Adaptive Resonance Theory

ART Networks

Stephen Grossberg &

Gail Carpenter

ART Network Varieties

- Several varieties:
  - ART1: Discrete (e.g. binary) patterns
  - ART2: Continuous patterns …
  - ARTMAP: supervised learning
  - Fuzzy Art: Fuzzy version of ART1

- All work by **competitive learning** and a type of **clustering**, represented by stored prototypes, with other nuances

Stability-Plasticity Dilemma

- **stability**: Recognized patterns should be insensitive to noise.

- **plasticity**: System should be capable of learning new patterns.

- The conflict between these is one of the things that ART hopes to resolve.

ART Networks

- Combine supervised and unsupervised (competitive, clustering)
- Dynamically create new categories
- Biologically motivated by an ODE model
- Models short- and long-term memory

Competitive Learning in ART

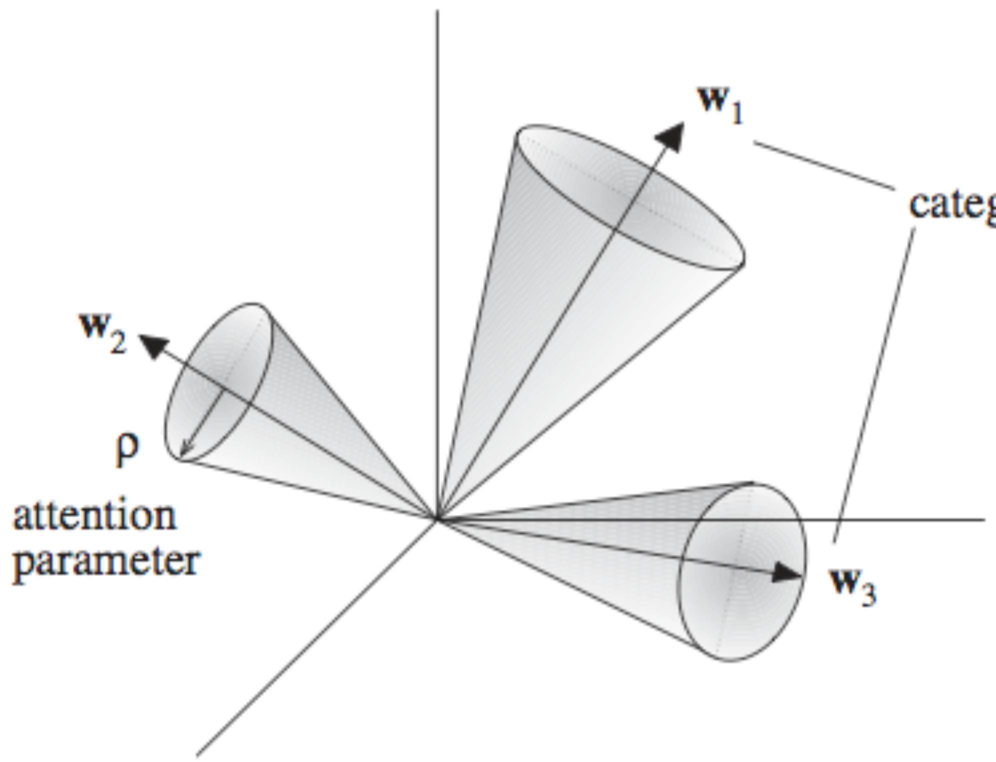
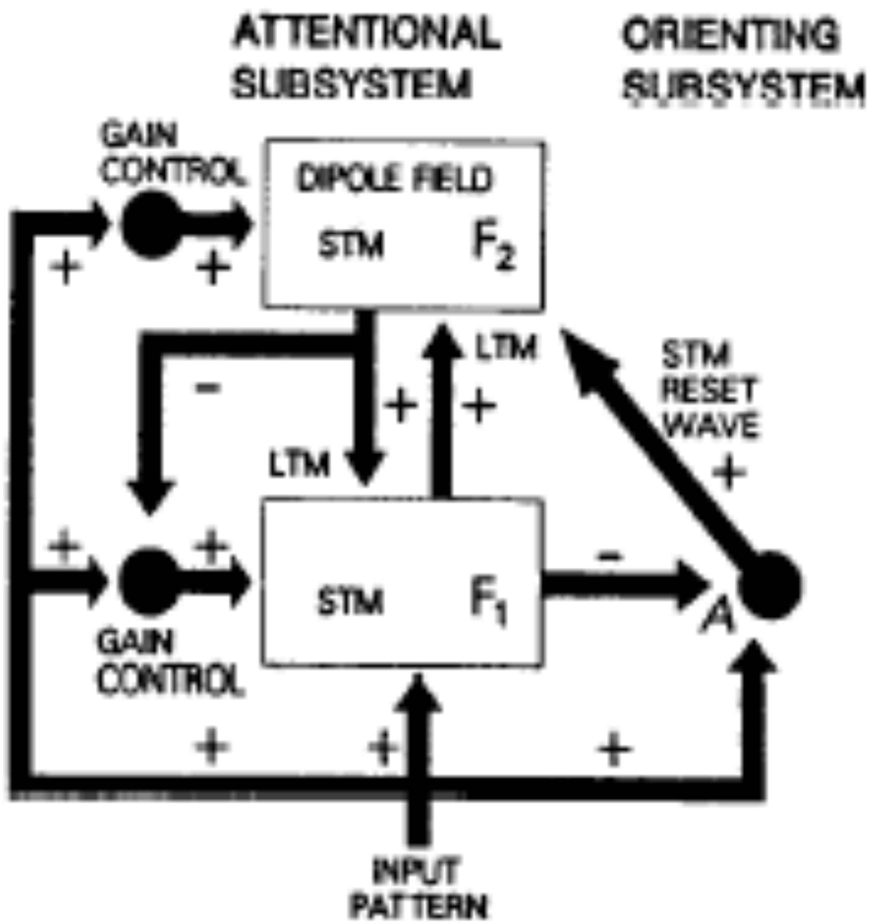


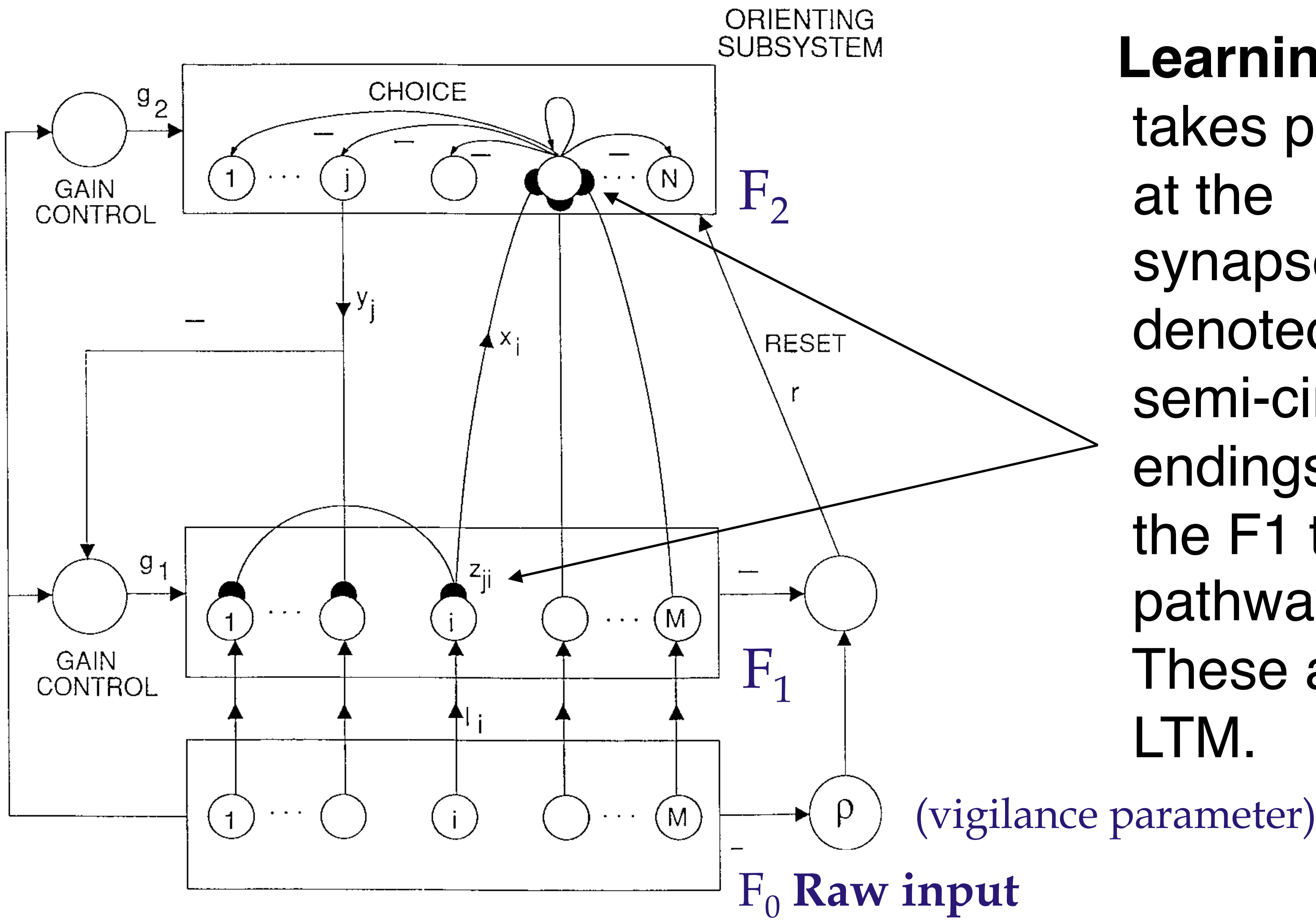
Fig. 16.2. Vector clusters and attention parameters

from Rojas' book

Diagram from Carpenter and Grossberg, 1988



ART1 Model



Learning takes place at the synapses denoted by semi-circular endings in the F1 to F2 pathways. These are LTM.

ART Structure Simplified

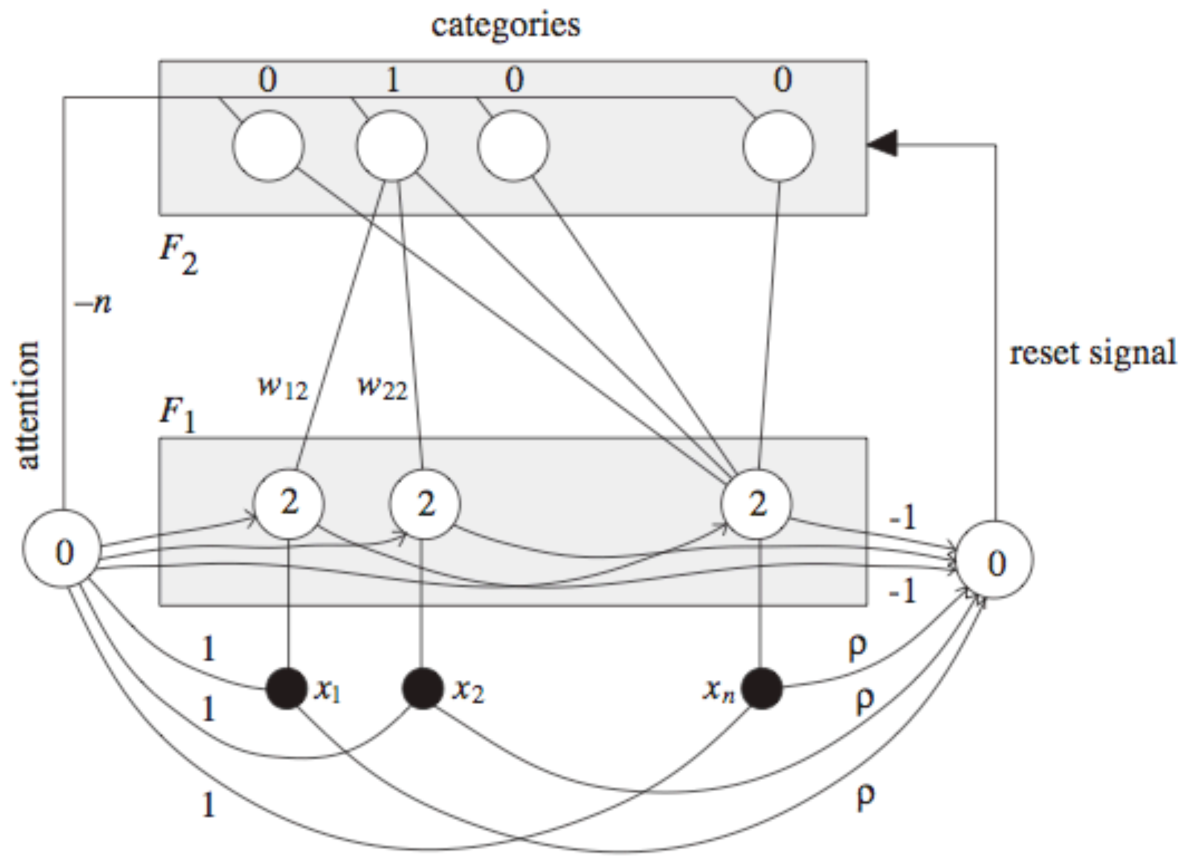


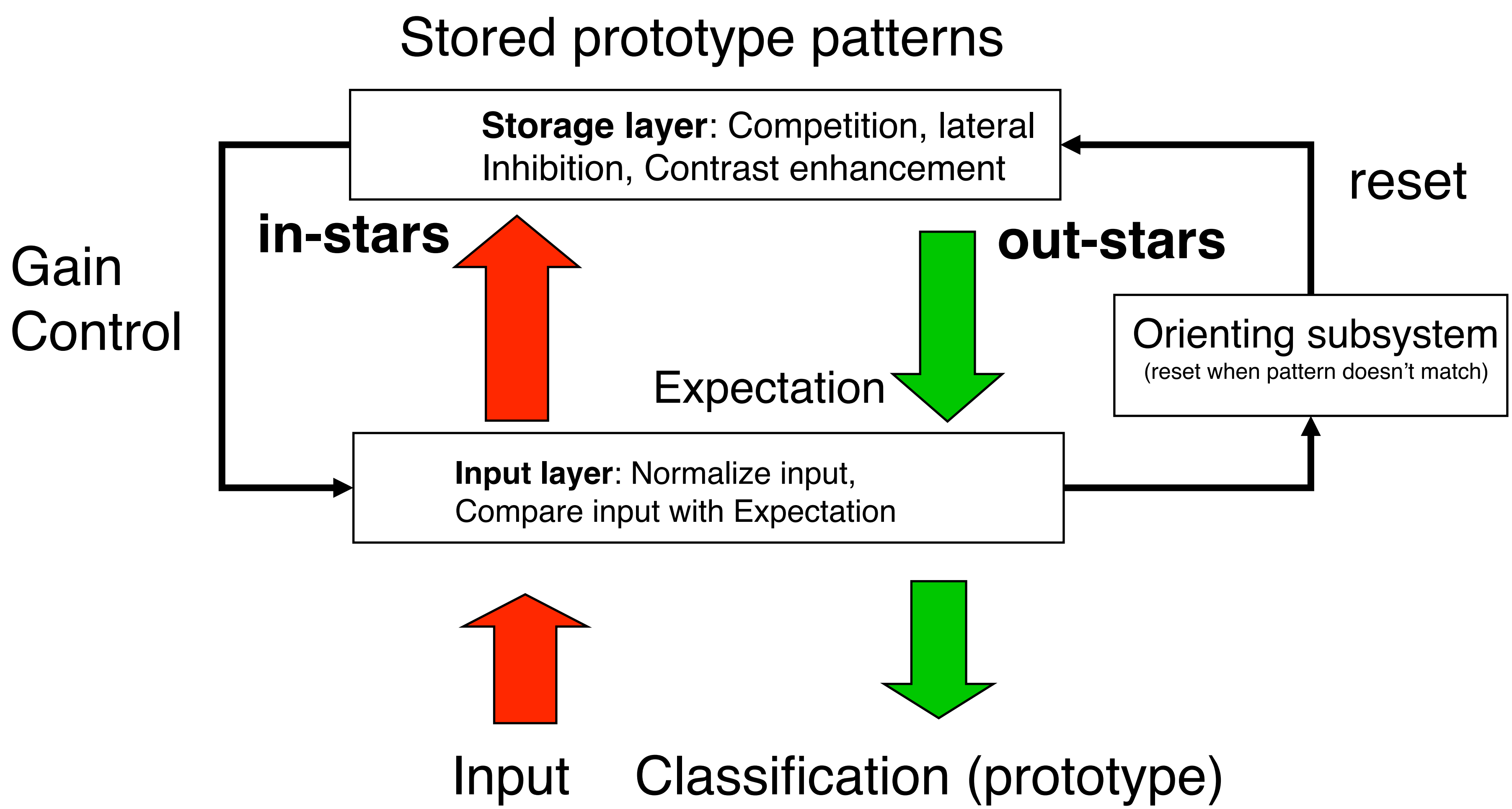
Fig. 16.3. The ART-1 architecture

from Rojas' book

ART General Workings

- F_2 activities can be interpreted as “making a hypothesis” about input at F_1.
- The LTM (Long-Term Memory)-gated signals from all the active F_2 nodes are added to generate the total **top-down feedback pattern** from F_2 to F_1.
- This pattern plays the role of a **learned expectation**. Activation of this expectation “tests the hypothesis” of the active F_2 category.
- The expected prototype of the category are compared with the bottom-up input pattern (exemplar) to F_1.

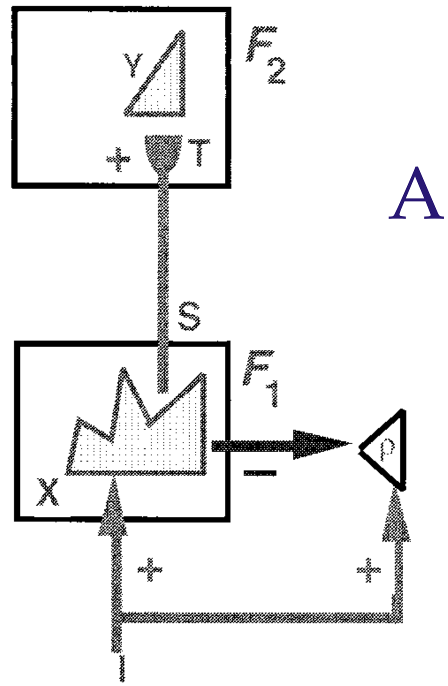
ART: instars and outstars



ART search for a recognition code (A of 4)

[Grossberg and Merrill (1996)]

- The input pattern I is instated across the feature detectors at level F_1 as a short-term memory (STM) activity pattern **X** represented by the hatched pattern across F_1.
- Input I also **nonspecifically** activates the orienting subsystem A
- Pattern X both inhibits A and generates the output pattern S from F_1.
- Pattern S is multiplied by long term memory (LTM) traces and added at F_2 nodes to form the input pattern T, which activates the STM pattern **Y** across the recognition categories coded at level F_2.

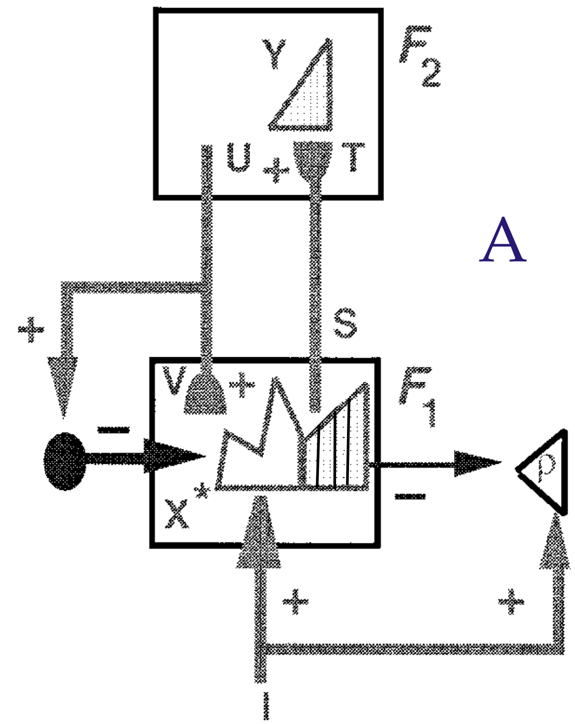


ART search for a recognition code (B of 4)

- Pattern Y generates the top-down output pattern U, which is multiplied by top-down LTM traces and added at F_1 nodes to form the **prototype pattern V** that encodes the **learned expectation** of the active F_2 nodes.

- If **V mismatches I** at F_1, then a new STM activity pattern X* is generated at F_1. X* is represented by the **hatched pattern** and includes the features of I that are **confirmed** by V. Inactivated nodes corresponding to unconfirmed features of X are unhatched.

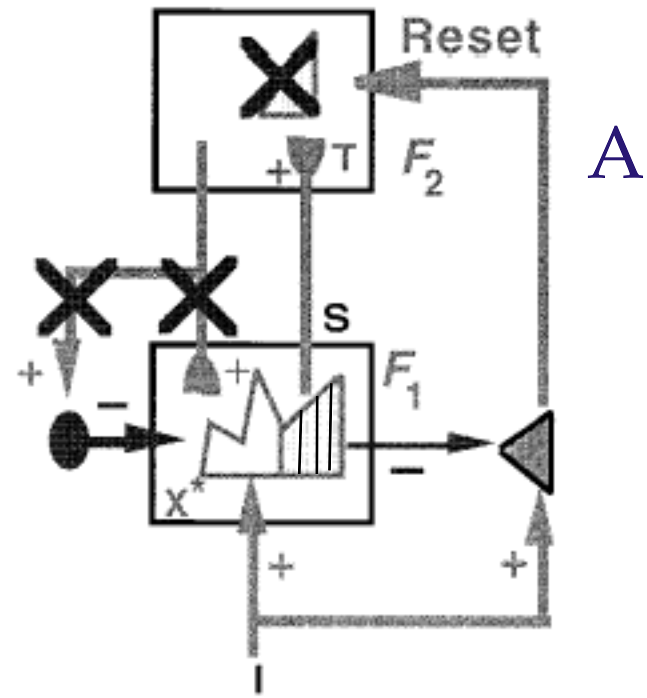
- The reduction in total STM activity, which occurs when X is transformed into X* causes a **decrease in the total inhibition** from F_1 to A.



ART search for a recognition code (C of 4)

- If inhibition decreases sufficiently

  [vigilance condition not met],
  A releases at a nonspecific arousal wave to F_2,
  which **resets** the STM pattern Y at F_2.



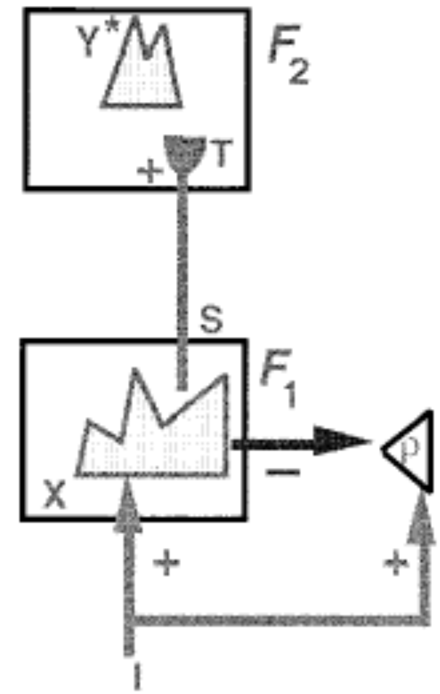
ART search for a recognition code (D of 4)

- After Y is inhibited, its top-down prototype signal is eliminated, and **X can be reinstated** at F_1.

- **Enduring traces** of the prior reset lead X to activate a **different** STM pattern Y* at F_2.

- If the top-down prototype due to Y* also mismatches I at F_1, then the **search** for an appropriate F_2 code continues ***until a more appropriate*** F_2 representation is selected.

- Then an attentive **resonance** develops and learning of the attended data is initiated.



Hypothesis Testing

- In an ART model, learning does not occur as soon as some winning F_2 activities are stored in STM. Instead, F_2 activities can be interpreted as **"making a hypothesis"** about input at F_1.

- The LTM-gated signals from all the active F_2 nodes are added to generate the total top-down feedback pattern from F_2 to F_1. This pattern plays the role of a learned expectation. Activation of this expectation **"tests the hypothesis"** of the active F_2 category.

- The expected prototype of the category are compared with the bottom-up input pattern (exemplar) to F_1.

ART Learning

- The **resonant state**, rather than bottom-up activation, drives the learning process (thus "adaptive resonance" theory):
  - "resonance" = **mutual reinforcement** between input and storage layers
  - "adaptive" = weights are adjusted when **resonance** occurs
- ART systems learn prototypes rather than exemplars because the attendant **feature vector *X**** rather than the input **exemplar itself** is learned.

ART1 Viewpoints

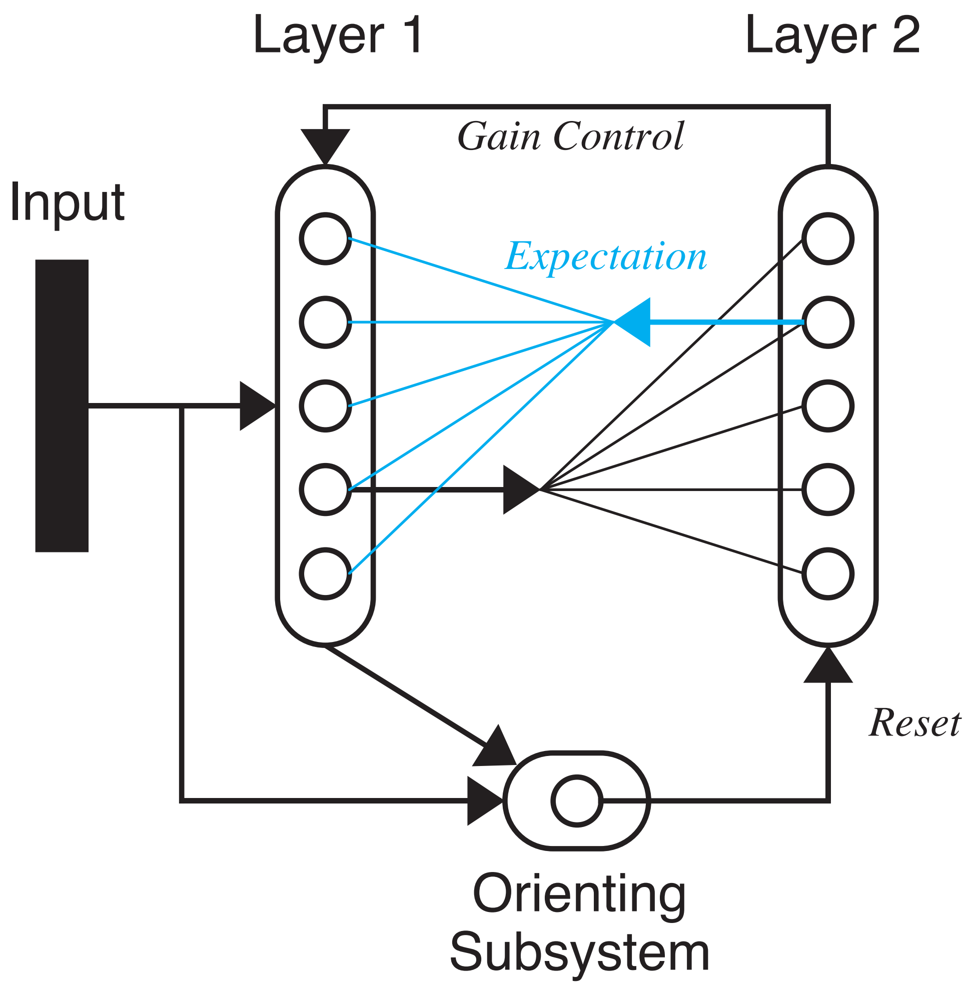
- Two kinds of explanations:
  - neural - as just seen
  - algorithmic

- The first is more complicated, since it involves neural explanations for the control aspects of the algorithmic approach.

Basic ART *Algorithm*

- Input pattern presented to input layer.
- Storage layer indicates tentative hypothetical classification.
- Input layer decides if hypothetical is **close enough**; if so, done.
- If not, storage layer indicates alternate hypothesis.
- The above two steps are repeated until the hypothetical classification is accepted.
- *All* hypotheses could be rejected; in this case, a **new** class is created in the storage layer.

Learning Laws: L1-L2 and L2-L1



The ART1 network has two separate learning laws: one for the L1-L2 connections (instars) and one for the L2-L1 connections (outstars).

Both sets of connections are updated at the same time - when the input and the expectation have an adequate match.

The process of matching, and subsequent adaptation is referred to as resonance.

ART1 Algorithm Summary (NND Book)

0) All elements of the initial $\mathbf{W}^{2:1}$ matrix are set to 1. All elements of the initial $\mathbf{W}^{1:2}$ matrix are set to $\zeta/(\zeta+S^1-1)$ [$\zeta > 1$, S^1 = # of neurons in layer 1].

1) Input pattern is presented. Since Layer 2 is not active,

$$\mathbf{a}^1 = \mathbf{p}$$

2) The input to Layer 2 is computed, and the neuron with the largest input is activated.

$$a_i^2 = \begin{cases} 1\ , & \text{if}\left(({}_i\mathbf{w}^{1:2})^T\mathbf{a}^1 = max[({}_k\mathbf{w}^{1:2})^T\mathbf{a}^1]\right) \\ 0\ , & \text{otherwise} \end{cases}$$

In case of a tie, the neuron with the smallest index is the winner.

3) The L2-L1 expectation is computed.

$$\mathbf{W}^{2:1}\mathbf{a}^2 = \mathbf{w}_j^{2:1}$$

Summary Continued

4) Layer 1 output is adjusted to include the L2-L1 expectation.

$$\mathbf{a}^1 = \mathbf{p} \cap \mathbf{w}_j^{2:1}$$

5) The orienting subsystem determines match between the expectation and the input pattern.

$$a^0 = \begin{cases} 1, & \text{if } \left[\left\| \mathbf{a}^1 \right\|^2 / \left\| \mathbf{p} \right\|^2 < \rho \right] \\ 0, & \text{otherwise} \end{cases}$$

6) If $a^0 = 1$, then set $a^2_j = 0$, **inhibit** it until resonance, and return to Step 1. If $a^0 = 0$, then continue with Step 7.

7) Resonance has occured. Update row j of $\mathbf{W}^{1:2}$.

$$_j\mathbf{w}^{1:2} = \frac{\zeta \mathbf{a}^1}{\zeta + \left\| \mathbf{a}^1 \right\|^2 - 1}$$

8) Update column j of $\mathbf{W}^{2:1}$.

$$\mathbf{w}_j^{2:1} = \mathbf{a}^1$$

9) Remove input, restore inhibited neurons, and return to Step 1.

ART1 Issues

- Subset-Superset dilemma:
  - If one pattern is **contained in** another, then a given input may have the same inner product with two different prototypes.
  - Can be resolved by allowing weights other than {0, 1} and **normalizing** the prototypes.

Subset/Superset Dilemma

Suppose that $\mathbf{W}^{1:2} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ so the prototypes are ${}_1\mathbf{w}^{1:2} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ ${}_2\mathbf{w}^{1:2} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

We say that ${}_1\mathbf{w}^{1:2}$ is a subset of ${}_2\mathbf{w}^{1:2}$, because ${}_2\mathbf{w}^{1:2}$ has a 1 wherever ${}_1\mathbf{w}^{1:2}$ has a 1.

If the output of layer 1 is $\mathbf{a}^1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ then the input to Layer 2 will be

$$\mathbf{W}^{1:2}\mathbf{a}^1 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

Both prototype vectors have the same inner product with $\mathbf{a}^1$, even though the first prototype is identical to $\mathbf{a}^1$ and the second prototype is not. This is called the *Subset/Superset* dilemma.

Subset/Superset Solution

Normalize the prototype patterns.

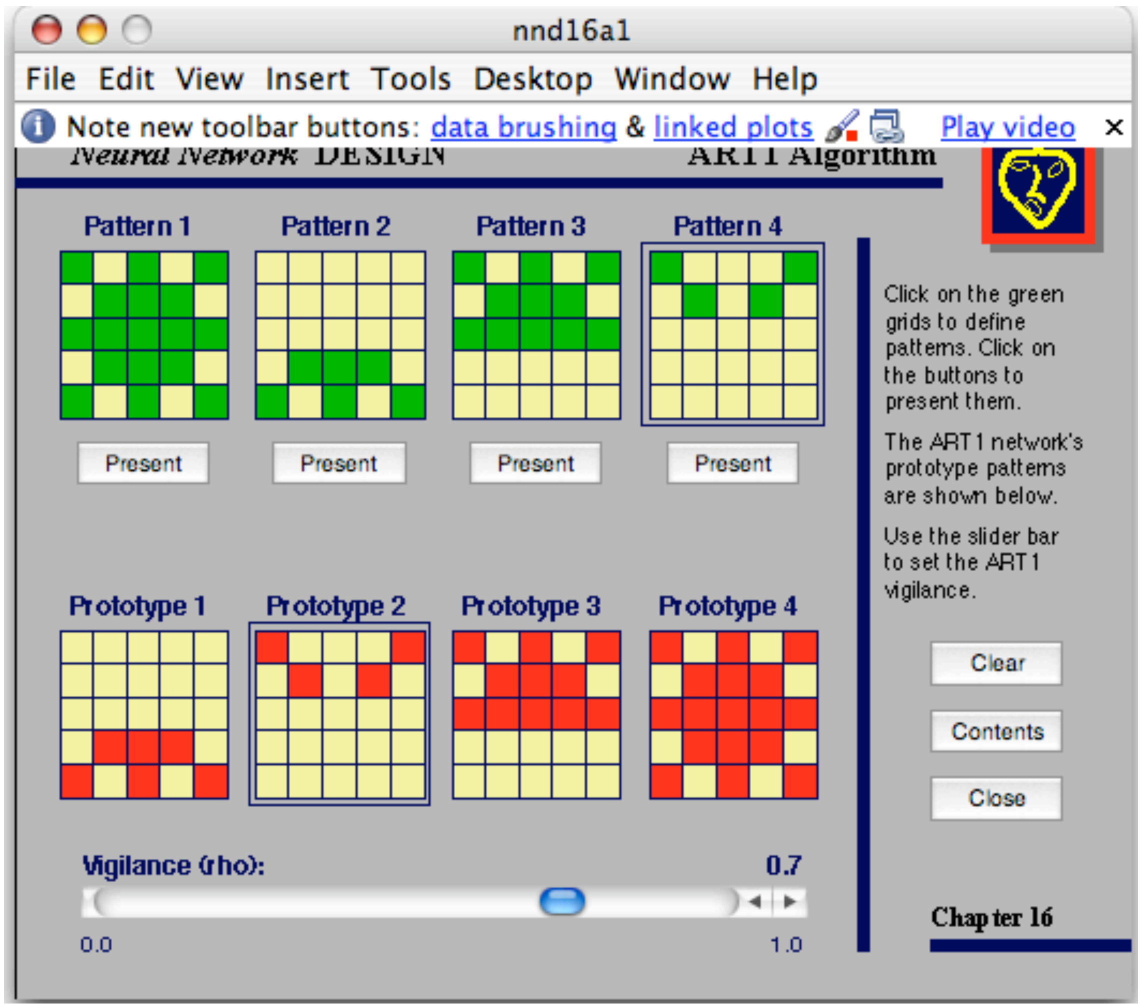
$$\mathbf{W}^{1:2} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

$$\mathbf{W}^{1:2}\mathbf{a}^{1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{2}{3} \end{bmatrix}$$

Now we have the desired result; the first prototype has the largest inner product with the input.

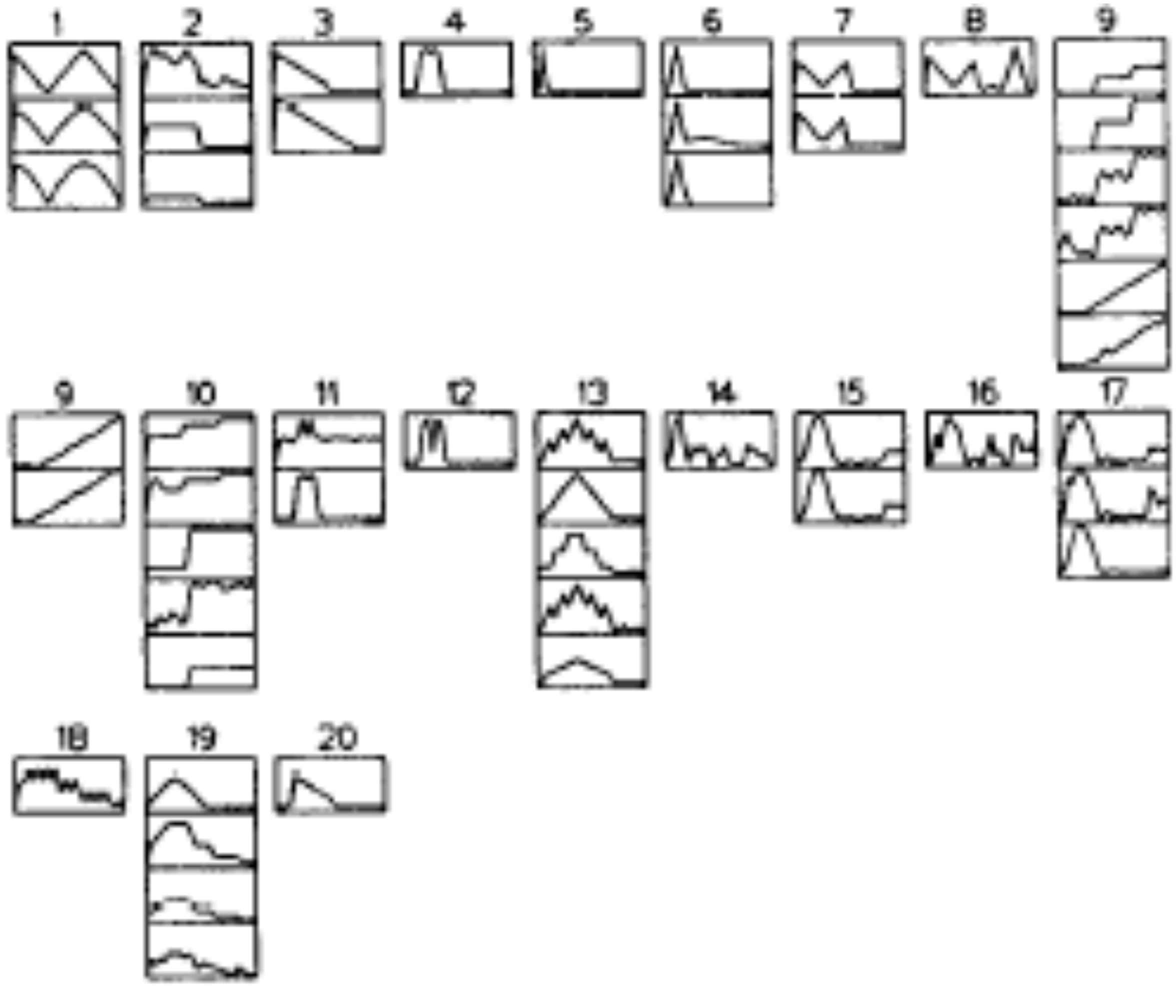
Matlab ART1 Demo (nnd16a1)

Increasing **vigilance** causes the network to be more selective, introducing a new prototype when the fit is not good.

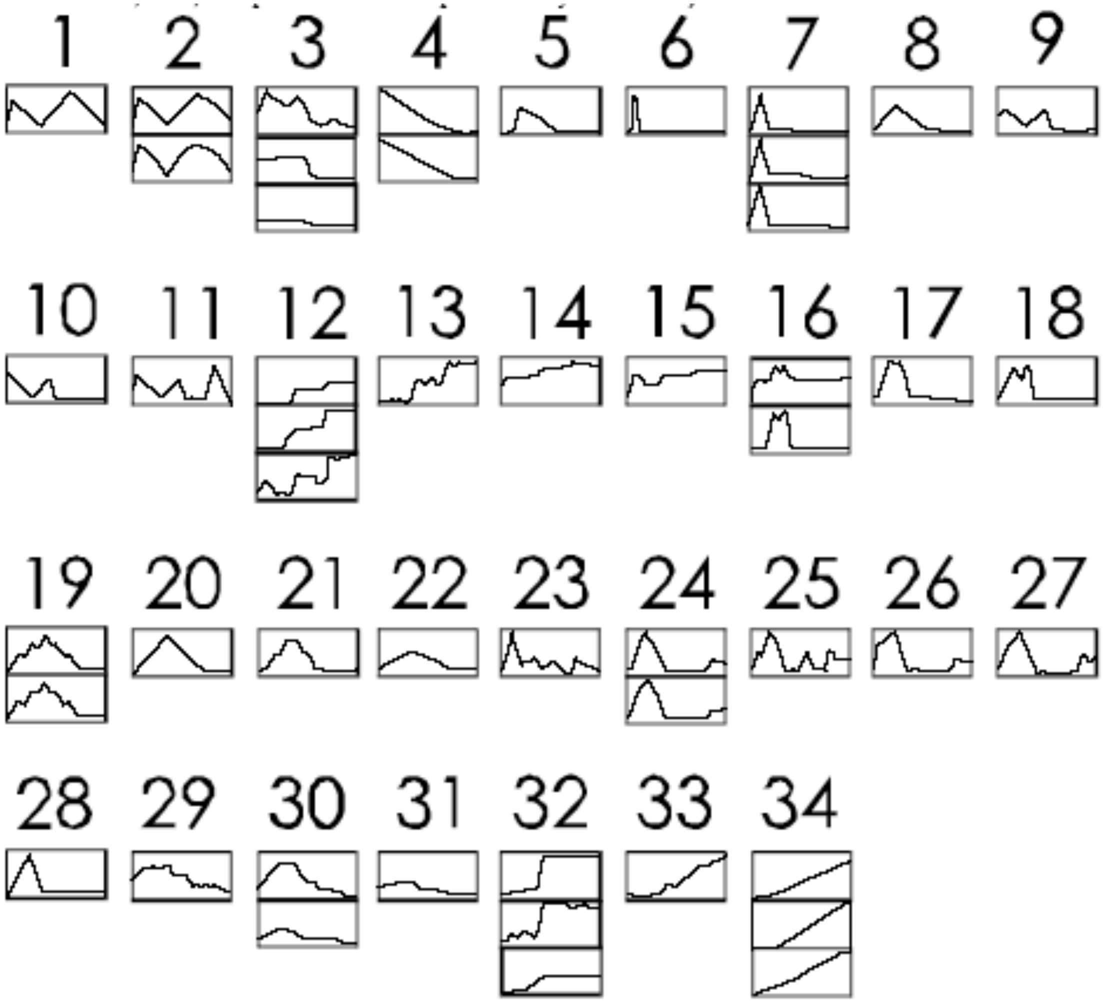


Try different patterns

ART2 Clustering (low vigilance)

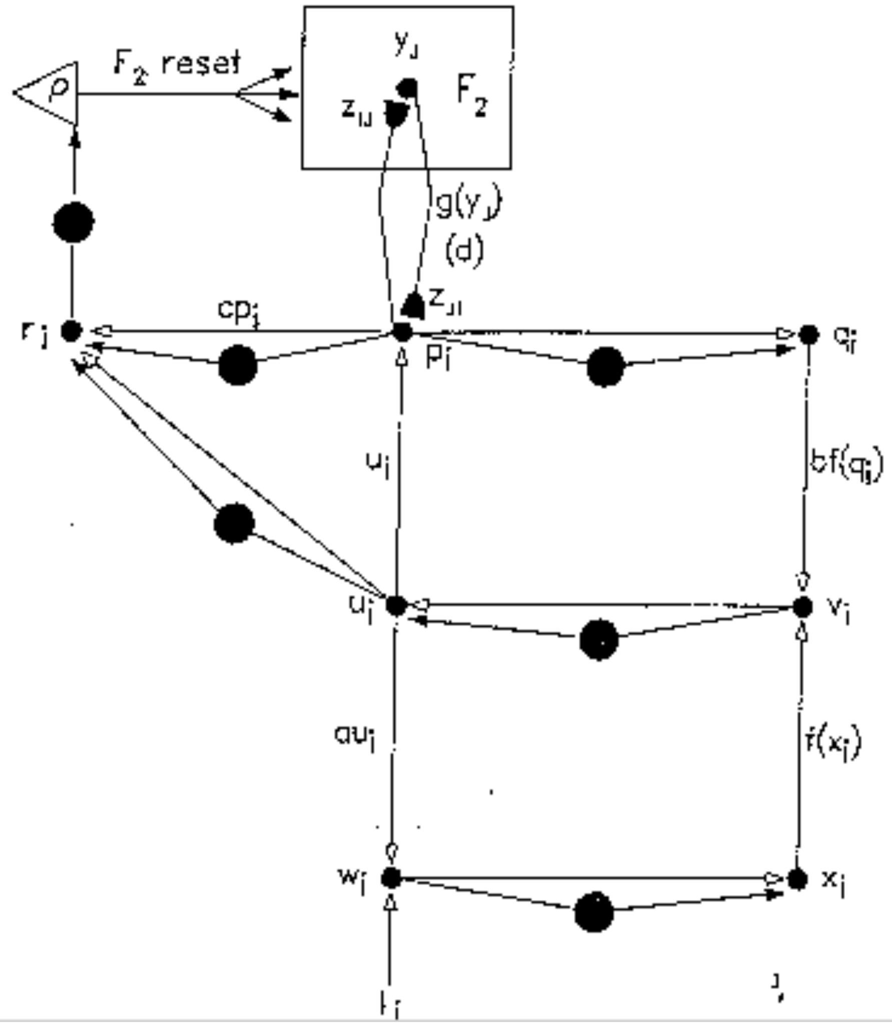


ART2 Clustering (higher vigilance)



ART2 replaced with Fuzzy Art

- ART2 tries to generalize the ART1 properties from binary patterns to analog patterns.
- Basically it expands the F_1 layer into three layers to incorporate bottom-up and top-down matching ability for analog patterns.
- ART2 is too complicated and is no longer alive since Fuzzy -ART does the binary-to -analog switch in a very straight-forward and simple way.



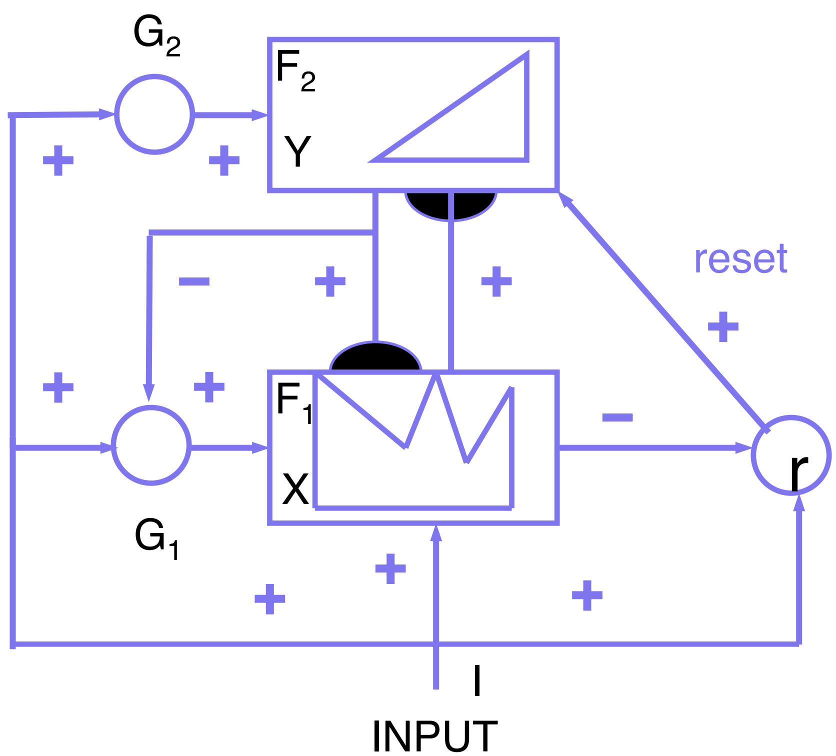
Fuzzy ART

Learning:
$z_J = I \wedge z_j^{old}$

$Z_J = \dfrac{(I \wedge z_j^{old})}{(\beta + |I \wedge z_j^{old}|)}$

Initialization:
$Z_{ij}(0) = \alpha_j$
where
$\alpha_1 > \alpha_2 > \ldots\ldots > \alpha_N$
where
$0 < \alpha_j < 1/(\beta + |I|)$ for
$\beta > 0$

$z_{ji}(0) = 1$



F_2 node activation:
$T_j = |I|\, \alpha_j$ if j is an uncommitted

$T_j = |I \wedge z_j| / (\beta + |z_j|)$ if j is committed

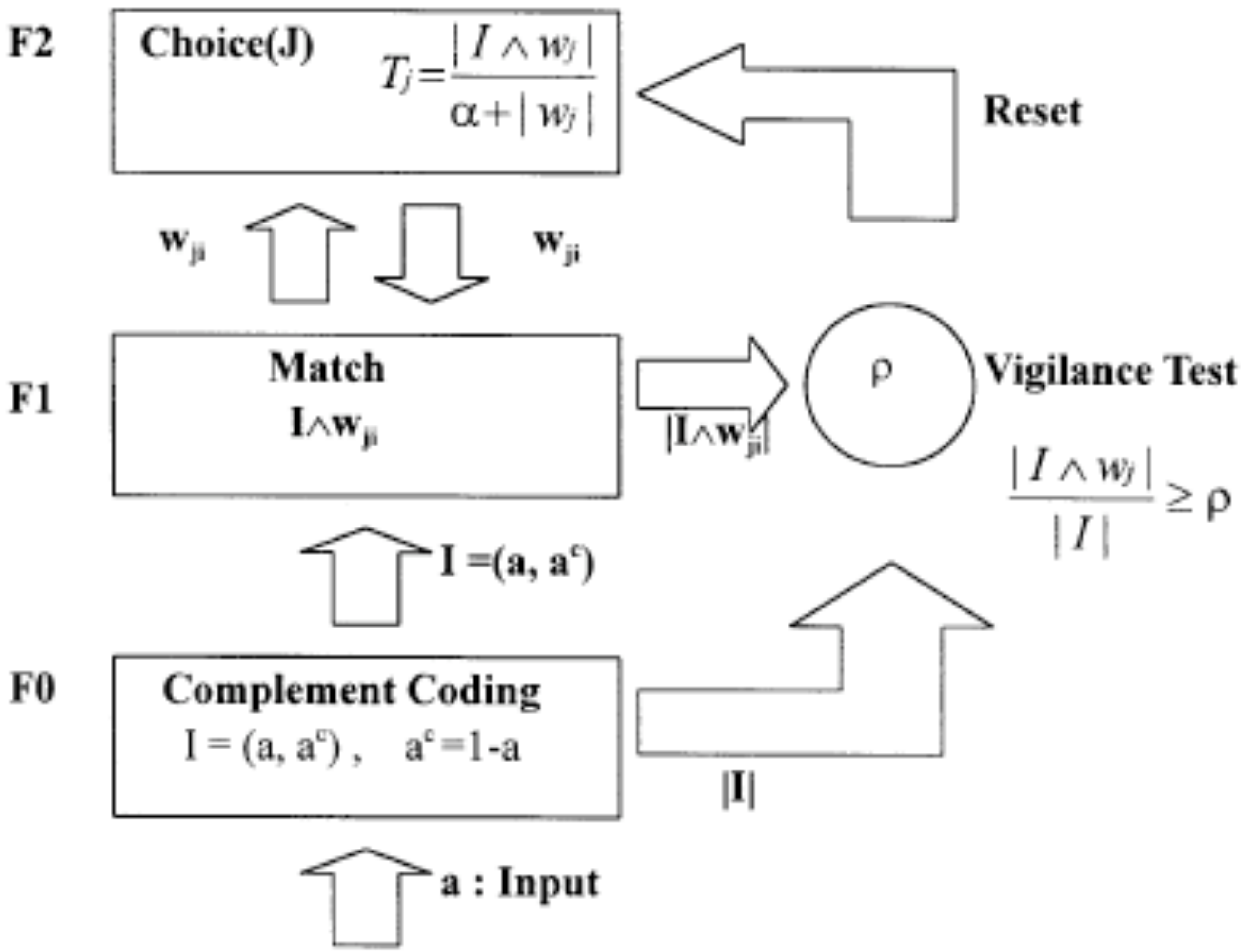
F_2 node choice:
$T_J = \max_j (T_j)$

F_1 node activation:

$X = I$ if F_2 is inactive
$X = I \wedge z_J$
if the J'th F_2 node is is the most active

Note the change of intersection operator to $\wedge$

Fuzzy ART



F2
Choice(J)
$T_j = \frac{|I \wedge w_j|}{\alpha + |w_j|}$
Reset
w_{ji}
w_{ji}
F1
Match
$I \wedge w_{ji}$
$|I \wedge w_{ji}|$
ρ
Vigilance Test
$\frac{|I \wedge w_j|}{|I|} \geq \rho$
$I = (a, a^c)$
F0
Complement Coding
$I = (a, a^c)$, $a^c = 1 - a$
$|I|$
a : Input

Step 1. Initalize all the parameters
Connection weights: $w_{ij}(0)=1$,
$0 \leq i \leq N$ (expected maximum number of clusters
$0 \leq j \leq M$ (number of classes)
select values for: $\alpha > 0$, $0 < \beta \leq 1$, $0 < \rho < 1$
(α: learning rate, β: choice parameter, ρ: vigilance parameter)
Step 2. Read new input vector I consisting of binary or analogue elements
Let I:= [next input vector]
Step 3. Compute choice function (T_j) for every input node.

$$T_j(I) = \frac{|I \wedge w_j|}{\alpha + |w_j|},$$

Where $\wedge$ is the fuzzy AND operator, defined as: $(x \wedge y) = \min(x_i, y_i)$
Step 4. Select the best matching exemplar:

$$T_J = \max\{T_j : j = 1, 2, \ldots, N\}$$

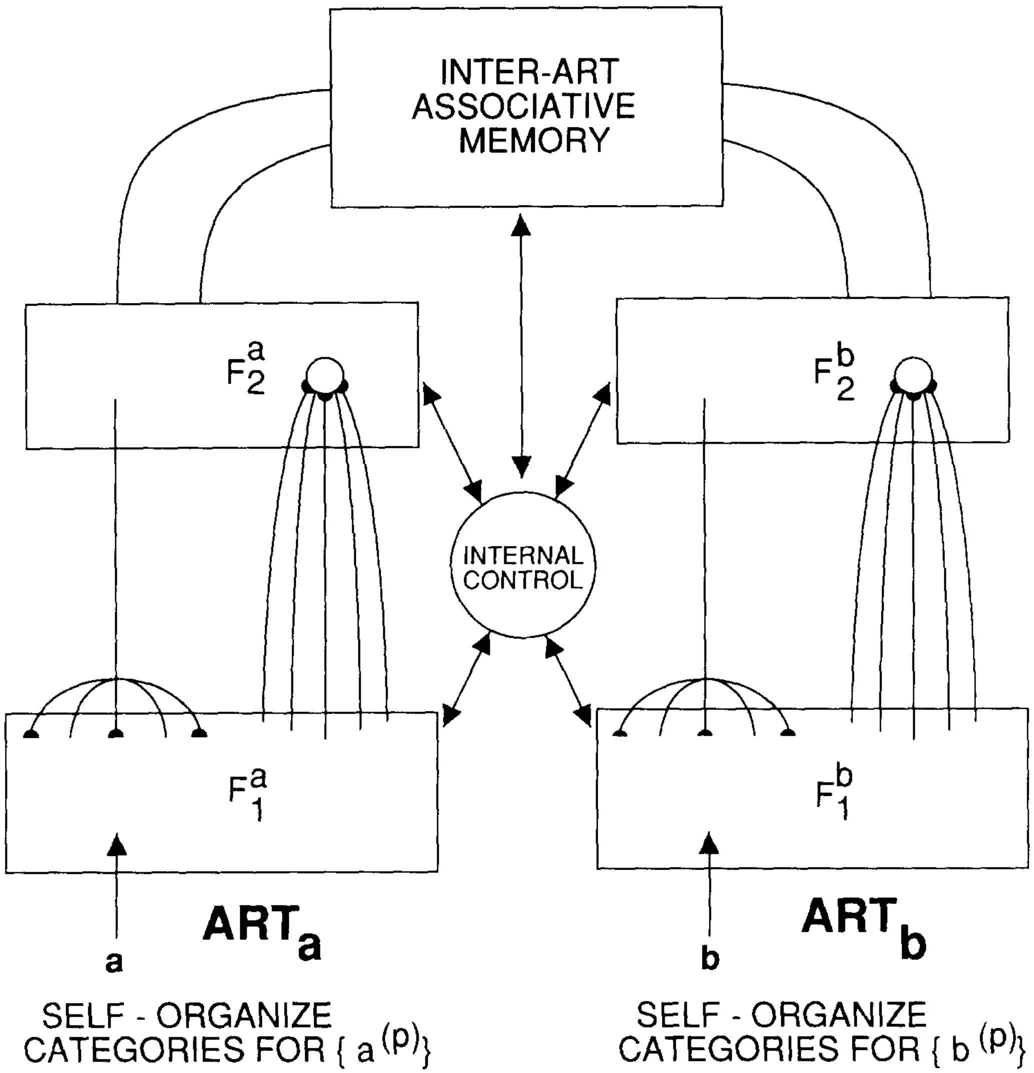
Step 5. Resonance test: (vigilance test)

If, $\frac{|I \wedge w_j|}{|I|} \geq \rho$, then go to step 7, else go to step 6
Step 6. Mismatch reset:
Set $T_j = -1$ and go to step 4
Step 7. Update best matching exemplar (learning law)

$$w_j^{(new)} = \beta (I \wedge w_J^{(old)}) + (1 - \beta) w_J^{(old)}.$$

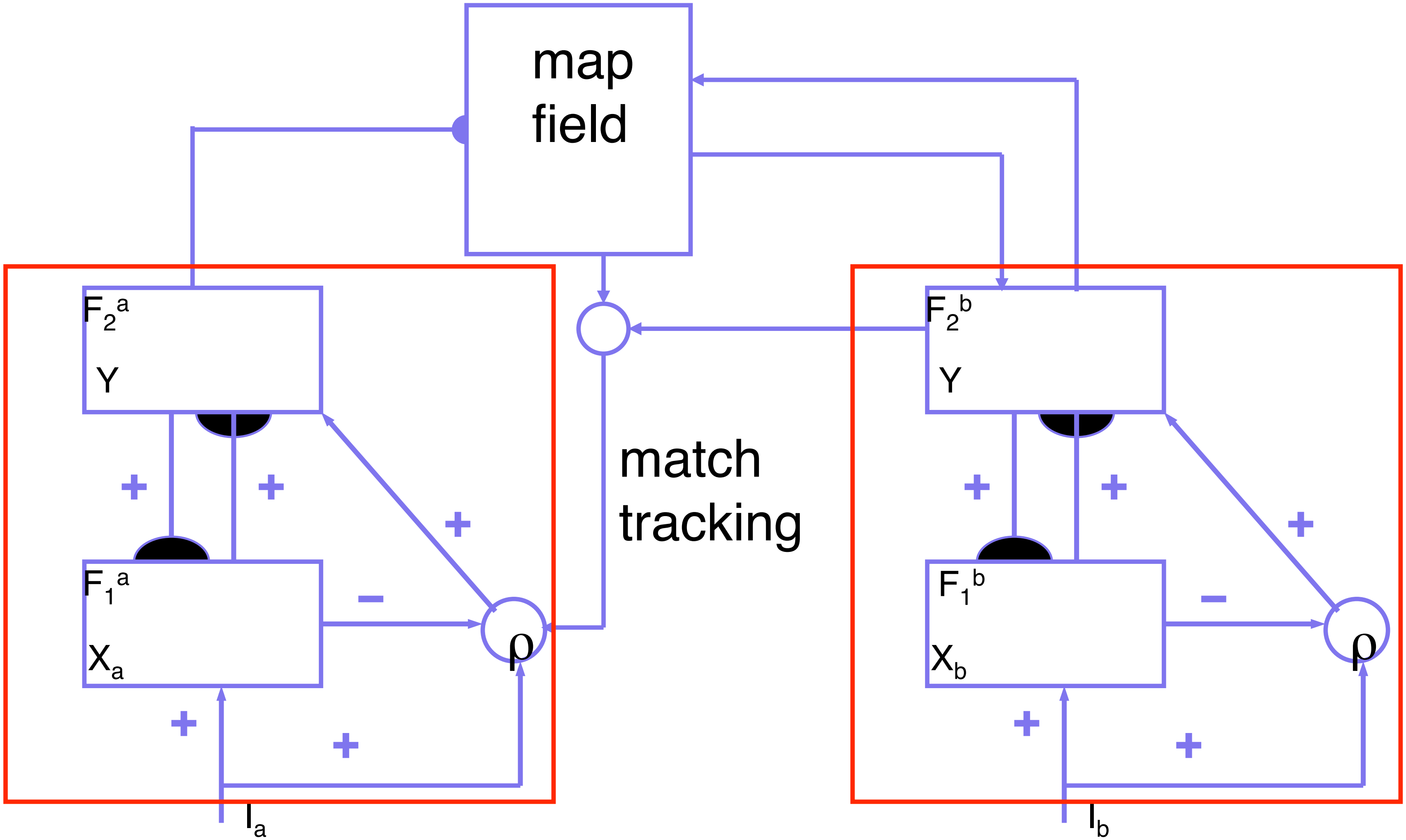
Step 8. Repeat: go to step 2

ART Map



INTER-ART ASSOCIATIVE MEMORY
F_2^a
F_2^b
INTERNAL CONTROL
F_1^a
F_1^b
ART_a
ART_b
a
b
SELF - ORGANIZE CATEGORIES FOR { a (p)}
SELF - ORGANIZE CATEGORIES FOR { b (p)}

Fuzzy ART-MAP



ART Map vs. BackProp

	Predictive ART	Back Propagation
supervised self-organizing real-time self-stabilizing	yes yes yes yes	yes no no no
learning:	fast or slow match	slow mismatch

Some ART Applications

- Disease identification (HMC Math Clinic)
- Tech support email automation (text similarity) (HMC CS Clinic)
- Satellite data anomaly detection (HMC CS Clinic)
- Music recognition
- Distinguishing poisonous vs. edible mushrooms
- Modeling biological neural processes