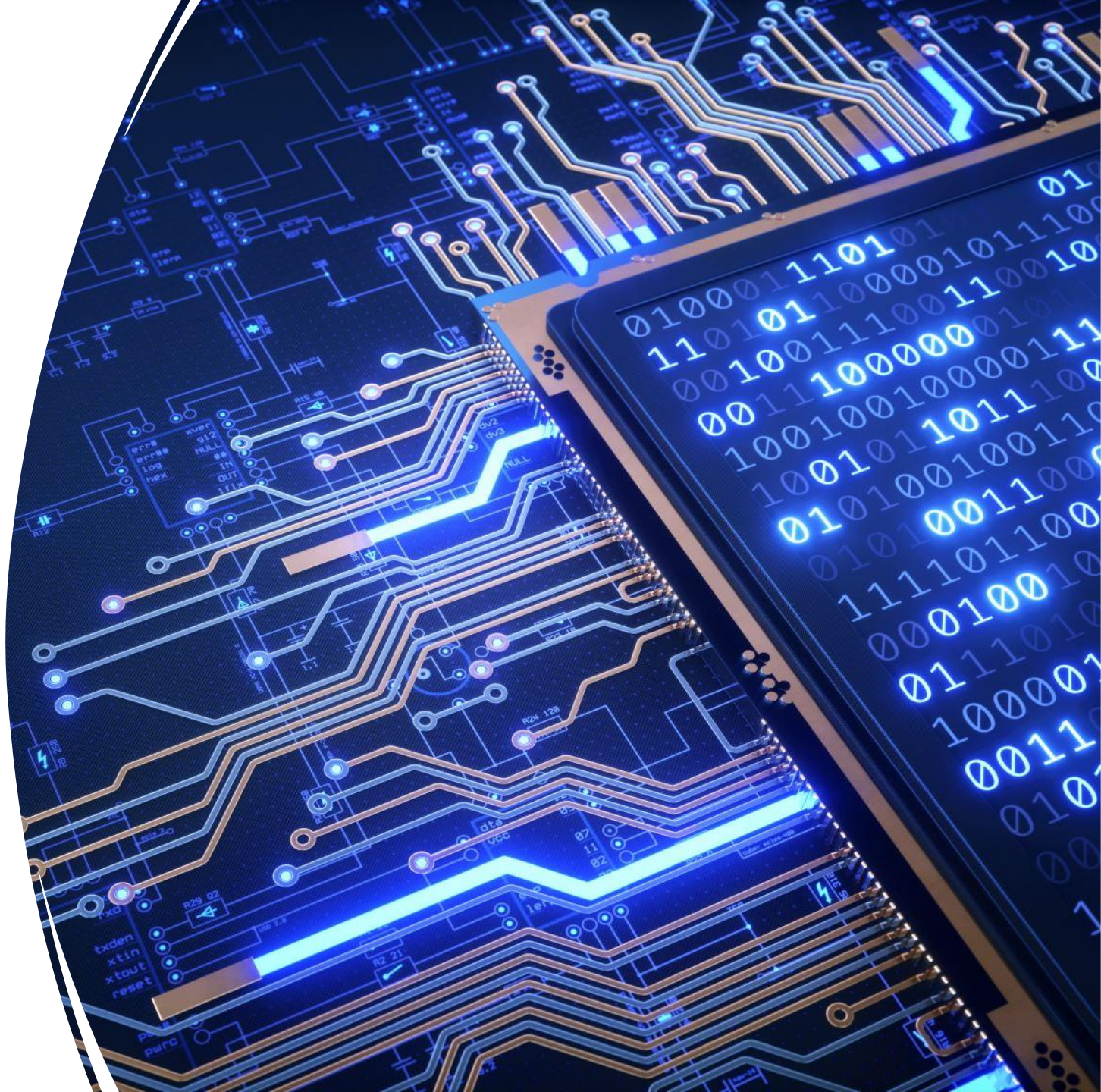

Proteins
Ultraviolet
pulse
Crosslink
RNA

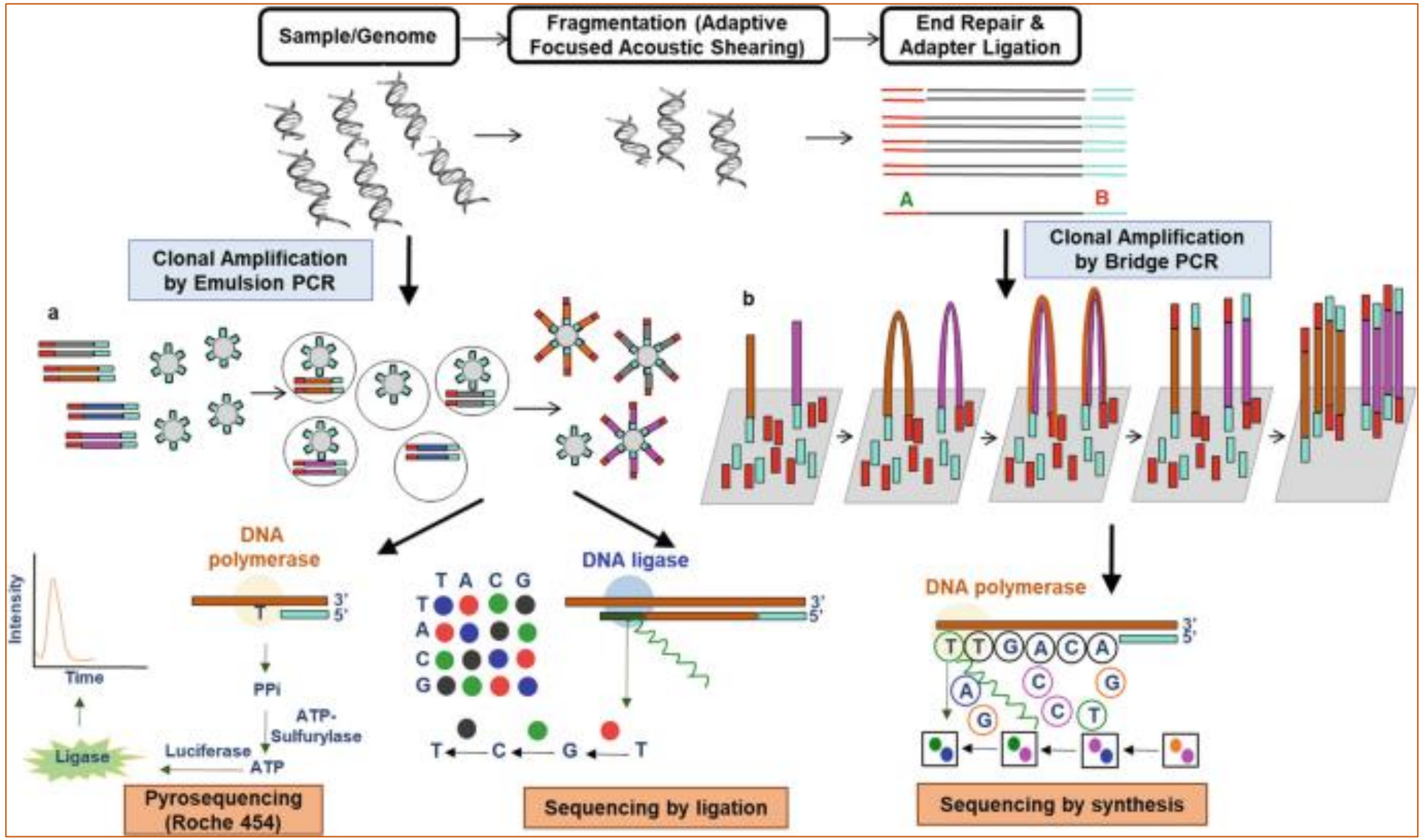
Potential applications

- Artificial intelligence
- Data storage
- Robotics
- Quantum
- Brain – computer interface

Reverse engineering

- Step 1: Identify the principal components and their fundamental functions.
- Step 2: Study the arrangement between fundamental functions.
- Step 3: Simulation, digital twin and validate the hypothesis with experiments.



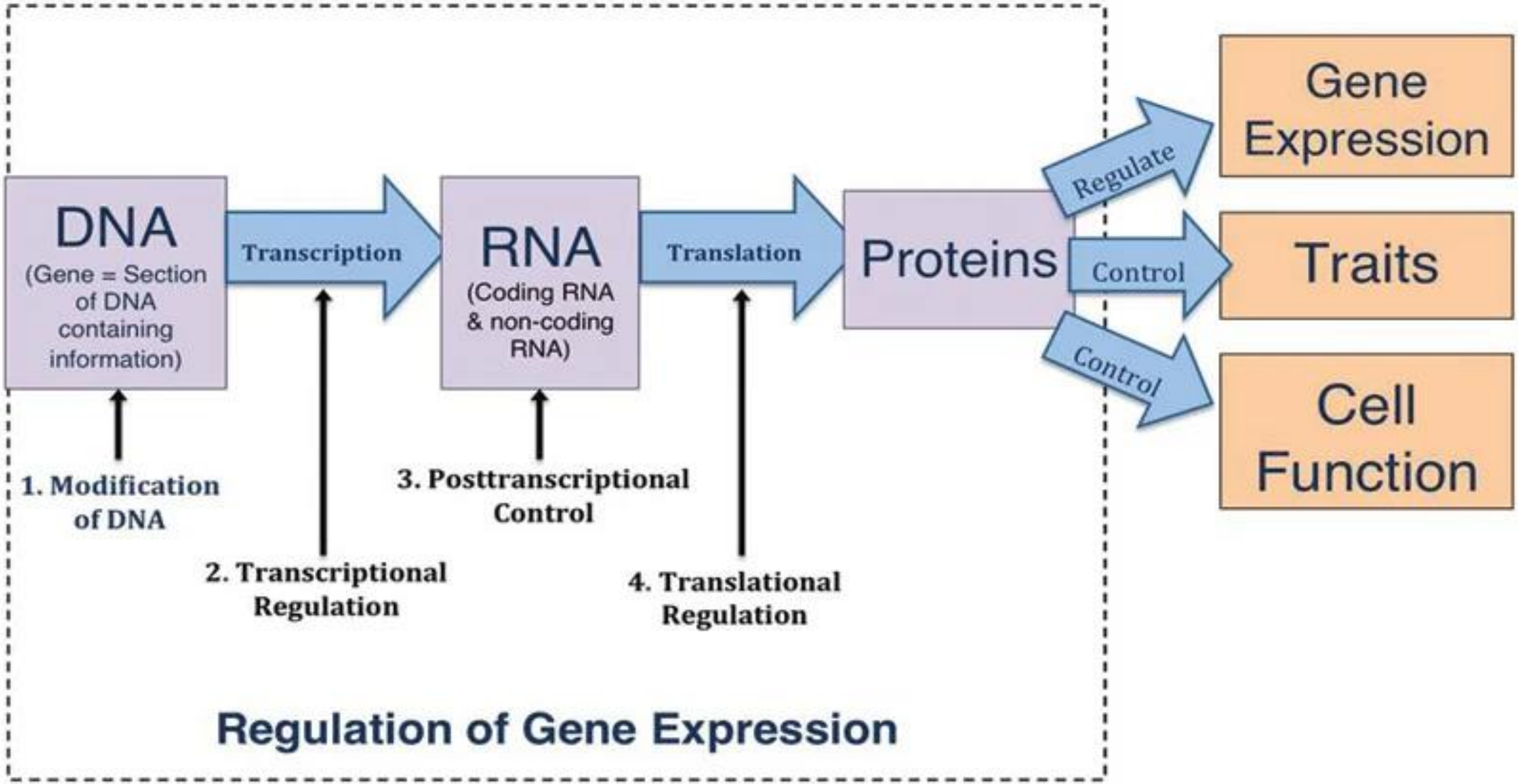

Sample/Genome
Fragmentation (Adaptive Focused Acoustic Shearing)
End Repair & Adapter Ligation
A
B
Clonal Amplification by Emulsion PCR
Clonal Amplification by Bridge PCR
a
b
DNA polymerase
DNA ligase
DNA polymerase
Intensity
Time
PPi
ATP-Sulfurylase
Luciferase
ATP
Ligase
T A C G
Pyrosequencing (Roche 454)
Sequencing by ligation
Sequencing by synthesis

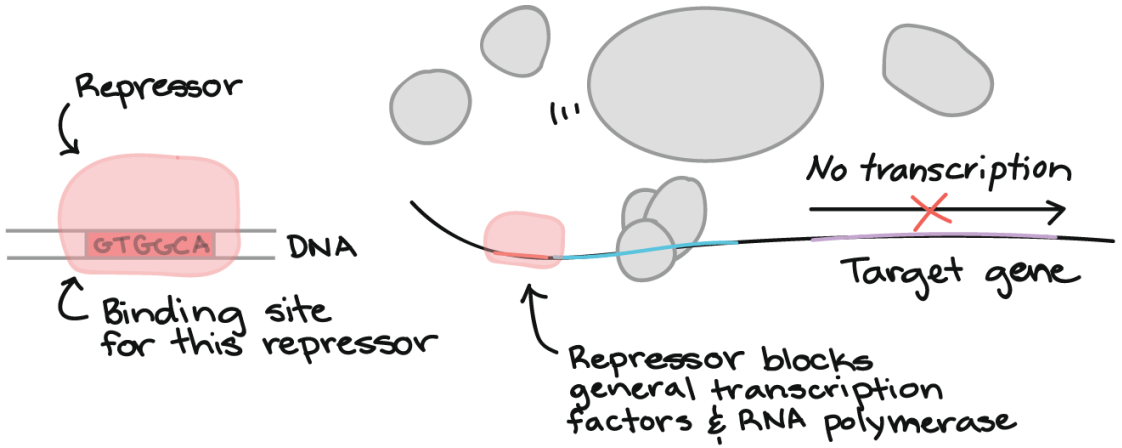
$$9 \quad 9 \quad 9 \quad 9 = 81$$

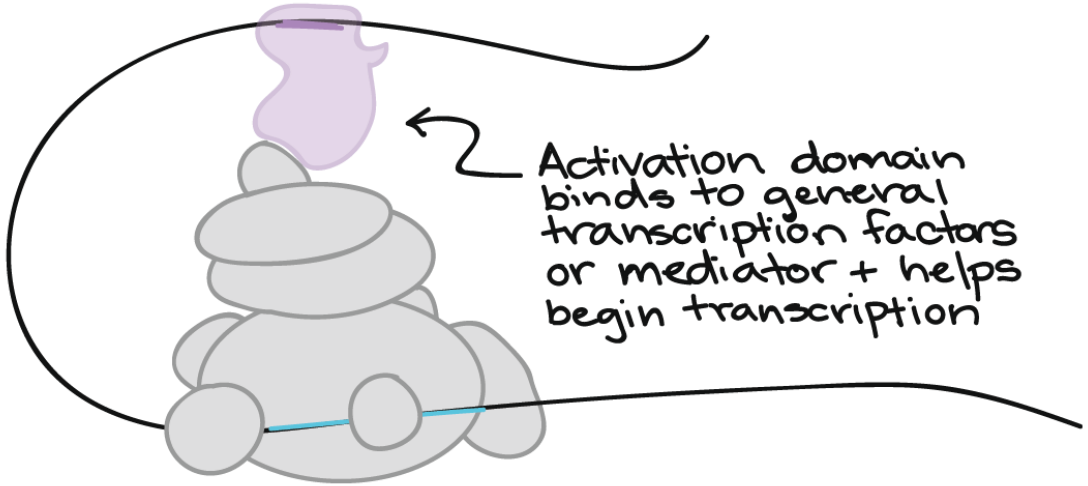
$$\{\times, +, -\}, \{+, -, \times\}$$

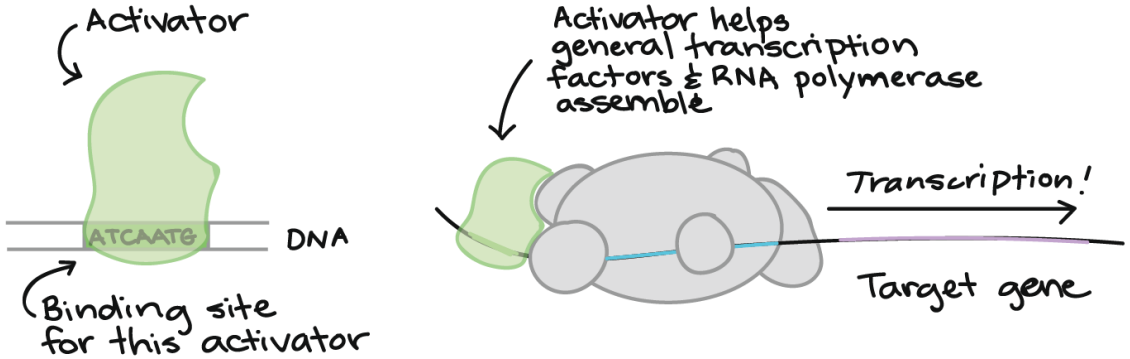
and

$$\{\times, \times, \div\}, \{\times, \div, \times\}, \{\div, \times, \times\}$$


DNA
(Gene = Section of DNA containing information)
Transcription
RNA
(Coding RNA & non-coding RNA)
Translation
Proteins
Regulate
Control
Control
Gene Expression
Traits
Cell Function
1. Modification of DNA
2. Transcriptional Regulation
3. Posttranscriptional Control
4. Translational Regulation
Regulation of Gene Expression

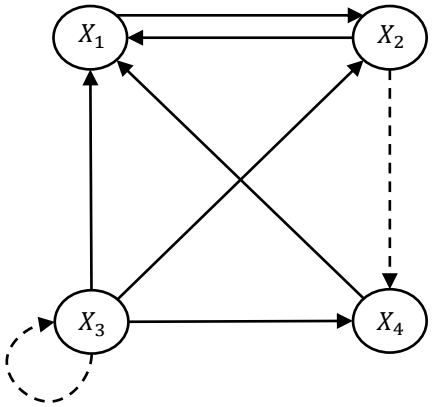

Repressor
GTGGCA
DNA
Binding site for this repressor
No transcription
Target gene
Repressor blocks general transcription factors & RNA polymerase


Activation domain binds to general transcription factors or mediator + helps begin transcription


Activator
ATCAATG
DNA
Binding site for this activator

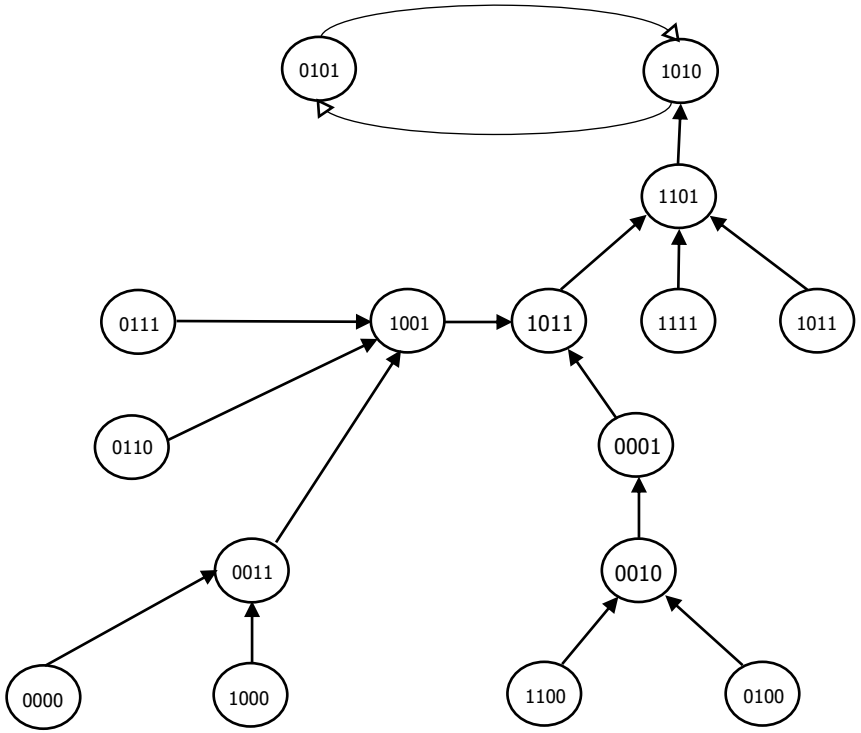
Activator helps general transcription factors & RNA polymerase assemble
Transcription!
Target gene

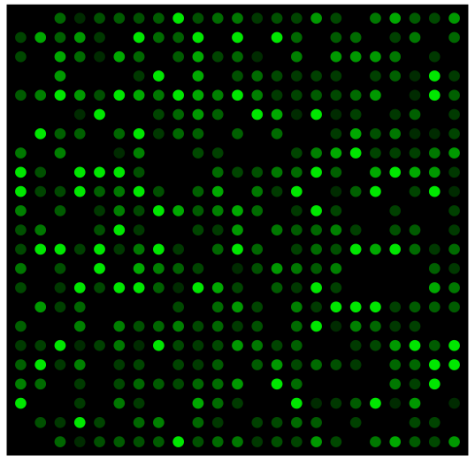
3. Boolean network model

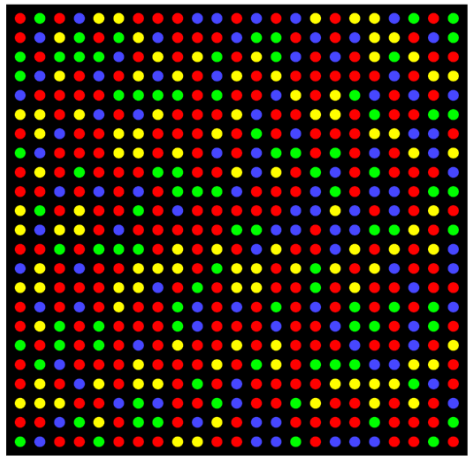


$$x_1 = (x_2 \wedge x_3) \vee x_4$$
$$x_2 = x_1 \wedge x_3$$
$$x_3 = \overline{x_3}$$
$$x_4 = \overline{x_2} \vee x_3$$

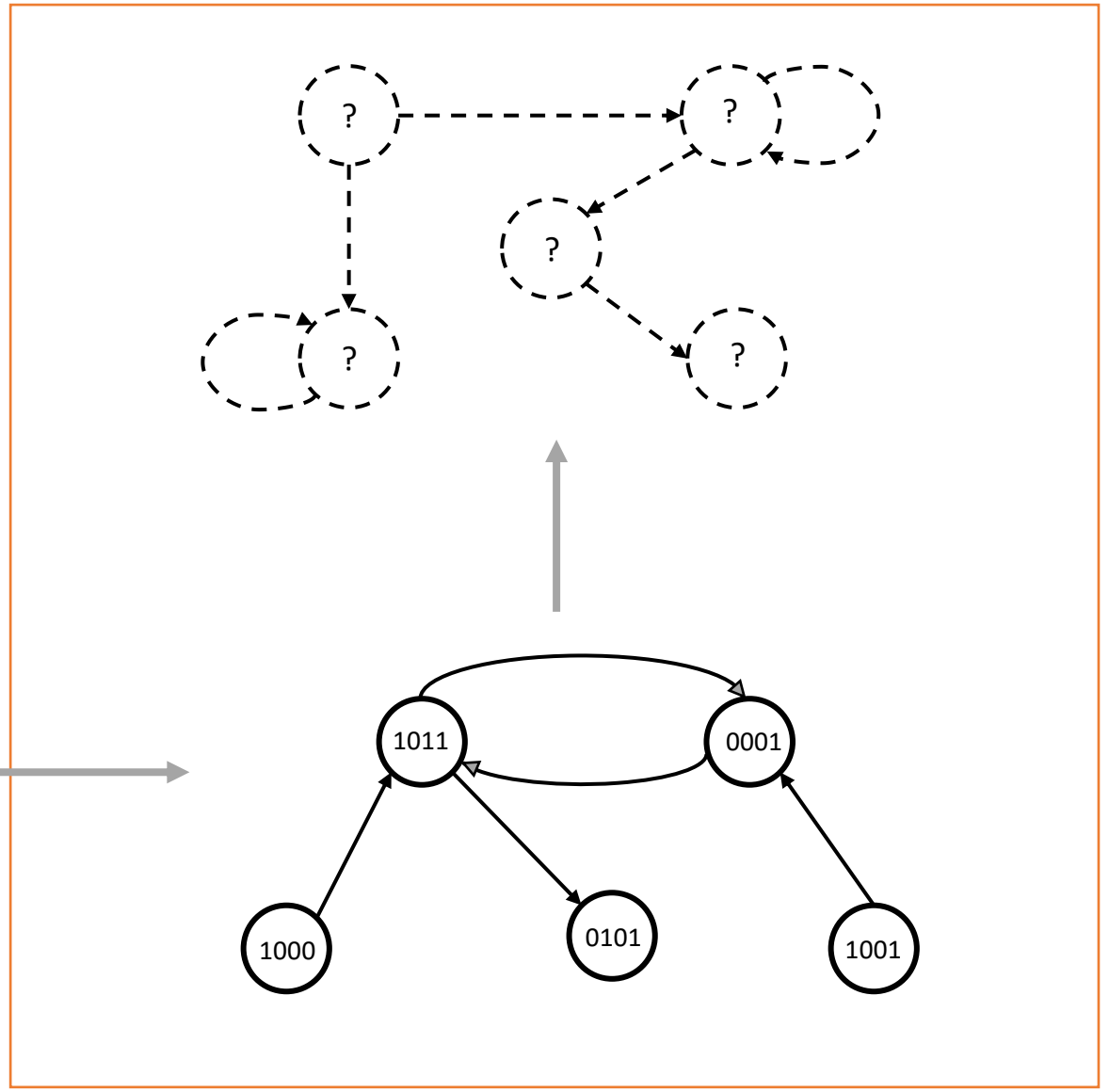
t				$t+1$			
x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4
0	0	0	0	0	0	1	1
0	0	0	1	1	0	1	1
0	0	1	0	0	0	0	1
0	0	1	1	1	0	0	1
0	1	0	0	0	0	1	0
0	1	0	1	1	0	1	0
0	1	1	0	1	0	0	1
0	1	1	1	1	0	0	1
1	0	0	0	0	0	1	1
1	0	0	1	1	0	1	1
1	0	1	0	0	1	0	1
1	0	1	1	1	1	0	1
1	1	0	0	0	0	1	0
1	1	0	1	1	0	1	0
1	1	1	0	1	1	0	1
1	1	1	1	1	1	0	1

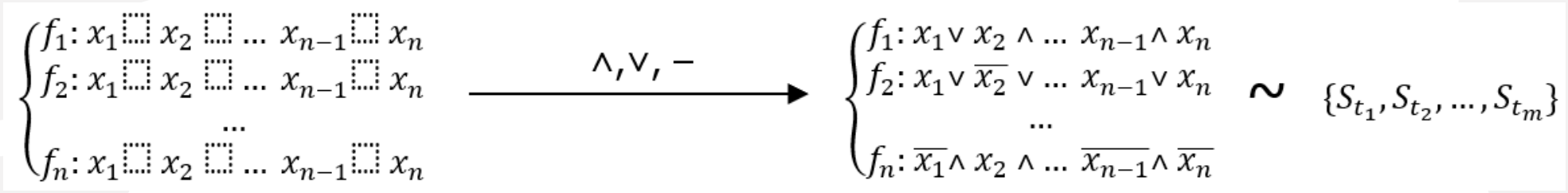






	Sample 1 (T1)	Sample 2 (T2)	...	Sample N (TN)
Gene 1	1	0	...	1
Gene 2	1	0	...	0
Gene 3	0	1	...	0
...	...	...	...	...
Gene M	1	0	...	1




$\begin{cases} f_1: x_1 \square x_2 \square \dots x_{n-1} \square x_n \\ f_2: x_1 \square x_2 \square \dots x_{n-1} \square x_n \\ \dots \\ f_n: x_1 \square x_2 \square \dots x_{n-1} \square x_n \end{cases}$
$\wedge, \vee, -$
$\begin{cases} f_1: x_1 \vee x_2 \wedge \dots x_{n-1} \wedge x_n \\ f_2: x_1 \vee \overline{x_2} \vee \dots x_{n-1} \vee x_n \\ \dots \\ f_n: \overline{x_1} \wedge x_2 \wedge \dots \overline{x_{n-1}} \wedge \overline{x_n} \end{cases}$
$\sim$
$\{S_{t_1}, S_{t_2}, \dots, S_{t_m}\}$

ITERATION OVER SEARCHING SPACE

Algorithm 1: *Inference of the Boolean network*

Input: *N, K, D, d, operation set, gene set V*

Output: *the prediction function set F*

1. Create tuples to store result;

$G \leftarrow \emptyset$;

$R \leftarrow \emptyset$;

$Result \leftarrow \emptyset$;

2. Create all possible Boolean function of gen x_i

For *all* x_i *in* V *:*

$L_i \leftarrow \emptyset$;

For $j = 0$ *to* K*:*

Create all possible Boolean function have j *variables* $f_j^{(i)}$;

$L_i \leftarrow L_i \cup f_j^{(i)}$;

$G \leftarrow G \cup L_i$;

3. Create all possible realizations

For *all* L_i *in* G*:*

Create all possible unique realizations r_i;

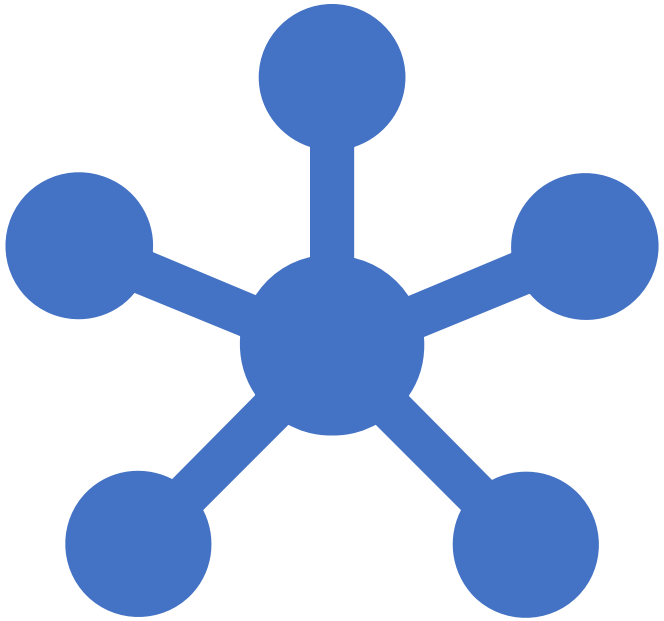
$R \leftarrow R \cup r_i$;

4. Compare

For *all* r_i *in* R *:*

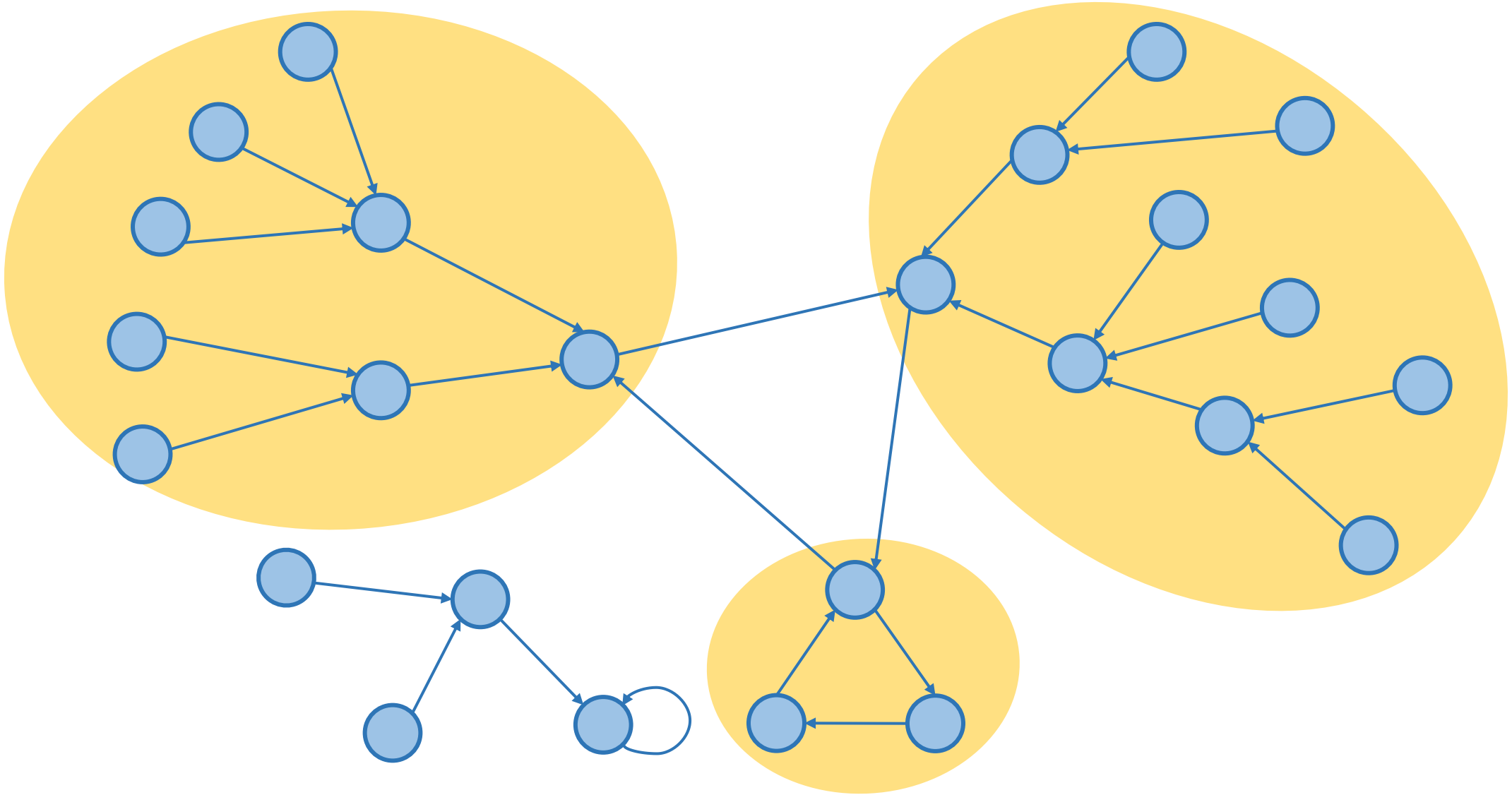
If r_i *compatible with all* $(X_i, Y_i) \in D$: $Result \leftarrow Result \cup r_i$;

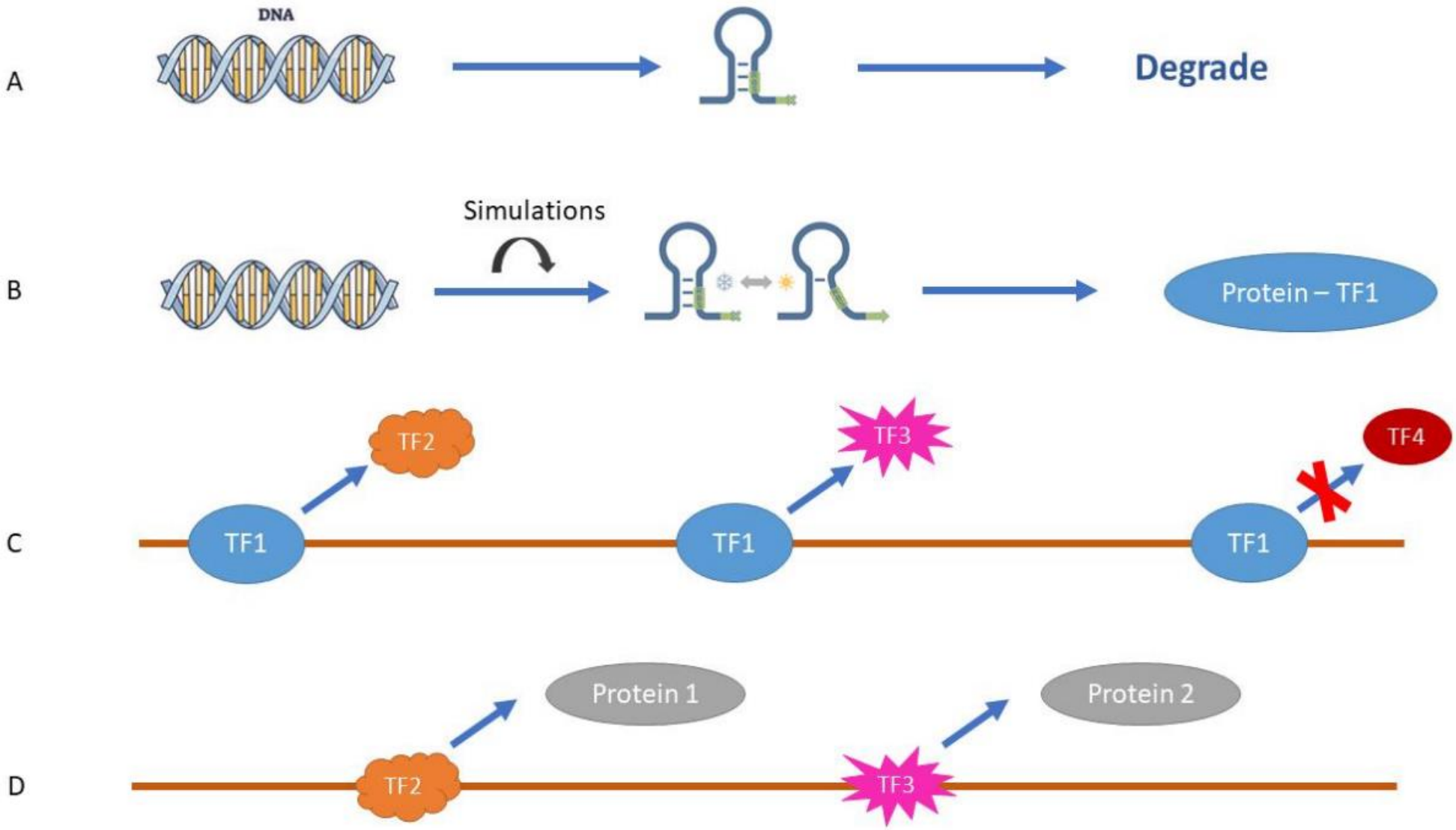
Return *Result*

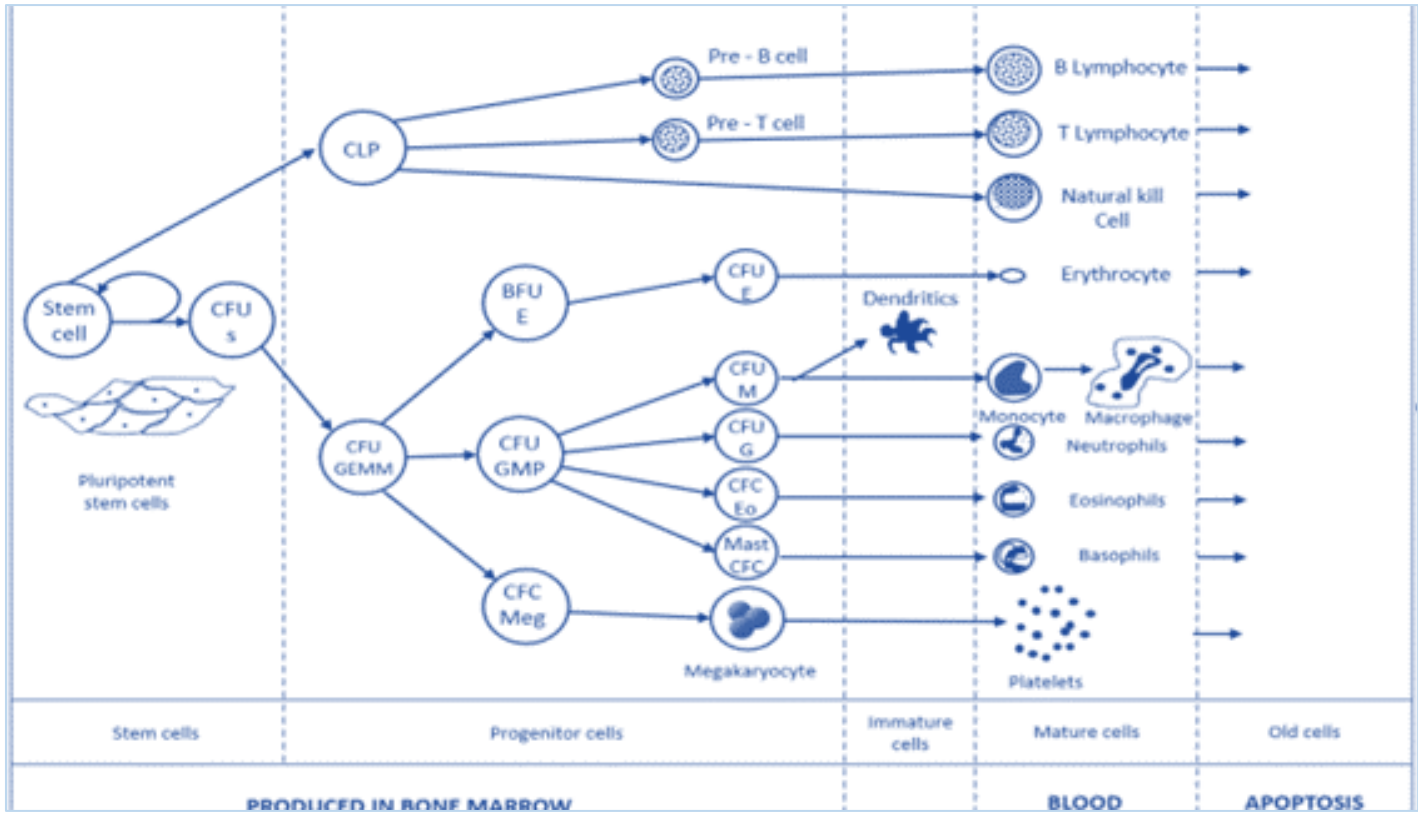


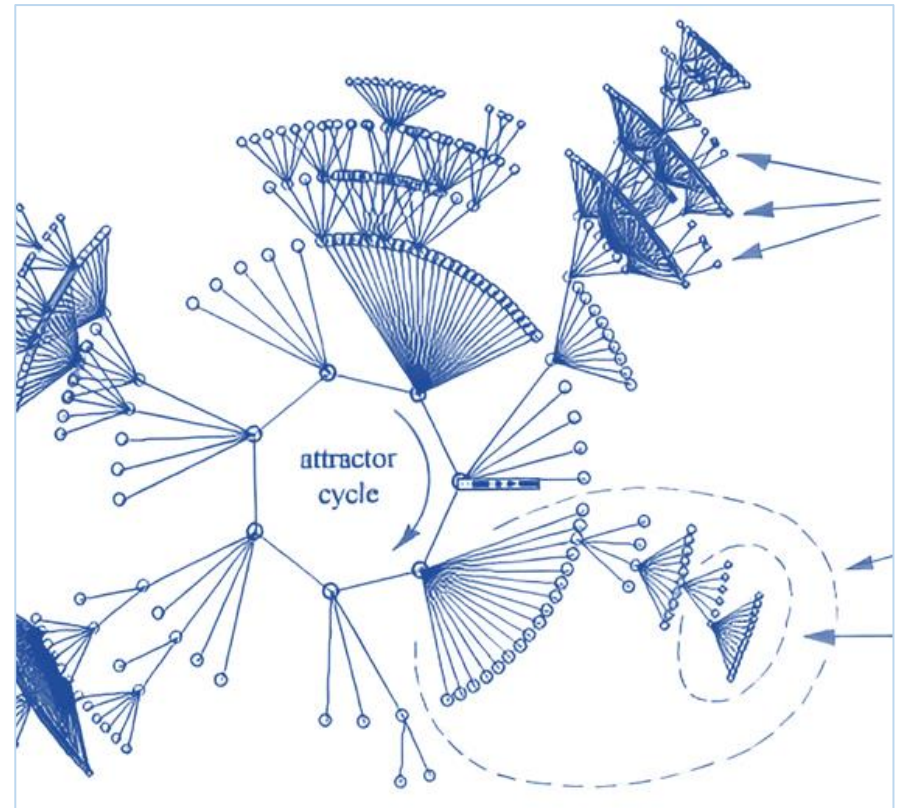
Boolean network characteristic

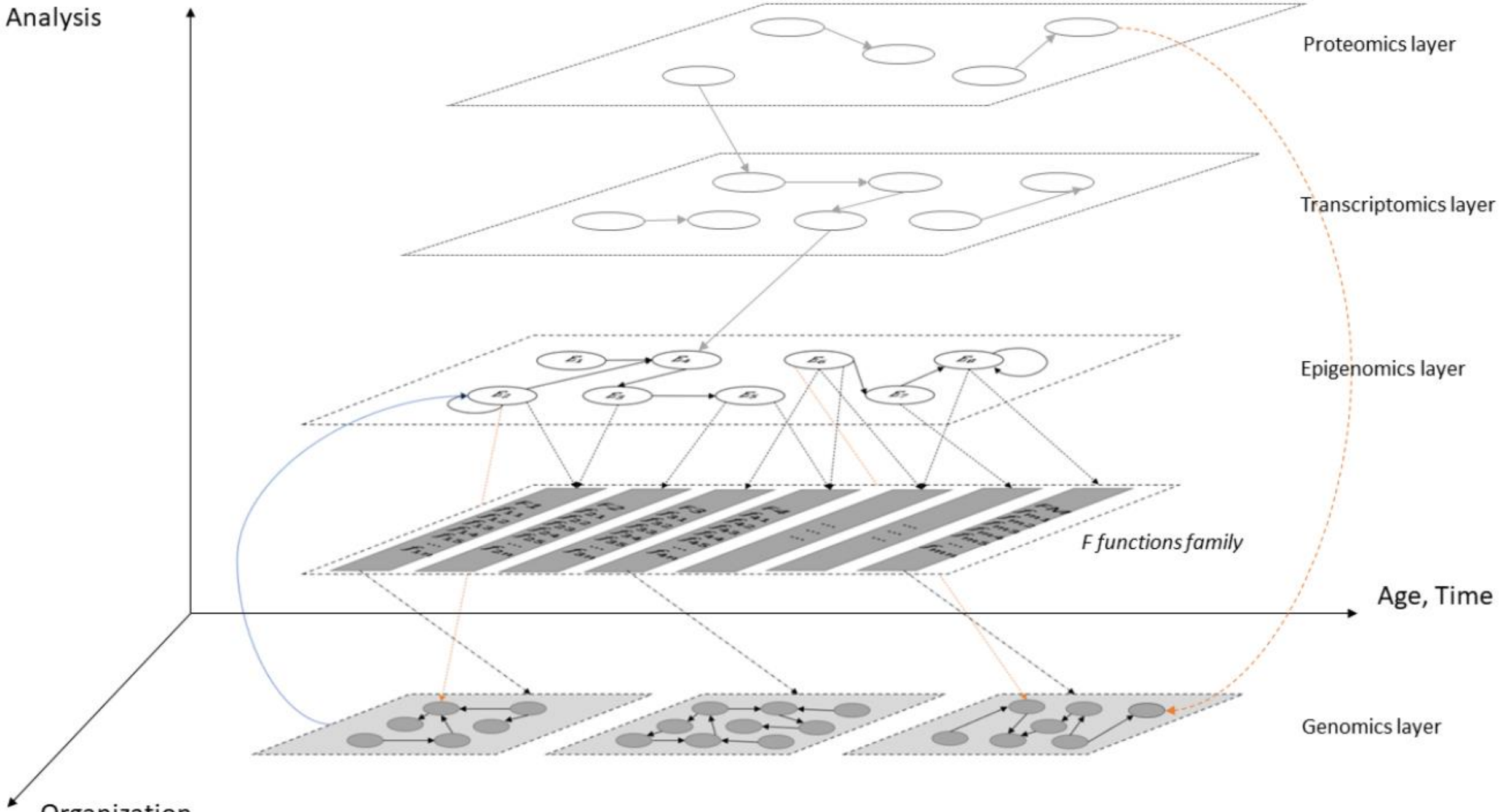
- Attractor and basin
- Canalizing functions
- Nested loops
- Convergence and divergence
- Communities



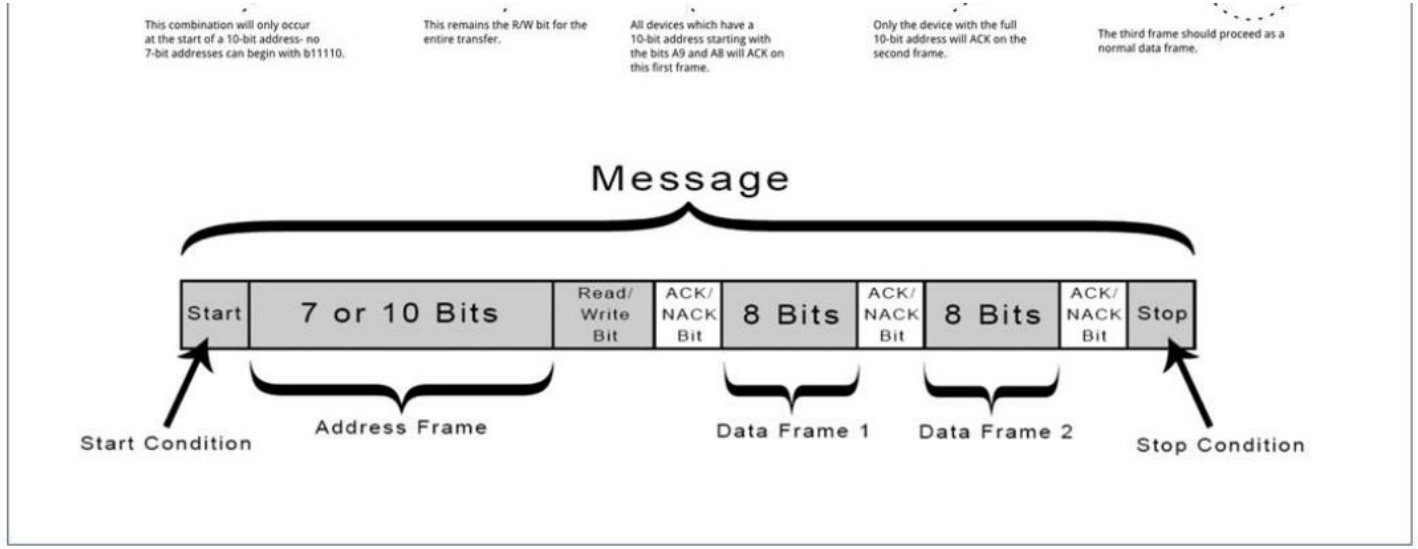

A
DNA
Degrade
B
Simulations
Protein – TF1
C
TF1
TF2
TF1
TF3
TF1
TF4
D
TF2
Protein 1
TF3
Protein 2

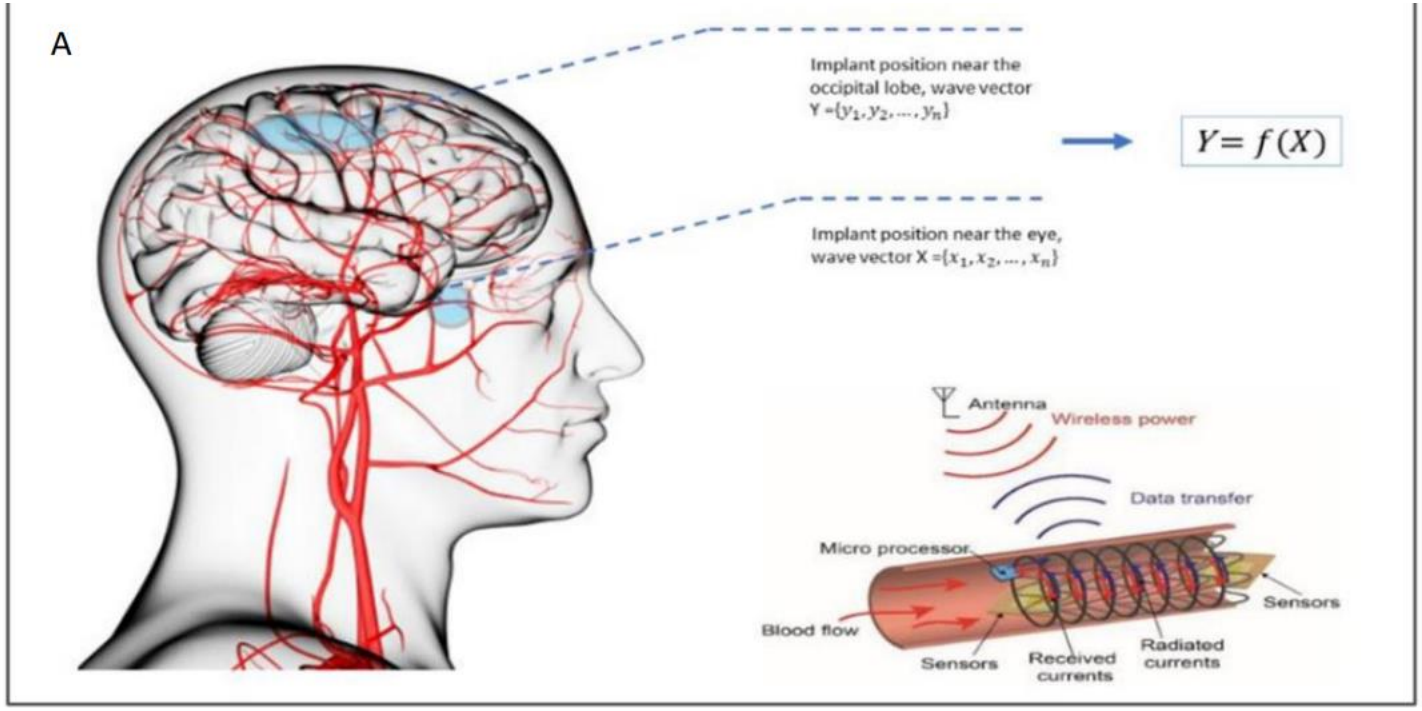

Pre - B cell
Pre - T cell
CLP
B Lymphocyte
T Lymphocyte
Natural kill Cell
Stem cell
CFU s
BFU E
CFU E
Erythrocyte
Dendritics
CFU M
Monocyte
Macrophage
CFU GEMM
CFU GMP
CFU G
Neutrophils
CFC Eo
Eosinophils
Mast CFC
Basophils
CFC Meg
Megakaryocyte
Platelets
Pluripotent stem cells
Stem cells
Progenitor cells
Immature cells
Mature cells
Old cells
PRODUCED IN BONE MARROW
BLOOD
APOPTOSIS


attractor cycle

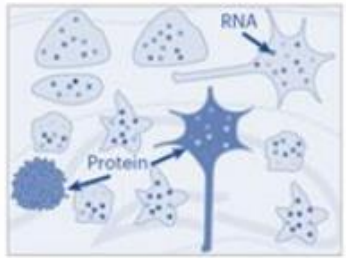

Analysis
Proteomics layer
Transcriptomics layer
Epigenomics layer
F functions family
Age, Time
Genomics layer
Organization

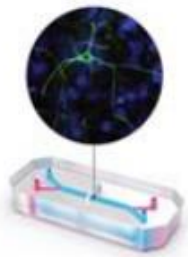
Communication protocol between brain and sensory organs





A



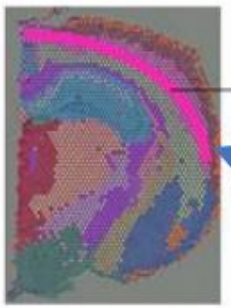


Cell type 1

Cell type 2

Cell type 3

B

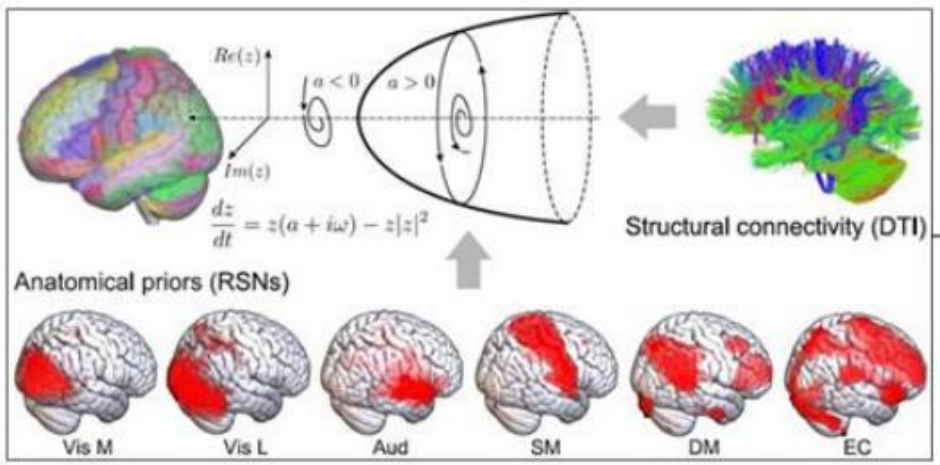


signal path

$Y = F(X)$

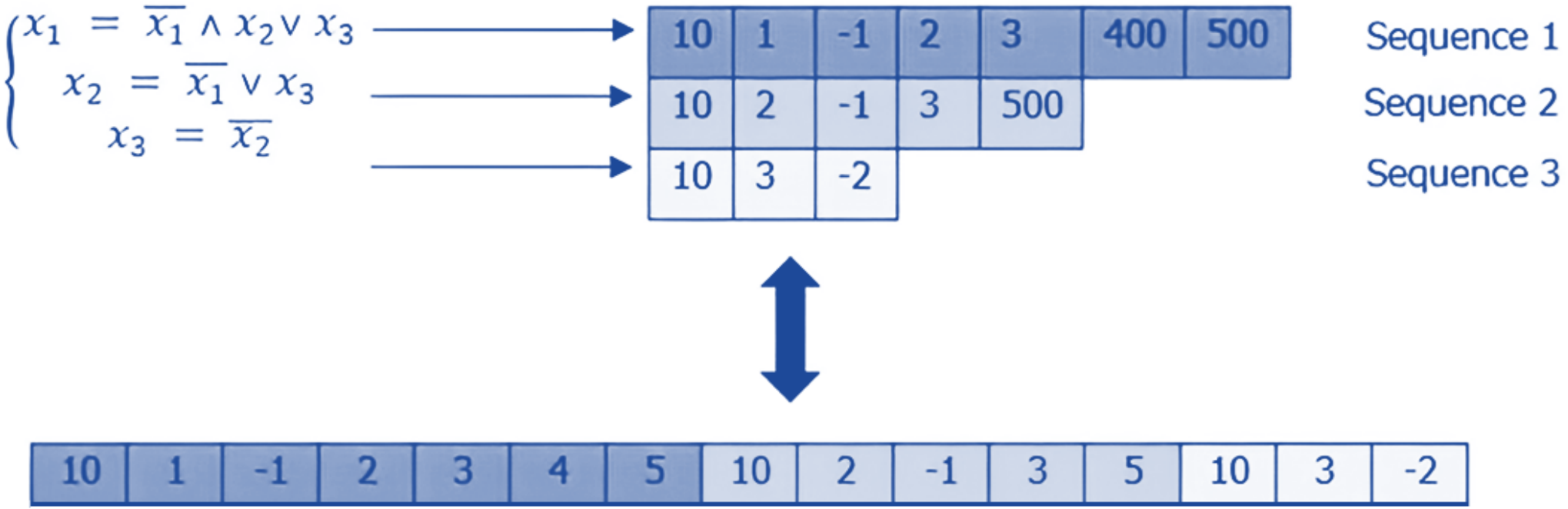
Whole brain functional connectivity

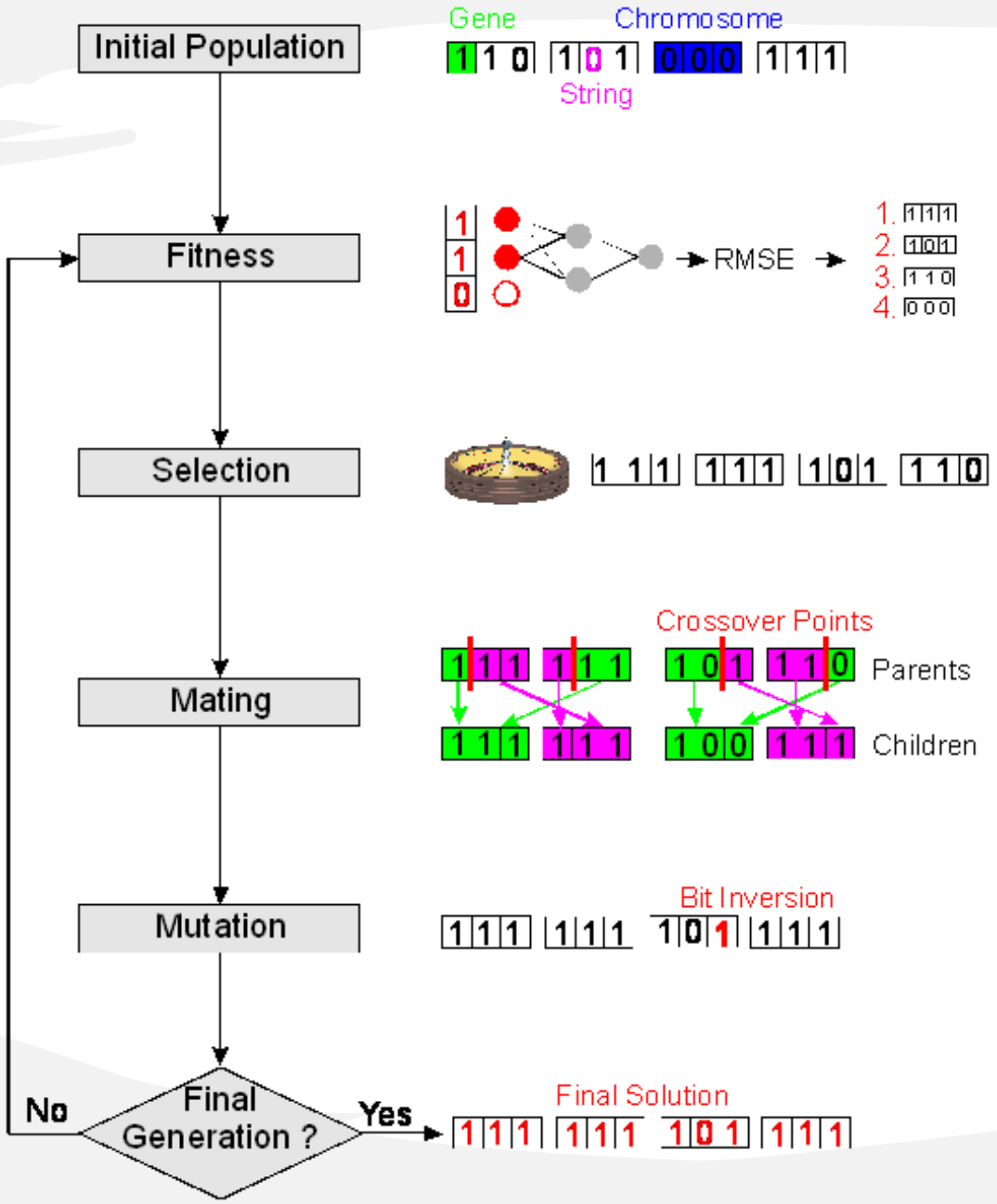
C



APLICATION OF AI

BOOLEAN NETWORK INFERENCE BY GENETICS ALGORITHM




Initial Population
Gene
Chromosome
String
Fitness
RMSE
Selection
Mating
Crossover Points
Parents
Children
Mutation
Bit Inversion
Final Generation ?
No
Yes
Final Solution

```
AXIN2*=  AXIN2 or TCF_LEF
Bcatenin_memb*=  Ecadherin and not Bcatenin_nuc
Bcatenin_nuc*=  not Dest_compl and not Bcatenin_memb and (not SUFU or not Ecadherin)
BTrCP*=  not Csn
CD44*=  TCF_LEF
CDC42*=  TGFBR or CHD1L
cfos*=  ERK
cMet*=  HGF or CD44
Csl*=  NOTCH_ic
Csn*=  NFKB
DELTA*=  RAS
DSH*=  Frizzled
Ecadherin*=  Bcatenin_memb and (not SNAI1 or not HEY1 or not ZEB1 or not ZEB2 or not FOXC2 or not TWIST1 or not SNAI2)
EGFR*=  EGF
EGR1*=  cfos
EMT*=   not Ecadherin
ERK*=  MEK
FGFR*=  FGF
FOXC2*=  Goosecoid or SNAI1 or TWIST1
Frizzled*=  Wnt
FUS*=  SMO
GLI*=  TCF_LEF or not SUFU
GSK3B*=  not DSH and not AKT and (not Csn or not ERK or not Dest_compl)
```

- Protein structure prediction.
- Protein–protein interactions.
- Methylation sites, DNA-protein binding sites.

Experimental:

X-ray crystallography.

Nuclear magnetic resonance spectroscopy.

Cryogenic transmission electron microscopy.

Computational:

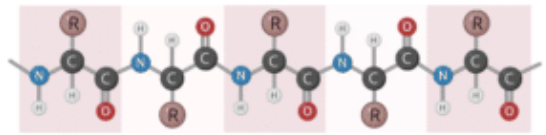
Explicit sequence-to-structure maps: physics-based molecules, ab initio.

Using co-evolution information: comparative, deep learning + assembly method.

End to end: RNN, CNN, LSTM, transformer architects.

Primary structure

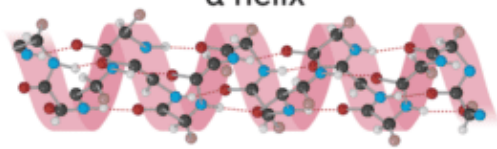
Polypeptide chain



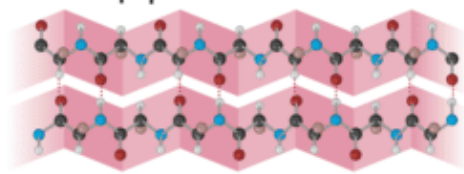
Amino acid

Secondary structure

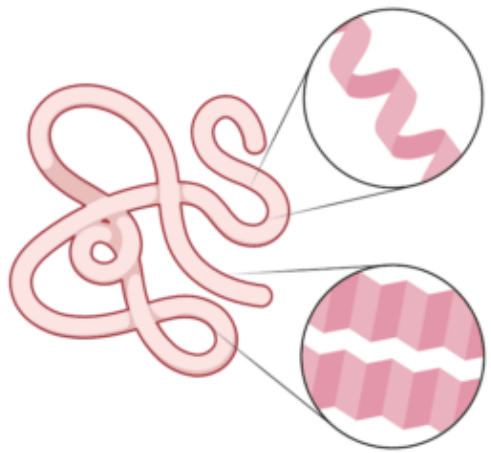
α-helix



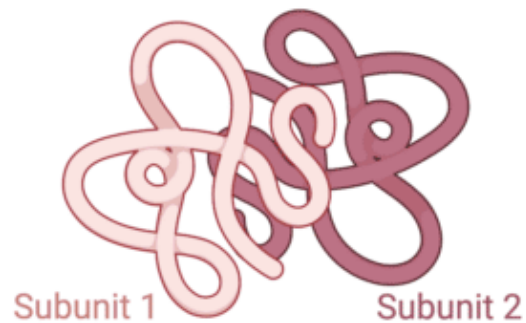
β-pleated sheet

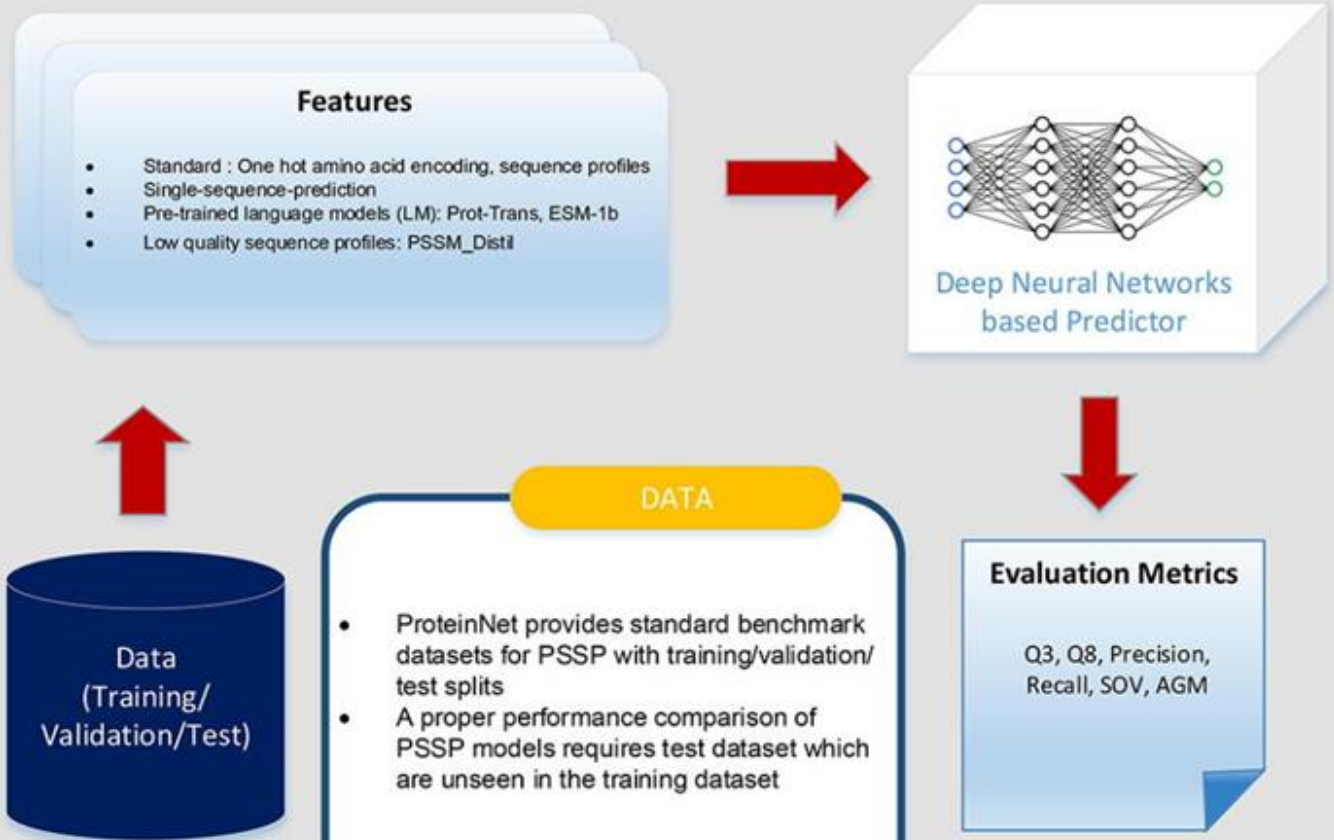


Tertiary structure



Quaternary structure





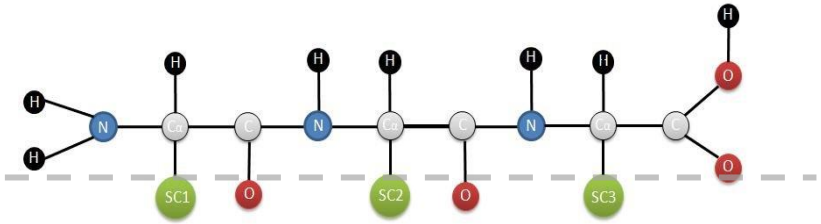
ARCHITECTURES

- Networks: CNN, LSTM, GRU, Inception Networks, Residual Networks, U-net, Graph Networks, Fractal Networks
- NLP methods: attention mechanism, Transformer, language models
- Ensemble and Multitask Learning
- Advanced Machine Learning methods: knowledge distillation, deep metric learning

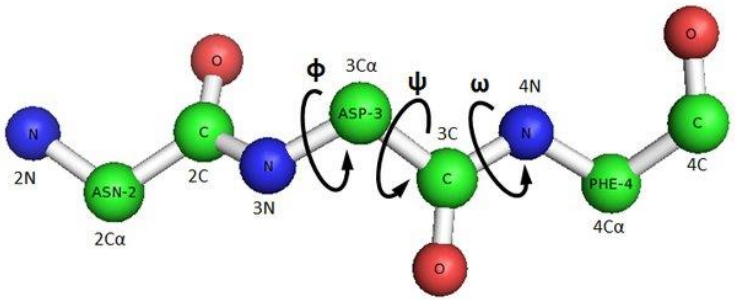
CHALLENGES

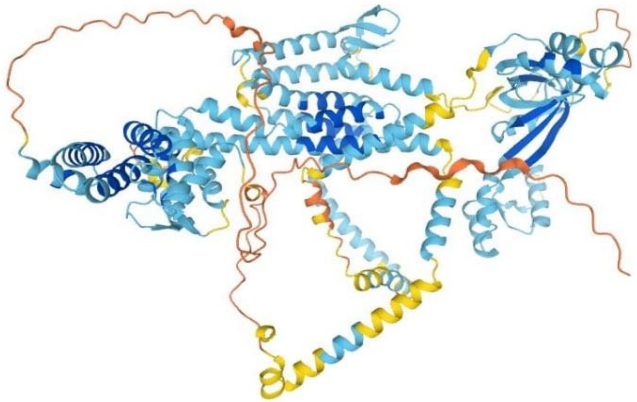
- The prediction accuracy of Sheet is not as high as of Helix
- The prediction accuracy of single-sequence-prediction falls far from the theoretical limit
- Structure prediction without using evolutionary information is needed for proteins without known homologs in the post-AlphaFold era

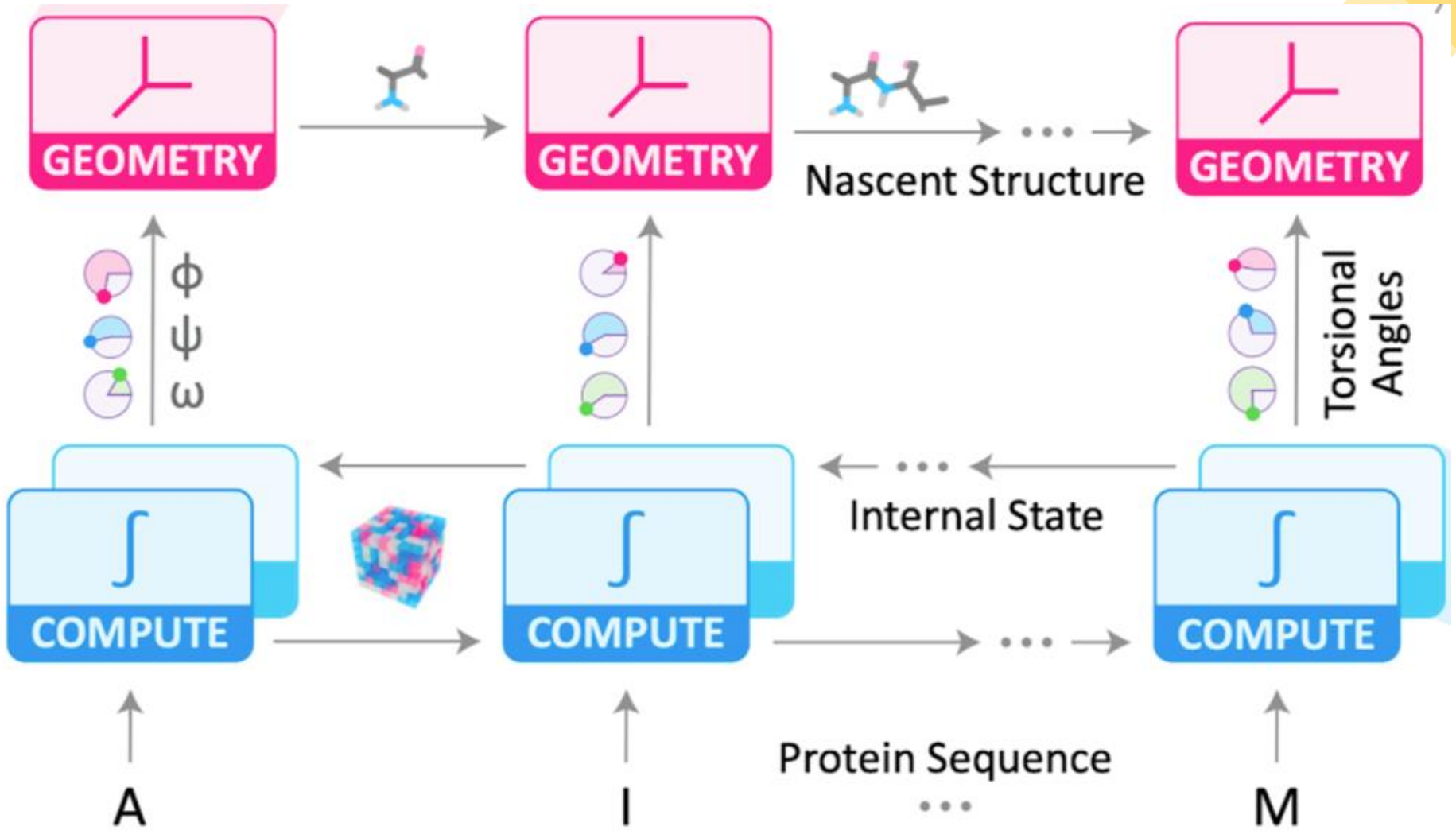
Backbone

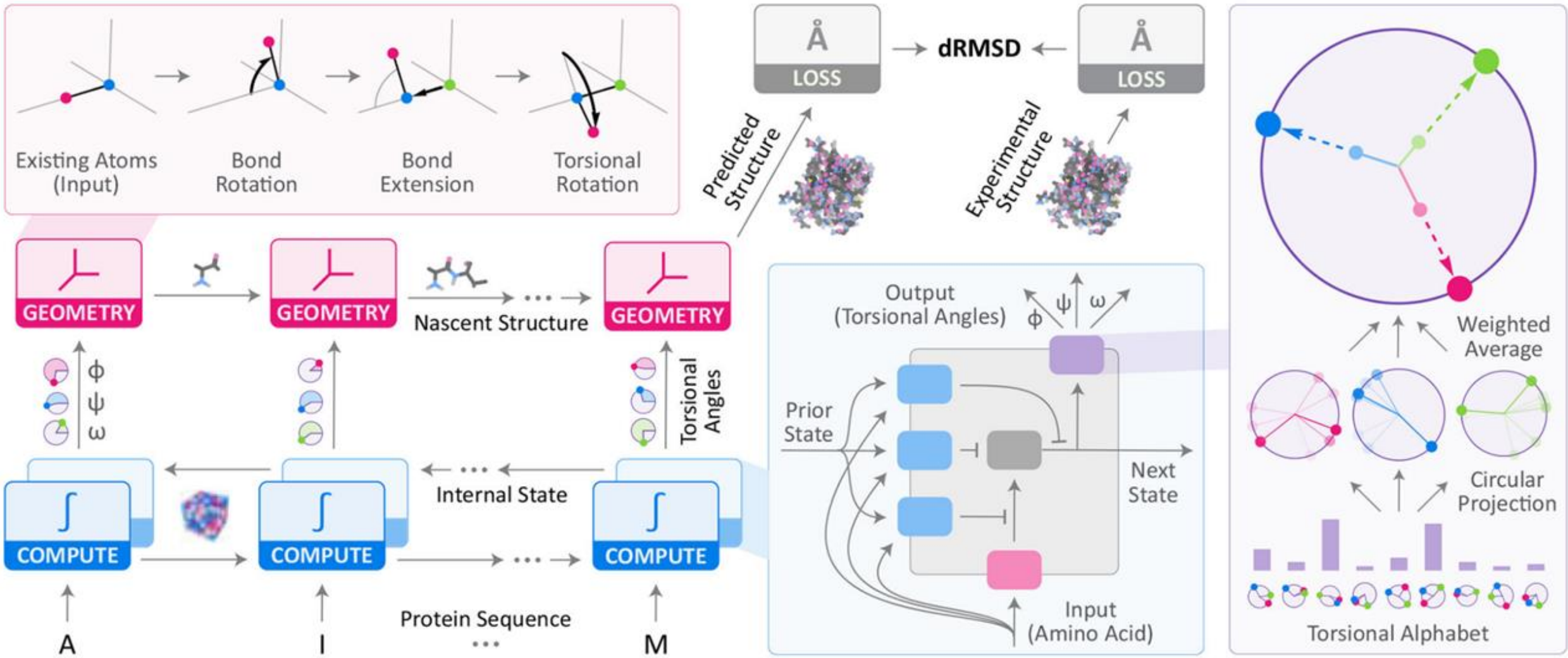


Side chain



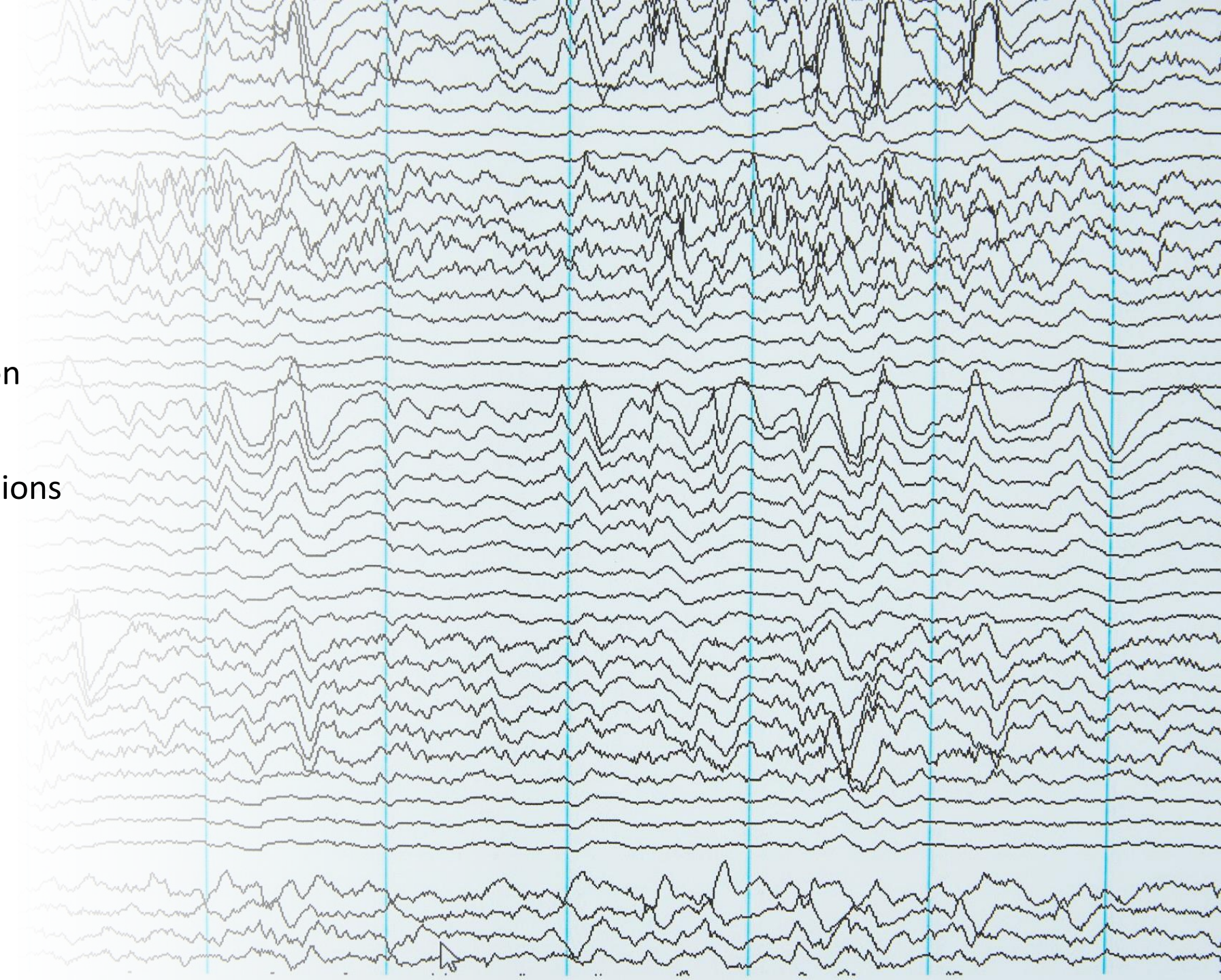


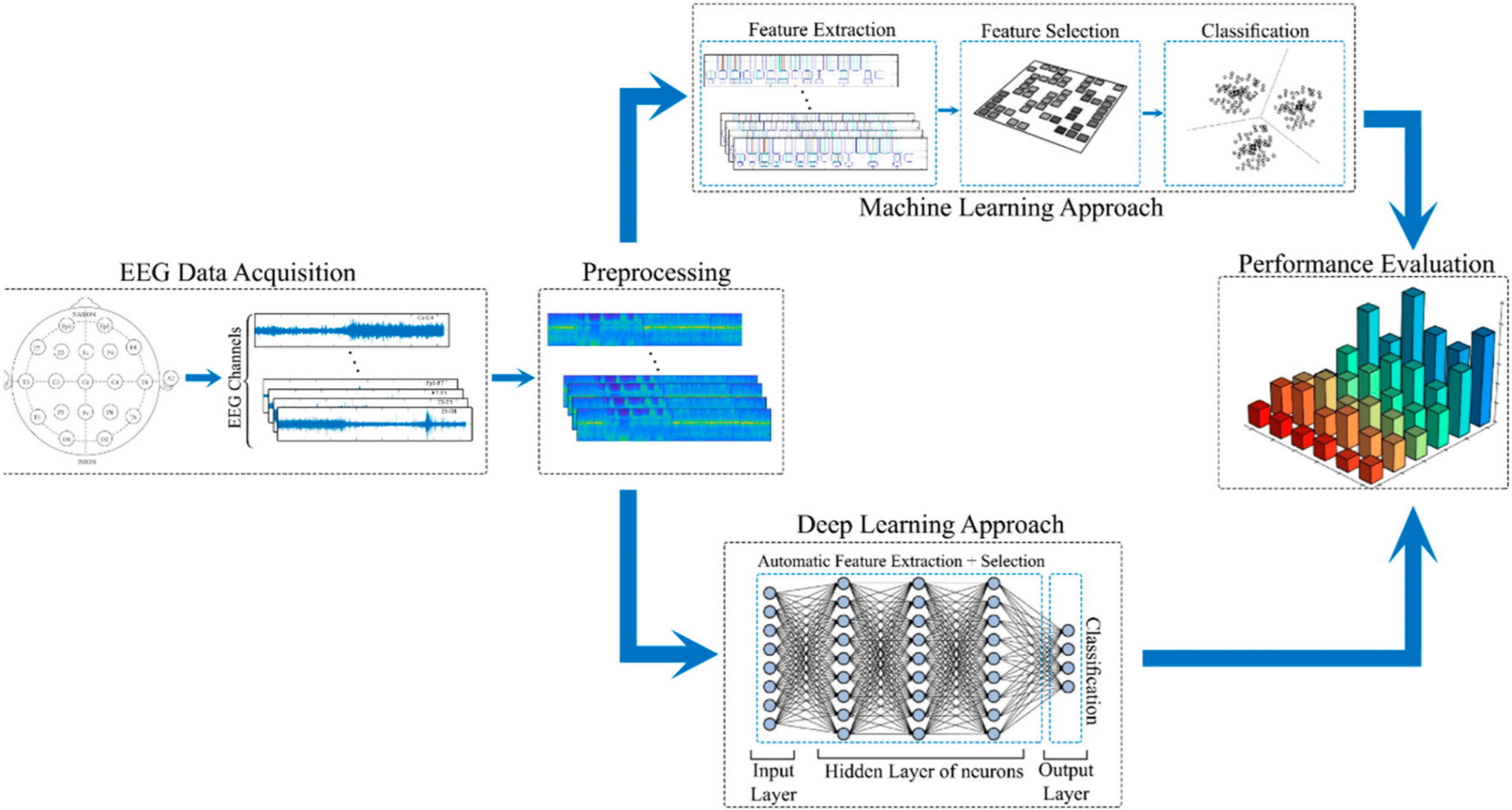

GEOMETRY
GEOMETRY
Nascent Structure
GEOMETRY
ϕ
ψ
ω
Torsional Angles
COMPUTE
COMPUTE
Internal State
COMPUTE
A
I
Protein Sequence
M


Existing Atoms (Input)
Bond Rotation
Bond Extension
Torsional Rotation
Å
LOSS
dRMSD
Å
LOSS
Predicted Structure
Experimental Structure
GEOMETRY
GEOMETRY
Nascent Structure
GEOMETRY
Torsional Angles
ϕ
ψ
ω
COMPUTE
COMPUTE
COMPUTE
Internal State
Protein Sequence
A
I
M
Output (Torsional Angles)
ϕ
ψ
ω
Prior State
Next State
Input (Amino Acid)
Weighted Average
Circular Projection
Torsional Alphabet

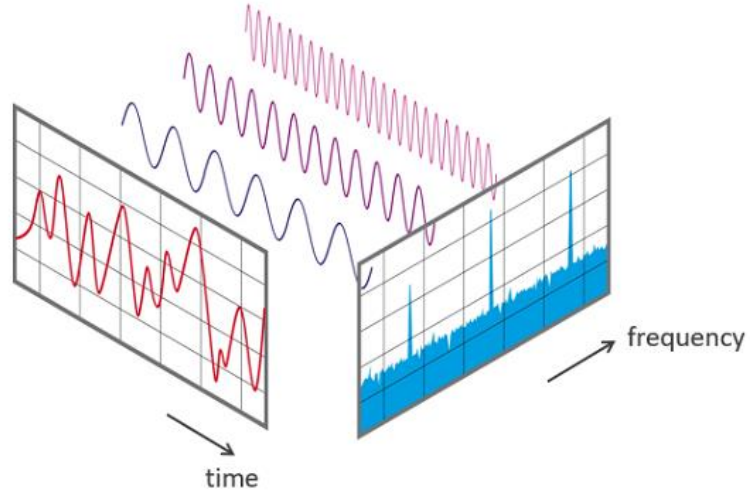
4. EEG Signal processing

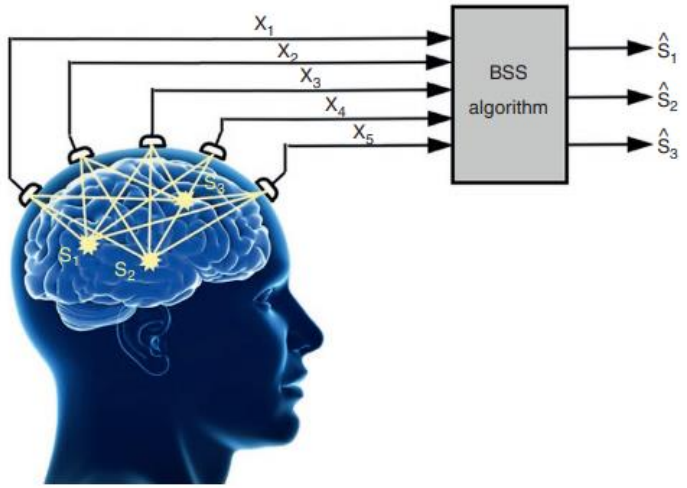
- Paresthesia phenomenon
- Electrical synapses
- Common mapping functions



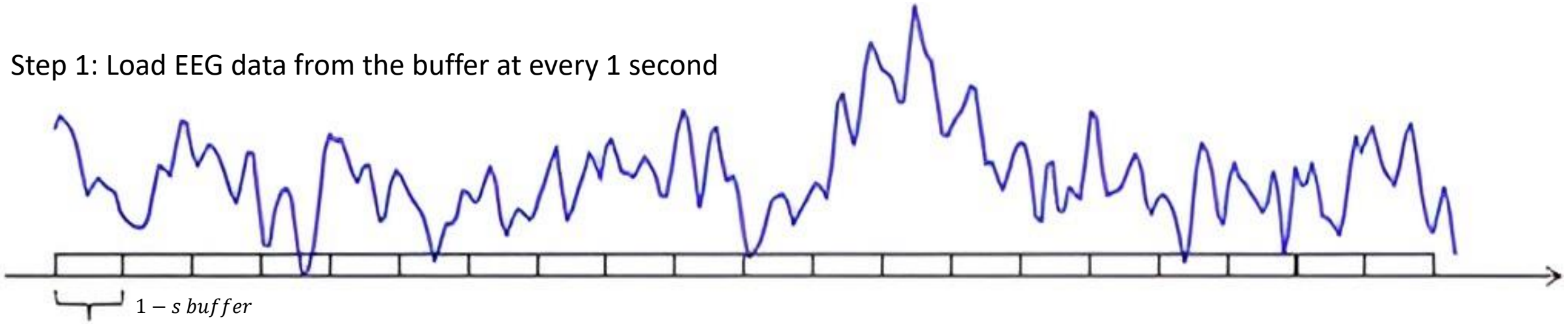

Feature Extraction
Feature Selection
Classification
Machine Learning Approach
EEG Data Acquisition
EEG Channels
Preprocessing
Performance Evaluation
Deep Learning Approach
Automatic Feature Extraction + Selection
Classification
Input Layer
Hidden Layer of neurons
Output Layer

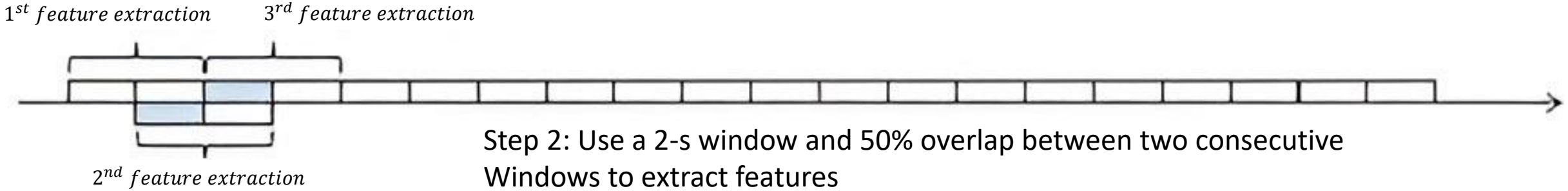
- FFT: Converts signals from the time domain to the frequency domain, filtering out signals outside the brain's frequency range.
- BSS: separate a set of source signals from a set of mixed signals without prior information about the source signals or the mixing process. The goal is to recover the original source signals from their observed mixtures.



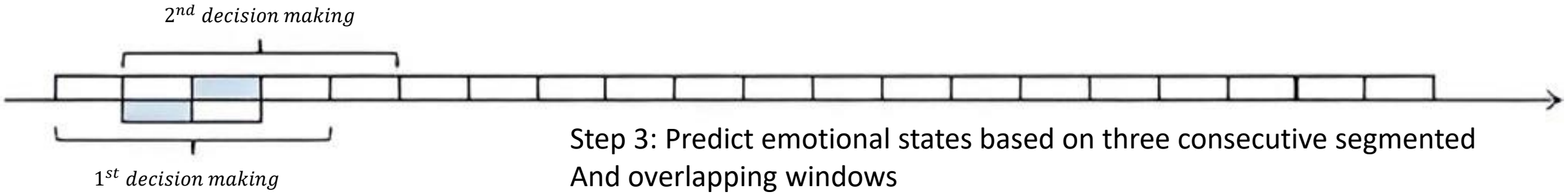


Step 1: Load EEG data from the buffer at every 1 second





Step 2: Use a 2-s window and 50% overlap between two consecutive Windows to extract features



Step 3: Predict emotional states based on three consecutive segmented And overlapping windows

- **Input Gate:**

Decides which information from the input should be stored in the memory cell.

- **Forget Gate:**

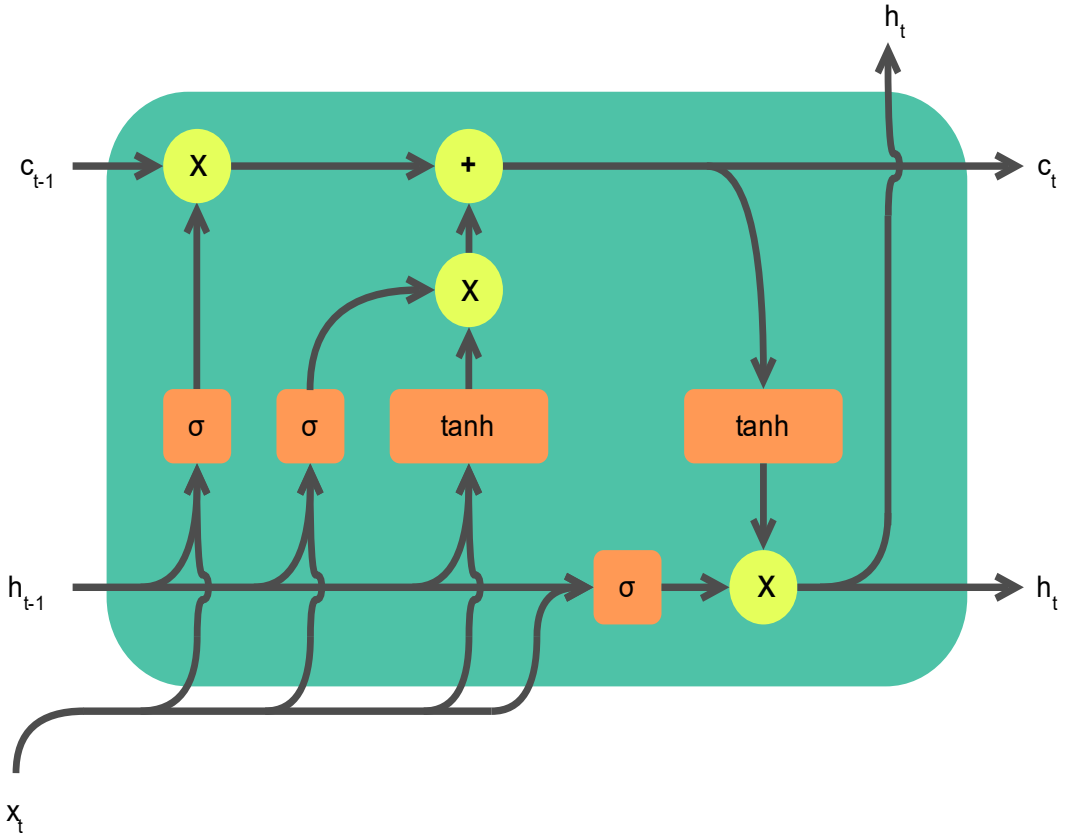
Decides which information stored in the memory cell should be discarded or kept.

- **Cell State:**

Represents the long-term memory and is modified by the input gate and forget gate.

- **Output Gate:**

Decides what information from the memory cell should be output as the final prediction.



REFFERENCE

- *Karl R. Gegenfurtner and Lindsay T. Sharpe. 1999. Color Vision: From Genes to Perception.*
- *E.R. Dougherty, I Shmulevich. 2010. Probabilistic Boolean Networks: The Modeling and Control of Gene Regulatory Networks.*
- *Michael L. Goldberg, Leland H. Hartwell, Janice A. Fischer, Leroy E. Hood. 2021. Genetic: From Genes to Genomes. 5th Edition.*
- *Jonathan Pevsner. 2015. Bioinformatics and Functional Genomics.*
- *Shi, Ning. 2021. Boolean network inference and control using metaheuristic algorithms.*
- *Mwape Chintu. 2021. Protein Structure and Function Prediction by Deep Learning.*
- *Patrick Bryant. 2022. Learning Protein Evolution and Structure.*
- *Aston Zhang, Zachary C. Lipton, Mu Li, Alexander J. Smola. 2022. Dive into Deep Learning.*
- *Mohammed AL Quraishi. 2019. End-to-end differentiable learning of protein structure.*


Thank you