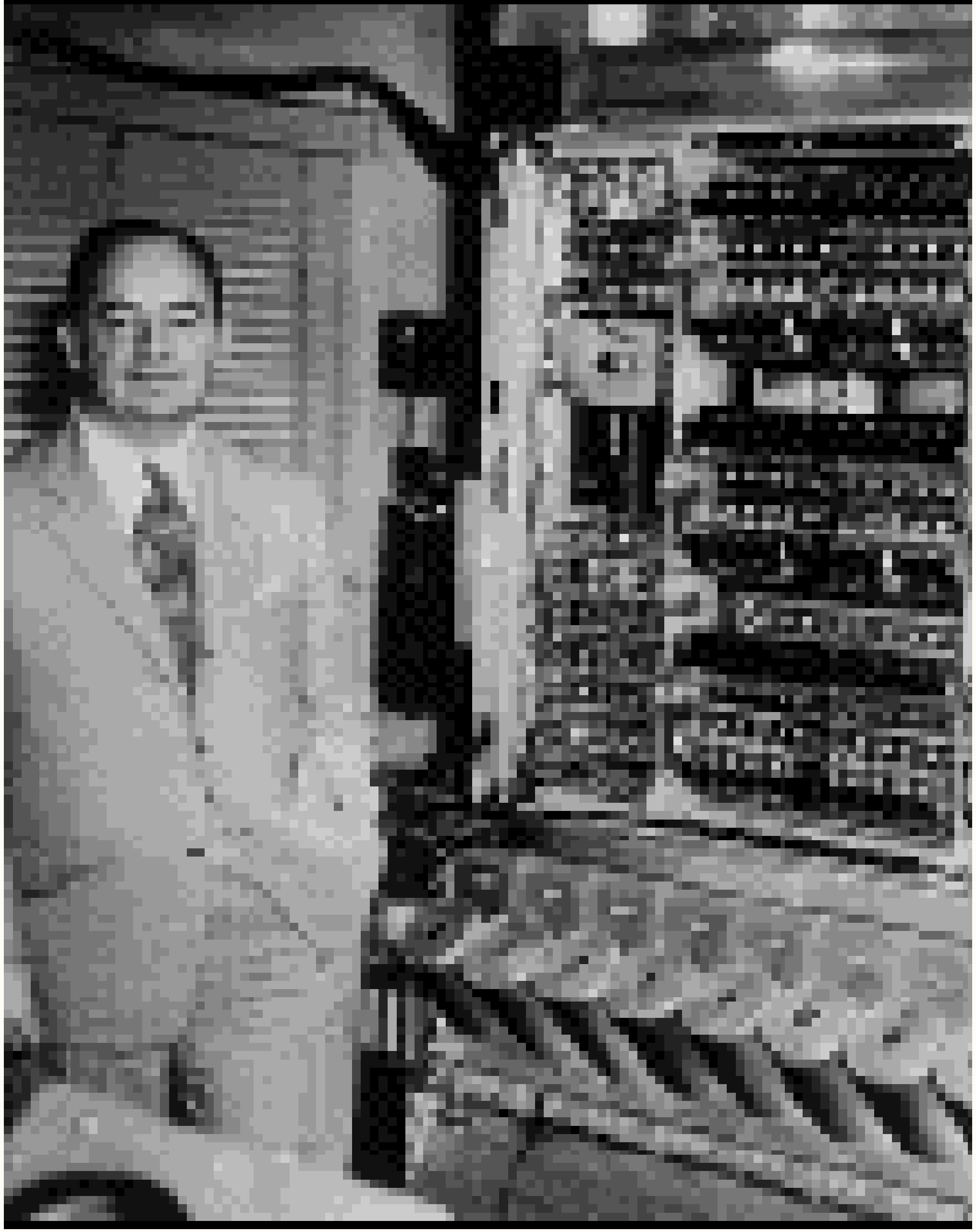
How do biological neural networks encode, learn, memorize, recall, and generalize as a "learning machine"?

James Ting-Ho Lo
Department of Mathematics and Statistics
University of Maryland Baltimore County
Baltimore, MD 21250
jameslo@umbc.edu



"We require exquisite numerical precision over many logical steps to achieve what brains accomplish in very few short steps."

The Computer and the Brain, 1958, p. 63.

John von Neumann, 1903-1957

A mathematical approach

biological postulates

$\downarrow$ mathematical inference

learning machine

$\downarrow$ biological interpretation

computational model of neural network

Results

1. An ANN or learning machine

   THPAM (Temporal Hierarchical Probabilistic Associative Memory)
   Cognitive Neurodynamics, 2010

2. A biologically plausible model

   LOM (Low-Order Model of Biological NNs)
   Neural Computation, 2011

3. An ANN or learning machine

   CIPAM (Clustering Interpreting Probabilistic Associative Memory)
   Neurocomputing, 2012

Low-Order Model (LOM)

Single model providing
logically coherent answers to
holy-grail questions:

- What information does a spike train carry?
- What computation do dendritic trees do?
- How is supervised learning performed?
- How is unsupervised learning performed?
- How is information stored in synapses?
- What computation do spiking and nonspiking somas do?
- How are corrupted, distorted and occluded patterns recognized?
- How are neurons organized into a neural network?

Single model providing logically coherent answers to holy-grail questions:

- **What information does a spike train carry?**
- What computation do dendritic trees do?
- How is supervised learning performed?
- How is unsupervised learning performed?
- How is information stored in synapses?
- What computation do spiking and nonspiking somas do?
- How are corrupted, distorted and occluded patterns recognized?
- How are neurons organized into a neural network?

What information should be communicated between spiking neurons?

Recognition of 26 Upper Case Letters:

A B C D E F G H I J K L M
N O P Q R S T U V W X Y Z

What letter is this? 

What is the best representation?

A W

Relative frequencies of the 26 letters:

A 8.17% (binary code 01000001)

W 2.36% (binary code 01010111)

8 somas needed to generate 8 bits.

Each bit assigned by the probability:

8.17/(8.17+2.36) or 2.36/(8.17+2.36)

Conjecture:

Neurons operate in groups generating labels.

Single model answering holy-grail questions:

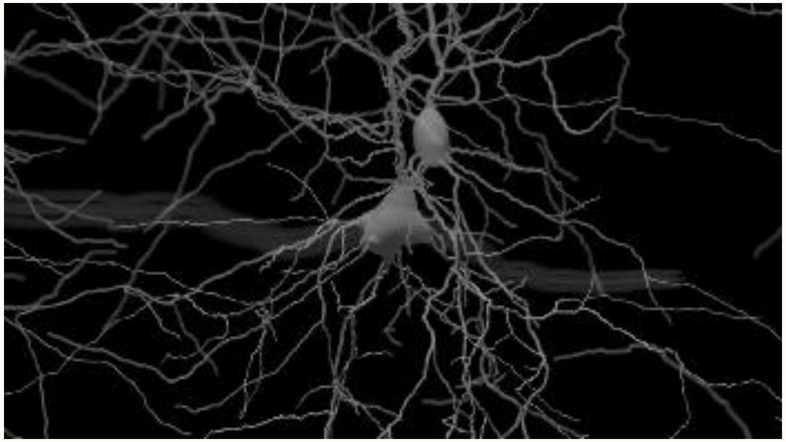
- What information does a spike train carry?
- **What computation do dendritic trees do?**
- How is supervised learning performed?
- How is unsupervised learning performed?
- How is information stored in synapses?
- What computation do spiking and nonspiking somas do?
- How corrupted, distorted and occluded patterns recognized?
- How are neurons organized into a neural network?

Dendrites

- Use more than 60% of the energy consumed by the brain
- Occupy more than 99% of the surface of some neurons
- Are the largest component of neural tissue in volume

Yet, dendrites are missing in all ANNs.

A model of a dendrite tree



Neurons

Outputs from 3 neurons:

$$v_\tau = \begin{bmatrix} v_{\tau 1} & v_{\tau 2} & v_{\tau 3} \end{bmatrix}' \quad \text{binary}$$

Take all the 2^3 combinations of the 3 entries:

$$\{\ \}, \{v_1\}, \{v_2\}, \{v_2, v_1\}, \{v_3\}, \{v_3, v_1\}, \{v_3, v_2\}, \{v_3, v_2, v_1\}$$

Apply a function ϕ to each combinations:

$$0, \phi(v_1), \phi(v_2), \phi(v_2, v_1), \phi(v_3), \phi(v_3, v_1), \phi(v_3, v_2), \phi(v_3, v_2, v_1)$$

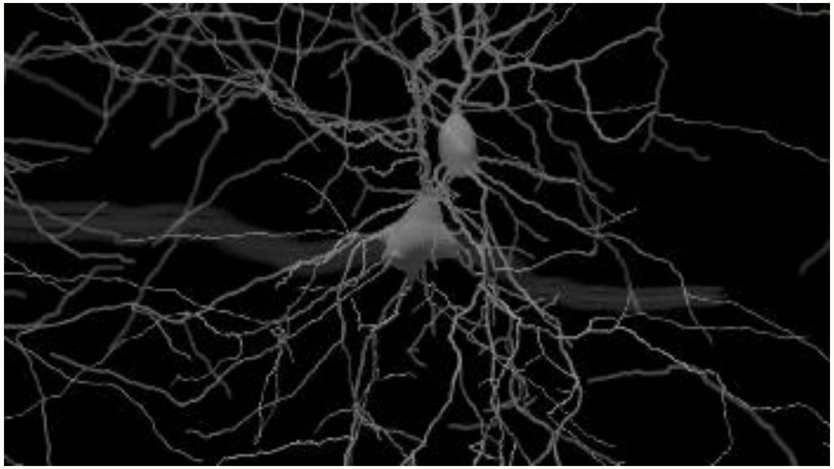
These are the outputs of the dendritic tree

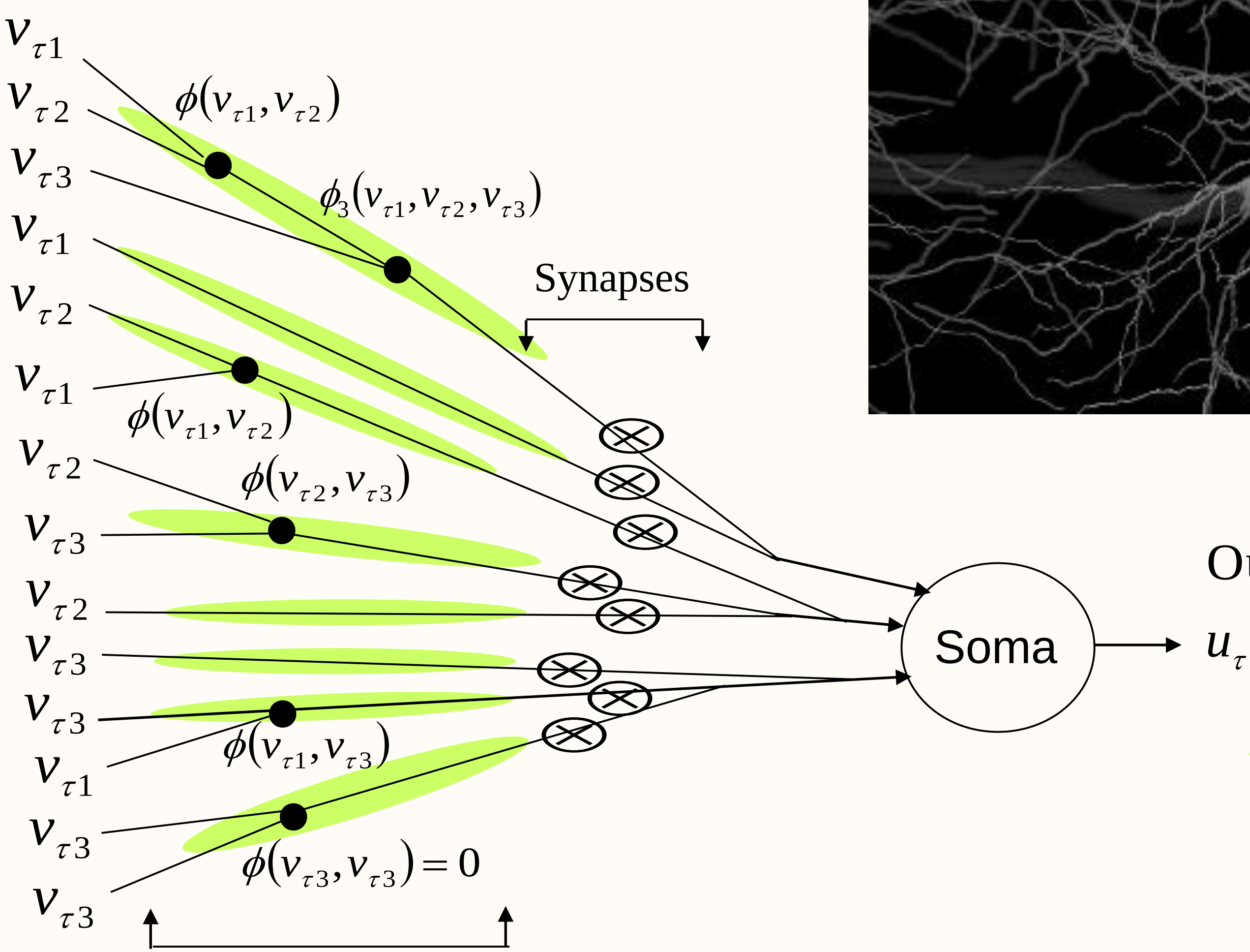
A model of a dendritic tree

Input from axons :

$v_\tau = [v_{\tau 1} \quad v_{\tau 2} \quad v_{\tau 3}]'$ binary (essentially)

Neurons





Output by an axon :

A dendritic encoder

A dendritic branch
is a pairity function

HYPOTHESIS

A dendritic branch is a parity function ϕ :

$$\phi\left(v_1, v_2, \cdots, v_n\right) = \begin{cases} 1 & \text{if } \; v_1 + v_2 + \cdots + v_n \; \text{ is odd} \\ 0 & \text{if } \; v_1 + v_2 + \cdots + v_n \; \text{ is even} \end{cases}$$

Experiment required to test the hypothesis:

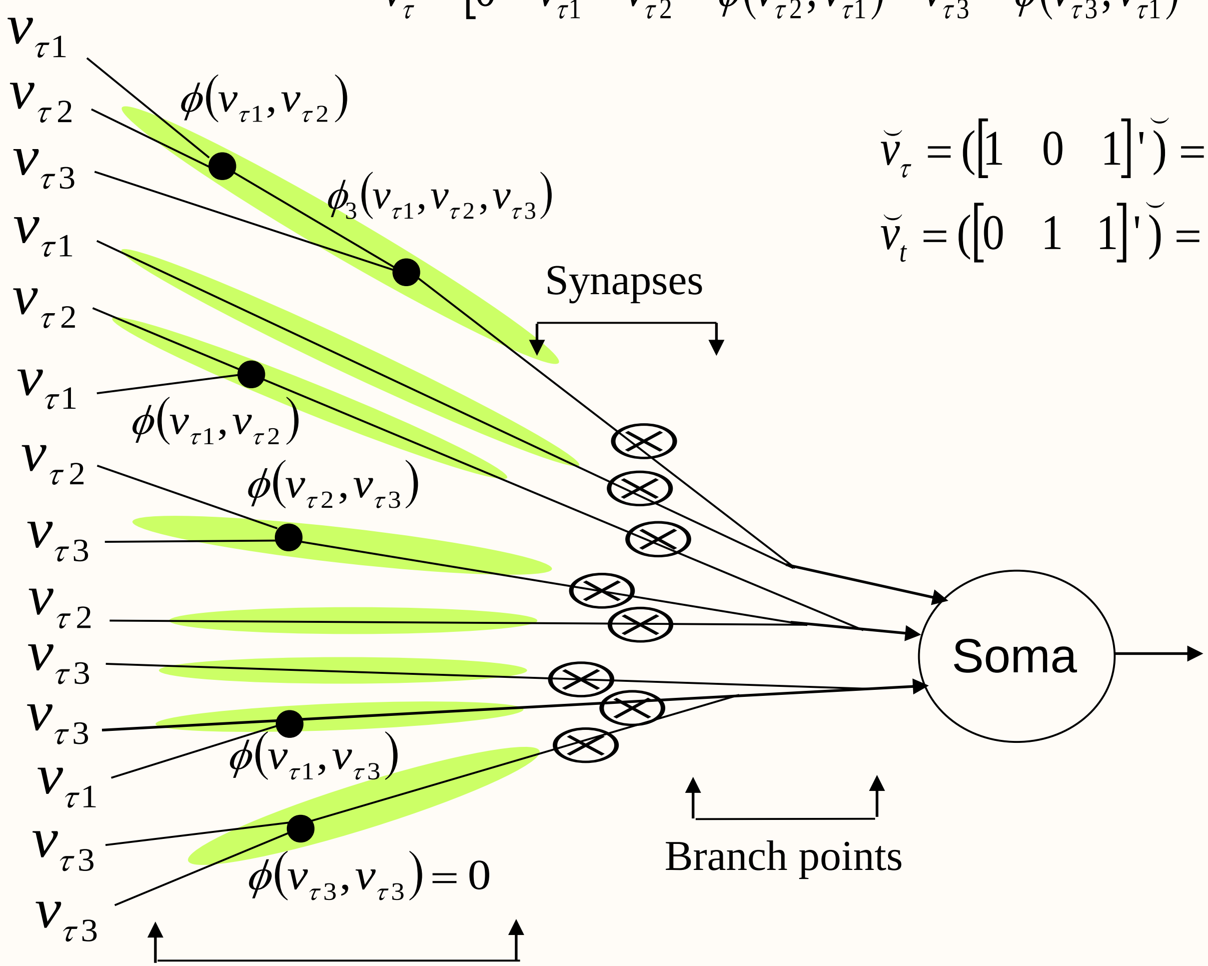
Measure the inputs and output of dendritic branches.

$$v_\tau = [v_{\tau 1} \quad v_{\tau 2} \quad v_{\tau 3}]'$$

$$\breve{v}_\tau = [0 \quad v_{\tau 1} \quad v_{\tau 2} \quad \phi(v_{\tau 2}, v_{\tau 1}) \quad v_{\tau 3} \quad \phi(v_{\tau 3}, v_{\tau 1}) \quad \phi(v_{\tau 3}, v_{\tau 2}) \quad \phi(v_{\tau 1}, v_{\tau 2}, v_{\tau 3})]'$$

$$\breve{v}_\tau = (\,[1 \quad 0 \quad 1]'\,)\breve{} = [0 \quad 1 \quad 0 \quad 1 \quad 1 \quad 0 \quad 1 \quad 0]'$$

$$\breve{v}_t = (\,[0 \quad 1 \quad 1]'\,)\breve{} = [0 \quad 0 \quad 1 \quad 1 \quad 1 \quad 1 \quad 0 \quad 0]'$$



A dendritic encoder

ORTHOGONALITY OF DENDRITIC CODES

$$\breve{v}_\tau = (\begin{bmatrix}1 & 0 & 1\end{bmatrix}')\breve{} = \begin{bmatrix}0 & 1 & 0 & 1 & 1 & 0 & 1 & 0\end{bmatrix}'$$

$$\breve{v}_t = (\begin{bmatrix}0 & 1 & 1\end{bmatrix}')\breve{} = \begin{bmatrix}0 & 0 & 1 & 1 & 1 & 1 & 0 & 0\end{bmatrix}'$$

$$\mathbf{1} = \begin{bmatrix}1 & 1 & 1 & 1 & 1 & 1 & 1 & 1\end{bmatrix}'$$

$$\left(\breve{v}_\tau - \frac{1}{2}\mathbf{1}\right)' = \begin{bmatrix}-\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2}\end{bmatrix}$$

$$= \frac{1}{2}\begin{bmatrix}-1 & 1 & -1 & 1 & 1 & -1 & 1 & -1\end{bmatrix}$$

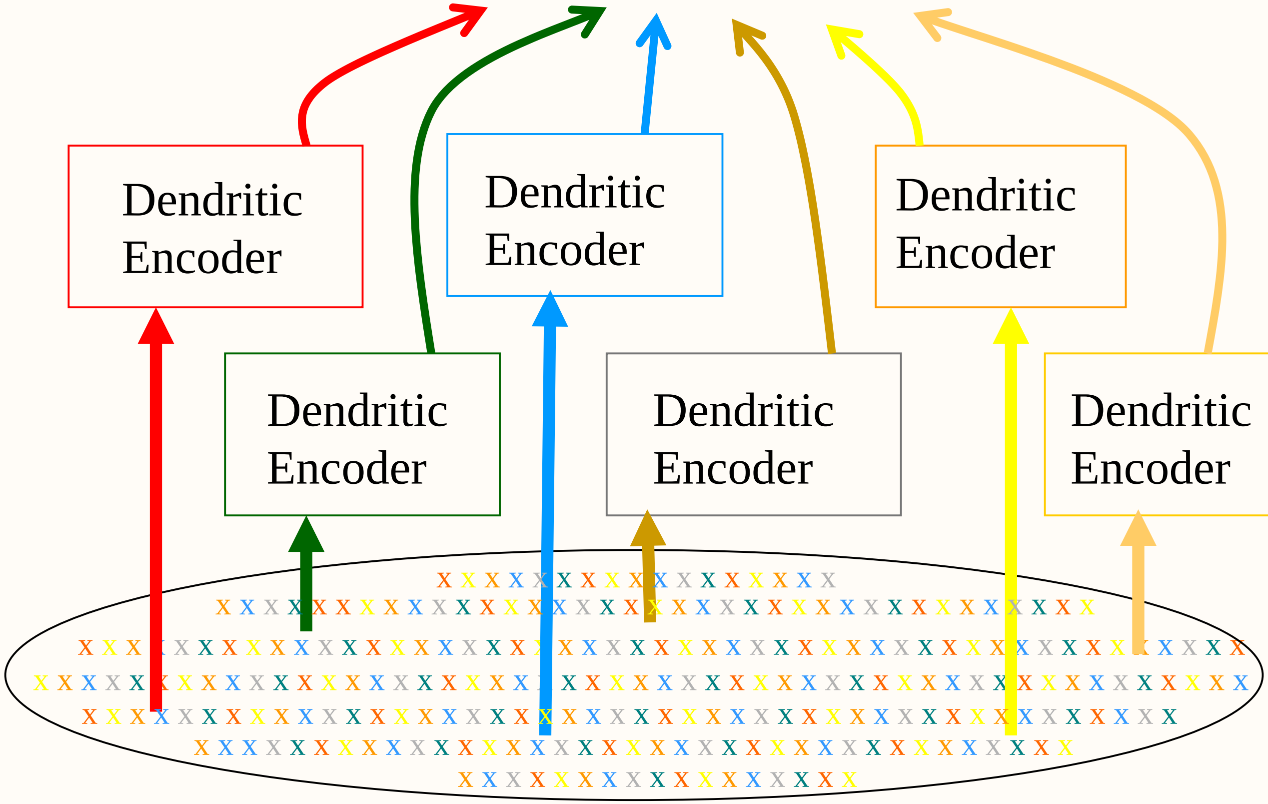
$$\left(\breve{v}_t - \frac{1}{2}\mathbf{1}\right)' = \begin{bmatrix}-\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2}\end{bmatrix}$$

$$= \frac{1}{2}\begin{bmatrix}-1 & -1 & 1 & 1 & 1 & 1 & -1 & -1\end{bmatrix}$$

$$\frac{1}{2}\left(\breve{v}_\tau - \frac{1}{2}\mathbf{1}\right)'\left(\breve{v}_t - \frac{1}{2}\mathbf{1}\right) = 0$$

$$\frac{1}{2}\left(\breve{v}_t - \frac{1}{2}\mathbf{1}\right)'\left(\breve{v}_t - \frac{1}{2}\mathbf{1}\right) = \frac{1}{2}\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)(8) = 1$$

General Dendritic Code



Compartmentalization of dendritic trees

Inputs to a neuron

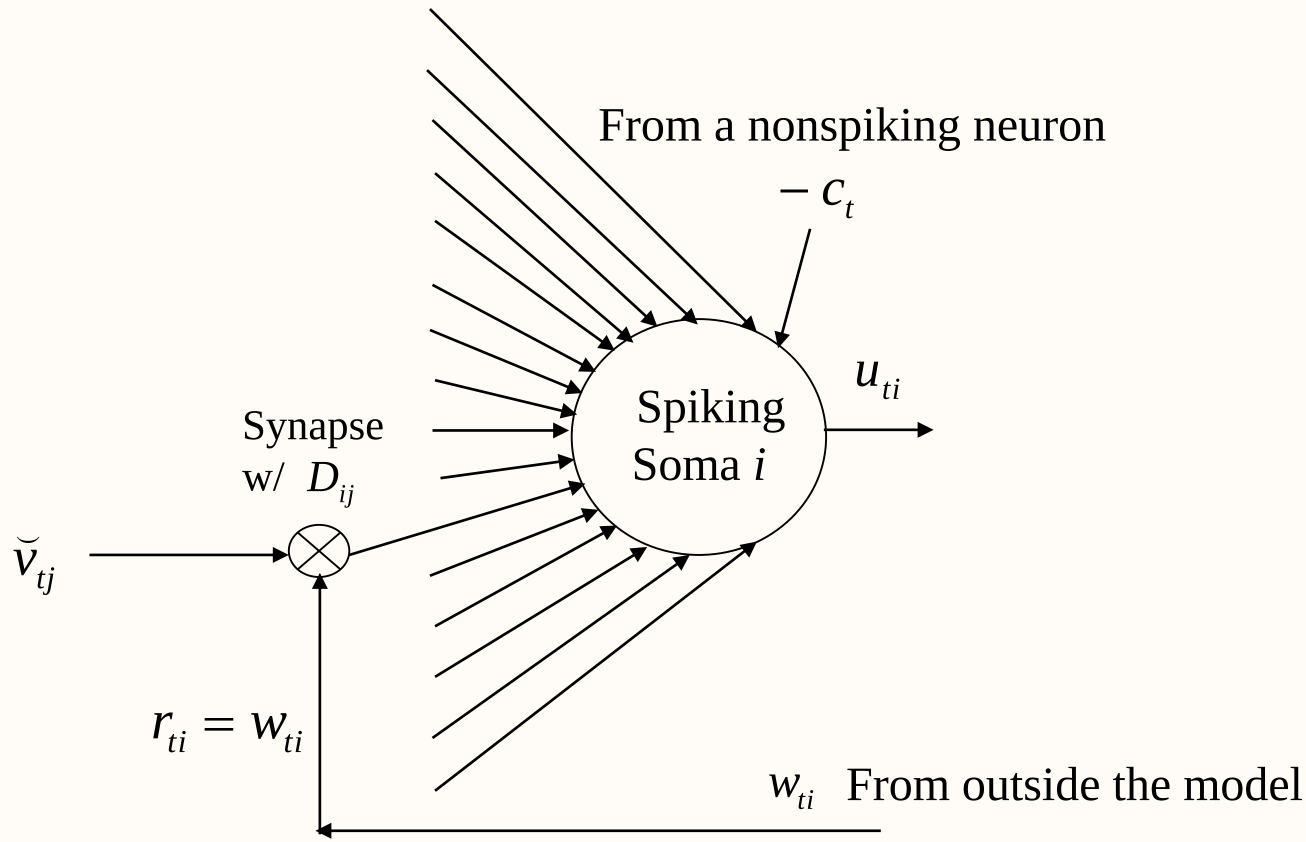
Each color-coded low-resolution input field is uniformly distributed

- Avoiding the curse of dimensionality
- Enhancing generalization

Single model answering holy-grail questions:

- What information does a spike train carry?
- What computation do dendritic trees do?
- **How is supervised learning performed?**
- How is unsupervised learning performed?
- How is information stored in synapses?
- What computation do spiking and nonspiking somas do?
- How corrupted, distorted and occluded patterns recognized?
- How are neurons organized into a neural network?

Supervised Covariance Learning



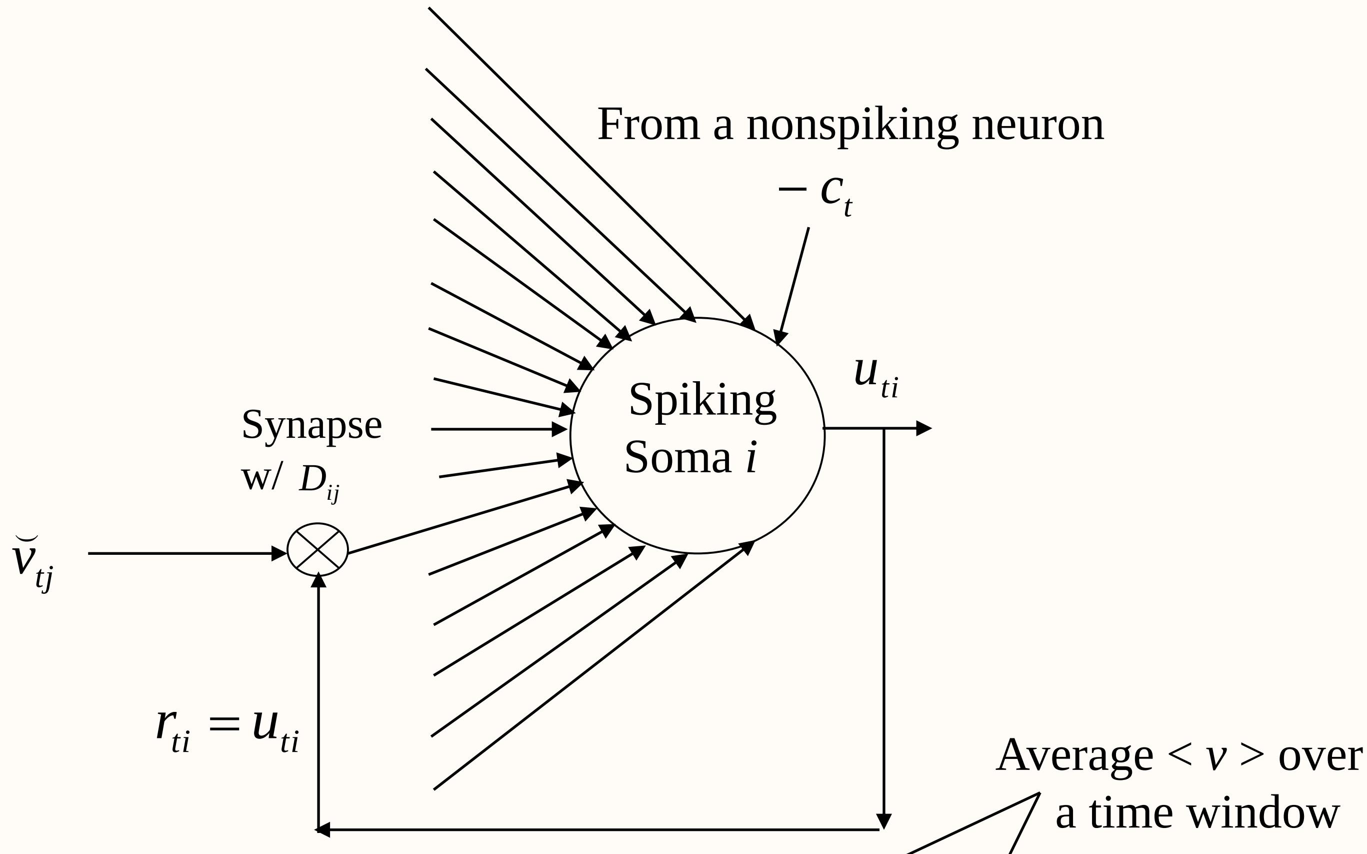
Hebbian-type rule:

$$D_{ij} \leftarrow D_{ij} + \left(w_{ti} - < w_{ti} >\right)\left(\breve{v}_{tj} - < \breve{v}_{tj} >\right)$$

Single model answering holy-grail questions:

- What information does a spike train carry?
- What computation do dendritic trees do?
- How is supervised learning performed?
- **How is unsupervised learning performed?**
- How is information stored in synapses?
- What computation do spiking and nonspiking somas do?
- How corrupted, distorted and occluded patterns recognized?
- How are neurons organized into a neural network?

Unsupervised Covariance Learning



Hebbian-type rule: $D_{ij} \leftarrow D_{ij} + \left(u_{ti} - < u_{ti} >\right)\left(\breve{v}_{tj} - < \breve{v}_{tj} >\right)$

T. J. Sejnowski 1977: $D_{ij} \leftarrow D_{ij} + \left(u_{ti} - < u_{ti} >\right)\left(v_{tj} - < v_{tj} >\right)$

Single model answering holy-grail questions:

- What information does a spike train carry?
- What computation do dendritic trees do?
- How is supervised learning performed?
- How is unsupervised learning performed?
- **How is information stored in synapses?**
- What computation do spiking and nonspiking somas do?
- How corrupted, distorted and occluded patterns recognized?
- How are neurons organized into a neural network?

Code Covariance Matrix D *learned by* supervised learning

$$D_{ij} \leftarrow D_{ij} + \left(w_{ti} - < w_{ti} >\right)\left(\breve{v}_{tj} - < \breve{v}_{tj} >\right)$$

$$D_{ij} = \sum_{t=1}^{T} \left(w_{ti} - < w_{ti} >\right)\left(\breve{v}_{tj} - < \breve{v}_{tj} >\right)$$

Code Covariance Matrix *D* *learned by* unsupervised learning

$$D_{ij} \leftarrow D_{ij} + \left(u_{ti} - \langle u_{ti} \rangle\right)\left(\breve{v}_{tj} - \langle \breve{v}_{tj} \rangle\right)$$

$$D_{ij} = \sum_{t=1}^{T} \left(u_{ti} - \langle u_{ti} \rangle\right)\left(\breve{v}_{tj} - \langle \breve{v}_{tj} \rangle\right)$$

Decovariance Retrieval

$$d_{\tau ij} = D_{ij}\left(\breve{v}_{\tau j} - < \breve{v}_{\tau j} >\right)$$

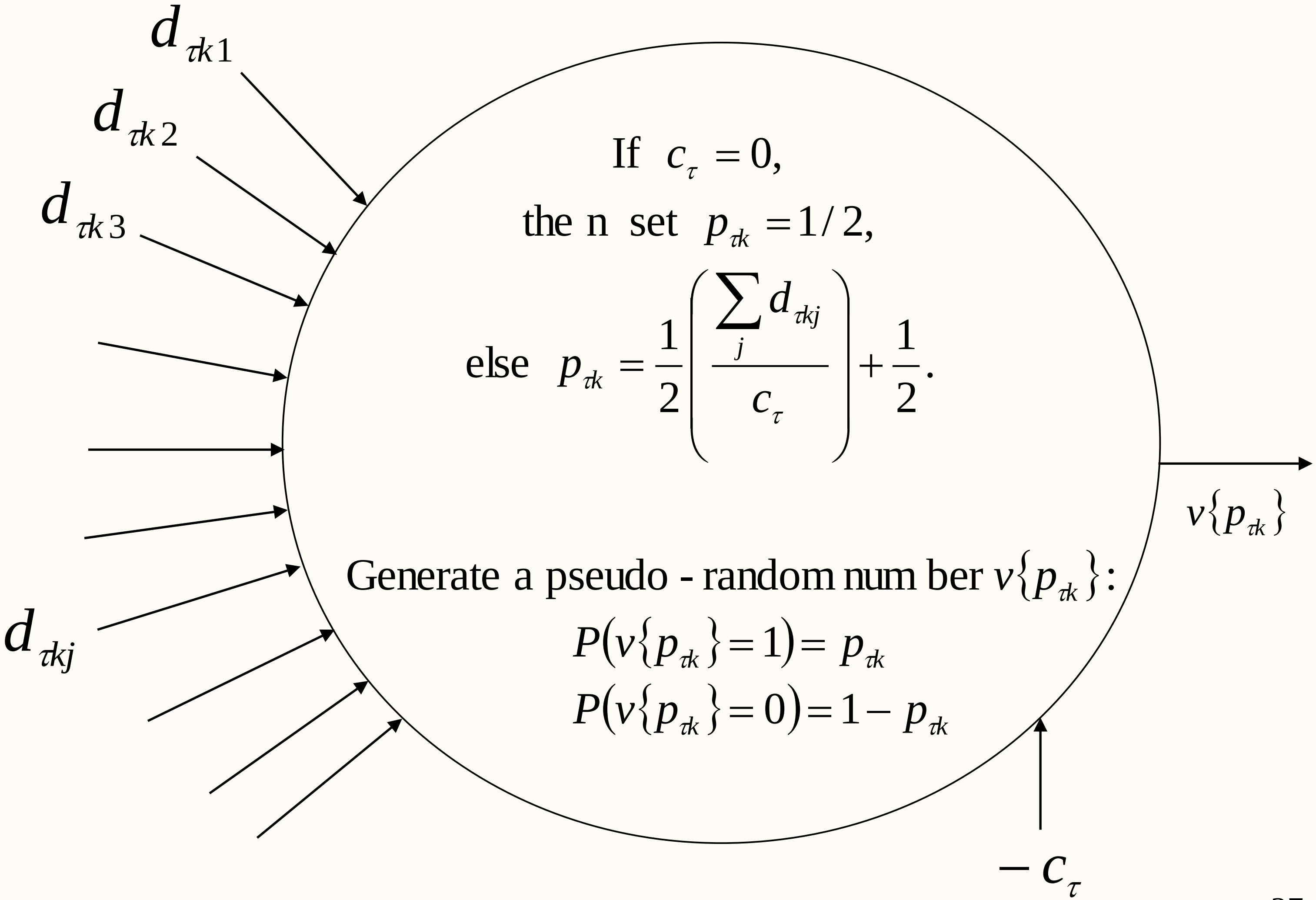
Synapse
with D_{ij}

$$\breve{v}_{\tau j} \longrightarrow \otimes \longrightarrow d_{\tau ij}$$

Single model answering holy-grail questions:

- What information does a spike train carry?
- What computation do dendritic trees do?
- How is supervised learning performed?
- How is unsupervised learning performed?
- How is information stored in synapses?
- **What computation do spiking and nonspiking somas do?**
- How corrupted, distorted and occluded patterns recognized?
- How are neurons organized into a neural network?

Spiking Soma k

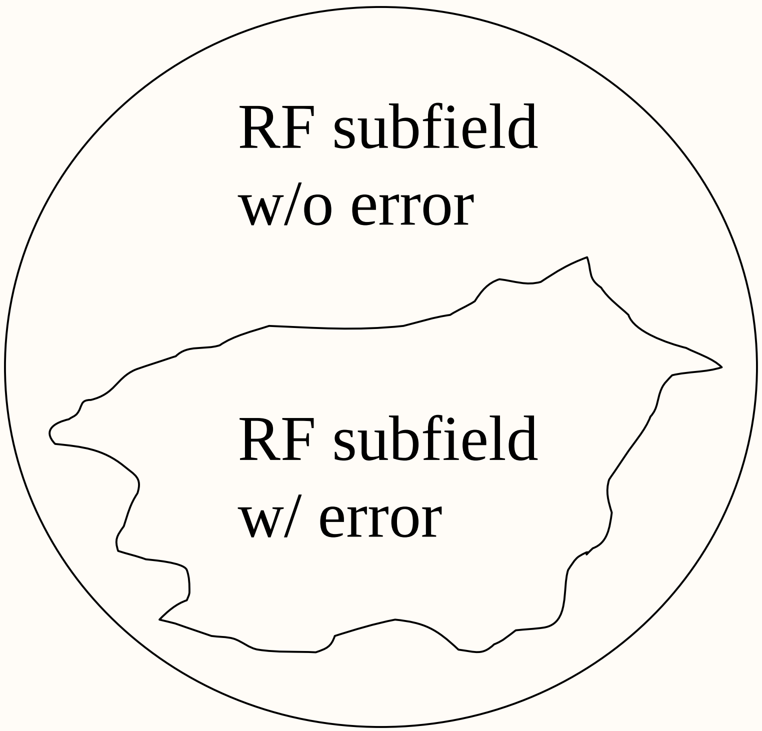


Single model answering holy-grail questions:

- What information does a spike train carry?
- What computation do dendritic trees do?
- How is supervised learning performed?
- How is unsupervised learning performed?
- How is information stored in synapses?
- What computation do spiking and nonspiking neurons do?
- **How are corrupted, distorted and occluded patterns recognized?**
- How are neurons organized into a neural network?

Maximal Generalization

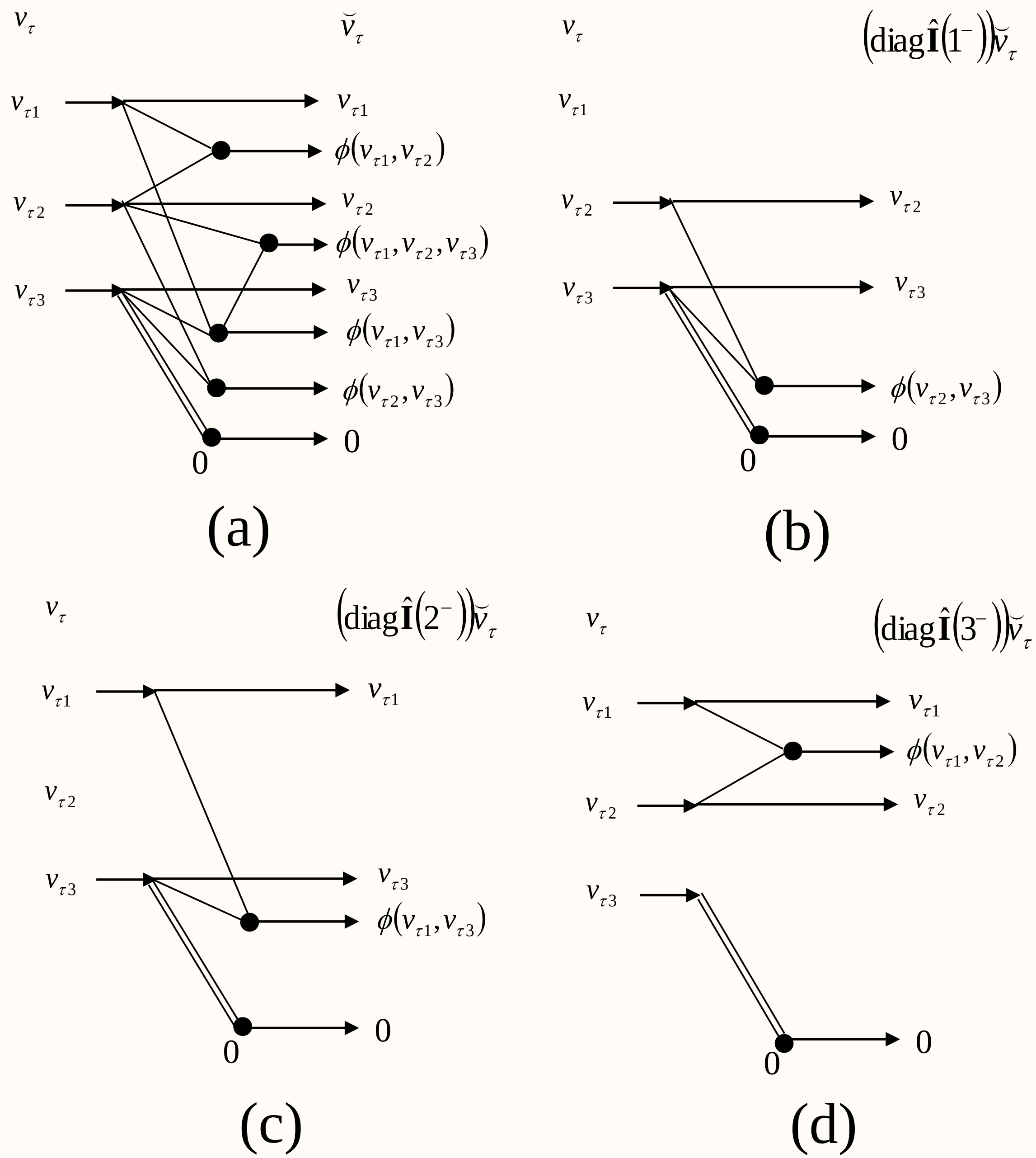
RF (Receptive Field)
of a PU



GENERALIZATION:
Using information in an RF subfield w/o error to estimate the label of the RF

MAXIMAL GENERALIZATION:
Generalization from the largest RF subfield w/o error

Maximal region <=> Best subjective probability available


v_τ
$\breve{v}_\tau$
$v_{\tau 1}$
$v_{\tau 2}$
$v_{\tau 3}$
$\phi(v_{\tau 1}, v_{\tau 2})$
$\phi(v_{\tau 1}, v_{\tau 2}, v_{\tau 3})$
$\phi(v_{\tau 1}, v_{\tau 3})$
$\phi(v_{\tau 2}, v_{\tau 3})$
0
(a)
v_τ
$\left(\mathrm{diag}\,\hat{\mathbf{I}}(1^-)\right)\breve{v}_\tau$
$v_{\tau 1}$
$v_{\tau 2}$
$v_{\tau 3}$
$\phi(v_{\tau 2}, v_{\tau 3})$
0
(b)
v_τ
$\left(\mathrm{diag}\,\hat{\mathbf{I}}(2^-)\right)\breve{v}_\tau$
$v_{\tau 1}$
$v_{\tau 2}$
$v_{\tau 3}$
$\phi(v_{\tau 1}, v_{\tau 3})$
0
(c)
v_τ
$\left(\mathrm{diag}\,\hat{\mathbf{I}}(3^-)\right)\breve{v}_\tau$
$v_{\tau 1}$
$v_{\tau 2}$
$v_{\tau 3}$
$\phi(v_{\tau 1}, v_{\tau 2})$
0
(d)

$$v = [v_1 \quad v_2 \quad v_3]'$$

$$\hat{v} = [1 \quad v_1 \quad v_2 \quad v_2 v_1 \quad v_3 \quad v_3 v_1 \quad v_3 v_2 \quad v_3 v_2 v_1]'$$

$$\hat{\mathbf{1}}(1^-) = [1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0]'$$

$$\hat{\mathbf{1}}(2^-) = [1 \quad 1 \quad 0 \quad 0 \quad 1 \quad 1 \quad 0 \quad 0]'$$

$$\hat{\mathbf{1}}(3^-) = [1 \quad 1 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0]'$$

$$\sum_{i=1}^{3} \hat{\mathbf{1}}(i^-) = [3 \quad 2 \quad 2 \quad 1 \quad 2 \quad 1 \quad 1 \quad 0]'$$

$$\operatorname{diag} \hat{\mathbf{1}}(1^-)\hat{v} = [1 \quad 0 \quad v_2 \quad 0 \quad v_3 \quad 0 \quad v_3 v_2 \quad 0]'$$

$$\operatorname{diag} \hat{\mathbf{1}}(2^-)\hat{v} = [1 \quad v_1 \quad 0 \quad 0 \quad v_3 \quad v_3 v_1 \quad 0 \quad 0]'$$

$$\operatorname{diag} \hat{\mathbf{1}}(3^-)\hat{v} = [1 \quad v_1 \quad v_2 \quad v_2 v_1 \quad 0 \quad 0 \quad 0 \quad 0]'$$

$\operatorname{diag} \hat{\mathbf{1}}(i_1^-, \cdots, i_j^-)\hat{v}$ sets the i_1th, $\cdots$, i_jth components of v in $\hat{v}$ set equal to 0

Example Masking Matrix M

$$M = I + 2^{-5} \sum_{i=1}^{3} \operatorname{diag} \hat{\mathbf{l}}\left(i^{-}\right)$$

$$M = I + 2^{-5} \begin{bmatrix} 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{a diagonal matrix}$$

Masking Matrix M

$$M = I + \sum_{j=1}^{J} 2^{-5j} \sum_{i_j=j}^{m} \cdots \sum_{i_2=2}^{i_3-1} \sum_{i_1=1}^{i_2-1} \mathrm{diag}\, \hat{\mathbf{I}}\left(i_1^-, \cdots, i_j^-\right)$$

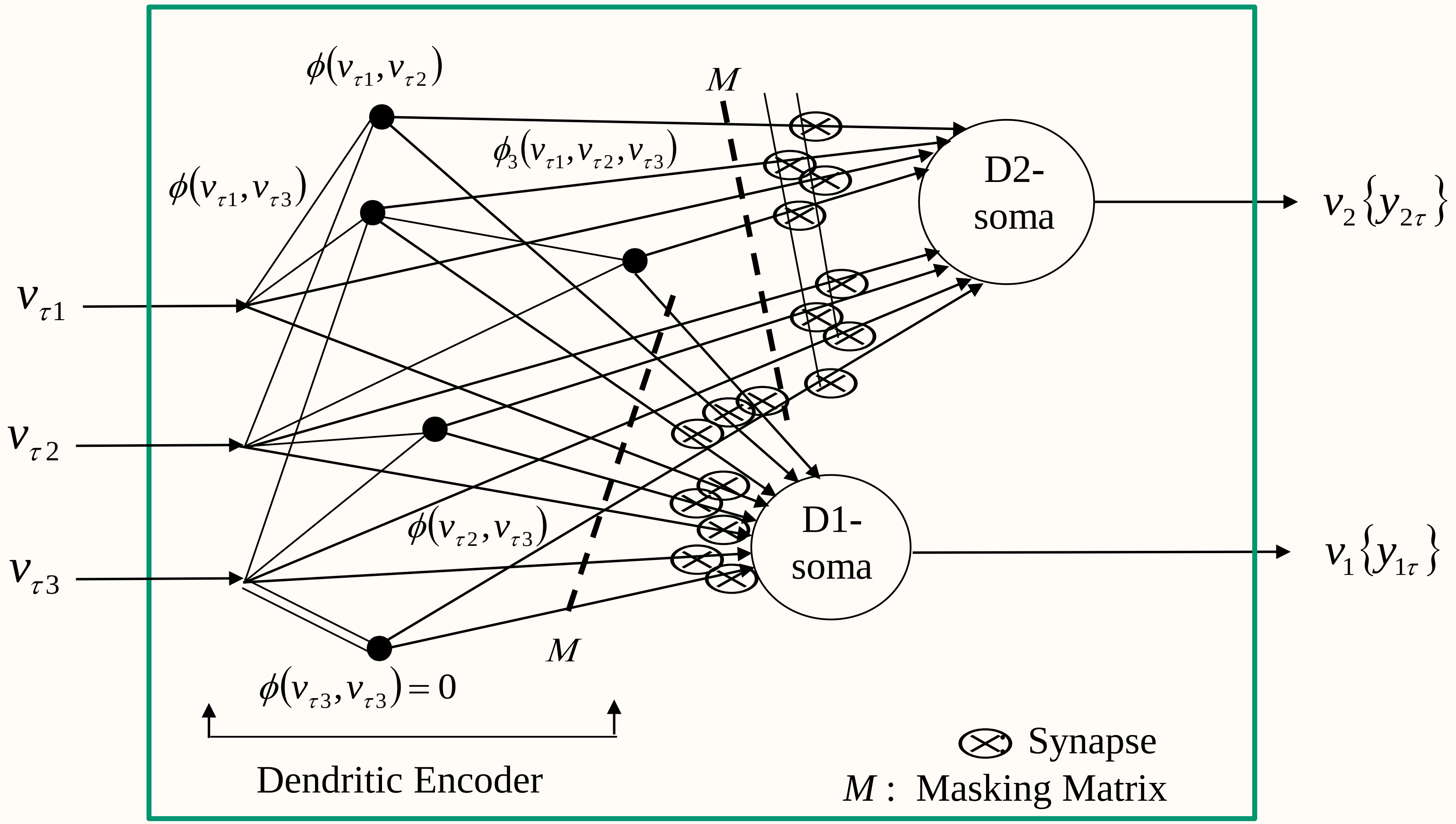
$$M = \begin{bmatrix} M_{11} & 0 & \cdots & 0 \\ 0 & M_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & M_{qq} \end{bmatrix}$$ a diagonal matrix

Generalization from the largest receptive subfield that matches a stored subfield

Single model answering holy-grail questions:

- What information does a spike train carry?
- What computation do dendritic trees do?
- How is supervised learning performed?
- How is unsupervised learning performed?
- How is information stored in synapses?
- What computation do spiking and nonspiking somas do?
- How corrupted, distorted and occluded patterns recognized?
- **How are neurons organized into a neural network?**

Processing Unit

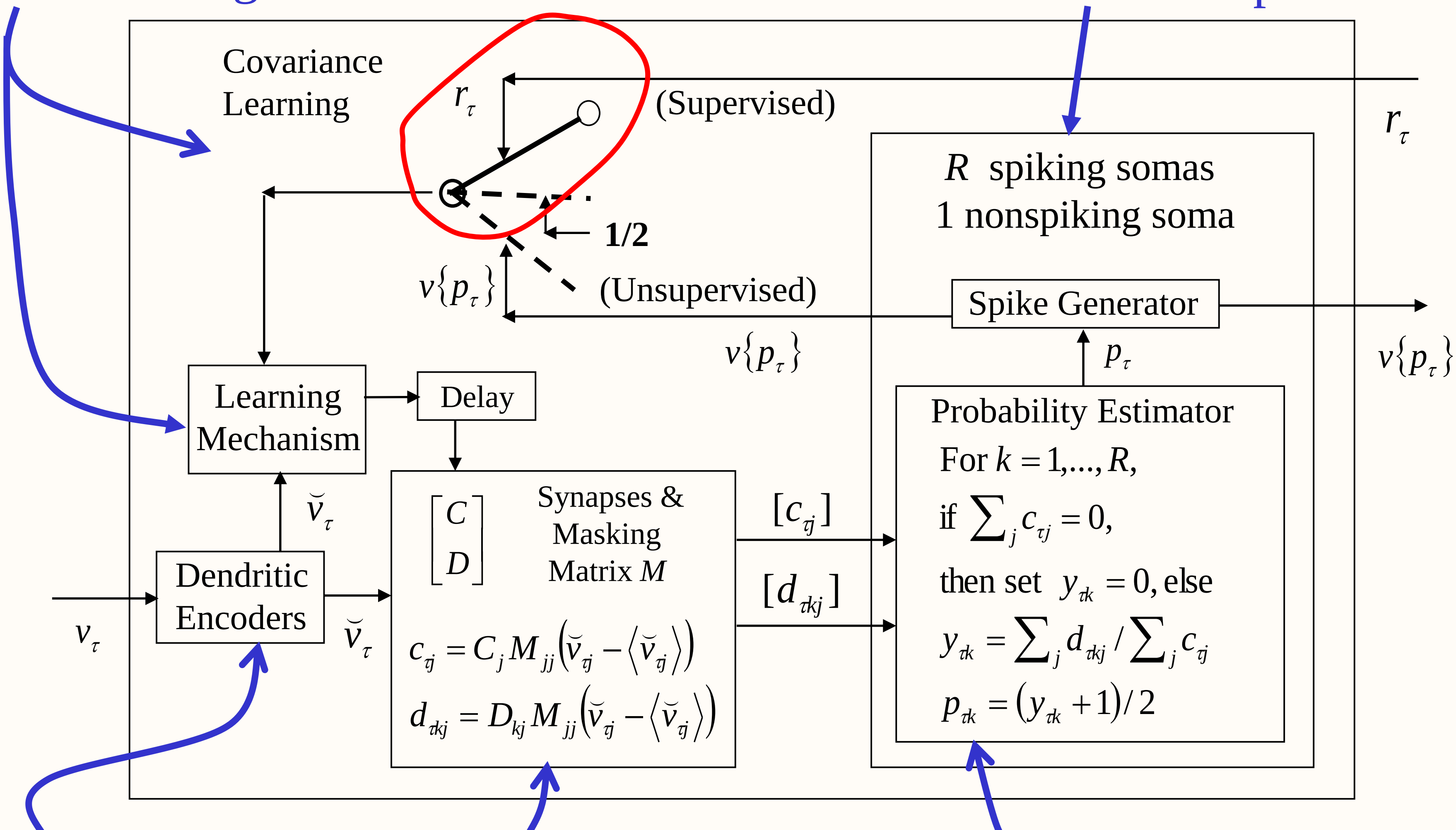


A pattern recognizer

Supervised Processing Unit (SPU)

2. Learning

6. Generation of spikes



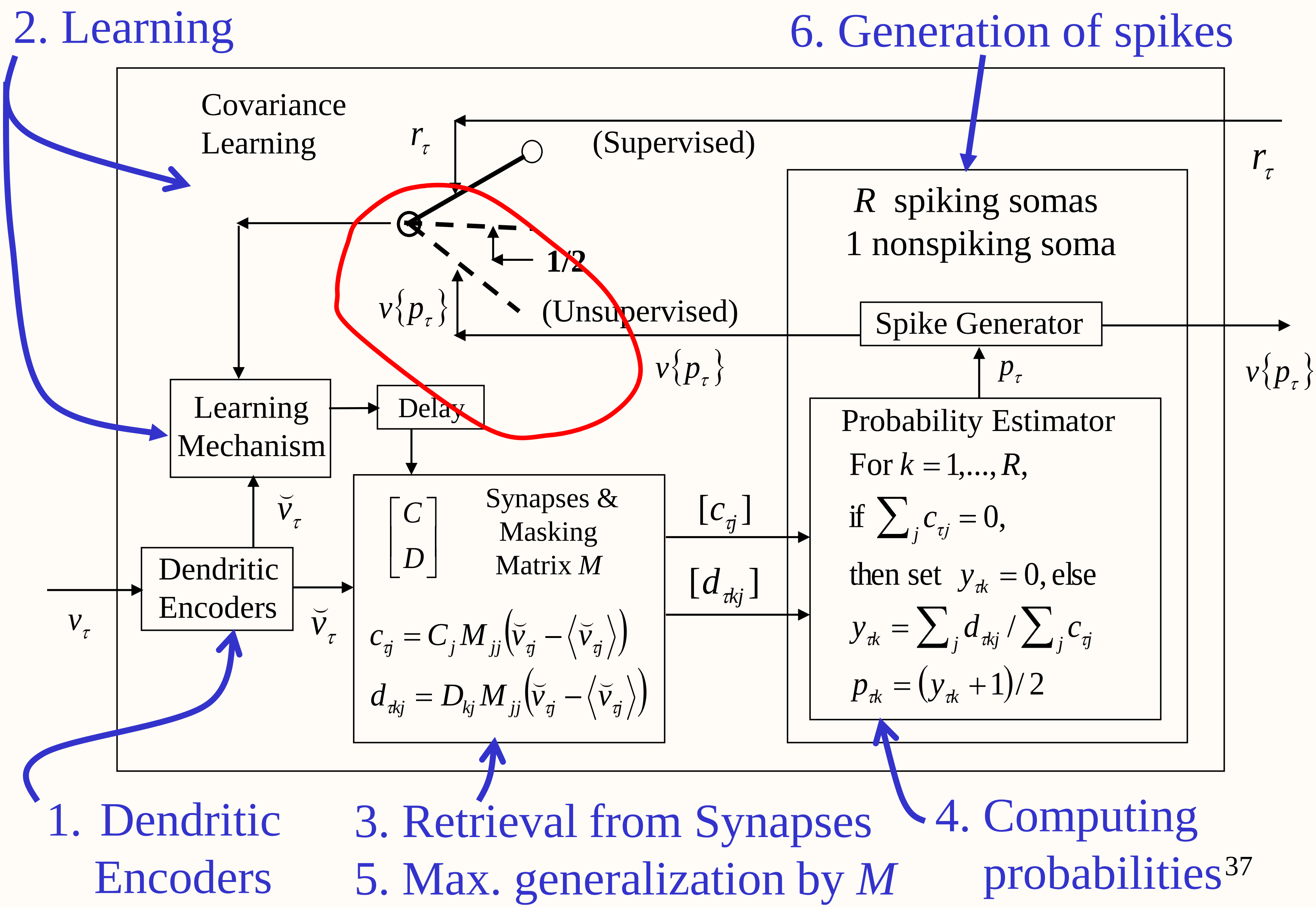
1. Dendritic Encoders

3. Retrieval from Synapses

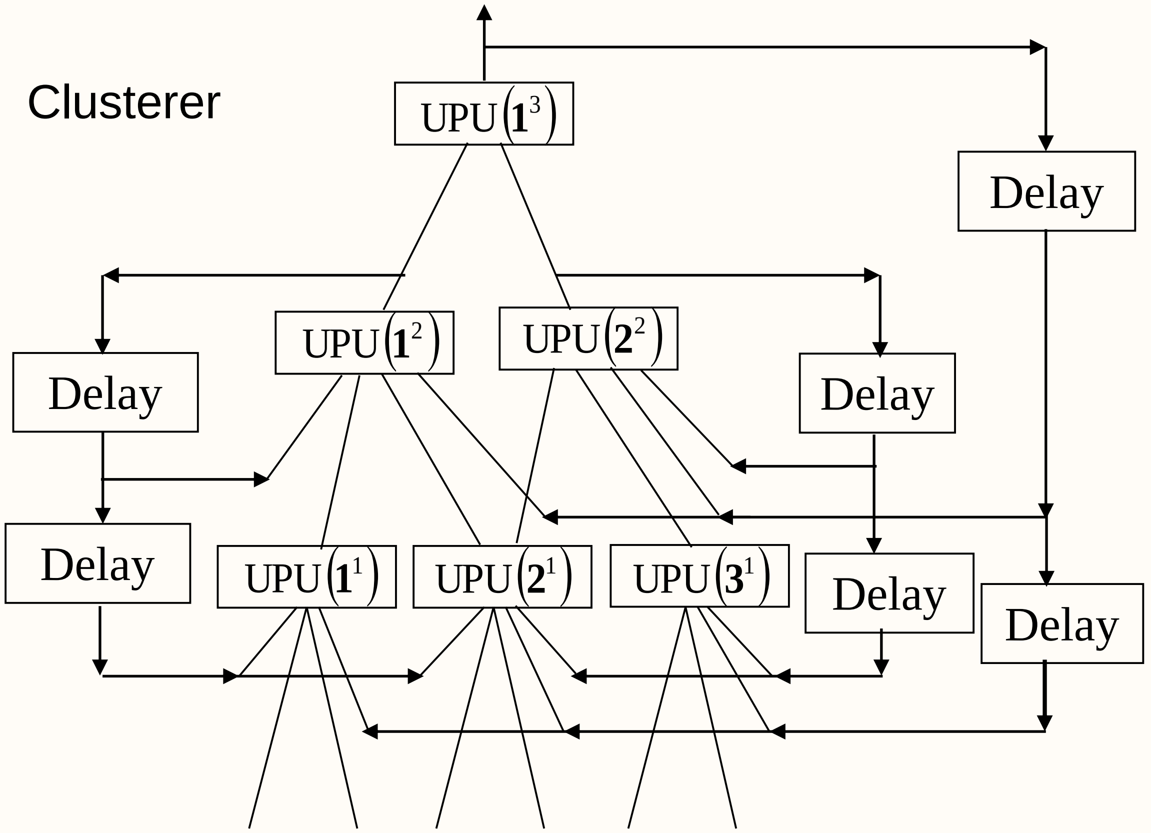
5. Max. generalization by *M*

4. Computing probabilities

Unsupervised Processing Unit (UPU)

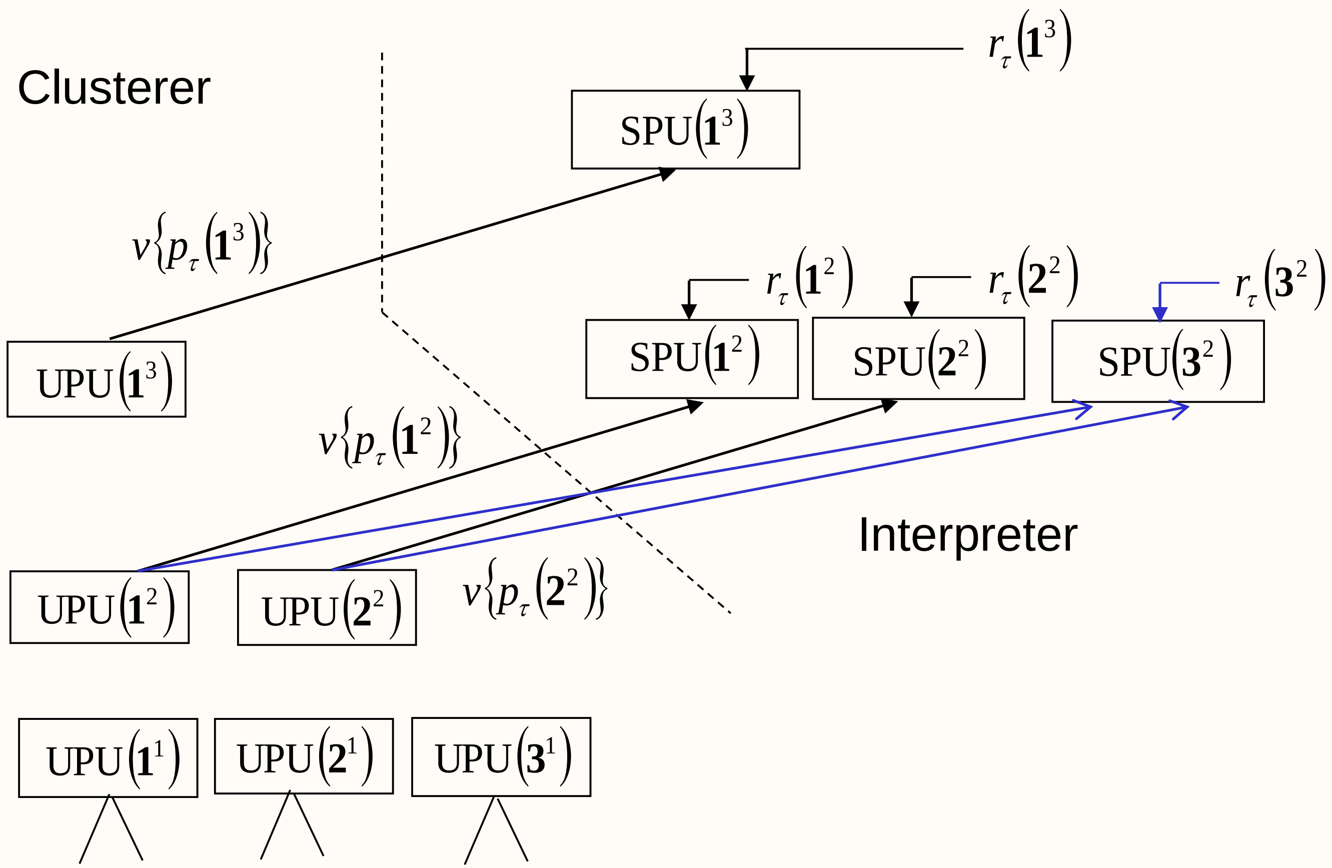


Clusterer



A Network of Unsupervised Processing Units (UPUs)

Offshoot Supervised Processing Units (SPUs)



Current and Future Work

- Applications

  - Spatial Pattern Recognition
    Handwriting, face, target, fingerprint, DNA, smell, taste, explosive/weapon (in baggage, containers)

  - Temporal Pattern Understanding/Classification
    Touch, speech, text, video (computer vision), financial data

- Theory

  - Extension to visual, auditory, gustatory, olfactory, somatosensory, and somatomotor systems

  - Motion detection, attention selection, prediction

Learning Machine for **BIG** Data

DLMs including CNN are present workhorses.

They are inadequate.

Capabilities needed for learning big data

1. Handcrafting labels impossible for big data
   ⟹ Learning without supervision
2. Big data too big for iterative optimization
   ⟹ Learning with photographic memory
3. Big data streaming in
   ⟹ Online learning
4. Big data not all conditioned for processing
   ⟹ Maximal generalization (treating noise, distortion, occlusion, translation, scaling, etc.)
5. Big data containing info about hierarchical worlds
   ⟹ Learning the hierarchical worlds (Recall the success of CNN.)
6. Big data containing temporal data (e.g., videos, texts)
   ⟹ Learning time series

Wish list!!

Capabilities of LOM

1. Learning data w/o handcrafted labels
2. Learning big data with photographic memory
3. Learning streaming data online
4. Maximal generalization (treating noise, distortion, occlusion, translation, scaling, etc.)
5. Learning the hierarchical worlds

   (Recall the success of CNN.)
6. Learning temporal big data (e.g., videos, texts)

Wish list!! **Fulfilled!!**

LOM for finding all the gene mutations
that cause each type of disease
and their empirical probabilities

jameslo@umbc.edu

Thank you!

Questions, comments, suggestions?

If you are interested, please talk to or email me.

jameslo@umbc.edu