

Adaptive resonance: an emerging neural theory of cognition

Gail A. Carpenter

Department of Cognitive and Neural Systems, 677 Beacon Street, Boston University, Boston, Massachusetts 02215, U.S.A.

Submitted to *Trends in Cognitive Sciences*, as a Review

September, 2000

Keywords: Adaptive Resonance Theory, ART, dART, neural network, distributed coding, fast learning, catastrophic forgetting, visual cortex

tel: +1 617 353 9483. fax: +1 617 353 7755. E-mail: gail@cns.bu.edu.

Technical Report CAS/CNS TR-2000-010
Boston, MA: Boston University

Abstract

Adaptive resonance is a theory of cognitive information processing which has been realized as a family of neural network models. In recent years, these models have evolved to incorporate new capabilities in the cognitive, neural, computational, and technological domains. Minimal models provide a conceptual framework, for formulating questions about the nature of cognition; an architectural framework, for mapping cognitive functions to cortical regions; a semantic framework, for precisely defining terms; and a computational framework, for testing hypotheses. These systems are here exemplified by the distributed ART (dART) model, which generalizes localist ART systems to allow arbitrarily distributed code representations, while retaining basic capabilities such as stable fast learning and scalability. Since each component is placed in the context of a unified real-time system, analysis can move from the level of neural processes, including learning laws and rules of synaptic transmission, to cognitive processes, including attention and consciousness. Local design is driven by global functional constraints, with each network synthesizing a dynamic balance of opposing tendencies. The self-contained working ART and dART models can also be transferred to technology, in areas that include remote sensing, sensor fusion, and content-addressable information retrieval from large databases.

When we go to the movies, we expect to relax. Here, nonetheless, even the adult moviegoer performs formidable feats of memorization. After leaving the theatre with friends, we can discuss details from all the scenes, and compare these with images from movies we saw only once years earlier. This nearly effortless blend of perception, attention, learning, and memory – the heart of cognitive science – is the subject of adaptive resonance theory (ART). In the mid-1970s, ART was introduced as a theory of human cognitive information processing^{1,2}. Starting in the mid-1980s, a series of neural network models have added new principles to the cognitive theory, and have embodied these principles in quantitative systems that have been applied to problems of learning, recognition, and prediction. Each network realizes a set of goals as a minimal real-time system, a design process that helps the user to define terms and ideas as well as suggesting explicit links between brain and behavior and allowing ready testing in applications. Analysis of the limitations of each system often leads to a new design stage, within a unified framework. The first such model, ART 1³, was an unsupervised learning system designed to categorize binary input patterns, with ART 2⁴ and fuzzy ART⁵ then extending the domain to include continuous-valued inputs. In recent years, the family of networks has continued to evolve in response to conceptual and computational demands of problems from cognitive and neural sciences and technology. This evolutionary development is here exemplified by the recently introduced distributed ART (dART) model^{6,7} (Fig. 1), with illustrative examples drawn from the movie-going experience.

Figure 1: Distributed ART network

Functional objectives of the ART variations developed over the years have ranged from the cognitive (e.g., variable-rate speech and word recognition⁸) to the technological (e.g., sensor

fusion in sonar applications⁹). Enduring goals in this series have included the design of a system that can, as needed, maintain permanent codes (stable learning) or incorporate large quantities of new information on one trial (fast learning); that can represent either prototypes or exemplars; that can focus attention on critical features or quickly reset an erroneously activated code before spurious associations are learned. Each model realizes a dialectical synthesis of such competing goals (Table 1). The resulting network embodies a dynamic balance, rather than a "correct" resolution, of opposites. For example, the system adjusts levels of generalization, from coarse to fine, according to context, just as we can recall details of a heroine's expressions while only vaguely remembering the appearance of a minor character. The degree of generalization is determined by a parameter ρ , called *vigilance*: low vigilance permits broad categories, while raising vigilance moves the system from prototype learning toward exemplar learning. Choosing a value for ρ may present a formidable problem for an unsupervised ART network, which does not, by itself, determine a useful level of generalization. However, when this module is embedded in a supervised ARTMAP network¹⁰⁻¹², the nature of ρ changes, from a fixed parameter to an internally controlled variable whose value may vary from moment to moment, based on the predictive success of the overall system. The larger network thereby balances the design goal of maximizing generalization, which is fostered by low ρ values, against the complementary goal of minimizing predictive error, which may require higher ρ values.

Table 1: Dynamic balance

Distributed coding

In traditional ART networks, localist, or winner-take-all (WTA), competitive activation supports stable coding by limiting learned changes to memory traces that project to or from the one active category node. However, this maximally compressed activation pattern may cause category proliferation when noisy inputs are trained with fast learning. In contrast, multilayer perceptrons (MLPs) feature distributed McCulloch-Pitts¹³ activation, which promotes noise tolerance and code compression, but which is prone to catastrophic forgetting. (See Ref. 14: French, *Trends in Cognitive Science*, 1999, for a review of catastrophic interference in neural networks; and see Ref. 15: Page, *Behavioral and Brain Sciences*, 2000, for a review and discussion of localist vs. distributed computation in neural modeling of cognition.)

Figure 2: Distributed coding in a traditional ART architecture

Distributed ART models seek to combine the best of these two worlds: distributed activation enhances noise tolerance and code compression while new system dynamics retain the stable fast learning capabilities of WTA ART systems. An obvious possible design solution would simply distribute activation across the ART coding field F_2 while retaining other features of the network architecture and dynamic laws (Fig. 2). However, this approach encounters serious problems at the outset. First, without slow learning, the system would suffer an unavoidable type of catastrophic forgetting: according to the gated steepest descent learning laws, all active nodes would code the same pattern. Second, feedback activation in the $F_1 \leftrightarrow F_2$ loop could keep the system from ever establishing orderly code representations.

The dART network¹⁶ solves the second problem by reconfiguring the network (Fig. 1) to eliminate the feedback loop while retaining primary ART computations, such as top-down/bottom-up pattern matching at F_1 . In fact, with WTA coding and fast learning, distributed ART reduces computationally to a fuzzy ART algorithm. With distributed coding, the dART network automatically apportions learned changes according to the degree of activation of each node, which permits fast as well as slow learning without catastrophic forgetting. A parallel distributed match-reset-search process also helps stabilize memory.

Figure 3: Local computation: dART coding neuron

The critical new element that allows dART to solve the catastrophic forgetting problem is the *dynamic weight* (C, 1994), which replaces the traditional neural network path weight. This quantity equals the rectified difference between coding node activation and an *adaptive threshold*, combining short-term memory (STM) with long-term memory (LTM) in the basic unit of memory (Fig. 3). Thresholds increase monotonically during learning according to a principle of *atrophy due to disuse*. However, in the code selection paths from F_0 to F_2 , monotonic change at the synaptic level manifests itself as bidirectional change at the dynamic level, where the result of adaptation resembles long-term potentiation (LTP) for single-pulse or low-frequency test inputs but can resemble long-term depression (LTD) for higher frequencies. This dynamic is traced to dual computational properties of frequency-dependent and frequency-independent components of the coding signal. During learning, the frequency-independent component increases nonspecifically, for all inputs, while the frequency-dependent component becomes more selective, maximally favoring the current input (Fig. 4). Seemingly paradoxical, the disappearance of LTP enhancement for high-frequency test inputs is similar to the phenomenon of *redistribution of synaptic efficacy*, as observed by Markram and Tsodyks¹⁷ in the neocortex. Analysis of the dART learning system indicates how these dynamics are related to the computational components needed to support stable coding in a real-time neural network.

Figure 4: Global computation: dART code selection

Features present and features absent

From *101 Dalmatians*, we recall that there were lots of dogs, but no elephants; and we probably have no clear recollection of the colors of the cars. Similarly, ART memories represent both critical features that are consistently present (dogs) with respect to a given code (that movie) and critical features that are consistently absent (elephants); and inconsistent features are treated as uninformative (car colors). To carry out this construction, ART and dART employ a preprocessing step called *complement coding*⁵, which presents to the learning system both the original input pattern and its complement. This formal device, which corresponds to on-cell/off-cell coding in the early visual system, allows the learning system to encode critical features that are normally absent in the input environment, in addition to features that are normally present. When features are represented formally as a vector of component values, a pattern of descending values in the original input would be presented to the learning system along with its mirrored pattern of ascending values, as in the F_0 activation pattern shown in Fig. 1. Note, then, that the input (I) to the learning system now has twice as many components as the original input. Note,

too, that the learned prototype pattern of critical features, which is a function of top-down as well as bottom-up inputs (the F_1 activation pattern) is no longer complement coded. In Figure 1, strong activation in the left portion of the matched pattern at F_1 denotes features that are not only present in the current input but are also encoded as having been consistently present in the prototype memory of the active code y ; and strong activation in the right portion of the F_1 pattern denotes features that are absent in the current input and encoded as having been consistently absent. The network inhibits representation of features of the current input that have been inconsistently present and absent during learning of the critical feature pattern of the active code. The system thereby carries out a type of large-amplitude noise suppression: car colors might be salient features of an environment, but we are hard pressed to remember or even notice them, unless we direct our attention to do so.

As part of the global system dynamics, complement coding solves a category proliferation problem¹⁸. It also suggests a computational solution to the tendency of redistribution of synaptic efficacy to enhance only low-frequency inputs: if an input component is consistently large with respect to a given code, then the network can embody this fact in the complementary component, which can be enhanced since it will be consistently small.

Cognitive and neural systems

ART models and concepts have provided a context for analyzing cognitive and neural data from many sources. Pollen¹⁹, in a wide-ranging review of the neural correlates of visual perception, resolves various past and current views of cortical function by placing them in a framework he calls "adaptive resonance theories." This unifying perspective postulates resonant feedback loops as the substrate of phenomenal experience. Interpreting ART network components in a cortico-hippocampal system integrates diverse studies of normal and amnesic learning and memory²⁰. Recent work concerning how the neocortex is organized into layers²¹ suggests how *laminar computing* leads to intelligent behavior by modeling how bottom-up, top-down, and horizontal interactions are organized within the cortical layers. Figure 4 of Ref. 22 (*Trends in Cognitive Sciences*, 2000) shows network design elements which are repeated in a hierarchical structure in this model system.

Figure 5: dART and cortical layers

Figure 5 shows how the laminar model of visual cortex might be augmented to include learning, by identifying key components of that model with corresponding components of distributed ART. The laminar model at V2 layer 4 matches signal patterns from V2 layer 6 and from V1 layer 2/3, just as the dART model at F_1 matches signal patterns from the coding field F_2 and from the input field F_0 . In the laminar model, signals from V1 layer 2/3 also project to V2 layer 4, just as the dART input pattern projects directly from F_0 to F_2 . A similar modular configuration appears in the laminar model between LGN and V1. These similarities suggest new roles, in learning, for the layers of the cortical architecture. In particular, since the dInstar and dOutstar laws permit fast, stable, distributed learning in this network configuration, these properties would be inherited by a cortical model that adopts them. The dART learning laws

therefore suggest how the laminar model may be extended to cortical areas that participate in recognition learning and categorization, including inferotemporal cortex.

Technology transfer

ART and dART systems are part of a rapidly growing family of attentive self-organizing systems that have evolved from the biological theory of cognitive information processing. These modules have found their way into such diverse applications as industrial design and manufacturing²³, the control of mobile robots^{24,25}, automatic target recognition²⁶⁻²⁸, medical imaging²⁹, electrocardiogram wave recognition^{30,31}, air quality monitoring³², strength prediction for concrete mixes³³, signature verification³⁴, tool failure monitoring^{35,36}, frequency selective surface design for electromagnetic system devices³⁷, analysis of musical scores³⁸, power transmission line fault diagnosis³⁹, and satellite mapping⁴⁰.

Figure 6: Supervised ARTMAP architectures

Many of these applications use supervised ART architectures, called ARTMAP systems (Fig. 6). These networks self-organize arbitrary mappings from input vectors, representing features such as spectral values and terrain variables, to output vectors, representing predictions such as vegetation classes in a remote sensing application. Recent research for technology transfer, as in the cognitive and neural domains, seeks to extend to earlier localist constructions to include the possibility of stable distributed coding in a network hierarchy⁴¹.

A moment of conscious experience

Adaptive resonance offers a core module for the representation of hypothesized processes underlying learning, memory, attention, search, recognition, and prediction. At the model's field of coding neurons, the continuous stream of information pauses for a moment, holding a fixed activation pattern long enough for attention and memory to proceed. Feedback loops fixing the moment are broken only by active reset, which flexibly segments the flow of experience according to the demands of perception, memory, and environmental feedback. As Pollen suggests (Ref. 19, pp. 15-16): "it may be the consensus of neuronal activity across ascending and descending pathways linking multiple cortical areas that in anatomical sequence subserves phenomenal visual experience and object recognition and that may underlie the normal unity of conscious experience."

Acknowledgements

The author thanks Stephen Grossberg for many years of discussion and collaboration on the topics presented in this review. This research was supported by grants from the Office of Naval Research and the Defense Advanced Research Projects Agency (ONR N00014-95-1-0409 and ONR N00014-1-95-0657) and the National Institutes of Health (NIH 20-316-4304-5).

Outstanding questions

- What might be the global anatomical representations of learning systems?
- Where does memory consolidation, on a time scale of hours or years, fit in a real-time learning system?
- What might be the local physiological representations of activation and learning laws, at presynaptic and postsynaptic sites?
- What rules of synaptic transmission support global computational goals in model systems and in their physiological counterparts?
- How should results of recent studies that demonstrate learning as a redistribution of synaptic efficacy, rather than a nonspecific gain increase, be incorporated into neural network models?
- In a distributed system, how can an efficient search process be designed to learn from its mistakes even though, when the system makes an error, it does not yet know where it should be heading?
- How should an on-line learning system distinguish between important rare cases and outliers?
- How can a network with distributed activations retain stable codes and fast learning without locking in early memories too soon?
- What feedback loop designs permit orderly information processing and learning in a neural network?

References

- 1 Grossberg S. (1976) Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors & II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics* 23, 121-134 & 187-202
- 2 Grossberg S. (1980) How does a brain build a cognitive code? *Psychological Review* 87, 1-51
- 3 Carpenter G.A. and Grossberg S. (1987) A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing* 37, 54-115
- 4 Carpenter G.A. and Grossberg S. (1987) ART 2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics* 26, 4919-4930
- 5 Carpenter G.A. et al. (1991) Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks* 4, 759-771
- 6 Carpenter, G.A. (1996) Distributed activation, search, and learning by ART and ARTMAP neural networks. *Proceedings of the International Conference on Neural Networks (ICNN'96): Plenary, Panel and Special Sessions* (pp. 244-249). Piscataway, N.J.: IEEE Press.
- 7 Carpenter, G.A. (1997) Distributed learning, recognition, and prediction by ART and ARTMAP neural networks. *Neural Networks* 10, 1473-1494
- 8 Grossberg, S. (1999) The link between brain learning, attention, and consciousness. *Consciousness and Cognition* 8, 1-44
- 9 Carpenter, G.A. and Streilein, W.W. (1998) ARTMAP-FTR: A neural network for fusion target recognition, with application to sonar classification. In *AeroSense: Proceedings of SPIE's 12th Annual Symposium on Aerospace/Defense Sensing, Simulation, and Control*. Orlando, April 13-17, 1998, Bellingham, WA: Society of Photo-Optical Instrumentation Engineers.
- 10 Carpenter G.A. et al. (1991) ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks* 4, 565-588
- 11 Carpenter, G.A. et al. (1992) Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks* 3, 698-713
- 12 Carpenter G.A. and Ross W.D. (1995) ART-EMAP: A neural network architecture for object recognition by evidence accumulation. *IEEE Transactions on Neural Networks* 6, 805-818
- 13 McCulloch, W.S., and Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 9, 127-147
- 14 French, R.M. (1999) Catastrophic forgetting in connectionist networks. *Trends Cognit. Sci.* 3, 128-135

- 15 Page, M. (2000) Connectionist modelling in psychology: a localist manifesto. *Behavioral and Brain Sciences* 23.
- 16 Carpenter, G.A. (1994) A distributed outstar network for spatial pattern learning. *Neural Networks* 7, 159-168
- 17 Markram, H. and Tsodyks, M. (1996) Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature* 382, 807-810
- 18 Moore, B. (1989) ART 1 and pattern clustering. In D. Touretzky, G. Hinton, and T. Sejnowski (Eds.), *Proceedings of the 1988 Connectionist Models Summer School*, (pp. 174-185). San Mateo, CA: Morgan Kaufmann Publishers
- 19 Pollen, D.A. (1999) On the neural correlates of visual perception. *Cerebral Cortex* 9, 4-19
- 20 Carpenter, G.A. and Grossberg, S. (1993) Normal and amnesic learning, recognition, and memory by a neural model of cortico-hippocampal interactions. *Trends in Neuroscience* 16, 131-137
- 21 Grossberg, S. (1999) How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex. *Spatial Vision* 12, 163-185
- 22 Grossberg, S. (2000) The complementary brain: unifying brain dynamics and modularity. *Trends Cognit. Sci.* 4, 233-246
- 23 Caudell T.P. et al. (1994) NIRS: Large scale ART-1 neural architectures for engineering design retrieval. *Neural Networks* 7, 1339-1350
- 24 Bachelder, I.A. et al. (1993) A neural system for mobile robot visual place learning and recognition. In *Proceedings of the World Congress on Neural Networks (WCNN'93)* (pp. I-512-517). Hillsdale, NJ: Lawrence Erlbaum Associates
- 25 Baloch, A.A. and Waxman, A.M. (1991) Visual learning, adaptive expectations, and behavioral conditioning of the mobile robot MAVIN. *Neural Networks* 4, 271-302
- 26 Bernardon, A.M. and Carrick, J.E. (1995) A neural system for automatic target learning and recognition applied to bare and camouflaged SAR targets. *Neural Networks* 8, 1103-1108
- 27 Koch, M.W. et al. (1995) Cueing, feature discovery, and one-class learning for synthetic aperture radar automatic target recognition. *Neural Networks* 8, 1081-1102
- 28 Waxman, A.M. et al. (1995) Neural processing of targets in visible, multispectral IR and SAR imagery. *Neural Networks* 8, 1029-1051
- 29 Soliz P. and Donohoe, G.W. (1996) Adaptive resonance theory neural network for fundus image segmentation. In *Proceedings of the World Congress on Neural Networks (WCNN'96)*, Erlbaum, Hillsdale, New Jersey, pp 1180-1183
- 31 Ham, F.M. and Han, S.W. (1993) Quantitative study of the QRS complex using fuzzy ARTMAP and the MIT/BIH arrhythmia database. In *Proceedings of the World Congress on Neural Networks (WCNN'93)* (pp. I-207-211). Hillsdale, NJ: Lawrence Erlbaum Associates
- 31 Suzuki, Y. et al. (1993) Self-organizing QRS wave recognition system in ECG using ART 2. In *Proceedings of the World Congress on Neural Networks (WCNN'93)* (pp. IV-39-42). Hillsdale, NJ: Lawrence Erlbaum Associates

- 32 Wienke, D. et al. (1994) An Adaptive Resonance Theory based artificial neural network (ART 2-A) for rapid identification of airborne particle shapes from their scanning electron microscopy images. *Chemometrics and Intelligent Laboratory Systems*.
- 33 Kasperkiewicz, J. et al. (1995) HPC strength prediction using artificial neural network. *Journal of Computing in Civil Engineering* 9, 279-284
- 34 Murshed, N.A. et al. (1995) Off-line signature verification, without *a priori* knowledge of class ω_2 . A new approach. In *Proceedings of ICDAR 95: The Third International Conference on Document Analysis and Recognition*
- 35 Ly, S., and Choi, J.J. (1994) Drill condition monitoring using ART-1. In *Proceedings of the 1994 IEEE International Conference on Neural Networks* (pp. II-1226-1229). Piscataway, NJ: IEEE
- 36 Tarn, Y.S. et al. (1994) Tool failure monitoring for drilling processes. In *Proceedings of the 3rd International Conference on Fuzzy Logic, Neural Nets and Soft Computing* (pp. 109-111), Iizuka, Japan
- 37 Christodoulou C.G. et al. (1995) Design of gratings and frequency selective surfaces using fuzzy ARTMAP neural networks. *Journal of Electromagnetic Waves and Applications* 9, 17-36
- 38 Gjerdingen, R.O. (1990) Categorization of musical patterns by self-organizing neuronlike networks. *Music Perception* 7, 339-370
- 39 Aggarwal, R.K. et al. (1999) A novel approach to fault diagnosis in multicircuit transmission lines using fuzzy ARTmap neural networks. *IEEE Transactions on Neural Networks* 10, 1214-1221
- 40 Gopal, S. et al. (1999) Fuzzy neural network classification of global land cover from a AVHRR data set. *Remote Sensing of Environment* 67, 230-243
- 41 Carpenter, G.A. et al. (1998) Distributed ARTMAP: a neural network for fast distributed supervised learning. *Neural Networks* 11, 793-813

Table 1: Dynamic balance of design elements

Each ART or dART model defines a host of design tradeoffs, and each network example embodies a set of choices that represent a synthesis of complementary properties.

System dynamics and design	
Bottom-up signals	Top-down signals
Feedforward inflow	Feedback outflow
Code learning	Prototype learning
Perception	Expectation
Localist activation	Distributed activation
Rules and symbols	Real-time processing
Specific signals	Nonspecific signals
Signal	Noise
Environmental input	Critical features
Prototypes	Exemplars
Generalization	Encoding rare cases
Present features	Absent features
On-cells	Off-cells
Search	
Attention	Orientation
Familiarity	Novelty
Match	Reset
Learning	
Stability	Plasticity
Unlimited capacity to learn essential new information	Catastrophic forgetting
Invariance	Change
Limited capacity of STM	Unlimited capacity of LTM
Dynamic weight	Fixed weight
On-line, incremental learning	Off-line, batch learning
Unsupervised learning	Supervised learning
Fast learning	Slow adaptation
Cognition	
Consciousness	Real-time flow of activation
Coding	Action
One-to-many maps	Many-to-one maps
Consistent worldview	Inconsistent perceptions
Lifetime memory	Amnesia

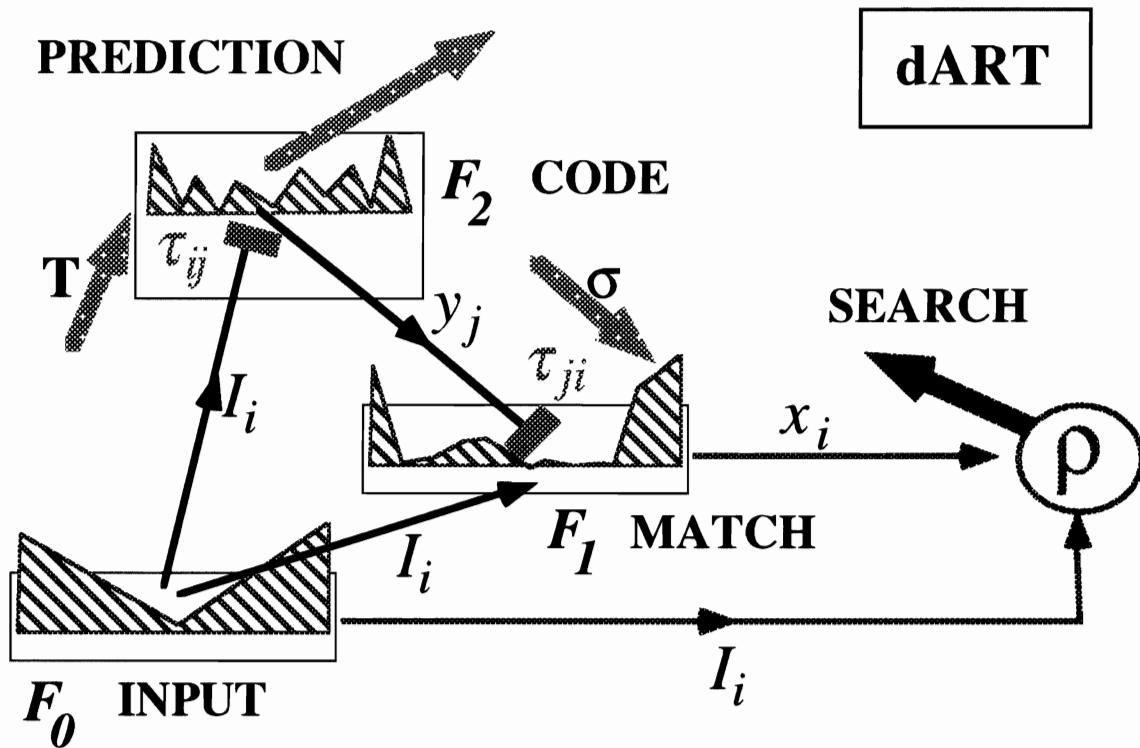


Fig. 1. dART model diagram. In the distributed ART (dART) network^{6,7}, the N nodes of a coding field F_2 receive a net signal pattern $\mathbf{T} = (T_1 \dots T_j \dots T_N)$ directly from an input field F_0 . A CAM (content-addressable memory) rule transforms these signals to an F_2 code, or activation pattern, $\mathbf{y} = (y_1 \dots y_j \dots y_N)$. Total F_2 activation is normalized ($\sum_{j=1}^N y_j = 1$), but may be distributed across arbitrarily many nodes. Activity $\mathbf{x} = (x_1 \dots x_i \dots x_M)$ at the field F_1 reflects a match between the input pattern $\mathbf{I} = (I_1 \dots I_i \dots I_M)$ from F_0 and the net signal pattern $\sigma = (\sigma_1 \dots \sigma_i \dots \sigma_M)$ from F_2 . The active code \mathbf{y} is reset when \mathbf{x} fails to meet the vigilance matching criterion, determined by parameter ρ . Long-term memory is stored as $F_0 \rightarrow F_2$ thresholds τ_{ij} , which adapt according to a distributed instar (dInstar) learning law; and as $F_2 \rightarrow F_1$ thresholds τ_{ji} , which adapt according to a distributed outstar (dOutstar) learning law¹⁶.

Design goals

FAST LEARNING

- on-line adaptation to rapid change
- encoding rare cases in one trial
- large databases

DISTRIBUTED CODING

- noise tolerance
- code compression

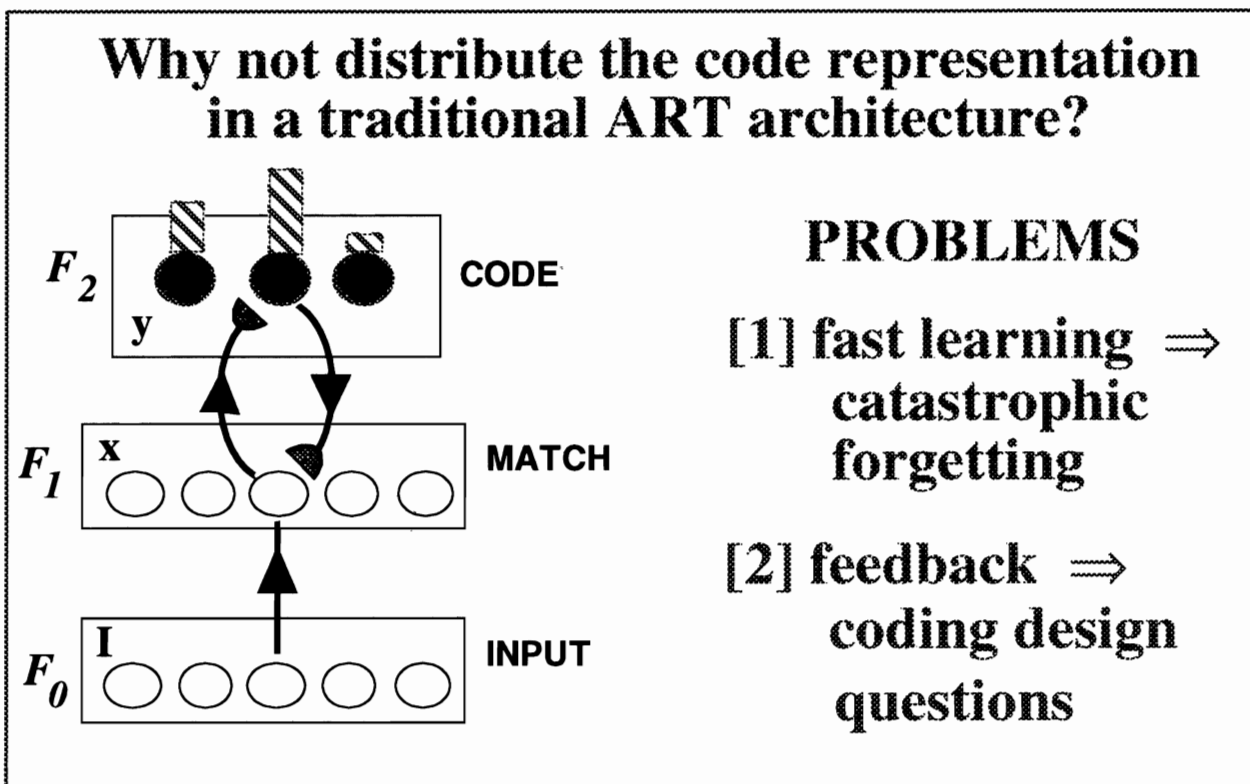


Figure 2

Fig. 2. Distributed coding causes problems in a traditional ART architecture. In ART models, the input, match, and coding fields are configured as $F_0 \rightarrow F_1 \leftrightarrow F_2$, respectively, in contrast to dART, where input signals flow directly from F_0 to F_2 (Fig. 1). In both ART and dART, the field F_1 matches the input pattern from F_0 against a learned expectation pattern generated by the code active at F_2 . In an ART network, winner-take-all (WTA) coding helps preserve code stability by restricting learned changes to paths projecting to and from the single active F_2 node. Moreover, instar learning laws, in paths from F_1 to F_2 , and outstar learning laws, in paths from F_2 to F_1 , ensure that the F_2 coding node chosen by the original bottom-up input I is the same as the one that would be chosen by the matched pattern x that becomes active at F_1 after F_2 sends top-down feedback³. This property, where top-down expectation confirms the original bottom-up choice, is essential for producing orderly real-time dynamics within the feedback loop. Code stability and consistent feedback may fail in a network that seeks to combine the advantages of fast learning and distributed coding simply by distributing activation across the ART coding field F_2 .

Local computation : dART coding neuron

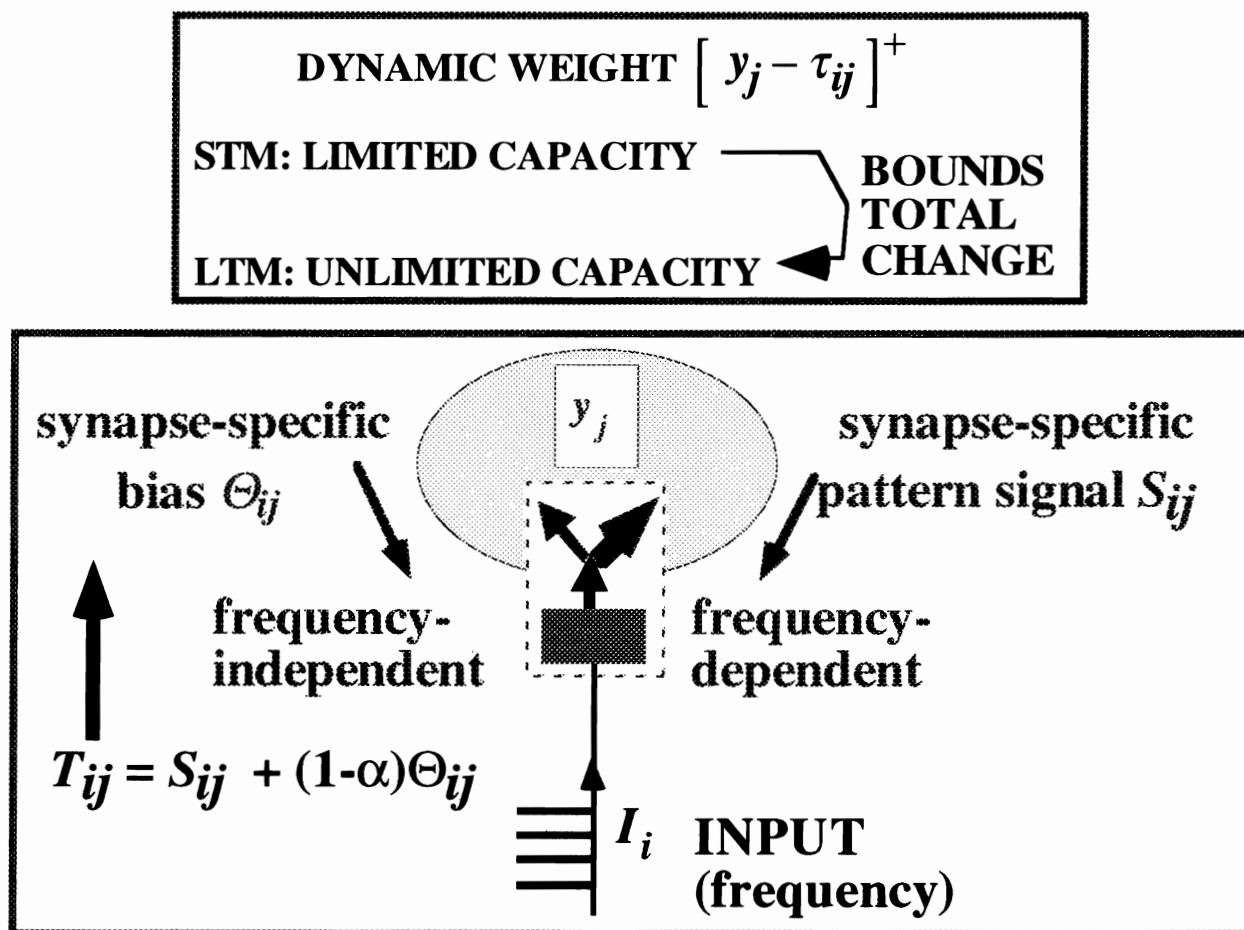


Figure 3

Fig. 3. A dART coding neuron. The dART architecture solves the feedback design problem (Fig. 2) by sidestepping it: the new dInstar and dOutstar learning laws allow the network to function without the $F_1 \leftrightarrow F_2$ feedback loop (Fig. 1), while retaining computational properties that are algorithmically equivalent to fuzzy ART when coding is winner-take-all. In dART, the dynamic weight $[y_j - \tau_{ij}]^+$ in the path from the i^{th} F_0 node to the j^{th} F_2 node equals the rectified difference between the target coding node activation y_j and the adaptive threshold τ_{ij} , i.e., $\max\{(y_j - \tau_{ij}), 0\}$. With WTA coding, where $y_J = 1$ at the single active F_2 node, a formal identification of the dynamic weight $[y_J - \tau_{iJ}]^+ = (1 - \tau_{iJ})$ with the fuzzy ART weight w_{iJ} reduces the dART algorithm to fuzzy ART, in the fast-learn limit. With distributed coding, dART solves the catastrophic forgetting problem by enlisting the limited capacity of STM at F_2 to bound learned changes. The $F_0 \rightarrow F_2$ signal T_{ij} is a weighted sum of a frequency-dependent component S_{ij} , which depends on the current input I_i ; and a frequency-independent component Θ_{ij} . Both S_{ij} and Θ_{ij} depend on the dynamic weight. In Ref. 7, the dART threshold τ_{ij} , and associated signal components S_{ij} and Θ_{ij} , are visualized as ligand-gated and voltage-gated membrane channels. In this representations, as τ_{ij} increases via dInstar learning, “disused” ligand-gated channels are converted to voltage-gated channels, which are weaker but input-independent. However, the dART model does not, by itself, uniquely specify either a presynaptic or a postsynaptic locus of the long-term memory trace.

Global computation: dART code selection

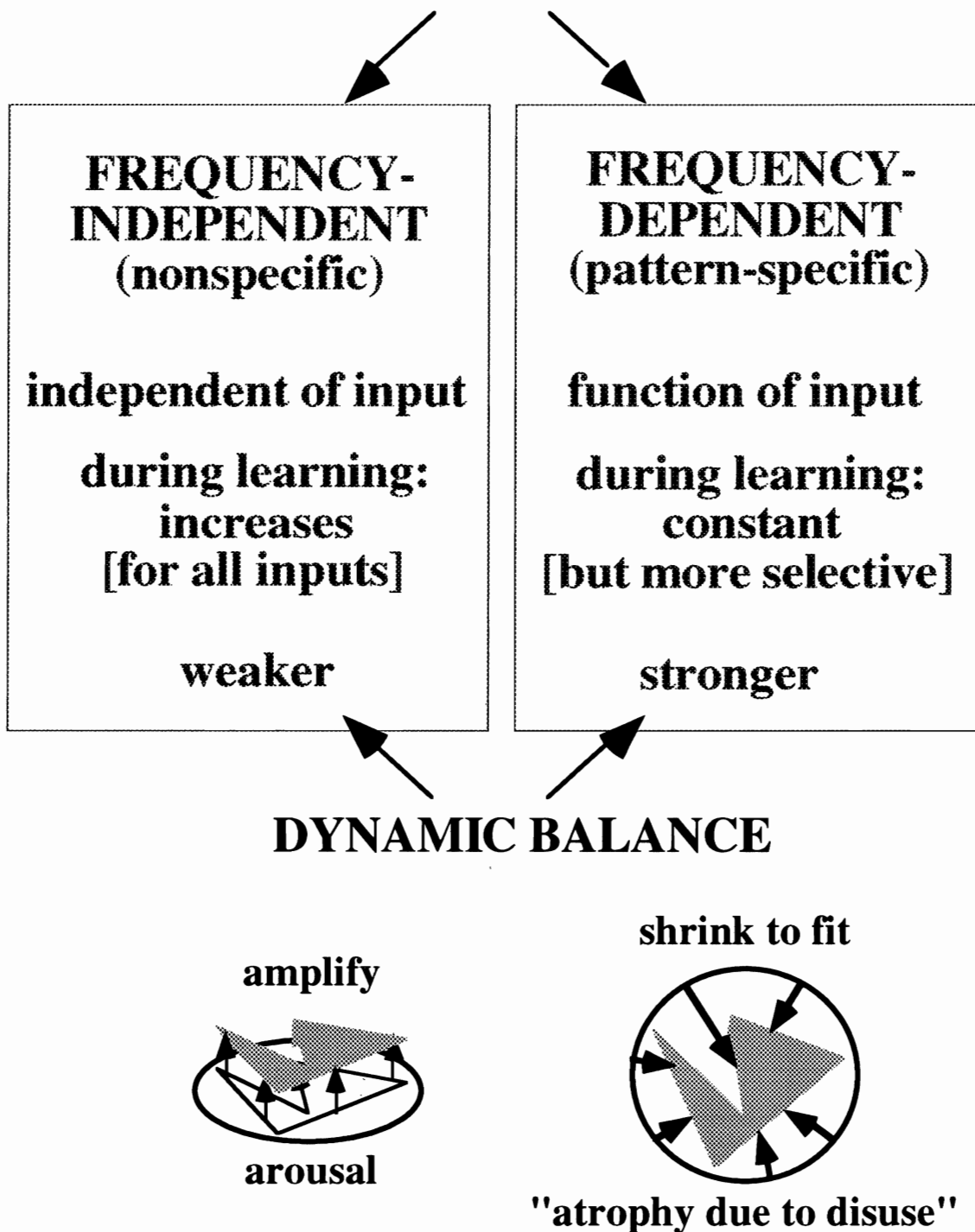


Figure 4

Fig. 4. dART code selection. The dynamic behavior of an individual dART synapse is viewed in the context of its role in stabilizing distributed pattern learning, rather than as a primary hypothesis. Redistribution of synaptic efficacy here reflects a local tradeoff between frequency-dependent and frequency-independent synaptic signal components which support a global tradeoff between pattern selectivity and a nonspecific path strengthening at the network level. Models that implement distributed coding via gain adaptation alone tend to suffer catastrophic forgetting and require slow or limited learning. In dART, each increase in frequency-independent synaptic efficacy is balanced by a corresponding decrease in frequency-dependent efficacy. With each frequency-dependent unit assumed to be stronger than each frequency-independent unit, the net result of learning is redistribution, rather than nonspecific enhancement, of synaptic efficacy. The system uses these complementary mechanisms to enhance network response to a given pattern while suppressing the response to mismatched patterns. At the same time, the dInstar learning law protects prior codes against catastrophic forgetting.

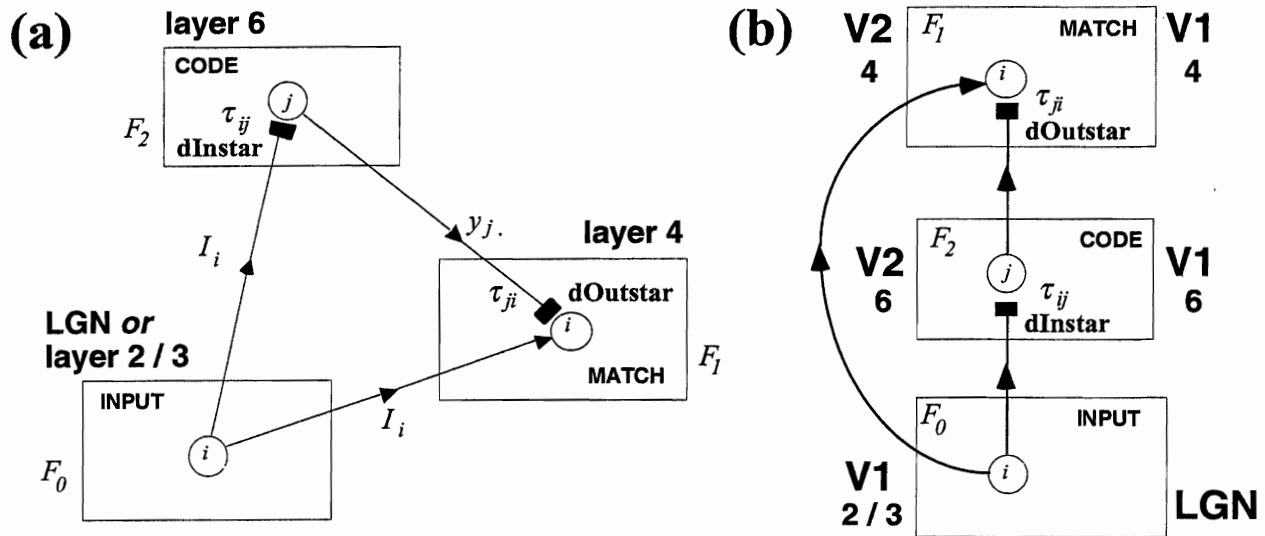


Fig. 5. dART and cortical layers. The dART network configuration (a) is isomorphic to modular components of a laminar model of visual cortex (b). (See Figure 4 of Ref. 20: *Trends in Cognitive Sciences*, 2000.) Comparing dART with the first level of the laminar computing model hierarchy, the input field F_0 may be identified with LGN, the coding field F_2 with V1 cortical layer 6, and the match field F_1 with the V1 cortical layer 4. This anatomical equivalence indicates how learning laws and other dynamic components of the dART network may be incorporated into the cortical model in such a way as to achieve key design goals, including code stability, and suggests new functional roles for the various layers. Since the laminar computing model features isomorphic structures in the cortical hierarchy, the dART function may be tested at each corresponding level. Note that the reconfiguration of the dART architecture from (a) to (b) blurs the distinction between “top-down” and “bottom-up” matching, expectation, and attentional focusing at F_1 . In addition, the laminar cortex model also includes other top-down attentional signals (e.g., from V2 layer 6 to V1 layer 6) as part of a “folded-feedback” circuit.

SUPERVISED LEARNING

ARTMAP \subseteq FUZZY ARTMAP \subseteq dARTMAP
(1991) (1992) (1998)

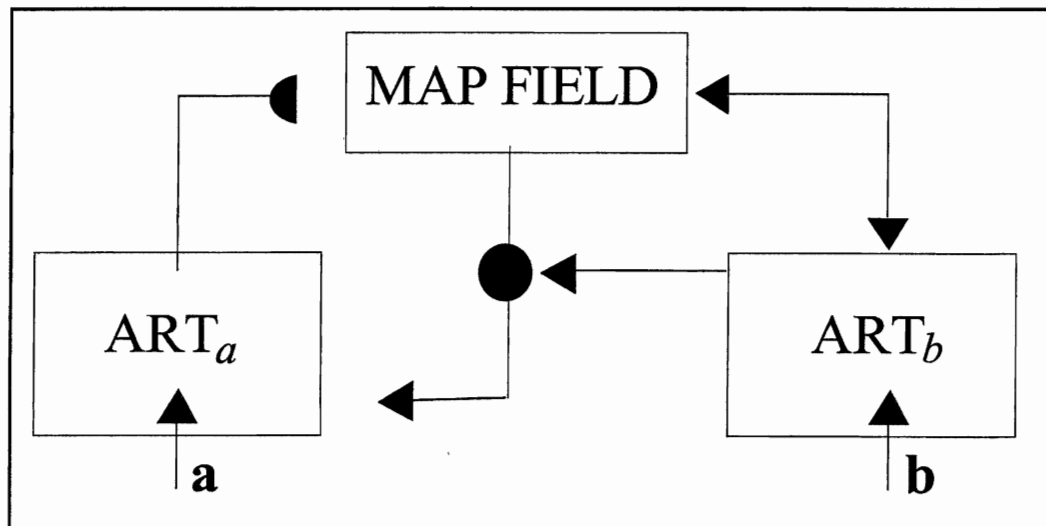


Fig. 6. Supervised ARTMAP. In most applications, unsupervised ART networks serve as modular components of a supervised system, where the system learns an input-to-output map ($a \rightarrow b$) during training. Following an evolutionary development like that of the unsupervised systems, distributed ARTMAP⁴¹ reduces to fuzzy ARTMAP¹¹ when coding is WTA, and further reduces to ARTMAP¹⁰ when inputs are binary.