LEARNING, CATEGORIZATION, RULE FORMATION, AND PREDICTION BY FUZZY NEURAL NETWORKS

Gail A. Carpenter† and Stephen Grossberg‡

Center for Adaptive Systems
and
Department of Cognitive and Neural Systems
Boston University
111 Cummington Street
Boston, Massachusetts 02215 USA

October 1994

To appear in
Fuzzy Logic and Neural Network Handbook
C.H. Chen (Editor)
New York: McGraw-Hill, 1995

Technical Report CAS/CNS-TR-94-028
Boston, MA: Boston University

† Supported in part by ARPA (ONR N00014-92-J-4015), the National Science Foundation (IRI 94-01659), and the Office of Naval Research (ONR N00014-91-J-4100)

‡ Supported in part by ARPA (AFOSR 90-0083 and N00014-92-J-4015) and the Office of Naval Research (ONR N00014-91-J-4100).

Acknowledgements: The authors wish to thank Cynthia E. Bradford and Diana J. Meyers for their valuable assistance in the preparation of the manuscript.

1. A Synthesis of Production Systems, Neural Networks, and Fuzzy Logic

For many years, the subjects of artificial intelligence, neural networks, and fuzzy logic were developed by separate intellectual communities. This was due more, perhaps, to social and institutional barriers to communication than to inherently different research goals and results. One manifestation of these barriers came in the form of claims, especially from the AI community, that the other approaches could not succeed at solving certain problems, despite progress to the contrary. This divisive period is fortunately substantially behind us. A growing number of models now computationally synthesize properties of expert production systems, neural networks, and fuzzy logic. Fuzzy ARTMAP, the topic of this chapter, is one such model. Fuzzy ARTMAP is a family of self-organizing neural architectures that are capable of rapidly learning to recognize, test hypotheses about, and predict consequences of analog or binary input patterns occurring in a nonstationary time series.

Fuzzy ARTMAP is the class of Adaptive Resonance Theory (ART) architectures designed for supervised learning. Since the introduction of ART as a cognitive and neural theory (Grossberg 1976a, 1976b), an ever expanding family of ART neural network architectures has been progressively developed at Boston University. These models include ART 1, ART 2, ART 2-A, ART 3, Fuzzy ART, ARTMAP, Fuzzy ARTMAP, and Fusion ARTMAP (Asfour, Carpenter, Grossberg, and Lesher, 1993; Carpenter and Grossberg, 1987a, 1987b, 1990, 1991; Carpenter, Grossberg, Markuzon, Reynolds, and Rosen, 1992; Carpenter, Grossberg, and Reynolds, 1991, 1993, 1994; Carpenter, Grossberg, and Rosen, 1991a, 1991b). Variants of these models have also been developed by a number of other investigators.

ART systems from the outset incorporated properties of production systems and fuzzy logic; e.g., see Grossberg (1980, 1982). However, Fuzzy ARTMAP provides a mathematically precise realization of ART concepts that is computationally powerful enough to outperform many expert systems, genetic algorithms, and other neural networks in benchmark studies (Tables 1–3) and to help solve outstanding technological problems (Bachelder, Waxman, and Seibert, 1993; Caudell, Smith, Johnson, Wunsch, and Escobedo, 1991; Dubrawski and Crow-

ley, 1994a, 1994b; Escobedo, Smith, and Caudell, 1993; Gan and Lua, 1992; Gopal, Sklarew, and Lambin, 1993; Ham and Han, 1993; Harvey, 1993; Kasperkiewicz, Racz, and Dubrawski, 1994; Keyvan, Durg, and Rabelo, 1993; Kumara, Merchawi, Karmarthi, and Thazhutaveetil, 1993; Johnson, 1993; Mehta, Vij, and Rabelo, 1993; Moya, Koch, and Hostetler, 1993; Suzuki, Abe, and Ono, 1993; Wienke, 1993, 1994; Wienke and Kateman, 1994; Wienke, Xie, and Hopke, 1994). A self-organizing neural architecture, called VIEWNET (Bradski and Grossberg, 1994), that can learn to recognize 3-D objects from sequences of their 2-D views is reviewed to show how Fuzzy ARTMAP can be embedded into larger systems. Another simulation example of 3-D object recognition illustrates how the ART-EMAP architecture (Carpenter and Ross, 1993, 1994a, 1994b) uses distributed network activity to improve noise tolerance while retaining the speed advantage of fast learning; and how temporal evidence accumulation can augment ARTMAP capabilities.

Tables 1–3

One useful property of Fuzzy ARTMAP is that it can learn recognition categories and predictions in an unsupervised mode, yet can also use predictive disconfirmations to supervise learning of categories that fit the statistics of the input-output environment. Such a supervised architecture is capable of learning complex mappings from m-dimensional Euclidean space to n-dimensional Euclidean space, given arbitrary finite dimensions m and n.

There are several ways in which Fuzzy ARTMAP embodies fuzzy concepts. The most general way concerns how the architecture autonomously calibrates the degree of compression, or generalization, that should occur in each category to fit the statistics of the environment. More general categories embody more "fuzziness" in the range of featural values accepted by the category. Category choice by the network embodies a decision process that discovers which learned ranges of fuzzy features best match the input pattern.

Fuzzy logic enters this learned recognition process in a more precise sense as well. The

fuzzy AND, or min, operator and the fuzzy OR, or max, operator are used to define the range of mathematical values that are tolerated by a category for each linguistic variable, or feature. The features of "hat on head" and "hair on head" are illustrative as applied to the heads of men and dogs. Men sometimes wear a hat and usually have at least some hair on their head. Dogs very rarely wear a hat but almost always have some hair on their head. To express this mathematically, let each feature's values lie in the interval [0,1], where value 0 means "never," value 1 means "always," and intermediate values range from "rarely" through "sometimes" to "frequently". Then "hair on head" is represented for a man's head by a wide interval [A,1], with A intermediate between 0 and 1. The variable of "hat on head" also translates into a wide interval [0,B] of expectations, with B intermediate between 0 and 1. For a dog, "hair on head" is represented by a narrow interval [C,1] with C close to 1, whereas "hat on head" becomes a narrow interval [0,D] with D close to 0. Simultaneous representation of both fuzzy features is achieved by category rectangles whose sides have a length and location that represent the fuzziness tolerated for the corresponding feature, as in Figure 1. Hyper-rectangles in $\Re^n$ code categories that represent n fuzzy features.

Figure 1

The min operator ($\wedge$) defines features that are "critically present" with intervals such as [C,1], whereas the max operator ($\vee$) defines features that are "critically absent" with intervals such as [0,D]. As shown below, the min operator may be realized by the on-cells of a neural network—namely, the nodes that are turned on by an input feature—whereas the max operator is realized by the off-cells of a neural network—namely, the nodes that are turned off by an input feature. Thus the duality between min and max in fuzzy logic translates into a duality between on and off in neural networks. The hypothesis that on and off cell pairs, or opponent processes, can represent fuzzy concepts has been part of ART heuristics since their inception (Grossberg, 1976b, 1980), and indeed part of the earliest modern tracts on neural networks (Grossberg, 1964, Section 170). Fuzzy ART and Fuzzy

ARTMAP develop fuzzy concepts into a computationally effective algorithm.

Fuzzy ARTMAP also computationally incorporates the operation of fuzzy subsethood (Kosko, 1986, Zadeh, 1965). Carpenter, Grossberg, and Rosen (1991b) observed that the ART 1 computations of choice, match, and search naturally map into corresponding fuzzy operations. (See Section 8.) In particular the ART 1 choice function that determines category selection can be interpreted in terms of the degree to which a learned category representation is a fuzzy subset of the current input.

Fuzzy logic provides a way for Fuzzy ARTMAP to adaptively categorize analog, as well as binary, input patterns. In particular, Fuzzy ARTMAP can autonomously learn, recognize, and make predictions with the following properties:

(A) Fast Learning of Rare Events

A successful autonomous agent must be able to learn about rare events that have important consequences, even if these rare events are similar to a surrounding cloud of frequent events that have different consequences. *Fast learning* is needed to categorize a rare event before it is supplanted by more frequent subsequent events. Many traditional learning schemes use a form of slow learning that tends to average over similar event occurrences. In contrast, Fuzzy ARTMAP can rapidly learn a rare event that predicts different consequences than a cloud of similar events in which it is embedded.

(B) Stable Learning of Large Nonstationary Data Bases

Rare events often occur in a nonstationary environment whose event statistics may change rapidly and unexpectedly through time. Individual events may also occur with variable probabilities and durations, and arbitrarily large numbers of events may need to be processed. Each of these factors tends to destabilize the learning process within traditional algorithms. New learning in such algorithms tends to unselectively wash away the memory traces of old, but still useful, knowledge. Using such an algorithm, for example, learning a new face-to-name association could erase the memory of a parent's face-to-name association.

More generally, learning new facts could erase the memory of previous expert knowledge. Fuzzy ARTMAP contains a *self-stabilizing memory* that permits accumulating knowledge to be stably stored in response to arbitrarily many events in a nonstationary environment under incremental learning conditions, until the algorithm's full memory capacity, which can be chosen arbitrarily large, is exhausted.

(C) Efficient Learning of Morphologically Variable Events

The morphological variability of information often changes through time, with some information coarsely defined whereas other information is precisely characterized. For example, it may only be necessary to recognize that an object is a face, or that it is the face of one's own father. It may be necessary to recognize that an object is a vehicle, or that it is a particular type of tank that is manufactured by a particular country. Under autonomous learning conditions, no teacher is generally available to instruct a system about how coarse its generalization, or compression, of particular types of data should be. Multiple scales of generalization, from fine to coarse, need to be available on an as-needed basis. Fuzzy ARTMAP automatically adjusts its scale of generalization to match the morphological variability of the data using a Minimax Learning Rule that conjointly *min*imizes predictive error and *max*imizes generalization using only information that is locally available under incremental learning conditions. The Minimax Learning Rule enables Fuzzy ARTMAP to autonomously calibrate how much fuzziness should be tolerated on each feature dimension of a category in order to achieve accurate predictive generalization.

(D) Associative Learning of Many-to-One and One-to-Many Maps

Many-to-one learning includes both categorization and associative prediction (Figure 2). For example, during categorization of printed letter fonts, many similar instances of the same printed letter may establish a single recognition category, or compressed representation. Different printed letter fonts or written versions of the letter may establish additional categories. Each of these categories carries out a many-to-one map of input into category.

During prediction, all of the categories that represent the same letter may be associatively mapped into the letter name, or prediction. This is a second, distinct, type of many-to-one map, since there need be no relationship between the visual features that define a printed letter A and a written letter A, yet both categories may have the same name for cultural reasons. The symbol that represents this name may, in turn, be transformed through learning into an arbitrary output pattern.

This two-stage many-to-one, or compressive, transformation, followed by a learned output transformation, is what enables Fuzzy ARTMAP to learn essentially arbitrary maps from $\Re^m$ to $\Re^n$. Individual recognition categories may be compared to the hidden units in the back propagation model (Parker, 1982; Rumelhart, Hinton, and Williams, 1986; Werbos, 1974). Fuzzy ARTMAP discovers on its own the number of categorical "hidden units" that it needs to achieve Minimax Learning, unlike back propagation where a human operator decides by trial and error how many hidden units are needed.

Figure 2

One-to-many learning is used to discover and accumulate expert knowledge about an object or event (Figure 3). For example, a computerized record of a patient's medical checkup may lead to a series of predictions about the patient's health. A chemical assay of a sample of coal or petroleum may lead to many predictions about its uses as an energy source or structural material. In many learning algorithms, including back propagation, the attempt to learn more than one prediction about an event leads to unselective forgetting of previously learned predictions, for the same reason that these algorithms become unstable in response to nonstationary data.

Figure 3

2. Attention, Hypothesis Testing, Match-Learning, and Confidence Estimation

Fuzzy ARTMAP achieves the properties (A)–(D) of Section 1 by implementing the following types of processes:

(E) Paying Attention and Top-Down Priming

A Fuzzy ARTMAP system can learn top-down expectations (also called primes, or queries) that can bias the system to ignore masses of irrelevant data. A large mismatch between a bottom-up input vector and a top-down expectation can suppress features in the input pattern that are not confirmed by the top-down prime and thereby drive an adaptive memory search that carries out a bout of hypothesis testing.

(F) Hypothesis Testing and Match-Learning

The system hereby selectively searches for recognition categories, or hypotheses, whose top-down expectations provide an acceptable match to bottom-up data. Each top-down expectation begins to focus attention upon, and bind, that cluster of input features that are part of the prototype which it has already learned, while suppressing features that are not. The search continues if the current match is not good enough. If no previously learned category, or hypothesis, provides a good enough match, then selection and learning of a new category and top-down expectation is automatically initiated. When the search discovers an old or new category that provides an acceptable match, the system locks into an attentive resonance during which learning occurs. The input pattern hereby refines the adaptive weights of the category prototype based on any new information that it contains.

Unlike many learning systems, such as back propagation, the Fuzzy ARTMAP system carries out match-learning, rather than mismatch-learning. A category modifies its previous learning only if its top-down expectation matches the input vector well enough to risk changing its defining characteristics. Otherwise, hypothesis testing selects a new category on which to base learning of a novel event.

(G) Choosing the Globally Best Answer without Recursive Search

In many learning algorithms, as learning proceeds, local minima or less than optimal solutions may be selected to represent the data. In Fuzzy ARTMAP, at any stage of learning, an input pattern first selects the category whose top-down expectation provides the globally

best match. It is in this sense that a top-down expectation acts as a *prototype* for the class of all the input patterns that its category represents. After learning self-stabilizes, every input directly selects and resonates with the globally best-matching category without triggering a recursive search. Familiar events directly resonate with the globally best category without recursive search, even if they are interspersed with unfamiliar events that drive recursive hypothesis testing for better matching categories.

(H) Learning Both Prototypes and Exemplars

The prototype represents the cluster of input features that the category learns based upon its past experience. The prototype represents the features to which the category "pays attention." In cognitive psychology, an input pattern is called an exemplar. A classical issue in cognitive psychology concerns whether the brain learns prototypes or exemplars. It sometimes seems that the brain learns prototypes, or abstract types of knowledge, such as being able to recognize that a particular object is a face or an animal. At other times, the brain appears to learn individual exemplars, or concrete types of knowledge, such as being able to recognize a particular face or a particular animal. What sort of hybrid system can learn both types of knowledge (Smith, 1990)? Fuzzy ARTMAP is such a hybrid system. It uses the Minimax Learning Rule to control how abstract or concrete—how fuzzy—a category can become in order to conjointly minimize predictive generalization and maximize predictive generalization. The next section indicates how this is accomplished.

(I) Controlling Vigilance to Calibrate Confidence

A confidence measure, called *vigilance*, calibrates how well an exemplar needs to match the prototype that it reads out in order for the corresponding category to resonate with it and be chosen. In other words, vigilance calibrates how well the hypothesis represented by the category matches the data. If vigilance is low, even poor matches are accepted. Many different exemplars can then be incorporated into one category, so compression and generalization by that category are high. If vigilance is high, then even good matches may

be rejected, and hypothesis testing may be initiated to select a new category. In this case, few exemplars activate the same category, so compression and generalization are low. A very high vigilance can select a unique category for a rare event that predicts an outcome different from that of any of the similar exemplars that surround it. In this limiting case, the prototype of the category learns the unique exemplar that the category represents, so that prototype learning reduces to exemplar learning.

The Minimax Learning Rule is realized by adjusting the vigilance parameter in response to a predictive error. Vigilance is increased just enough to initiate hypothesis testing to discover a better category, or hypothesis, with which to match the data. In this way, a minimum amount of generalization is sacrificed to correct the error. This process is called *match tracking* because vigilance tracks the degree of match between exemplar and prototype in response to a predictive error. How this is computed is described below.

(J) Rule Extraction and Fuzzy Reasoning

At any stage of learning, a user can translate the state of a Fuzzy ARTMAP system into a fuzzy set of rules. These rules evolve as the system is exposed to new inputs. Suppose, for example, that n categories are associated with the m^{th} prediction of the network. Backtrack from prediction m along the associative pathways whose adaptive weights have learned to connect the n categories to this prediction (Figure 2). Each of these categories codes a "reason" for making the m^{th} prediction. The prototype of each category embodies the set of features, or constraints, whose binding together constitutes that category's "reason". The IF-THEN rule takes the form: IF the features of any of these n categories are found bound together, within the fuzzy constraints that lead to selection of that category, THEN the m^{th} prediction holds. An example of such a rule is: if feature 1 falls in interval $[a_1, b_1]$ *and* feature 7 falls in interval $[a_7, b_7]$, *or* feature 3 falls in interval $[a_3, b_3]$, feature 8 falls in interval $[a_8, b_8]$, *and* feature 94 falls in interval $[a_{94}, b_{94}]$, then prediction 7 is made. Fuzzy ARTMAP is thus a type of self-organizing fuzzy production system (Laird, Newell, and Rosenbloom, 1987) that evolves adaptively from its unique input-output experiences.

The IF-THEN rules of Fuzzy ARTMAP can be read off from the learned adaptive weights of the system at any stage of the learning process. This property is particularly important in applications such as medical diagnosis from a large database of patient records, where doctors may want to study the rules by which the system reaches its diagnostic decisions. Tables 1–3 summarize some medical and other benchmark studies that compare the performance of Fuzzy ARTMAP with alternative recognition and prediction models. These and other benchmarks are described elsewhere in greater detail (Carpenter, Grossberg, and Iizuka, 1992; Carpenter, Grossberg, Markuzon, Reynolds, and Rosen, 1992; Carpenter, Grossberg, and Reynolds, 1991).

(K) Properties Scale to Arbitrarily Large Databases

One of the most serious weaknesses of traditional Artificial Intelligence algorithms is that their desirable properties tend to break down as small toy problems are generalized to large-scale problems. In contrast, all of the desirable properties of Fuzzy ARTMAP scale to arbitrarily large problems. On the other hand, Fuzzy ARTMAP helps to solve only learned categorization and prediction problems. These problems are, however, core problems in many intelligent systems, and have been technology bottlenecks for many alternative approaches.

A summary is now given of Adaptive Resonance Theory, or ART, networks for unsupervised learning and categorization. The connection between ART systems and fuzzy logic is noted in an exposition of Fuzzy ART networks for unsupervised learning and categorization. Fuzzy ART modules are then combined into a Fuzzy ARTMAP system that is capable of supervised learning, recognition, and prediction.

3. Unsupervised Self-Organizing Feature Map and ART Systems

As noted above, Adaptive Resonance Theory was introduced as a theory of human cognitive information processing (Grossberg, 1976b, 1980). Theoretical development has continued to explain and predict ever-larger cognitive and neural data bases to the present day; see Carpenter and Grossberg (1991, 1993); Grossberg (1987a, 1987b, 1988, 1994); Grossberg

and Merrill (1992); and Grossberg, Mingolla and Ross (1994a) for illustrative contributions. In addition, an evolving series of self-organizing neural network models have been developed for applications to adaptive pattern recognition and prediction. These self-organizing models can operate in either an unsupervised or a supervised mode. Unsupervised learning occurs when network predictions do not generate environmental feedback. Supervised learning occurs when prediction-contingent feedback is available. This option does not occur in many supervised learning algorithms, such as back propagation, which can learn only when feedback is available. Unsupervised ART models learn stable recognition categories in response to arbitrary input sequences with either fast or slow learning. These model families include ART 1 (Carpenter and Grossberg, 1987a), which can stably learn to categorize binary input patterns presented in an arbitrary order; ART 2, ART2-A, and Fuzzy ART (Carpenter and Grossberg, 1987b; Carpenter, Grossberg, and Rosen, 1991a, 1991b), which can stably learn to categorize either analog or binary input patterns presented in an arbitrary order; and ART 3 (Carpenter and Grossberg, 1990), which can carry out parallel search, or hypothesis testing, of distributed recognition codes in a multi-level network hierarchy. Variations of these models adapted to the demands of individual applications have been developed by a number of authors.

Figure 4

Figure 4 illustrates one example from the family of ART 1 models, and Figure 5 illustrates a typical ART search cycle. Level F_1 in Figure 4 contains a network of nodes, each of which represents a particular combination of sensory features. Level F_2 contains a network of nodes that represent recognition codes which are selectively activated by patterns of activation across F_1. The activities of nodes in F_1 and F_2 are also called short term memory (STM) traces. STM is the type of memory that can be rapidly reset without leaving an enduring trace. For example, it is easy to reset the STM of a list of numbers that a person has just heard once by distracting the person with an unexpected event. STM is distinct from LTM,

or long term memory, which is the type of memory that we usually ascribe to learning. For example, we do not forget our parents' names when we are distracted by an unexpected event.

Figure 5

As shown in Figure 5a, an input vector **I** registers itself as a pattern **X** of activity across level F_1. The F_1 output vector **S** is then transmitted through the multiple converging and diverging adaptive filter pathways emanating from F_1. This transmission event multiplies the vector **S** by a matrix of adaptive weights, or LTM traces, to generate a net input vector **T** to level F_2. The internal competitive dynamics of F_2 contrast-enhance vector **T**. Whereas many F_2 nodes may receive inputs from F_1, competition or lateral inhibition between F_2 nodes allows only a much smaller set of F_2 nodes to store their activation in STM. A compressed activity vector **Y** is thereby generated across F_2. In ART 1, the competition is tuned so that the F_2 node that receives the maximal $F_1 \rightarrow F_2$ input is selected. Only one component of **Y** is nonzero after this choice takes place. Activation of such a winner-take-all node defines the category, or symbol, of the input pattern **I**. Such a category represents all the inputs **I** that maximally activate the corresponding node. So far, these are the rules of a self-organizing feature map, also called competitive learning, self-organizing feature maps, or learned vector quantization.

In a self-organizing feature map, only the F_2 nodes that win the competition and store their activity in STM can influence the learning process. STM activity opens a learning gate at the LTM traces that abut the winning nodes. These LTM traces can then approach, or track, the input signals in their pathways by a process of steepest descent. This learning law is thus often called *gated steepest descent*, or *instar learning*. It was introduced into neural network models in Grossberg (1969) and is the learning law that was used to introduce ART (Grossberg, 1976b). Such an LTM trace can either increase or decrease to track the signals in its pathway. Since it is thus not a strictly Hebbian associative law, which allows traces only

to increase, the instar law is sometimes said to be both Hebbian and anti-Hebbian. It has been used to model neurophysiological data about hippocampal LTP (Levy, 1985; Levy and Desmond, 1985) and adaptive tuning of cortical feature detectors during the visual critical period (Rauschecker and Singer, 1979; Singer, 1983), lending support to ART predictions that these two systems would employ such a learning law (Grossberg, 1976b).

Self-organizing feature map models were introduced and computationally characterized in the early 1970s (Grossberg, 1972, 1976a, 1978; Malsburg, 1973). In brief, a model for learned classification by an instar network was introduced in Grossberg (1972). Malsburg (1973) modified the model equations used in that paper to introduce the first biologically-motivated self-organizing feature map (SOFM). The Malsburg model used a non-local adaptive weight equation and did not operate in real-time. The Grossberg (1976a) model (Figure 6) showed how to modify the Malsburg (1973) model to introduce the first locally defined real-time SOFM, whose rules have been used in essentially all such subsequent models. This model achieves L^1 normalization of its input patterns and adaptive weights. Rumelhart and Zipser (1985) used this version of the SOFM model in their computer simulations of competitive learning. Grossberg (1978) noted how a Euclidean, or L^2, norm achieves unbiased learning, and more generally considered L^P norms. Kohonen (1984/1989) popularized the L^2 version through his influential book. Other contributions and applications of the SOFM model include those of Amari and Takeuchi (1978), Bienenstock, Cooper, and Munro (1982), Cohen and Grossberg (1987), Grossberg, and Kuperstein (1986/1989), Linsker, (1986), and Willshaw and Malsburg (1976).

Figure 6

By now, there are many hundreds of papers that develop variants of the SOFM model, either mathematically or in applications. That is because they exhibit many useful properties, especially if not too many input patterns, or clusters of input patterns, perturb level F_1 relative to the number of categorizing nodes in level F_2. Under these sparse environmen-

tal conditions, category learning is provably stable: the LTM traces track the statistics of the environment, are self-normalizing, and oscillate a minimum number of times; and the classifier has Bayesian properties (Grossberg, 1976a, 1978). These are the key properties that Kohonen (1984/1989) discussed and applied. It was also proved, however, that under arbitrary environmental conditions, learning becomes unstable (Grossberg, 1976b). Such a model could forget its parents' faces. Although a gradual switching off of plasticity, or very slow learning, can partially overcome this problem, such a mechanism cannot work in a recognition learning system whose plasticity is maintained throughout adulthood.

This memory instability is due to basic properties of feedforward associative learning networks. An analysis of this instability, as well as of behavioral and neural data about categorization, learning, and attention, led to the introduction of ART top-down feedback, or expectation, that stabilizes the memory of SOFM models in response to an arbitrary stream of input patterns (Grossberg, 1976b). Thus ART models were introduced as examples of how internal control mechanisms can *self-stabilize* SOFM learning. Learning in SOFM models has also been partially stabilized by external control of the learning rate. One theme of considerable research interest still today is whether, and how, external or internal controllers are used to stabilize SOFM learning.

4. Hypothesis Testing, Attention, and Resonance

The ART scheme for self-stabilizing its embedded SOFM model incorporates heuristics that are also used in expert production systems and fuzzy systems. In particular, ART systems carry out a form of hypothesis testing to discover new recognition categories and to stabilize learning. Thus in an ART model (Carpenter and Grossberg, 1987a, 1991), learning does not occur whenever some winning F_2 activities are stored in STM. Instead activation of F_2 nodes may be interpreted as "making a hypothesis" about an input **I**. When **Y** is activated (Figure 5a), it generates an output vector **U** that is sent top-down through the second adaptive filter. After multiplication by the adaptive weight matrix of the top-down

filter, a net vector **V** inputs to F_1 (Figure 5b). Vector **V** plays the role of a learned top-down expectation. Activation of **V** by **Y** may be interpreted as "testing the hypothesis" **Y**, or "reading out the category prototype" **V**. The ART 1 network is designed to match the "expected prototype" **V** of the category against the active input pattern, or exemplar, **I**. Nodes that are activated by **I** are suppressed if they do not correspond to large LTM traces in the prototype pattern **V** (Figure 5c). Thus F_1 features that are not "expected" by **V** are suppressed. Expressed in a different way, the matching process may change the F_1 activity pattern **X** by suppressing activation of all the feature detectors in **I** that are not "confirmed" by hypothesis **Y**. The resultant pattern **X*** encodes the cluster of features in **I** that the network deems relevant to the hypothesis **Y** based upon its past experience. Pattern $\mathbf{X}^*$ encodes the pattern of features to which the network "pays attention."

If the expectation **V** is close enough to the input **I**, then a state of *resonance* develops as the attentional focus takes hold. The pattern **X*** of attended features reactivates hypothesis **Y** which, in turn, reactivates **X***. The network locks into a resonant state through the mutual positive feedback that dynamically links **X*** with **Y**. In ART, the resonant state, rather than bottom-up activation, drives the learning process. The resonant state persists long enough, at a high enough activity level, to activate the slower learning process; hence the term *adaptive resonance* theory. ART systems learn prototypes, rather than exemplars, because the attended feature vector $\mathbf{X}^*$, rather than the input **I** itself, is learned. These prototypes may, however, also be used to encode individual exemplars, as described below.

5. Pattern Matching, Stable Learning, and Phonemic Restoration

The ART attentive matching process may be realized in several ways (Carpenter and Grossberg, 1987a). In one instantiation, three different types of inputs are combined at level F_1 (Figure 4): bottom-up inputs, top-down expectations, and attentional gain control signals. The attentional gain control channel sends the same signal to all F_1 nodes; it is a "nonspecific", or modulatory, channel. This sort of attentive matching is said to obey a 2/3

Rule (Carpenter and Grossberg, 1987a): an F_1 node can be fully activated only if two of the three input sources that converge upon it send positive signals at a given time.

The 2/3 Rule allows F_1 nodes to generate supratheshold outputs in response to bottom-up inputs, since an input directly activates its target F_1 features and indirectly activates them via the nonspecific gain control channel to satisfy the 2/3 Rule (Figure 5a). After the input instates itself at F_1, leading to selection of a hypothesis $\mathbf{Y}$ and a top-down prototype $\mathbf{V}$, the 2/3 Rule ensures that only those F_1 nodes that are confirmed by the top-down prototype can remain active and be attended at F_1 after an F_2 category is selected, since top-down feedback shuts off the attentional gain control signals.

ART matching rules like the 2/3 Rule enable an ART network to realize a self-stabilizing learning process. Carpenter and Grossberg (1987a) proved that ART learning and memory are stable in arbitrary environments, but become unstable when the ART matching rule is eliminated. They also defined several circuits that generate the desired matching properties. Thus a type of matching that guarantees stable learning also enables the network to selectively pay attention to feature combinations that are confirmed by a top-down expectation.

ART matching in the brain is illustrated by experiments on phonemic restoration (Repp, 1991; Samuel, 1981a, 1981b; Warren, 1984; Warren and Sherman, 1974). Suppose that a noise spectrum replaces a letter sound in a word heard in an otherwise unambiguous context. Then subjects hear the correct letter sound, not the noise, to the extent that the noise spectrum includes the letter formants. If silence replaces the noise, then only silence is heard. Top-down expectations thus amplify expected input features while suppressing unexpected features, but do not create activations not already in the input.

ART matching rules also show how an ART system can be primed, or biased, to respond differently based upon prior short-term activation. This property has been used to explain paradoxical reaction time and error data from priming experiments during lexical decision and letter gap detection tasks (Grossberg and Stone, 1986; Schvaneveldt and MacDonald,

1981). Although priming is often thought of as a residual effect of previous bottom-up activation, a combination of bottom-up activation and top-down ART matching was needed to explain the complete data pattern. This analysis combined bottom-up priming with a type of top-down priming; namely, the top-down activation that prepares a network for an expected event that may or may not occur. ART matching rules clarify why top-down priming, by itself, has subthreshold (and in the brain unconscious) effects, even though it can facilitate suprathreshold processing of a subsequent expected event.

6. Vigilance, Hypothesis Testing, and the Control of Category Generalization

The criterion of an acceptable match is defined by a parameter ρ called *vigilance* (Carpenter and Grossberg, 1987a, 1991). The vigilance parameter is computed in the orienting subsystem $\mathcal{A}$. Vigilance weighs how similar an input exemplar must be to a top-down prototype in order for resonance to occur. Resonance occurs if $\rho|\mathbf{I}| - |\mathbf{X}^*| \leq 0$, where $0 \leq \rho \leq 1$. This inequality says that the F_1 attentional focus $\mathbf{X}^*$ inhibits $\mathcal{A}$ more than the input $\mathbf{I}$ excites it. If $\mathcal{A}$ remains quiet, then an $F_1 \leftrightarrow F_2$ resonance can develop.

Vigilance calibrates how much novelty the system can tolerate before activating $\mathcal{A}$ and searching for a different category. If the top-down expectation and the bottom-up input are too different to satisfy the resonance criterion, then hypothesis testing, or memory search, is triggered. Memory search leads to selection of a better category at level F_2 with which to represent the input features at level F_1. During search, the orienting subsystem interacts with the attentional subsystem, as in Figures 5c and 5d, to rapidly reset mismatched categories and to select other F_2 representations with which to learn about novel events, without risking unselective forgetting of previous knowledge. Search may select a familiar category if its prototype is similar enough to the input to satisfy the vigilance criterion. The prototype may then be refined by top-down attentional focussing. If the input is too different from any previously learned prototype, then an uncommitted population of F_2 cells is selected and learning of a new category is initiated.

Because vigilance can vary across learning trials, recognition categories capable of encoding widely differing degrees of generalization or abstraction can be learned by a single ART system. Low vigilance leads to broad generalization and abstract prototypes since then $\rho|\mathbf{I}| - |\mathbf{X}^*| \leq 0$ for all but the poorest matches. High vigilance leads to narrow generalization and to prototypes that represent fewer input exemplars, even a single exemplar. Thus a single ART system may be used, say, to recognize abstract categories of faces and dogs, as well as individual faces and dogs. A single system can learn both, during supervised learning, by increasing vigilance just enough to activate $\mathcal{A}$ if a previous categorization leads to a predictive error (Carpenter and Grossberg, 1992; Carpenter, Grossberg, and Reynolds, 1991; Carpenter, Grossberg, Markuzon, Reynolds, and Rosen, 1992). ART systems hereby provide a new answer to whether the brain learns prototypes or exemplars. Various authors have realized that neither one nor the other alternative is satisfactory, and that a hybrid system is needed (Smith, 1990). ART systems can perform this hybrid function in a manner that is sensitive to environmental demands.

7. Memory Consolidation and Direct Access to the Globally Best Category

As inputs are practiced over learning trials, the search process eventually converges upon stable categories. The process whereby search is automatically disengaged may be interpreted as a form of memory consolidation. Inputs familiar to the network access their correct category directly, without the need for search. The category selected is the one whose prototype provides the globally best match to the input pattern. If both familiar and unfamiliar events are experienced, familiar inputs can directly activate their learned categories, while unfamiliar inputs continue to trigger adaptive memory searches for better categories, until the network's memory capacity is fully utilized (Carpenter and Grossberg, 1987a).

These ART properties have been used to explain and predict of cognitive and brain data that have, as yet, received no other theoretical explanation (Carpenter and Grossberg, 1991;

Grossberg, 1987a, 1987b). For example, a formal lesion of the orienting subsystem creates a memory disturbance that remarkably mimics properties of medial temporal amnesia (Carpenter and Grossberg, 1987c, 1993; Grossberg and Merrill, 1992). These and related data correspondences to orienting properties (Grossberg and Merrill, 1992) have led to a neurobiological interpretation of the orienting subsystem in terms of the hippocampal formation of the brain. In applications to visual object recognition, the interactions within the F_1 and F_2 levels of the attentional subsystem are interpreted in terms of data concerning the prestriate visual cortex and the inferotemporal cortex (Desimone, 1992), with the attentional gain control pathway interpreted in terms of the pulvinar region of the brain.

Figure 7

8. Natural Link between ART Systems and Fuzzy Logic

Fuzzy ART is a generalization of ART 1 that incorporates operations from fuzzy logic (Carpenter, Grossberg, and Rosen, 1991b). Although ART 1 can learn to classify only binary input patterns, Fuzzy ART can learn to classify both analog and binary input patterns. Moreover, Fuzzy ART reduces to ART 1 in response to binary input patterns. As shown in Figure 7, the generalization to learning both analog and binary input patterns is achieved simply by replacing appearances of the binary intersection operator ($\cap$) in ART 1 by the analog MIN operator ($\wedge$) of fuzzy set theory. The MIN operator reduces to the intersection operator in the binary case. Of particular interest is the fact that, as parameter α approaches 0, the function T_j which controls category choice through the bottom-up filter (Figure 5a) then measures the degree to which the adaptive weight vector $\mathbf{w}_j$ is a fuzzy subset (Kosko, 1986) of the input vector $\mathbf{I}$. The network first chooses the category j that maximizes T_j.

Figure 8

In Fuzzy ART, input vectors are L^1 normalized at a preprocessing stage (Figure 8). This normalization procedure, called complement coding, leads to a symmetric theory in which the MIN operator ($\wedge$) and the MAX operator ($\vee$) of fuzzy set theory (Zadeh, 1965)

play complementary roles. Geometrically, the categories formed by Fuzzy ART are then hyper-rectangles. Figure 9 illustrates how MIN and MAX define these rectangles in the 2-dimensional case, with the MIN and MAX values defining the acceptable range of feature variation in each dimension. Complement coding uses on-cell (with activity **a** in Figure 8) and off-cell (with activity $\mathbf{a}^c$ in Figure 8) opponent processes to represent the input pattern. This representation preserves individual feature amplitudes while normalizing the total on-cell/off-cell vector. The on-cell portion of a prototype encodes features that are critically present in category exemplars, while the off-cell portion encodes features that are critically absent (Figure 8). Each category is then defined by an interval of expected values for each input feature (Figure 9). Thus, as noted in Section 1, for the category "man", Fuzzy ART would encode the feature of "hair on head" by a wide interval ([A, 1]) and the feature "hat on head" by a wide interval ([0, B]) (Figure 1). For the category "dog", two narrow intervals, [C, 1] for hair and [0, D] for hat correspond to narrower ranges of expectations for these two features.

Figure 9

Learning in Fuzzy ART is stable because all adaptive weights can only decrease in time. Decreasing weights correspond to increasing sizes of category "boxes". Smaller vigilance values lead to larger category boxes. Learning stops when the input space is covered by boxes. The use of complement coding works with the property of increasing box size to prevent a proliferation of categories. With fast learning, constant vigilance, and a finite input set of arbitrary size and composition, learning stabilizes after just one presentation of each input pattern. A fast-commit slow-recode option combines fast learning with a forgetting rule that buffers system memory against noise. Using this option, rare events can be rapidly learned, yet previously learned memories are not rapidly erased in response to statistically unreliable input fluctuations. The equations that define the Fuzzy ART and ARTMAP algorithms are listed in the Appendix.

When the supervised learning of Fuzzy ARTMAP cont[illegible] ordered set of dictive error can force the creation of new categories that [illegible]sted with Supervised learning permits the creation of complex cate[illegible]tabase stability.

Figure 10

9. Fuzzy ARTMAP

Each Fuzzy ARTMAP system includes a pair of Fuzzy ART modules (ART_a and ART_b) as in Figure 10. During supervised learning, ART_a receives a stream $\{\mathbf{a}^{(p)}\}$ of input patterns and ART_b receives a stream $\{\mathbf{b}^{(p)}\}$ of input patterns, where $\mathbf{b}^{(p)}$ is the correct prediction given $\mathbf{a}^{(p)}$. These modules are linked by an associative learning network and an internal controller that ensures autonomous system operation in real time. The controller is designed to create the minimal number of ART_a recognition categories, or "hidden units," needed to meet accuracy criteria. As noted above, this is accomplished by realizing a Minimax Learning Rule that conjointly minimizes predictive error and maximizes category generalization. This scheme automatically links predictive success to category size on a trial-by-trial basis using only local operations. It works by increasing the vigilance parameter ρ_a of ART_a by the minimal amount needed to correct a predictive error at ART_b (Figure 11).

Figure 11

Parameter ρ_a calibrates the minimum confidence that ART_a must have in a recognition category, or hypothesis, that is activated by an input $\mathbf{a}^{(p)}$ in order for ART_a to accept that category, rather than search for a better one through an automatically controlled process of hypothesis testing. As in ART 1, lower values of ρ_a enable larger categories to form. These lower ρ_a values lead to broader generalization and higher code compression. A predictive failure at ART_b increases the minimal confidence ρ_a by the least amount needed to trigger hypothesis testing at ART_a, using a mechanism called *match tracking* (Carpenter, Grossberg, and Reynolds, 1991). Match tracking sacrifices the minimum amount of generalization

necessary to correct the predictive error. Speaking intuitively match track embodies the idea that the criterion confidence level that permitted selection of the active hypothesis needs to be raised to satisfy the demands of the current environment. Match tracking increases the criterion confidence just enough to trigger hypothesis testing. Hypothesis testing leads to the selection of a new ART_a category, which focuses attention on a new cluster of $\mathbf{a}^{(p)}$ input features that is better able to predict $\mathbf{b}^{(p)}$. The combination of match tracking and fast learning allows a single ARTMAP system to learn a different prediction for a rare event than for a cloud of similar frequent events in which it is embedded. The equations for Fuzzy ART and Fuzzy ARTMAP are given in the Appendix in algorithmic form.

The next two sections illustrate how variants of Fuzzy ARTMAP can be embedded into larger neural architectures.

10. VIEWNET: A Neural Architecture for Learning to Recognize 3-D Objects from Sequences of 2-D Views

This section shows how a Fuzzy ARTMAP network can be incorporated into a self-organizing neural architecture for invariant 3-D object recognition (Bradski and Grossberg, 1994). This architecture is called VIEWNET because it uses View Information Encoded With NETworks (Figure 12). VIEWNET accumulates evidence across sequences of possibly noisy or incomplete 2-D views of a 3-D object in order to generate more accurate object identifications than would otherwise be possible. VIEWNET processes individual 2-D views of 3-D objects using the CORT-X 2 filter (Carpenter, Grossberg, and Mehanian, 1989; Grossberg and Wyse, 1991), which discounts the illuminant, regularizes and completes figural boundaries, and removes noise from the images. A log-polar transform is taken with respect to the centroid of the resulting figure and then re-centered to achieve 2-D scale and rotation invariance. The invariant images are coarse coded to further reduce noise, reduce foreshortening effects, and increase generalization. These compressed codes are input into a variant of the Fuzzy ARTMAP algorithm which learns 2-D view categories. Evidence from sequences

of 2-D view categories is stored in a working memory. Voting based on the unordered set of stored categories determines object recognition. Recognition performance was tested with noisy and clean images using slow and fast learning on an MIT Lincoln Laboratory database of 2-D views of aircraft with and without additive noise. A recognition rate of up to 90% was achieved with one 2-D view category and of up to 98.5% correct with three 2-D view categories.

Figure 12

Seibert and Waxman (1990a, 1990b, 1991, 1992) pioneered the use of 2-D view sequences for 3-D object recognition by adaptive neural networks. Seibert and Waxman developed their neural architecture based on the Koenderink and van Doorn (1979) concept of Aspect Graphs. The VIEWNET model is inspired by the Seibert-Waxman model, but uses a different preprocessor, adaptive pattern classifier, and evidence accumulation scheme to achieve better compression and higher accuracy on the same data base from MIT Lincoln Laboratory, which they generously shared.

Seibert and Waxman used an unsupervised ART 2 network to classify coarse-coded maximal curvature image data. This approach generated unambiguous recognition based on one category only 25% of the time, using 41 categories, as compared to the VIEWNET 90% accuracy using 33 categories.

To compensate for the general categories formed in their model, the Seibert-Waxman model used ordered view transitions between pairs of view categories to supplement individual recognition categories. For example, to recognize multiple views of an F-16 aircraft, 70 view transitions were learned. VIEWNET used only the unordered set of 2-D view categories, stored in a working memory, to vote for a best object. No view transitions were used, and it was shown that they do not improve accuracy. Two views achieved up to 94% accuracy, and three views up to 98.5% accuracy. These results suggest that the selection of preprocessor, classifier, and evidence accumulation scheme may substantially alter predictive accuracy. A

number of issues, including the possible utility of ordered view transitions, remain open for further study.

The image database used to test the VIEWNET architecture consists of multiple 2-D images of three jets. Video images were taken of 3 airplane models: an F-16, an F-18, and an HK-1. Each airplane was painted black and suspended by string against a light background to aid in segmentation. The camera was mounted anywhere in an arc around the jets that started at 0.0 degrees above horizontal and went in increments of 4.5 degrees to a maximum of 72.0 degrees above horizontal. For each camera angle, the airplanes were spun and frames covering one full revolution (an average of 88 frames) were retained resulting in 1200 to 1400 images per object. The images themselves were 128x128 pixel gray scale. The images were then thresholded and binarized into a SUN raster format to form the "raw" database. For our processing, data was turned into a floating point format scaled between 0.0 and 1.0 and an additive noise process was introduced. The noise consisted of a 128×128 pixel images with each pixel taken from a uniform distribution between 0.0 and 1.0 scaled by a constant $C \geq 0.0$. These scaled, 128×128 noise images were then added to the 128×128 jet images prior to preprocessing. Thus, both noise-free and noisy 2-D views covering a half-sphere surrounding the 3-D object were collected, keeping their spatial relationships intact.

Even numbered rotation images from each camera angle were taken as the training set with the odd numbered images forming the test set. The system was trained using random walks over the half-sphere of training images. Testing was done using random walks over the half-sphere of test images so that the paths taken and views seen were never the same between the training and test sets.

Figure 13

The CORT-X 2 filter (Grossberg and Wyse, 1991, 1992) discounts the illuminant and normalizes image contrasts, regularizes and completes figural boundaries, and suppresses image noise. Alternative filters could have been used on the present data. CORT-X 2 was

used because it is also capable of preprocessing more complex noisy images. The processing stages of CORT-X 2 are shown in Figure 13. They embody a one-shot feedforward simplification of a model of the boundary segmentation process that takes place in visual cortex (Grossberg, Mingolla, and Todorović, 1989).

The 2-D boundary segmentation is centered by dividing its 1^{st} moments by its 0^{th} moment to find the figure centroid, subtracting off the center of the image and then shifting the figure by this amount. A log-polar transform is then taken with respect to the center of the image. Each point (x, y) is represented as $re^{i\theta}$. Taking the logarithm yields coordinates of log radial magnitude and angle. As is well known (Schwartz, 1977) figural sizes and rotations of a centered image are converted into figural shifts under log-polar transformation. Using these shift parameters to center the log-polar transformed image leads to a figural representation that is invariant under 2-D changes in position, size and rotation.

Coarse coding, or data reduction, reduces memory requirements as it helps to compensate for inaccuracies of figural alignment, 3-D viewpoint specific foreshortening, and self-occlusions. Too much coarse coding can, however, obscure critical input features and thereby harm recognition performance. These effects were balanced to maximize the benefits of coarse coding.

Coarse coded used a spatial averaging method that convolved the original image I with a function Ψ and then sampled the resultant image with delta functions spaced every T pixels: $\delta(x - nT, y - kT)$. For simplicity, in 1D this is

$$(I * \Psi) \cdot \sum_{n=-\infty}^{\infty} \delta(x - nT). \tag{1}$$

If the Fourier transform of I is $\hat{I}$, and that of Ψ is $\hat{\Psi}$, then the Fourier transform of equation is

$$(\hat{I} \cdot \hat{\Psi}) * \frac{2\pi}{T} \sum_{k=-\infty}^{\infty} \delta(\Omega - k\Omega_s), \tag{2}$$

where $\Omega_s = 2\pi/T$, and T is the sampling period in pixels. If Ω_N is the highest frequency in the image, then for the image to be uniquely determined by its samples, we must have by the Nyquist sampling theorem that

$$\Omega_s = \frac{2\pi}{T} > 2\Omega_N. \tag{3}$$

Two simple spatial averaging functions Ψ are: (1) uniform averaging of the input image so that all pixels in a window of some width are summed and divided by the number of pixels in the window; (2) Gaussian averaging of the input image so that a normalized, Gaussian weighted sum of all pixels is taken over a window of some width. Both approaches were investigated by Bradski and Grossberg (1994). Method (1) has the problem that uniform averaging is a rectangular filter in the space domain and a sinc function in the frequency domain which introduces high frequency aliasing ("ringing") in the resultant image. The Gaussian function of method (2) is a "smoother" low pass filter and so does not suffer from this problem. A Gaussian is also an eigenfunction of a Fourier transform, which simplifies calculation.

Figure 14

To best set the standard deviation σ of the Gaussians, we define two standard deviations away from the Gaussian midpoint to be essentially zero. The cutoff frequency of such a low pass filter is then $\pi/2\sigma$, which by equation (3) yields at equality:

$$\sigma = \frac{T}{2}. \tag{4}$$

Thus, the zero point of each Gaussian just touches the center of the next Gaussian. Figure 14 summarizes the preprocessing: 14(a) shows the output of CORT-X 2, 14(b) the centered log

polar transform of (a), 14(c) depicts Gaussian coarse coding according to equation (4), and 14(d-f) show coarse coding down to 16 × 16, 8 × 8, and 4 × 4 pixels. The best results, as summarized in Table 4, were achieved with 16 x 16 coarse coding.

Table 4

A modified version of Fuzzy ARTMAP was used in the VIEWNET architecture. First the network was simplified, as in Carpenter, Grossberg, and Iizuka (1992), to consist of a Fuzzy ART module (Carpenter, Grossberg, and Rosen, 1991b) ART_a and a field of output nodes F^b, rather than an ART_b module, linked to ART_a by an associative memory F^{ab} that is called the *map field* (see Appendix). Fuzzy ARTMAP was modified to allow for on-line slow learning from ART_a 2-D view category nodes F_2^a to the map field nodes (Carpenter, Grossberg, and Reynolds, 1993, 1994). A maximal ART_a vigilance level, $\bar{\rho}_{max}$ was introduced such that an error at the map field triggers match tracking only if match tracking leads to a vigilance $\rho_a \leq \bar{\rho}_{max}$. If $\rho_a > \bar{\rho}_{max}$, learning takes place instead of memory search. By setting the map field learning rate β_{ab}, baseline ($\bar{\rho}_a$) and maximal ($\bar{\rho}_{max}$) vigilance levels appropriately, weights from F_2^a nodes to the map field approximate the conditional probability of the true class given the selected F_2^a category (Table 4).

For the airplane data set as processed by VIEWNET, the average overall length of an error sequence was 1.31 2-D views with a standard deviation of 0.57 views. Thus, when an error occurs, collecting evidence from, and voting over, two more views will usually be sufficient to correct the error. This can be done most simply in VIEWNET by adding an integration field (F^{int}), or working memory, to accumulate evidence between the map field (F^{ab}) and the winner-take-all field (F^{wta}). The equation for the integrator field is stepped once each time ART_a chooses a category:

$$(x_k^{int})^{new} = \beta_{int} x_k^{ab} + (1 - \beta_{int})(x_k^{int})^{old}, \tag{5}$$

where x_k^{int} is an integrator node for the k^{th} object, β_{int} is the integration rate each time

the equation is stepped, and x_k^{ab} is the k^{th} map field category. The maximal integration node is chosen by the winner-take-all field (F^{wta}) as the network's identification of the 3-D object. Thus VIEWNET recognition is based upon the unordered winner of a temporal voting scheme.

The Fuzzy ARTMAP architecture computes goodness of fit information that may be used to enhance its power in various applications. For example, the algorithm's match or choice equation in may be used to measure to the *quality* of the recognition. Thus if VIEWNET recognizes a 3-D object, but its ART_a category prototype provides a poor fit to the input vector, then the goodness of fit information could be used to cause VIEWNET to collect more data before a final recognition decision is made. Likewise, if VIEWNET is embedded in an active vision system, then a poorly fitting view could be used to trigger the system to move to get a better perspective.

11. ART-EMAP: Object Recognition by Spatial and Temporal Evidence Accumulation

ART-EMAP (Figure 15) uses spatial and temporal evidence accumulation to recognize target objects and pattern classes in noisy or ambiguous input environments (Carpenter and Ross, 1993, 1994b). During performance, ART-EMAP integrates spatial evidence distributed across recognition categories to predict a pattern class. During training, ART-EMAP is equivalent to fuzzy ARTMAP and so inherits the advantages of fast, on-line, incremental learning, such as speed, stability, and the ability to encode rare cases (Section 1). Distributed activation during performance also endows the network with the advantages of slow learning, including noise tolerance and error correction. When a decision criterion determines the pattern class choice to be ambiguous, additional input from the same unknown class may be sought. Evidence from multiple inputs accumulates until the decision criterion is satisfied and the system makes a high confidence prediction. Accumulated evidence can also fine-tune performance during unsupervised rehearsal learning.

Figure 15

In four incremental stages, ART-EMAP improves predictive accuracy of fuzzy ARTMAP and extends its domain to include spatio-temporal recognition and prediction. ART-EMAP applications include a vision system that samples 2-D perspectives of 3-D objects. In this scenario, a sensor generates an organized database of inputs that are views of each object from different perspectives or noisy samples of fixed views. Evidence accumulation has been successfully used in neural network machine vision applications, as in the aspect network (Baloch and Waxman, 1991; Seibert and Waxman, 1990b). ART-EMAP further develops this strategy.

Figure 16

3-D object recognition: Simulations illustrate performance of fuzzy ARTMAP and ART-EMAP (Stages 1–4), on a recognition problem that requires a system to identify three similar 3-D objects (pyramid, prism, house). Inputs consist of ambiguous 2-D views taken from various angles (Figure 16). The problem is made difficult by the similarity of views across objects and by several test set views that do not resemble any training set view of the same object. Fuzzy ARTMAP correctly identifies only 65% of the objects from noise-free test set images. Stage 1 ART-EMAP raises performance accuracy to 71%, while Stage 2 and Stage 3 both boost performance to 98.0% (Figure 17).

Figure 17

Database inputs: The simulation database was constructed using Mathematica to generate shaded 2-D projections of 3-D objects illuminated by an achromatic point light source. For each of the three objects, 24 training set views were obtained from perspectives spaced 30° to 60° apart around a viewing hemisphere (Figure 16a). For each object, 17 test set views, spaced at 45° intervals, were obtained from perspectives between those of the training set (Figure 16b). Each 2-D view was then preprocessed, using Gabor filters (Gabor,

1946; Daugman, 1988) to recover boundaries, competitive interactions to sharpen boundary locations and orientations (Grossberg and Mingolla, 1985), and coarse coding, to yield a 100-component input vector **a**. The preprocessing algorithm is a typical feature extractor, chosen to illustrate comparative performance of different recognition systems, and was not selected to optimize performance of any one of these systems.

Training regime: Fuzzy ARTMAP and ART-EMAP Stage 1 through Stage 4 were evaluated using both a noise-free test set and a noisy test set. The noisy test set was constructed by adding Gaussian noise (SD = 0.2) to each input component. Each system was initially trained under one standard supervised learning protocol, with the training set presented once. Since the training set views were selected to be sparse and nonredundant, a situation of minimal code compression was simulated during training. This was achieved by assigning a high value to the ARTMAP baseline vigilance ($\bar{\rho}_a = 0.9$), which established 58 ART_a recognition categories for the 72 training set pairs.

Fuzzy ARTMAP simulation: Performance measures of fuzzy ARTMAP and ART-EMAP on the 3-D object recognition database are summarized in Figure 17, for noise-free test set inputs (plots a-c) and for noisy test set inputs (plots d-f). The prediction of each test set view is represented graphically, on shaded viewing hemispheres. Each hemisphere shows 17 faces, which correspond to the 17 test set viewing angles (Figure 16b). For each simulation, three hemispheres show object class predictions made by the system in response to the corresponding input, with shading of a face indicating a prediction of pyramid (black), prism (gray), or house (white).

Fuzzy ARTMAP made only 64.7% correct object class predictions on the noise-free test set (Figure 17a), and 60.8% correct predictions on the noisy test set (Figure 17d). This poor performance indicates the difficult nature of the problem when prediction must be made on the basis of a single view. Note, for example, that many of the test set inputs from the lower left part of the pyramid view hemisphere were incorrectly identified as prism views. The reason for these errors can be inferred from the correspondence between certain pyramid

test set views and similar prism training set views (Figure 16).

ART-EMAP Stage 1: Spatial evidence accumulation

ART-EMAP employs a spatial evidence accumulation process that integrates a distributed pattern of activity across coded category nodes to help disambiguate a noisy or novel test set input. In contrast, previous ART and ARTMAP simulations chose only the most highly activated, winner-take-all category node at the field F_2^a as the basis for recognition and prediction.

In the fast-learn fuzzy ARTMAP system, the input from F_1^a to the j^{th} F_2^a node (Figure 15) is given by:

$$T_j^a = \frac{|\mathbf{A} \wedge \mathbf{w}_j^a|}{\alpha + |\mathbf{w}_j^a|}, \tag{6}$$

as in Figure 7. Fuzzy ARTMAP uses a binary choice rule:

$$y_J^a = \begin{cases} 1 & \text{if } T_J^a > T_j^a \text{ for all } j \neq J \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Then, only the F_2^a category J that receives maximal $F_1^a \rightarrow F_2^a$ input predicts the ART_b output.

ART-EMAP also uses the binary choice rule (7) during the initial period of supervised training. However, during performance, F_2^a output $\mathbf{y}^a$ is determined by less extreme contrast enhancement of the $F_1^a \rightarrow F_2^a$ input pattern $\mathbf{T}^a$. Limited contrast enhancement extracts more information from the relative activations of F_2^a categories than does the all-or-none choice rule (7).

Power rule: Raising the input T_j^a of the j^{th} F_2^a category to a power $p > 1$ is a simple way to implement contrast enhancement. Equation (8) defines a normalized power rule:

$$y_j^a = \frac{(T_j^a)^p}{\sum_{n=1}^{N_a} (T_n^a)^p}. \tag{8}$$

Normalization constrains the F_2^a output values to a manageable range without altering relative values or subsequent predictions. The power rule (8) approximates the dynamics of a shunting competitive short-term-memory (STM) network that contrast-enhances its input pattern (Grossberg, 1973). The power rule is equivalent to the choice rule (7) when p is large. For smaller p, the distributed activity pattern (8) uses information from the relative F_2^a category activations to improve test set predictive performance at ART_b. In all ART-EMAP 3-D object simulations, $p = 24$.

After contrast enhancement, the F_2^a output $\mathbf{y}^a$ is filtered through the weights w_{jk}^{ab} to activate the EMAP field F_1^{ab}. The input S_k^{ab} from F_2^a to the $k^{th} F_1^{ab}$ node obeys the equation:

$$S_k^{ab} = \sum_{j=1}^{N_a} w_{jk}^{ab} y_j^a. \tag{9}$$

Since distributed F_2^a activity generally determines distributed EMAP field F_1^{ab} input, some means of choosing a winning prediction at the EMAP field is required. The simplest method is to choose the EMAP category K that receives maximal input from F_2^a. This can be implemented by letting $x_k^{ab} = S_k^{ab}$ and defining F_2^{ab} activity by:

$$y_K^{ab} = \begin{cases} 1 & \text{if } x_K^{ab} > x_k^{ab} \text{ for all } k \neq K \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

Other methods for predicting an ART_b category will be discussed below.

Stage 1 simulation: Like fuzzy ARTMAP, Stage 1 ART-EMAP, with its spatially distributed activity pattern at F_2^a, is required to make a prediction from each single test set view. Nevertheless, predictive accuracy improves significantly, from 64.7% to 70.6% on the noise-free test set (Figure 17b) and from 60.8% to 64.7% on the noisy test set (Figure 17e).

Stage 2: EMAP predictive decision criterion

An alternative to the Stage 1 predictive choice rule (10) uses a *decision criterion* (DC) at the EMAP field F_2^{ab}. The decision criterion permits ART_b choice only when the most active category K becomes a minimum proportion more active than the next most active EMAP category. Thus:

$$y_K^{ab} = \begin{cases} 1 & \text{if } x_K^{ab} > (DC)x_k^{ab} \text{ for all } k \neq K \\ 0 & \text{otherwise,} \end{cases} \tag{11}$$

where DC ≥ 1. With DC $= 1$, the Stage 2 decision criterion rule (11) reduces to the Stage 1 F_2^{ab} choice rule (10). With DC > 1, the decision criterion prevents prediction when multiple EMAP categories are about equally activated at F_1^{ab}, representing ambiguous predictive evidence. As the DC increases, both accuracy and the number of required input samples per decision tend to increase. When the decision criterion fails, and (11) implies that $y_k^{ab} = 0$ for all k, additional input is sought to resolve the perceived ambiguity. In an application, additional inputs might correspond to multiple views or to multiple samples of a single view.

Stage 2 simulation: Stage 1 spatial evidence accumulation improves performance by causing a novel view to activate categories of two or more nearby training set views, which then strongly predict the correct object. However, many single view errors, caused by similar views across different objects, remain. Stage 2 or Stage 3 corrects most of these errors, when multiple views of the unknown object are available. With a high fixed decision criterion (DC=5.0) and an average of 4.8 test set views, Stage 2 ART-EMAP achieves 98.0% accuracy on the noise-free test set. Even on the noisy test set, object identification remains at 90.2% accurate, with an average of 6.8 test set views. Both performance and the average number of views decrease as the fixed decision criterion decreases from 5 to 1.

Stage 3: Temporal evidence accumulation

The predictive decision criterion strategy (Stage 2 ART-EMAP) searches multiple views or samples until one input satisfies the decision criterion. However any single noisy input

vector **a** might produce map field activity that satisfies a given decision criterion but still make an incorrect prediction. The Stage 2 strategy does not benefit from the partial evidence provided by all the views that failed to meet the decision criterion. Further performance improvement in a noisy input environment is achieved through the application of a decision criterion to *time-integrated* predictions that are generated by multiple inputs. Stage 3 ART-EMAP accumulates evidence at a *map evidence accumulation field* F_E^{ab} (Figure 15). The time scale of this medium-term memory (MTM) process is longer than that of the STM field activations resulting from the presence of a single view, but shorter than the long-term memory (LTM) stored in adaptive weights.

Additive evidence integration: A straightforward way to implement evidence accumulation at the EMAP module is to sum a sequence of F_1^{ab} map activations at the evidence accumulation field F_E^{ab}:

$$(T_k^{ab})^{(new)} = (T_k^{ab})^{(old)} + x_k^{ab}. \tag{12}$$

At F_E^{ab}, evidence accumulating MTM (T_k^{ab}) starts at zero and is reset to zero when the decision criterion is met. Activities y_k^{ab} at field F_2^{ab} obey:

$$y_K^{ab} = \begin{cases} 1 & \text{if } T_K^{ab} > (DC)T_k^{ab} \text{ for all } k \neq K \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

A decision will eventually be made if the DC starts large and gradually decreases toward 1. As in Stage 2, larger DC values tend to covary with both greater accuracy and longer input sequences. In simulations, the DC decreased exponentially from 6 to 1:

$$DC(l) = 5(1.0 - r)^{l-1} + 1, \tag{14}$$

where $\mathbf{a}(l)$ is the l^{th} input in a same-class sequence (l = 1,2, ...). The decay rate (r) was set equal to 0.2. Additive integration is equivalent to applying the decision criterion to a

running average of map field activations $\mathbf{x}^{ab}$ rather than to $\mathbf{x}^{ab}$ itself.

Stage 3 simulation: For a two-class prediction problem, evidence accumulation improves performance primarily by averaging across noisy inputs. Stage 3 ART-EMAP becomes increasingly useful as the number of predicted classes increases, since evidence accumulation can also help solve the difficult problem of disambiguating nearly identical views of different objects. With three or more object classes, when equal predictive evidence may exist for both the correct object and an incorrect one, the identity of the erroneous class tends to vary from one input to the next. As the sequence of views grows, erroneous evidence is quickly overwhelmed by evidence for the correct object. In the Stage 3 ART-EMAP three-object simulations, with the decreasing DC function (14), an average of 9.2 views were needed to reach 98.0% correct performance on the noise-free test set (Figure 17c). On the noisy test set, an average of 11.3 views allowed the system to reach 92.2% correct performance (Figure 17f).

Stage 4: Unsupervised rehearsal learning

Temporal evidence accumulation allows the Stage 3 ART-EMAP system to recognize objects from a series of ambiguous views. However the system learns nothing from the final outcome of this decision process. If, for example, an input sequence $\mathbf{a}^{(1)}$, ..., $\mathbf{a}^{(L)}$ predicts an ART_b category K, by (12)-(13), the entire sequence would need to be presented again before the same prediction would be made.

Unsupervised rehearsal learning (Stage 4) fine-tunes performance by feeding back to the system knowledge of the final prediction. Specifically, after input $\mathbf{a}^{(L)}$ allows ART-EMAP to choose the ART_b category K, the sequence $\mathbf{a}^{(1)}$, ..., $\mathbf{a}^{(L)}$ is re-presented, or rehearsed. Weights in an adaptive filter u_{jk}^{ab} from F_2^a to F_E^{ab} are then adjusted, shifting category decision boundaries so that each input $\mathbf{a}^{(l)}$ in the sequence becomes more likely, on its own, to predict category K.

Stage 4 simulation: Unsupervised rehearsal learning improves single view test set

performance only marginally on the 3-D object simulations. Stage 4 rehearsal learning was conducted on the 51 noise-free test set views. Temporal evidence accumulation drew from an enlarged test set that included 72 additional views. Accessing exemplars from this larger test set allows stable fine-tuning by decreasing the percentage of ambiguous test views. After this fine-tuning, performance on individual views from the original 51 test set inputs was 73%, compared to 70.6% at Stage 1 (Figure 17b).

Spatial and temporal evidence accumulation by ART-EMAP have been shown to improve fuzzy ARTMAP performance on both the ARPA benchmark circle-in-the-square problem (Carpenter and Ross, 1993, 1994b; Wilensky, 1990) and on the 3-D object recognition problem described here. Unsupervised rehearsal learning illustrates how self-training can fine-tune system performance. ART-EMAP is a general purpose algorithm for pattern class prediction based on the temporal integration of predictive evidence resulting from distributed recognition across a small set of trained categories. The system promises to be of use in a variety of applications, including spatio-temporal image analysis and prediction as well as recognition of 3-D objects from ambiguous 2-D views.

12. Concluding Remarks

ARTMAP systems illustrate how neural networks can incorporate properties of fuzzy logic and expert production systems into a unified computational framework. Such algorithms are helping to overcome previously arbitrary boundaries between these disciplines. They exhibit combinations of properties that have not been attainable by more traditional approaches, which helps to explain the rapidly growing number of diverse applications in which they are being successfully used.

13. Appendix: Fuzzy ART Algorithm and Fuzzy ARTMAP

Fuzzy ART activity vectors: Each ART system includes a field F_0 of nodes that represent a current input vector; a field F_1 that receives both bottom-up input from F_0 and top-down input from a field F_2 that represents the active code, or category. The F_0 activity

vector is denoted $\mathbf{I} = (I_1, \ldots, I_M)$, with each component I_i in the interval [0,1], $i = 1, \ldots, M$. The F_1 activity vector is denoted $\mathbf{x} = (x_1, \ldots, x_M)$ and the F_2 activity vector is denoted $\mathbf{y} = (y_1, \ldots, y_N)$. The number of nodes in each field is arbitrary.

Weight vector: Associated with each F_2 category node $j (j = 1, \ldots, N)$ is a vector $\mathbf{w}_j \equiv (w_{j1}, \ldots, w_{jM})$ of adaptive weights, or LTM traces. Initially

$$w_{j1}(0) = \ldots = w_{jM}(0) = 1; \tag{A1}$$

then each category is said to be *uncommitted*. After a category is selected for coding it becomes *committed*. As shown below, each LTM trace w_{ji} is monotone nonincreasing through time and hence converges to a limit. The Fuzzy ART weight vector $\mathbf{w}_j$ subsumes both the bottom-up and top-down weight vectors of ART 1.

Parameters: Fuzzy ART dynamics are determined by a choice parameter $\alpha > 0$; a learning rate parameter $\beta \in [0, 1]$; and a vigilance parameter $\rho \in [0, 1]$.

Category choice: For each input $\mathbf{I}$ and F_2 node j, the *choice function* T_j is defined by

$$T_j(\mathbf{I}) = \frac{|\mathbf{I} \wedge \mathbf{w}_j|}{\alpha + |\mathbf{w}_j|}, \tag{A2}$$

where the fuzzy AND (Zadeh, 1965) operator $\wedge$ is defined by

$$(\mathbf{p} \wedge \mathbf{q})_i \equiv \min(p_i, q_i) \tag{A3}$$

and where the norm $|\cdot|$ is defined by

$$|\mathbf{p}| \equiv \sum_{i=1}^{M} |p_i|. \tag{A4}$$

for any M-dimensional vectors $\mathbf{p}$ and $\mathbf{q}$. For notational simplicity, $T_j(\mathbf{I})$ in (A2) is often written as T_j when the input $\mathbf{I}$ is fixed.

The system is said to make a *category choice* when at most one F_2 node can become active at a given time. The category choice is indexed by J, where

$$T_J = \max\{T_j : j = 1 \ldots N\}. \tag{A5}$$

If more than one T_j is maximal, the category j with the smallest index is chosen. In particular, nodes become committed in order $j = 1, 2, 3, \ldots$. When the J^{th} category is chosen, $y_J = 1$; and $y_j = 0$ for $j \neq J$. In a choice system, the F_1 activity vector $\mathbf{x}$ obeys the equation

$$\mathbf{x} = \begin{cases} \mathbf{I} & \text{if } F_2 \text{ is inactive} \\ \mathbf{I} \wedge \mathbf{w}_J & \text{if the } J^{th} \ F_2 \text{ node is chosen.} \end{cases} \tag{A6}$$

Resonance or reset: *Resonance* occurs if the *match function* $|\mathbf{I} \wedge \mathbf{w}_J|/|\mathbf{I}|$ of the chosen category meets the vigilance criterion:

$$\frac{|\mathbf{I} \wedge \mathbf{w}_J|}{|\mathbf{I}|} \geq \rho; \tag{A7}$$

that is, by (A6), when the J^{th} category is chosen, resonance occurs if

$$|\mathbf{x}| = |\mathbf{I} \wedge \mathbf{w}_J| \geq \rho |\mathbf{I}|. \tag{A8}$$

Learning then ensues, as defined below. *Mismatch reset* occurs if

$$\frac{|\mathbf{I} \wedge \mathbf{w}_J|}{|\mathbf{I}|} < \rho; \tag{A9}$$

that is, if

$$|\mathbf{x}| = |\mathbf{I} \wedge \mathbf{w}_J| < \rho |\mathbf{I}|. \tag{A10}$$

Then the value of the choice function T_J is set to 0 for the duration of the input presentation to prevent the persistent selection of the same category during search. A new index J is then chosen, by (A5). The search process continues until the chosen J satisfies (A7).

Learning: Once search ends, the weight vector $\mathbf{w}_J$ is updated according to the equation

$$\mathbf{w}_J^{(\text{new})} = \beta(\mathbf{I} \wedge \mathbf{w}_J^{(\text{old})}) + (1 - \beta)\mathbf{w}_J^{(\text{old})}. \tag{A11}$$

Fast learning corresponds to setting $\beta = 1$.

Fast-commit slow-recode option: For efficient coding of noisy input sets, it is useful to set $\beta = 1$ when J is an uncommitted node, and then to take $\beta < 1$ after the category

is committed. Then $\mathbf{w}_J^{(\text{new})} = \mathbf{I}$ the first time category J becomes active. Moore (1989) introduced the learning law (A11), with fast commitment and slow recoding, to investigate a variety of generalized ART 1 models. Some of these models are similar to Fuzzy ART, but none includes the complement coding option. Moore described a category proliferation problem that can occur in some analog ART systems when a large number of inputs erode the norm of weight vectors. Complement coding solves this problem.

Input normalization/complement coding option: Proliferation of categories is avoided in Fuzzy ART if inputs are normalized. *Complement coding* is a normalization rule that preserves amplitude information. Complement coding represents both the on-response and the off-response to an input vector **a** (Figure 8). To define this operation in its simplest form, let **a** itself represent the on-response. The complement of **a**, denoted by $\mathbf{a}^c$, represents the off-response, where

$$a_i^c \equiv 1 - a_i. \tag{A12}$$

The complement coded input **I** to the field F_1 is the 2M-dimensional vector

$$\mathbf{I} = (\mathbf{a}, \mathbf{a}^c) \equiv (a_1, \ldots, a_M, a_1^c, \ldots, a_M^c). \tag{A13}$$

Note that

$$\begin{aligned} |\mathbf{I}| &= |(\mathbf{a}, \mathbf{a}^c)| \\ &= \sum_{i=1}^{M} a_i + (M - \sum_{i=1}^{M} a_i) \\ &= M, \end{aligned} \tag{A14}$$

so inputs preprocessed into complement coding form are automatically normalized. Where complement coding is used, the initial condition (A1) is replaced by

$$w_{j1}(0) = \ldots = w_{j,2M}(0) = 1. \tag{A15}$$

Fuzzy ARTMAP Algorithm

The Fuzzy ARTMAP system incorporates two Fuzzy ART modules ART_a and ART_b that are linked together via an inter-ART module F^{ab} called a *map field* (Figure 10). The

map field is used to form predictive associations between categories and to realize the *match tracking rule* whereby the vigilance parameter of ART_a increases in response to a predictive mismatch at ART_b. The interactions mediated by the map field F^{ab} may be operationally characterized as follows.

ART_a and ART_b: Inputs to ART_a and ART_b are in the complement code form: for ART_a, $\mathbf{I} = \mathbf{A} = (\mathbf{a}, \mathbf{a}^c)$; for ART_b, $\mathbf{I} = \mathbf{B} = (\mathbf{b}, \mathbf{b}^c)$. Variables in ART_a or ART_b are designated by subscripts or superscripts "a" or "b". For ART_a, let $\mathbf{x}^a \equiv (x_1^a \dots x_{2M_a}^a)$ denote the F_1^a output vector; let $\mathbf{y}^a \equiv (y_1^a \dots y_{N_a}^a)$ denote the F_2^a output vector; and let $\mathbf{w}_j^a \equiv (w_{j1}^a, w_{j2}^a, \dots, w_{j,2M_a})$ denote the j^{th} ART_a weight vector. For ART_b, let $\mathbf{x}^b \equiv (x_1^b \dots x_{2M_b}^b)$ denote the F_1^b output vector; let $\mathbf{y}^b \equiv (y_1^b \dots y_{N_b}^b)$ denote the F_2^b output vector; and let $\mathbf{w}_k^b \equiv (w_{k1}^b, w_{k2}^b, \dots, w_{k,2M_b}^b)$ denote the k^{th} ART_b weight vector. For the map field, let $\mathbf{x}^{ab} \equiv (x_1^{ab}, \dots, x_{N_b}^{ab})$ denote the F^{ab} output vector, and let $\mathbf{w}_j^{ab} \equiv (w_{j1}^{ab}, \dots, w_{jN_b}^{ab})$ denote the weight vector from the j^{th} F_2^a node to F^{ab}. Vectors $\mathbf{x}^a, \mathbf{y}^a, \mathbf{x}^b, \mathbf{y}^b$, and $\mathbf{x}^{ab}$ are set to $\mathbf{0}$ between input presentations.

Map field activation: The map field F^{ab} is activated whenever one of the ART_a or ART_b categories is active. If node J of F_2^a is chosen, then its weights $\mathbf{w}_J^{ab}$ activate F^{ab}. If node K in F_2^b is active, then the node K in F^{ab} is activated by 1–to–1 pathways between F_2^b and F^{ab}. If both ART_a and ART_b are active, then F^{ab} becomes active only if ART_a predicts the same category as ART_b via the weights $\mathbf{w}_j^{ab}$. The F^{ab} output vector $\mathbf{x}^{ab}$ obeys

$$\mathbf{x}^{ab} = \begin{cases} \mathbf{y}^b \wedge \mathbf{w}_J^{ab} & \text{if the Jth } F_2^a \text{ node is active and } F_2^b \text{ is active} \\ \mathbf{w}_J^{ab} & \text{if the Jth } F_2^a \text{ node is active and } F_2^b \text{ is inactive} \\ \mathbf{y}^b & \text{if } F_2^a \text{ is inactive and } F_2^b \text{ is active} \\ \mathbf{0} & \text{if } F_2^a \text{ is inactive and } F_2^b \text{ is inactive.} \end{cases} \tag{A16}$$

By (A16), $\mathbf{x}^{ab} = \mathbf{0}$ if the prediction $\mathbf{w}_J^{ab}$ is disconfirmed by $\mathbf{y}^b$. Such a mismatch event triggers an ART_a search for a better category, as follows.

Match tracking: At the start of each input presentation the ART_a vigilance parameter ρ_a equals a baseline vigilance $\overline{\rho_a}$. The map field vigilance parameter is ρ_{ab}. If

$$|\mathbf{x}^{ab}| < \rho_{ab}|\mathbf{y}^b|, \tag{A17}$$

then ρ_a is increased until it is slightly larger than $|\mathbf{A} \wedge \mathbf{w}_J^a||\mathbf{A}|^{-1}$, where $\mathbf{A}$ is the input to F_1^a, in complement coding form. Then

$$|\mathbf{x}^a| = |\mathbf{A} \wedge \mathbf{w}_J^a| < \rho_a|\mathbf{A}|, \tag{A18}$$

where J is the index of the active F_2^a node, as in (A10). When this occurs, ART$_a$ search leads either to activation of another F_2^a node J with

$$|\mathbf{x}^a| = |\mathbf{A} \wedge \mathbf{w}_J^a| \geq \rho_a|\mathbf{A}| \tag{A19}$$

and

$$|\mathbf{x}^{ab}| = |\mathbf{y}^b \wedge \mathbf{w}_J^{ab}| \geq \rho_{ab}|\mathbf{y}^b|; \tag{A20}$$

or, if no such node exists, to the shut-down of F_2^a for the remainder of the input presentation.

Map field learning: Learning rules determine how the map field weights w_{jk}^{ab} change through time, as follows. Weights w_{jk}^{ab} in $F_2^a \rightarrow F^{ab}$ paths initially satisfy

$$w_{jk}^{ab}(0) = 1. \tag{A21}$$

During resonance with the ART$_a$ category J active, $\mathbf{w}_J^{ab}$ approaches the map field vector $\mathbf{x}^{ab}$. With fast learning, once J learns to predict the ART$_b$ category K, that association is permanent; i.e., $w_{JK}^{ab} = 1$ for all time.

REFERENCES

Amari, S.-I. and Takeuchi, A. (1978). Mathematical theory on formation of category detecting nerve cells. *Biological Cybernetics*, **29**, 127–136.

Asfour, Y.R., Carpenter, G.A., Grossberg, S. and Lesher, G.W. (1993). Fusion ARTMAP: A neural network architecture for multi-channel data fusion and classification. In **Proceedings of the world congress on neural networks (WCNN-93)**, Hillsdale, NJ: Lawrence Erlbaum Associates, **II**-210–215. Technical Report CAS/CNS-TR-93-006, Boston, MA: Boston University.

Bachelder, I.A., Waxman, A.M., and Seibert, M. (1993). A neural system for mobile robot visual place learning and recognition. In **Proceedings of the world congress on neural networks (WCNN-93)**, Hillsdale, NJ: Lawrence Erlbaum Associates, **I**-512–517.

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.

Bienenstock, E.L., Cooper, L.N., and Munro, P.W. (1982). Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*, **2**, 32–48.

Bradski, G. and Grossberg, S. (1994). Recognition of 3-D objects from multiple 2-D views by a self-organizing neural architecture. In V. Cherkassky, J.H. Friedman, and H. Wechsler (Eds.), **From statistics to neural networks: Theory and pattern recognition**. New York: Springer-Verlag.

Carpenter, G.A. and Grossberg, S. (1987a). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, **37**, 54–115.

Carpenter, G.A. and Grossberg, S. (1987b). ART 2: Stable self-organization of pattern recognition codes for analog input patterns. *Applied Optics*, **26**, 4919–4930.

Carpenter, G.A. and Grossberg, S. (1987c). Neural dynamics of category learning and recognition: Attention, memory consolidation, and amnesia. In S. Grossberg (Ed.), **The adaptive brain, I: Cognition, learning, reinforcement, and rhythm**. Amsterdam: Elsevier/North Holland, pp. 238–286.

Carpenter, G.A. and Grossberg, S. (1990). ART 3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural Networks*, **3**, 129–152.

Carpenter, G.A. and Grossberg, S. (Eds.) (1991). **Pattern recognition by self-organizing neural networks**. Cambridge, MA: MIT Press.

Carpenter, G.A. and Grossberg, S. (1992). Fuzzy ARTMAP: Supervised learning, recognition, and prediction by a self-organizing neural network. *IEEE Communications Magazine*, **30**, 38–49.

Carpenter, G.A. and Grossberg, S. (1993). Normal and amnesic learning, recognition, and memory by a neural model or cortico-hippocampal interactions. *Trends in Neurosciences*, **16**, 131–137.

Carpenter, G.A. and Grossberg, S. (1994). Fuzzy ARTMAP: A synthesis of neural networks and fuzzy logic for supervised categorization and nonstationary prediction. In R.R. Yager and L.A. Zadeh (Eds.), **Fuzzy sets, neural networks, and soft computing**. New York, NY: Van Nostrand Reinhold, 126–165.

Carpenter, G.A., Grossberg, S., and Iizuka, K. (1992). Comparative performance measures of Fuzzy ARTMAP, learned vector quantization, and back propagation for handwritten character recognition. In **Proceedings of the international joint conference on neural networks (WCNN-93)**, Piscataway, NJ: IEEE Service Center, **I**-794–799.

Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H., and Rosen, D.B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks*, **3**, 698–713.

Carpenter, G.A., Grossberg, S. and Mehanian, C. (1989). Invariant recognition of cluttered

scenes by a self-organizing ART architecture: CORT-X boundary segmentation. *Neural Networks*, **2**, 169–181.

Carpenter, G.A., Grossberg, S., and Reynolds, J.H. (1991). ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*, **4**, 565–588.

Carpenter, G.A., Grossberg, S., and Reynolds, J.H. (1993). Fuzzy ARTMAP, slow learning, and probability estimation. In **Proceedings of the world congress on neural networks (WCNN-93)**, Hillsdale, NJ: Erlbaum Associates, **II**-26–30.

Carpenter, G.A., Grossberg, S., and Reynolds, J.H. (1994). A fuzzy ARTMAP nonparametric probability estimator for nonstationary pattern recognition problems. *IEEE Transactions on Neural Networks*, in press. Technical Report CAS/CNS-TR-93-047. Boston, MA: Boston University.

Carpenter, G.A., Grossberg, S., and Rosen, D.B. (1991a). ART2-A: An adaptive resonance algorithm for rapid category learning and recognition. *Neural Networks*, **4**, 493–504.

Carpenter, G.A., Grossberg, S., and Rosen, D.B. (1991b). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*, **4**, 759–771.

Carpenter, G.A. and Ross, W.D. (1993). ART-EMAP: A neural network architecture for learning and prediction by evidence accumulation. **Proceedings of the world congress on neural networks (WCNN-93)**, Hillsdale, NJ: Lawrence Erlbaum Publishing, **III**-649–656.

Carpenter, G.A. and Ross, W.D. (1994a). 3-D object recognition by the ART-EMAP evidence accumulation network. In **Proceedings of the world congress on neural networks (WCNN-94)**, Hillsdale, NJ: Lawrence Erlbaum Publishing, **I**-749–758.

Carpenter, G.A. and Ross, W.D. (1994b). ART-EMAP: A neural network architecture for object recognition by evidence accumulation. *IEEE Transactions on Neural Networks*, in

press. Technical Report CAS/CNS-TR-93-035, Boston, MA: Boston University.

Carpenter, G.A. and Tan, A.-H. (1994). Rule extraction: From neural architecture to symbolic representation. *Connection Science*. in press. Technical Report CAS/CNS-TR-94-005, Boston, MA: Boston University.

Caudell, T., Smith, S., Johnson, C., Wunsch, D., and Escobedo, R. (1991). An industrial application of neural networks to reusable design. **Adaptive neural systems**, Technical Report BCS-CS-ACS-91-001, Seattle, WA: The Boeing Company, pp. 185–190.

Cohen, M. and Grossberg, S. (1987). Masking fields: A massively parallel architecture for learning, recognizing, and predicting multiple groupings of patterned data. *Applied Optics*, **26**, 1866–1891.

Daugman, J.G. (1988). Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **36**, 1169–1179.

Desimone, R. (1992). Neural circuits for visual attention in the primate brain. In G.A. Carpenter and S. Grossberg (Eds.), **Neural networks for vision and image processing**. Cambridge, MA: MIT Press, pp. 343–364.

Dubrawski, A. and Crowley, J.L. (1994a). Learning locomotion reflexes: A self-supervised neural system for a mobile robot. *Robotics and Autonomous Systems*, **12**, 133–142.

Dubrawski, A. and Crowley, J.L. (1994b). Self-supervised neural system for reactive navigation. In **Proceedings of the IEEE international conference on robotics and automation**, Los Alamitos, CA: IEEE Computer Society Press, 2076–2081.

Escobedo, R., Smith, S.D.G., and Caudell, T.P. (1993). The ART of design retrieval. **Adaptive neural systems**, Technical Report BCS-CS/ACS-93-008, Seattle, WA: The Boeing Company, pp. 149–160.

Feng, C., Sutherland, A., King, S., Muggleton, S., and Henery, R. (1993). Comparison of machine learning classifiers to statistics and neural networks. **Proceedings of the fourth**

international workshop on artificial intelligence and statistics, 363–368.

Gabor, D. (1946). A theory of communication. *Journal of the Institute of Electrical Engineers*, **93**, 429–457.

Gan, K.W. and Lua, K.T. (1992). Chinese character classification using an adaptive resonance network. *Pattern Recognition*, **25**, 877–882.

Gopal, S., Sklarew, D.M., and Lambin, E. (1993). Fuzzy neural networks in multi-temporal classification of landcover change in the Sahel. **Proceedings of the DOSES workshop on new tools for spatial analysis**, Lisbon, Portugal.

Grossberg, S. (1964). The theory of embedding fields with applications to psychology and neurophysiology. Ph.D. Thesis. New York: Rockefeller Institute for Medical Research.

Grossberg, S. (1969). On learning and energy-entropy dependence in recurrent and nonrecurrent signed networks. *Journal of Statistical Physics*, **1**, 319–350.

Grossberg, S. (1972). Neural expectation: Cerebellar and retinal analogs of cells fired by learnable or unlearned pattern classes. *Kybernetik*, **10**, 49–57.

Grossberg, S. (1973). Contour enhancement, short term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, **LII**, 213–257.

Grossberg, S. (1976a). Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors. *Biological Cybernetics*, **23**, 121–134.

Grossberg, S. (1976b). Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics*, **23**, 187–202.

Grossberg, S. (1978). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plan. In R. Rosen and F. Snell (Eds.), **Progress in theoretical biology**, Volume 5. New York: Academic Press, 233–374. [Reprinted in S. Grossberg, **Studies of mind and brain: Neural principles of learning, perception, development, cognition, and motor control**. Boston: Reidel Press, 1982.]

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, **1**, 1–51.

Grossberg, S. (1982). **Studies of mind and brain: Neural principles of learning, perception, development, cognition, and motor control**. Boston: Reidel Press.

Grossberg, S. (Ed.) (1987a). **The adaptive brain, I: Cognition, learning, reinforcement, and rhythm**. Amsterdam: Elsevier/North-Holland.

Grossberg, S. (Ed.) (1987b). **The adaptive brain, II: Vision, speech, language, and motor control**. Amsterdam: Elsevier/North-Holland.

Grossberg, S. (Ed.) (1988). **Neural networks and natural intelligence**. Cambridge, MA: MIT Press.

Grossberg, S. (1994). 3-D vision and figure-ground separation. *Perception and Psychophysics*, **55**, 48–120.

Grossberg, S. and Kuperstein, M. (1986/1989). **Neural dynamics of adaptive sensory-motor control: Expanded edition**. Elmsford, NY: Pergamon Press.

Grossberg, S. and Merrill, J.W.L. (1992). A neural network model of adaptively timed reinforcement learning and hippocampal dynamics. *Cognitive Brain Research*, **1**, 3–38.

Grossberg, S. and Mingolla, E. (1985). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review*, **92**, 173–211.

Grossberg, S., Mingolla, E., and Ross, W.D. (1994). A neural theory of attentive visual search: Interactions of visual, spatial, and object representations. *Psychological Review*, **101**, 470–789.

Grossberg, S., Mingolla, E., and Todorović, D. (1989). A neural network architecture for preattentive vision. *IEEE Transactions on Biomedical Engineering*, **36**, 65–84.

Grossberg, S. and Stone, G.O. (1986). Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review*, **93**, 46–74.

Grossberg, S. and Wyse, L. (1991). Invariant recognition of cluttered scenes by a self-organizing ART architecture: Figure-ground separation. *Neural Networks*, **4**, 723-742.

Grossberg, S. and Wyse, L. (1992). Figure-ground separation of connected scenic figures: Boundaries, filling-in, and opponent processing. In G.A. Carpenter and S. Grossberg (Eds.), **Neural networks for vision and image processing**. Cambridge, MA: MIT Press, 161–194.

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 congess on neural networks (WCNN-93)**, Hillsdale, NJ: Lawrence Erlbaum Associates, **I**-207–211.

Harvey, R.M. (1993). Nursing diagnosis by computers: An application of neural networks. *Nursing Diagnosis*, **4**, 26–34.

Johnson, C. (1993). Agent learns user's behavior. *Electrical Engineering Times*, June 28, pp. 43, 46.

Kasperkiewicz, J., Racz, J., and Dubrawski, A. (1994). HPC strength prediction using an artificial neural network. *ASCE Journal of Computing in Civil Engineering*, submitted.

Keyvan, S., Durg, A., and Rabelo, L.C. (1993). Application of artificial neural networks for development of diagnostic monitoring system in nuclear plants. *American Nuclear Society Conference Proceedings*, April 18–21.

Koenderink, J.J. and van Doorn, A.J. (1979). The internal representation of solid shape with respect to vision. *Biological Cybernetics*, **32**, 211–216.

Kohonen, T. (1984/1989). **Self-organization and associative memory** (Third edition). New York: Springer-Verlag.

Kosko, B. (1986). Fuzzy entropy and conditioning. *Information Sciences*, **40**, 165–174.

Kumara, S.R.T., Merchawi, N.S., Karmarthi, S.V., and Thazhutaveetil, M. (1994). **Neural networks in design and manufacturing**. Chapman and Hall Publishers.

Laird, J.E., Newell, A., and Rosenbloom, P.S. (1987). SOAR: An architecture for general intelligence. *Artificial Intelligence*, **33**, 1–64.

Levy, W.B. (1985). Associative changes at the synapse: LTP in the hippocampus. In W.B. Levy, J. Anderson and S. Lehmkuhle (Eds.), **Synaptic modification, neuron selectivity, and nervous system organization**. Hillsdale, NJ: Erlbaum Associates, pp. 5–33.

Levy, W.B. and Desmond, N.L. (1985). The rules of elemental synaptic plasticity. In W.B. Levy, J. Anderson, and S. Lehmkuhle (Eds.), **Synaptic modification, neuron selectivity, and nervous system organization**. Hillsdale, NJ: Erlbaum Associates, pp. 105–121.

Linsker, R. (1986). From basic network principles to neural architecture: Emergence of spatial-opponent cells. *Proceedings of the National Academy of Sciences*, **83**, 8779–8783.

Malsburg, C. von der (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, **14**, 85–100.

Mehta, B.V., Vij, L., and Rabelo, L.C. (1993). Prediction of secondary structures of proteins using fuzzy ARTMAP. In **Proceedings of the world congress on neural networks (WCNN-93)**, Hillsdale, NJ: Lawrence Erlbaum Associates, **I**-228–232.

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**, San Mateo, CA: Morgan Kaufmann, pp. 174–185.

Moya, M.M., Koch, M.W., and Hostetler, L.D. (1993). One-class classifier networks for target recognition applications. In **Proceedings of the world congress on neural networks (WCNN-93)**, Hillsdale, NJ: Lawrence Erlbaum Associates, **III**-797–801.

Parker, D.B. (1982). Learning-logic. Invention Report 581-64, File 1, Office of Technology Licensing, Stanford University, October.

Rauschecker, J.P. and Singer, W. (1979). Changes in the circuitry of the kitten's visual cortex are gated by postsynaptic activity. *Nature*, **280**, 58–60.

Repp, B.H. (1991). Perceptual restoration of a "missing" speech sound: Auditory induction or illusion? **Haskins Laboratories Status Report on Speech Research**, SR-107/108, 147–170.

Rumelhart, D.E., Hinton, G., and Williams, R. (1986). Learning internal representations by error propagation. In D.E. Rumelhart and J.L. McClelland (Eds.), **Parallel distributed processing**. Cambridge, MA: MIT Press, 318–362.

Rumelhart, D.E. and Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive Science*, **9**, 75–112.

Samuel, A.G. (1981a). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, **110**, 474–494.

Samuel, A.G. (1981b). The rule of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance*, **7**, 1124–1131.

Schvaneveldt, R.W. and MacDonald, J.E. (1981). Semantic context and the encoding of words: Evidence for two modes of stimulus analysis. *Journal of Experimental Psychology: Human Perception and Performance*, **7**, 673–687.

Schwartz, E. (1977). Spatial mapping in primate sensory projection: Analytic structure and relevance to perception. *Biological Cybernetics*, **25**, 181–194.

Seibert, M. and Waxman, A. (1990a). Learning aspect graph representations of 3-D objects in a neural network. In **Proceedings of the international joint conference on neural networks (IJCNN-90)**, Washington, D.C., **2**-233–236.

Seibert, M. and Waxman, A. (1990b). Learning aspect graph representations from view sequences. In Touretzky, D. (Ed.), **Advances in neural information processing systems, 2**, San Mateo, CA: Morgan Kaufmann Publishing, pp. 258–265.

Seibert, M. and Waxman, A.M. (1991). Learning and recognizing 3D objects from multiple views in a neural system. In H. Wechsler (Ed.), **Neural networks for perception, Volume 1**. New York: Academic Press.

Seibert, M. and Waxman, A. (1992). Adaptive 3-D object recognition from multiple views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 107–124.

Singer, W. (1983). Neuronal activity as a shaping factor in the self-organization of neuron

assemblies. In E. Basar, H. Flohr, H. Haken, and A.J. Mandell (Eds.), **Synergetics of the brain.** New York: Springer-Verlag, pp. 89–101.

Smith, E.E. (1990). In D.O. Osherson and E.E. Smith (Eds.), **An invitation to cognitive science.** Cambridge, MA: MIT Press.

Smith, J.W., Everhart, J.E., Dickson, W.C., Knowler, W.C., and Johannes, R.S. (1988). Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In **Proceedings of the symposium on computer applications and medical care.** Piscataway, NJ: IEEE Computer Society Press, pp. 261–265.

Suzuki, Y., Abe, Y., and Ono, K. (1993) Self-organizing QRS wave recognition system in ECG using ART 2. In **Proceedings of the world congress on neural networks (WCNN-93)**, Hillsdale, NJ: Lawrence Erlbaum Associates, **IV**-39–42.

Warren, R.M. (1984). Perceptual restoration of obliterated sounds. *Psychological Bulletin*, **96**, 371–383.

Warren, R.M. and Sherman, G.L. (1974). Phonemic restorations based on subsequent context. *Perception and Psychophysics*, **16**, 150–156.

Werbos, P. (1974). Beyond regression: New tools for prediction and analysis in the behavioral sciences. PhD Thesis, Harvard University, Cambridge, Massachusetts.

Wienke, D. (1993). ART pattern recognition software for chemists. User Documentation, University of Nijmegen.

Wienke, D. (1994). Neural resonance and adaptation—Towards nature's principles in artificial pattern recognition. In L. Buydens and W. Melssen (Eds.), **Chemometrics: Exploring and exploiting chemical information.** Nijmegen, The Netherlands: University Press.

Wienke, D. and Kateman, G. (1994). Adaptive resonance theory based artificial neural networks for treatment of open-category problems in chemical pattern recognition—Application to UV-Vis and IR spectroscopy. *Chemometrics and Intelligent Laboratory Systems.*

Wienke, D., Xie, Y., and Hopke, P.K. (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.*

Wilensky, G. (1990). Analysis of neural network issues: Scaling, enhanced nodal processing, comparison with standard classification. DARPA Neural Network Program Review, October 29-30, 1990.

Willshaw, D.J. and von der Malsburg, C. (1976). How patterned neural connections can be set up by self-organization. *Proceedings of the Royal Society of London (B)*, **194**, 431–445.

Zadeh, L. (1965). Fuzzy sets. *Information Control*, **8**, 338–353.

FIGURE CAPTIONS

Figure 1. Representation of fuzzy categories using feature rectangles and, in higher dimensions, hyper-rectangles. [Reprinted with permission from Carpenter, G.A. and Grossberg, S. (1994). Fuzzy ARTMAP: A synthesis of neural networks and fuzzy logic for supervised categorization and nonstationary prediction. In R.R. Yager and L.A. Zadeh (Eds.), **Fuzzy sets, neural networks, and soft computing**. New York, NY: Van Nostrand Reinhold, p. 128.]

Figure 2. Many-to-one learning combines categorization of many exemplars into one category, and labelling of many categories with the same name. [Reprinted with permission from Carpenter, G.A. and Grossberg, S. (1994). Fuzzy ARTMAP: A synthesis of neural networks and fuzzy logic for supervised categorization and nonstationary prediction. In R.R. Yager and L.A. Zadeh (Eds.), **Fuzzy sets, neural networks, and soft computing**. New York, NY: Van Nostrand Reinhold, p. 131.]

Figure 3. One-to-many learning enables one input vector to be associated with many output vectors. If the system predicts an output that is disconfirmed at a given stage of learning, the predictive error drives a memory search for a new category to associate with the new prediction, without degrading its previous knowledge about the input vector. [Reprinted with permission from Carpenter, G.A. and Grossberg, S. (1994). Fuzzy ARTMAP: A synthesis of neural networks and fuzzy logic for supervised categorization and nonstationary prediction. In R.R. Yager and L.A. Zadeh (Eds.), **Fuzzy sets, neural networks, and soft computing**. New York, NY: Van Nostrand Reinhold, p. 132.]

Figure 4. Interactions between the attentional and orienting subsystems of an adaptive resonance theory (ART) circuit: Level F_1 encodes a distributed representation of an event to be recognized via a short-term memory (STM) activation pattern across a network of

feature detectors. Level F_2 encodes the event to be recognized using a more compressed STM representation of the F_1 pattern. Learning of these recognition codes takes place at the long-term memory (LTM) traces within the bottom-up and top-down pathways between levels F_1 and F_2. The top-down pathways can read-out learned expectations whose prototypes are matched against bottom-up input patterns at F_1. Mismatches in response to novel events activate the orientation subsystem $\mathcal{A}$, thereby resetting the recognition codes that are active in STM at F_2 and initiating a memory search for a more appropriate recognition code. Output from subsystem $\mathcal{A}$ can also trigger an orienting response. (a) Block diagram of circuit. (b) Individual pathways of circuit, including the input level F_0 that generates inputs to level F_1. The gain control input to level F_1 helps to instantiate the 2/3 Rule (see text). Gain control to level F_2 is needed to instate a category in STM. [Reprinted with permission from Carpenter, G.A. and Grossberg, S. (1987a). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, **37**, p. 56.]

Figure 5. ART search for an F_2 recognition code: (a) The input pattern **I** generates the specific STM activity pattern **X** at F_1 as it nonspecifically activates the orienting subsystem $\mathcal{A}$. **X** is represented by the hatched pattern across F_1. Pattern **X** both inhibits $\mathcal{A}$ and generates the output pattern **S**. Pattern **S** is transformed by the LTM traces into the input pattern **T**, which activates the STM pattern **Y** across F_2. (b) Pattern **Y** generates the top-down output pattern **U** which is transformed into the prototype pattern **V**. If **V** mismatches **I** at F_1, then a new STM activity pattern $\mathbf{X}^*$ is generated at F_1. $\mathbf{X}^*$ is represented by the hatched pattern. Inactive nodes corresponding to **X** are unhatched. The reduction in total STM activity which occurs when **X** is transformed into $\mathbf{X}^*$ causes a decrease in the total inhibition from F_1 to $\mathcal{A}$. (c) If the vigilance criterion fails to be met, $\mathcal{A}$ releases a nonspecific arousal wave to F_2, which resets the STM pattern **Y** at F_2. (d) 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 $\mathbf{X}$ to activate a different STM pattern $\mathbf{Y}^*$ at F_2. If the top-down prototype due to $\mathbf{Y}^*$ also mismatches $\mathbf{I}$ 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. [Reprinted with permission from Carpenter, G.A. and Grossberg, S. (1987a). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, **37**, p. 61.]

Figure 6. The computational rules for the self-organizing feature map model were established in 1976. [Reprinted with permission from Grossberg, S. (1976a). Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors. *Biological Cybernetics*, **23**, p. 122.]

Figure 7. Comparison of ART 1 and Fuzzy ART. [Reprinted with permission from Carpenter,G.A., Grossberg, S. and Rosen, D.B. (1991b). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*, **4**, p. 762.]

Figure 8. Complement coding uses on-cell and off-cell pairs to normalize input vectors. [Reprinted with permission from Carpenter, G.A., Grossberg, S. and Rosen, D.B. (1991b). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*, **4**, p. 764.]

Figure 9. Fuzzy AND and OR operations generate category hyper-rectangles. [Reprinted with permission from Carpenter, G.A., Grossberg, S. and Rosen, D.B. (1991b). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*, **4**, p. 763.]

Figure 10. Fuzzy ARTMAP architecture. The ART_a complement coding preprocessor transforms the M_a-vector $\mathbf{a}$ into the $2M_a$-vector $\mathbf{A} = (\mathbf{a}, \mathbf{a}^c)$ at the ART_a field F_0^a. $\mathbf{A}$ is the

input vector to the ART_a field F_1^a. Similarly, the input to F_1^b is the $2M_b$-vector $(\mathbf{b}, \mathbf{b}^c)$. When a prediction by ART_a is disconfirmed at ART_b, inhibition of map field activation induces the match tracking process. Match tracking raises the ART_a vigilance ρ_a to just above the F_1^a to F_0^a match ratio $|\mathbf{x}^a|/|\mathbf{A}|$. This triggers an ART_a search which leads to activation of either an ART_a category that correctly predicts $\mathbf{b}$ or to a previously uncommitted ART_a category node. [Reprinted with permission from Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H., and Rosen, D.B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks*, **3**, p. 699.]

Figure 11. Match tracking: (a) A prediction is made by ART_a when the baseline vigilance ρ_a is less than the analog match value. (b) A predictive error at ART_b increases the baseline vigilance value of ART_a until it just exceeds the analog match value, and thereby triggers hypothesis testing that searches for a more predictive bundle of features to which to attend. [Reprinted with permission from Carpenter, G.A. and Grossberg, S. (1994). Fuzzy ARTMAP: A synthesis of neural networks and fuzzy logic for supervised categorization and nonstationary prediction. In R.R. Yager and L.A. Zadeh (Eds.), **Fuzzy sets, neural networks, and soft computing**. New York, NY: Van Nostrand Reinhold, p. 148.]

Figure 12. The image processing flow chart of the VIEWNET system, from presenting a 2-D image in the image database until the read-out of the predicted 3-D object.

Figure 13. CORT-X 2 flow chart.

Figure 14. Preprocessing summary. (a) Output of CORT-X 2 preprocessing. (b) Centered log-polar image. (c) Gaussian coarse coding pattern. (d-f) Coarse coding reduction from 128×128 pixels down to 16×16, 8×8, and 4×4 pixels.

Figure 15. ART-EMAP synthesizes adaptive resonance theory (ART) and spatial and temporal evidence integration for dynamic predictive mapping (EMAP). The network extends

the capabilities of fuzzy ARTMAP in four incremental stages. Stage 1 introduces distributed pattern representation at a view category field. Stage 2 adds a decision criterion to the mapping between view and object categories, delaying identification of ambiguous objects when faced with a low confidence prediction. Stage 3 augments the system with a field where evidence accumulates in medium-term memory (MTM). Stage 4 adds an unsupervised learning process to fine-tune performance after the limited initial period of supervised network training. Simulations of the four ART-EMAP stages demonstrate performance on a difficult 3-D object recognition problem. [Reprinted with permission from Carpenter, G.A. and Ross, W.D. (1993a). ART-EMAP: A neural network architecture for learning and prediction by evidence accumulation. In **Proceedings of the world congress on neural networks (WCNN-93)**, Hillsdale, NJ: Lawrence Erlbaum Publishing, **III**, p. 650.]

Figure 16. 3-D object recognition training and test set images for ART-EMAP simulations. Circled exemplars indicate that single-view object identification may be difficult because a training set view of one object can be the same as a test set view of a different object; and because test set views of two different objects can be the same. [Reprinted with permission from Carpenter, G.A. and Ross, W.D. (1994). 3-D object recognition by the ART-EMAP evidence accumulation network. In **Proceedings of the world congress on neural networks (WCNN-94)**, Hillsdale, NJ: Lawrence Erlbaum Publishing, **I**, p. 751.]

Figure 17. 3-D object simulations. Response viewing hemispheres for each object show predictions from each test set view. A window in the hemisphere corresponds to one of the 17 test views (Figure 16b). Plots (a), (b), and (c) show noise-free test set results and plots (d), (e), and (f) show noisy test set results. Plots (a) and (d) show fuzzy ARTMAP performance, using the F_2^a choice rule. Plots (b) and (e) show Stage 1 ART-EMAP performance using the power rule (3) with $p = 24$. Plots (c) and (f) show Stage 3 ART-EMAP performance with $p = 24$ plus temporal evidence accumulation with the decreasing decision criterion (14) and multiple views. [Reprinted with permission from Carpenter, G.A. and Ross, W.D. (1994). 3-

D object recognition by the ART-EMAP evidence accumulation network. In **Proceedings of the world congress on neural networks (WCNN-94)**, Hillsdale, NJ: Lawrence Erlbaum Publishing, **I**, p. 753.]

TABLE CAPTIONS

Table 1. ARTMAP benchmark studies. [Reprinted with permission from Carpenter, G.A. and Grossberg, S. (1993). Normal and amnesic learning, recognition, and memory by a neural model or cortico-hippocampal interactions. *Trends in Neurosciences*, **16**, p. 134.]

Table 2. Fuzzy ARTMAP applied to the Landsat image database (Feng *et al.*, 1993). With the exception of K-N-N, Fuzzy ARTMAP test set performance exceeded that of other neural network and machine learning algorithms. Compared to K-N-N, Fuzzy ARTMAP showed a 6:1 code compression ratio.

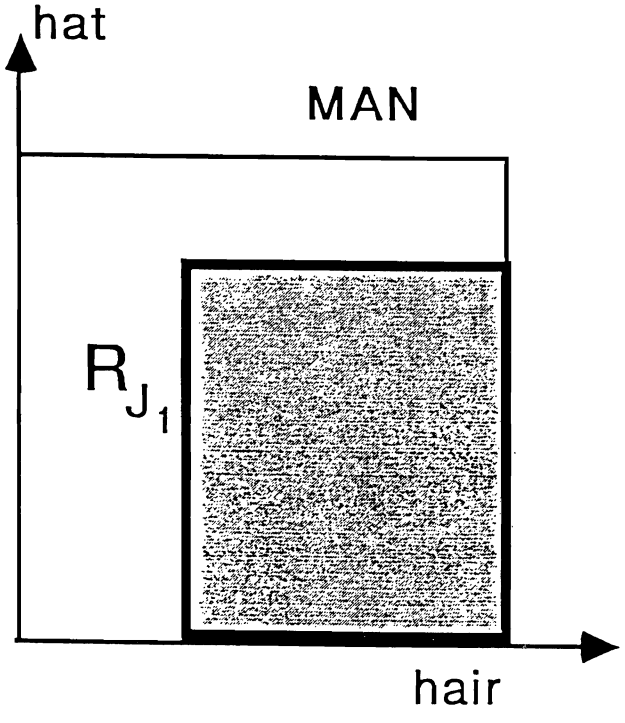
Table 3. On the Pima Indian Diabetes (PID) database, Fuzzy ARTMAP test set performance was similar to that of the ADAP algorithm (Smith *et al.*, 1988) but with far fewer rules and faster training. An ARTMAP pruning algorithm (Carpenter and Tan, 1994) further reduces the number of rules by an order of magnitude, and also boosts test set accuracy to 79%.

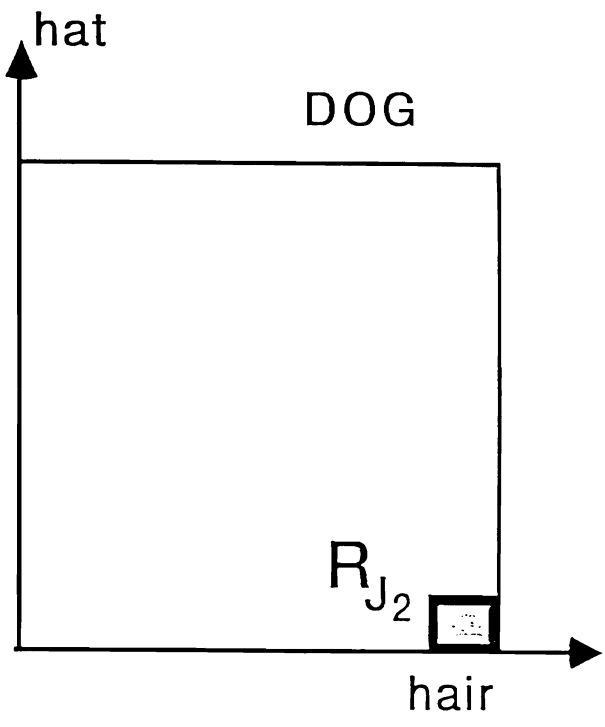
Table 4. VIEWNET recognition results on noisy data ($C = 1$) with slow learning to the map field ($\beta_{ab} = 0.2$, $\bar{\rho}_{max} = 0.95$). Due to the low levels of noise surviving CORT-X 2 preprocessing, the recognition results here are not substantially different than those found using fast learning in noise. As noise increases, slow learning becomes more important for maintaining good recognition scores.

EXAMPLE: Features = hair, hat

man's head (category J_1)
hair: usually
hat: sometimes

dog's head (category J_2)
hair: almost always
hat: almost never





COMPLEMENT CODING:
a - present features
a^c - absent features

WITHOUT complement coding
hat feature (usually absent)
→ 0 in both cases.

Figure 1

MANY-TO-ONE MAP

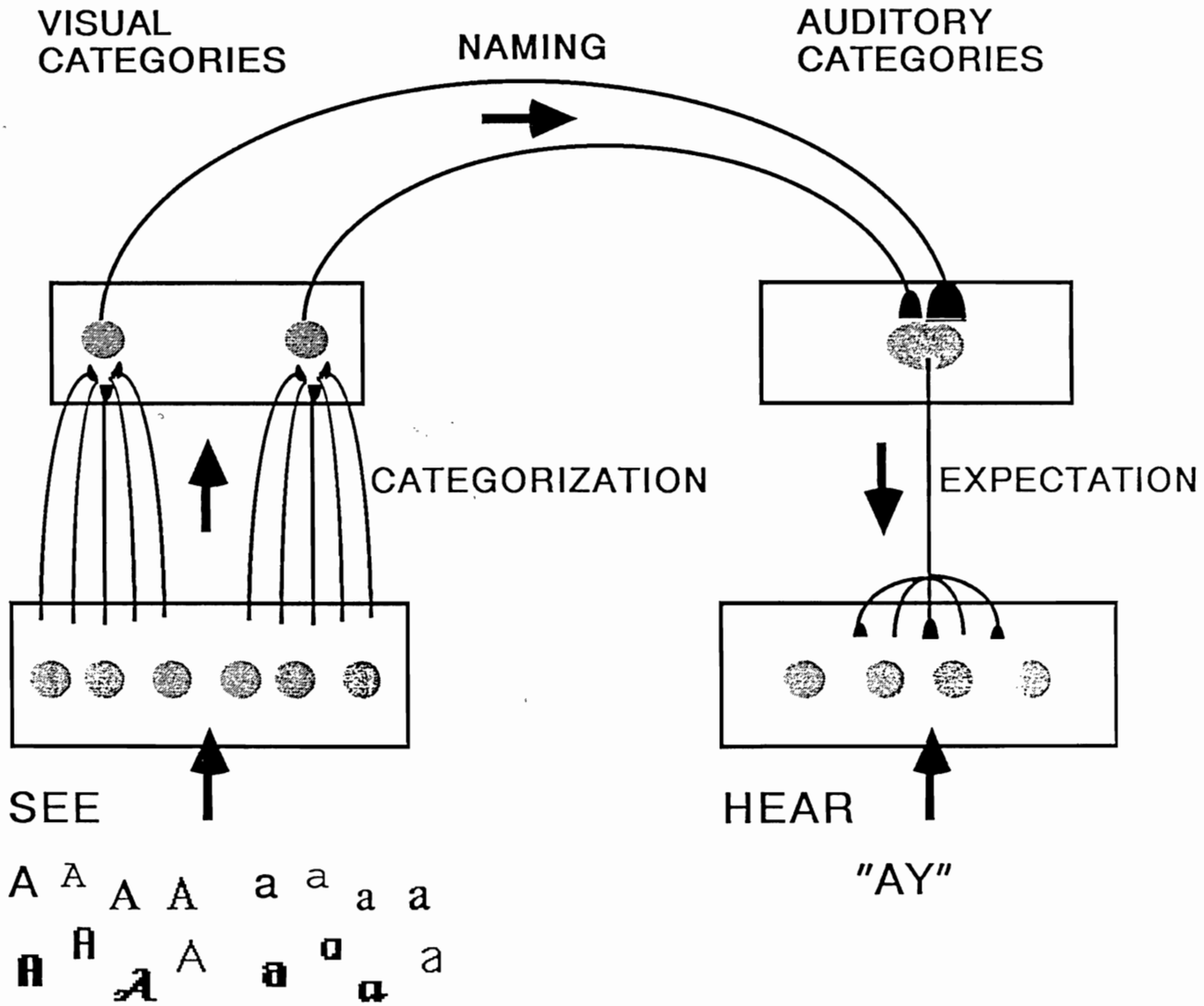


Figure 2

ONE-TO-MANY MAP

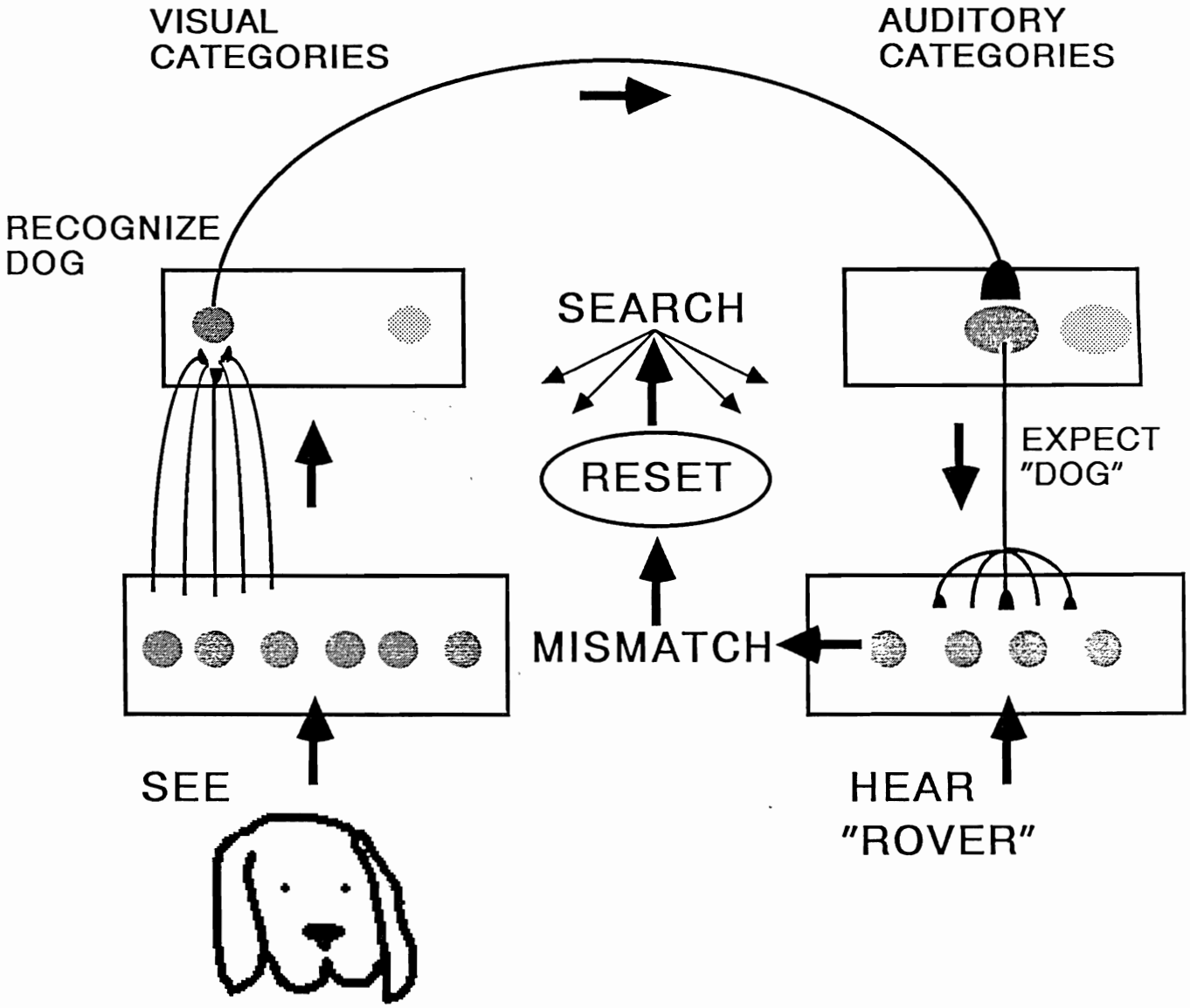


Figure 3

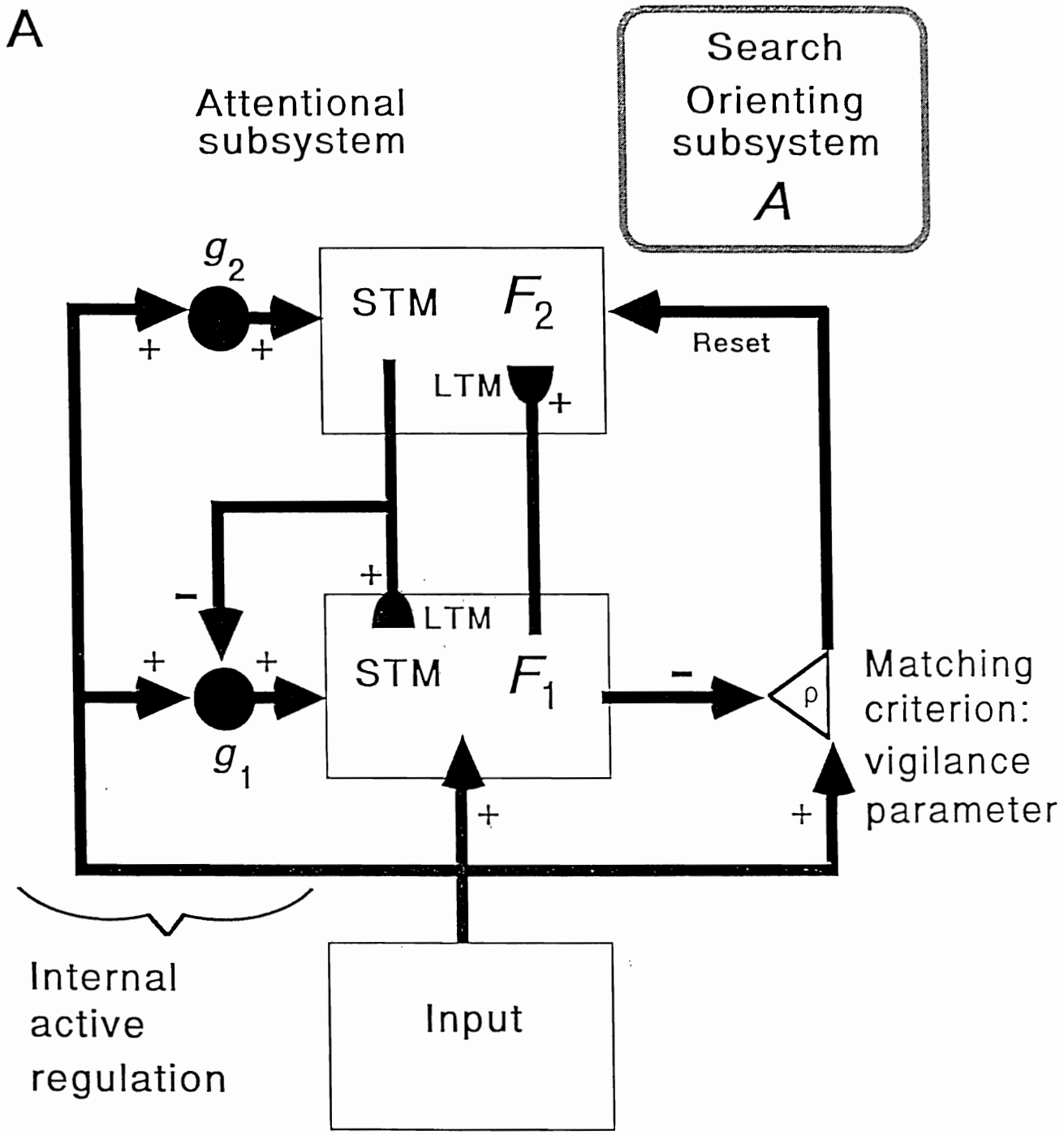

A
Attentional subsystem
Search Orienting subsystem
A
g_2
STM F_2
LTM
Reset
LTM
STM F_1
g_1
ρ
Matching criterion: vigilance parameter
Internal active regulation
Input

Figure 4a

B

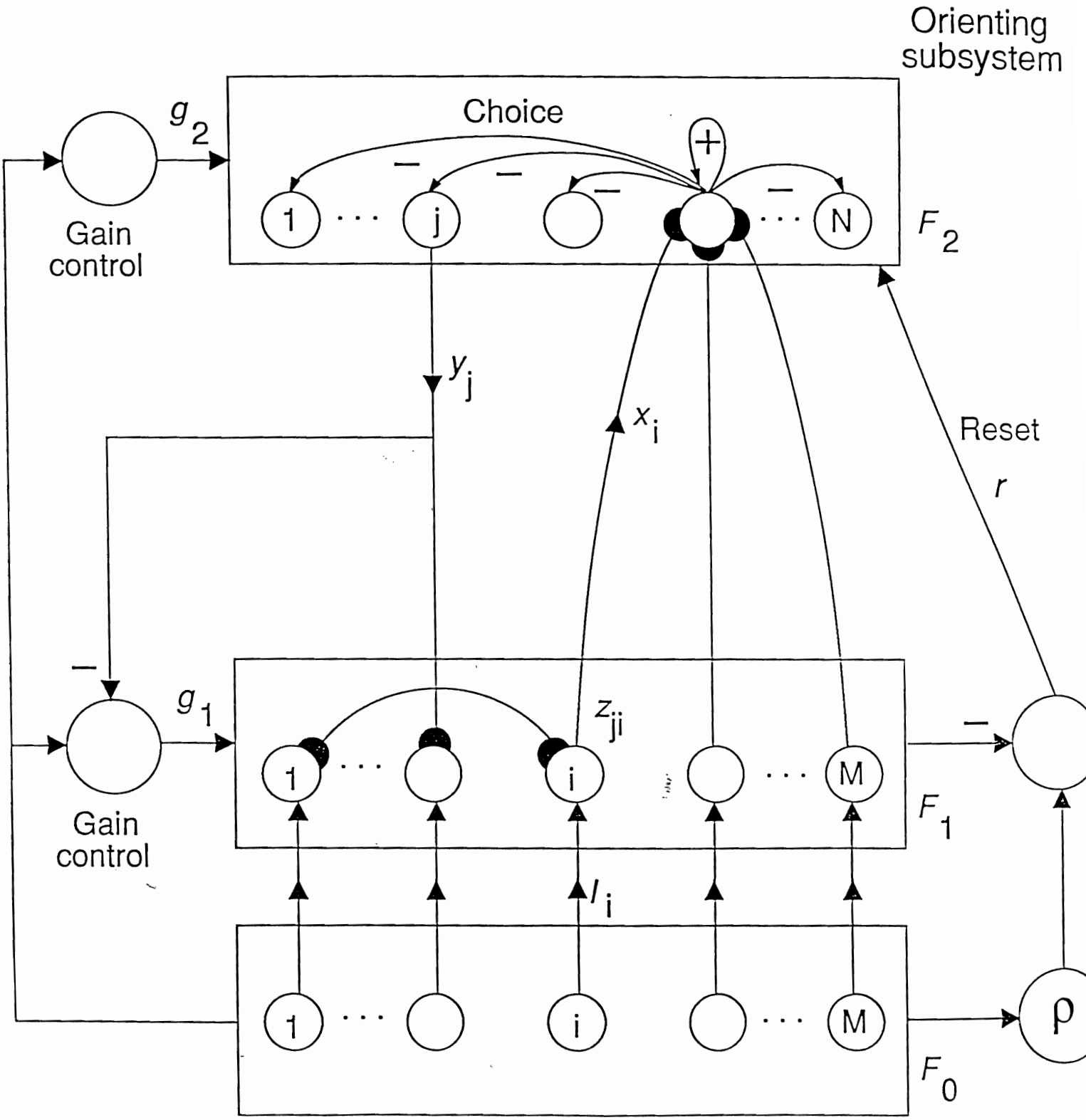

Orienting subsystem
Choice
g_2
Gain control
1
j
N
F_2
y_j
x_i
Reset
r
g_1
z_{ji}
i
M
F_1
Gain control
I_i
F_0
ρ

Figure 4b

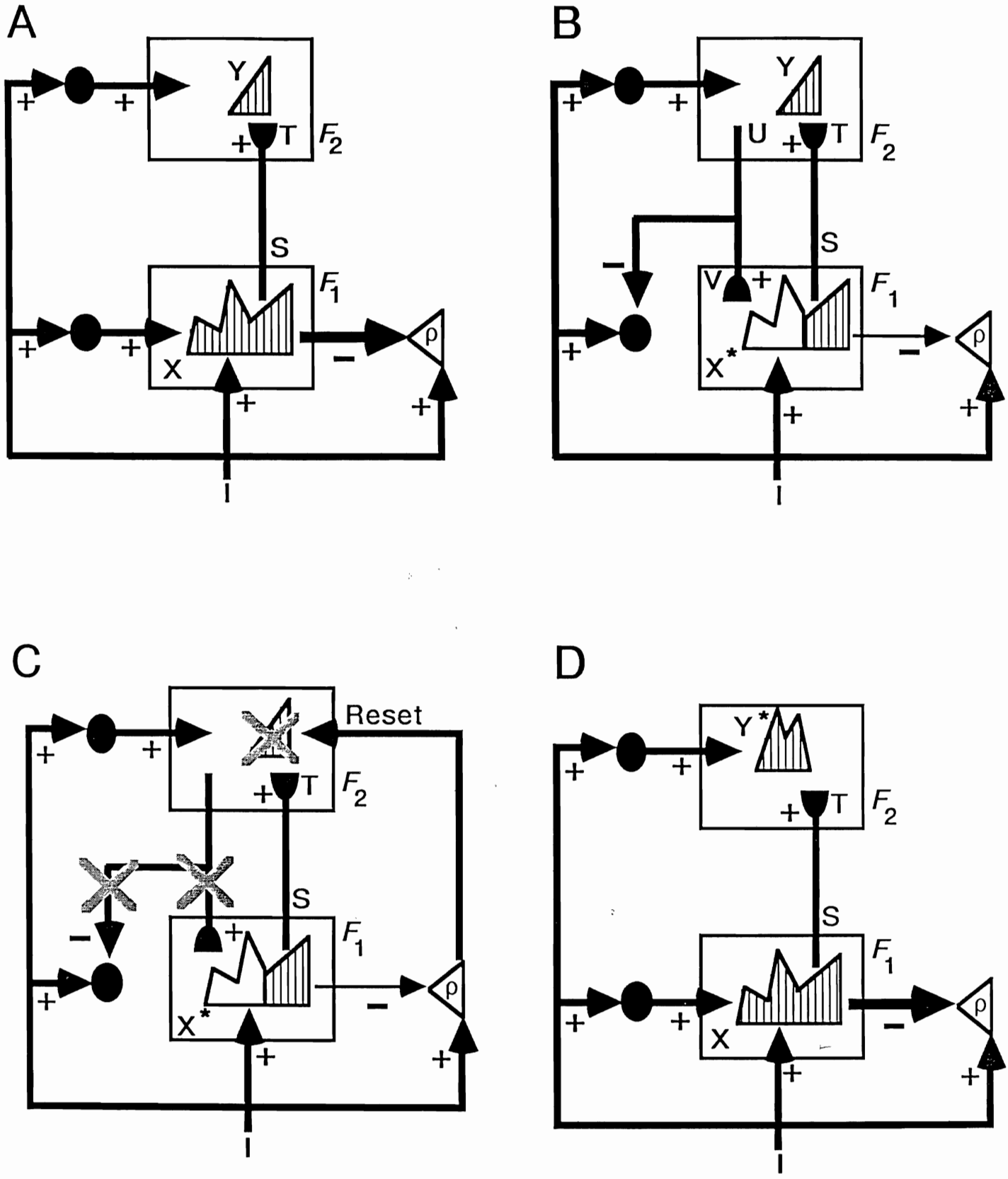
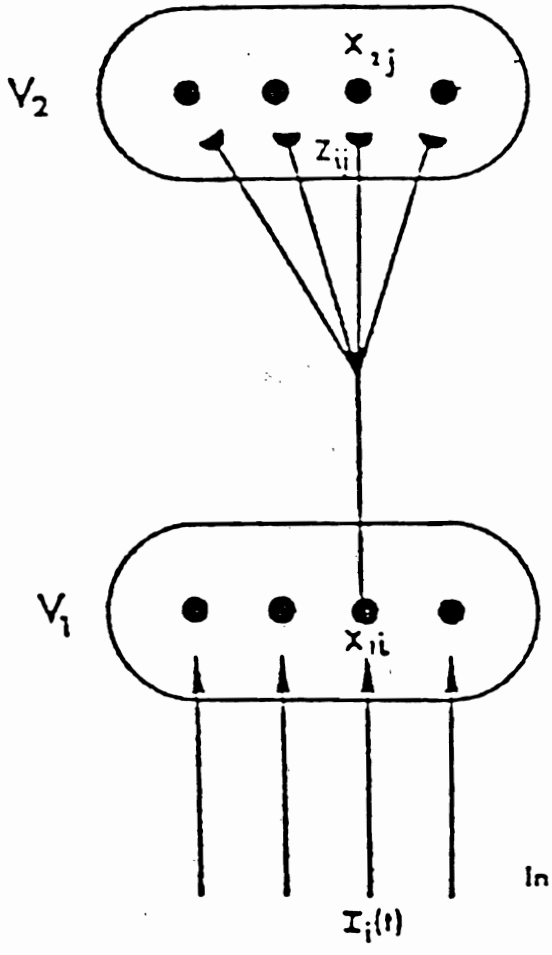

A
Y
T
F_2
S
F_1
X
ρ
I
B
Y
U
T
F_2
S
V
F_1
X^*
ρ
I
C
Reset
T
F_2
S
F_1
X^*
ρ
I
D
Y^*
T
F_2
S
F_1
X
ρ
I

Figure 5



1. Normalize Total Activity
2. Contrast Enhance
3. STM

LTM IN PLASTIC SYNAPTIC STRENGTHS

1. Compute Time-Average of Presynaptic Signal and Postsynaptic STM Trace Product
2. Multiplicatively Gate Signals

1. Normalize Total Activity

Input Pattern

Figure 6

ART 1 (BINARY)	FUZZY ART (ANALOG)

CATEGORY CHOICE

$$T_j = \frac{|I \cap w_j|}{\alpha + |w_j|} \qquad T_j = \frac{|I \wedge w_j|}{\alpha + |w_j|}$$

MATCH CRITERION

$$\frac{|I \cap w_J|}{|I|} \geq \rho \qquad \frac{|I \wedge w_J|}{|I|} \geq \rho$$

FAST LEARNING

$$w_J^{(new)} = I \cap w_J^{(old)} \qquad w_J^{(new)} = I \wedge w_J^{(old)}$$

$\cap$ = logical AND intersection

$\wedge$ = fuzzy AND minimum

Figure 7

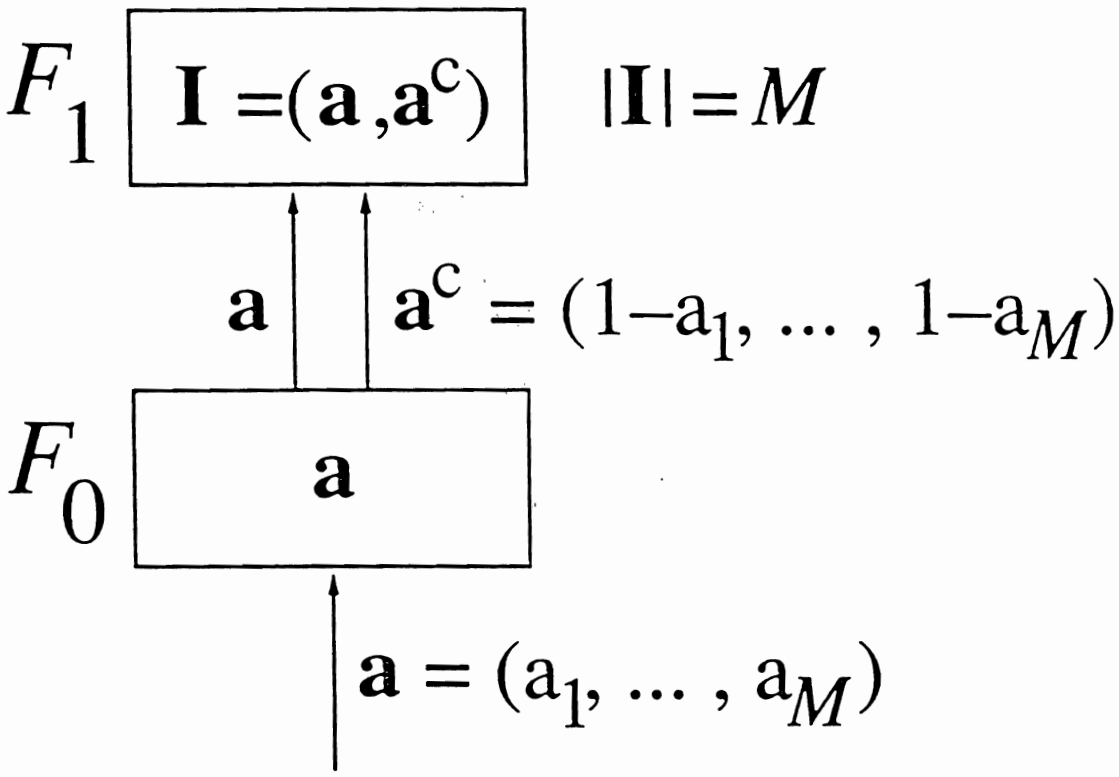
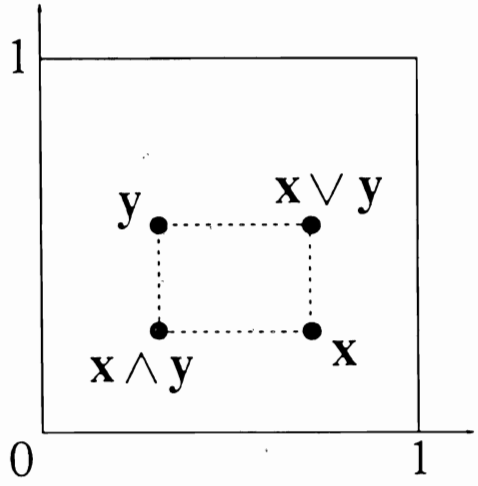

F_1
$\mathbf{I} = (\mathbf{a}, \mathbf{a}^c)$
$|\mathbf{I}| = M$
$\mathbf{a}$
$\mathbf{a}^c = (1-a_1, \ldots, 1-a_M)$
F_0
$\mathbf{a}$
$\mathbf{a} = (a_1, \ldots, a_M)$

Figure 8

$\wedge$ Fuzzy AND (conjunction)

$\vee$ Fuzzy OR (disjunction)



$$\mathbf{x} = (x_1, x_2) \qquad \mathbf{y} = (y_1, y_2)$$

$$(\mathbf{x} \wedge \mathbf{y})_1 = \min(x_1, y_1) \qquad (\mathbf{x} \wedge \mathbf{y})_2 = \min(x_2, y_2)$$

$$(\mathbf{x} \vee \mathbf{y})_1 = \max(x_1, y_1) \qquad (\mathbf{x} \vee \mathbf{y})_2 = \max(x_2, y_2)$$

Figure 9

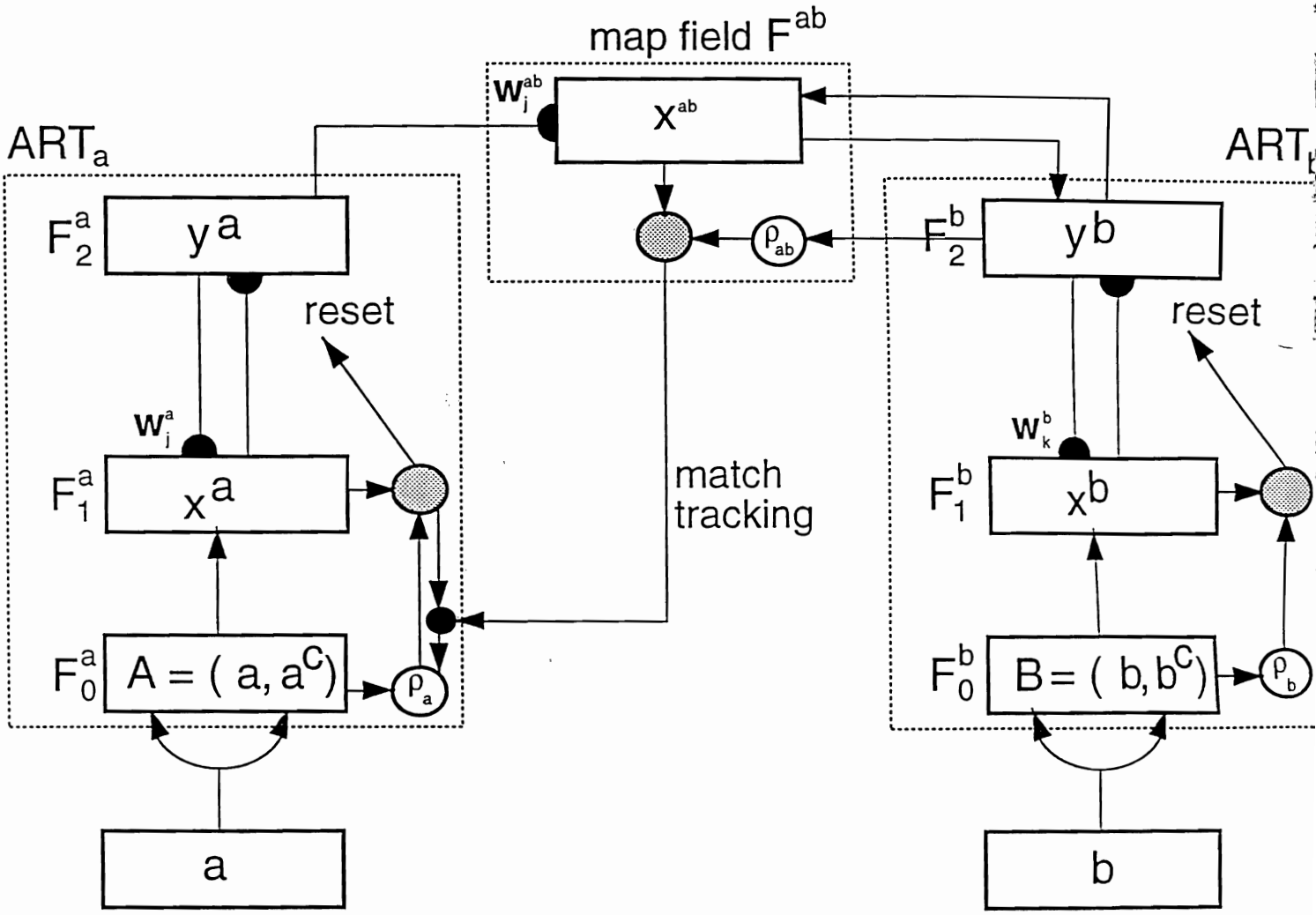

map field F^{ab}
w_j^{ab}
x^{ab}
ART_a
ART_b
F_2^a
y^a
F_2^b
y^b
ρ_{ab}
reset
reset
w_j^a
w_k^b
F_1^a
x^a
F_1^b
x^b
match
tracking
F_0^a
$A = (a, a^c)$
ρ_a
F_0^b
$B = (b, b^c)$
ρ_b
a
b

Figure 10

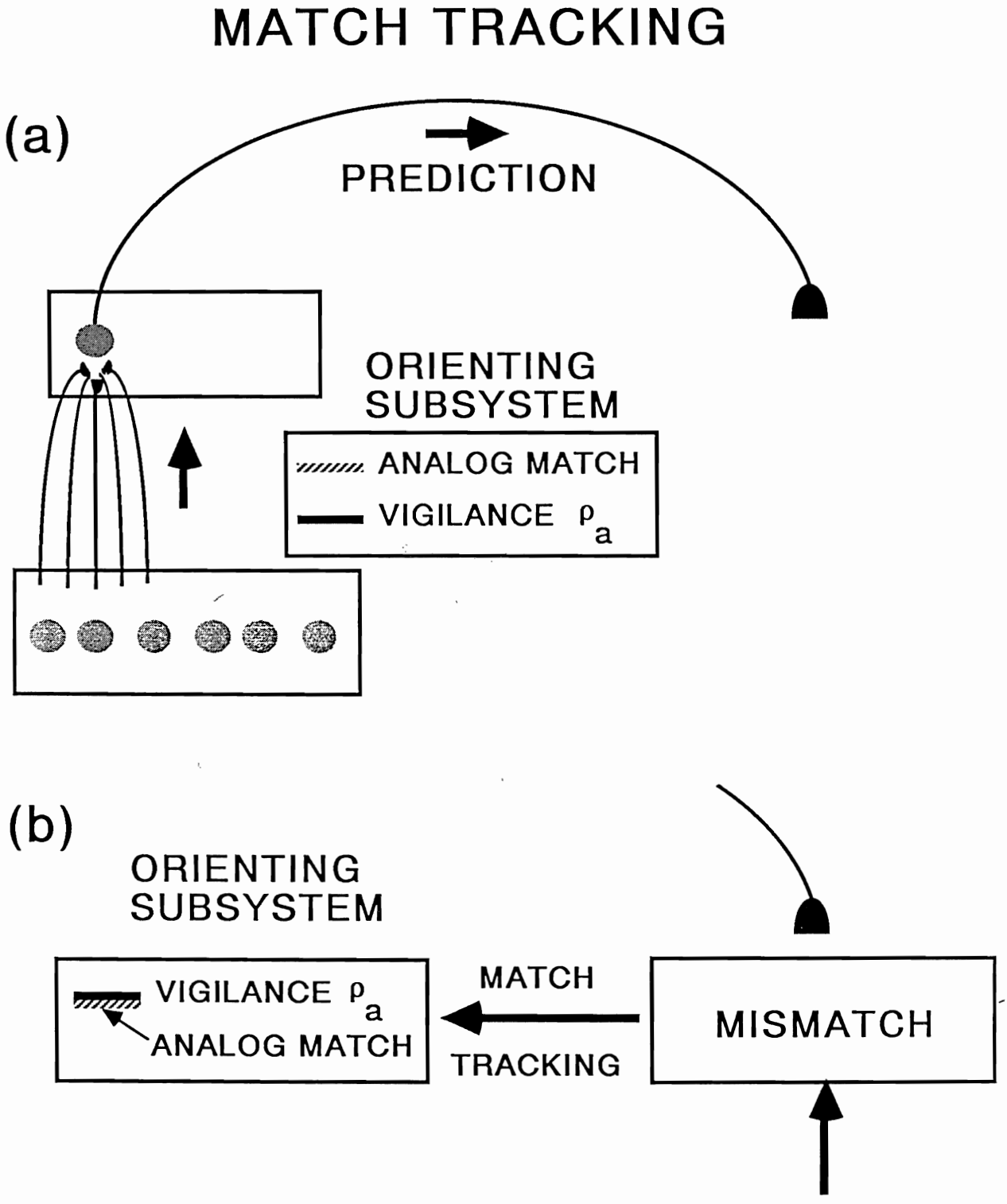

MATCH TRACKING
(a)
PREDICTION
ORIENTING
SUBSYSTEM
ANALOG MATCH
VIGILANCE ρ_a
(b)
ORIENTING
SUBSYSTEM
VIGILANCE ρ_a
ANALOG MATCH
MATCH
TRACKING
MISMATCH

Figure 11

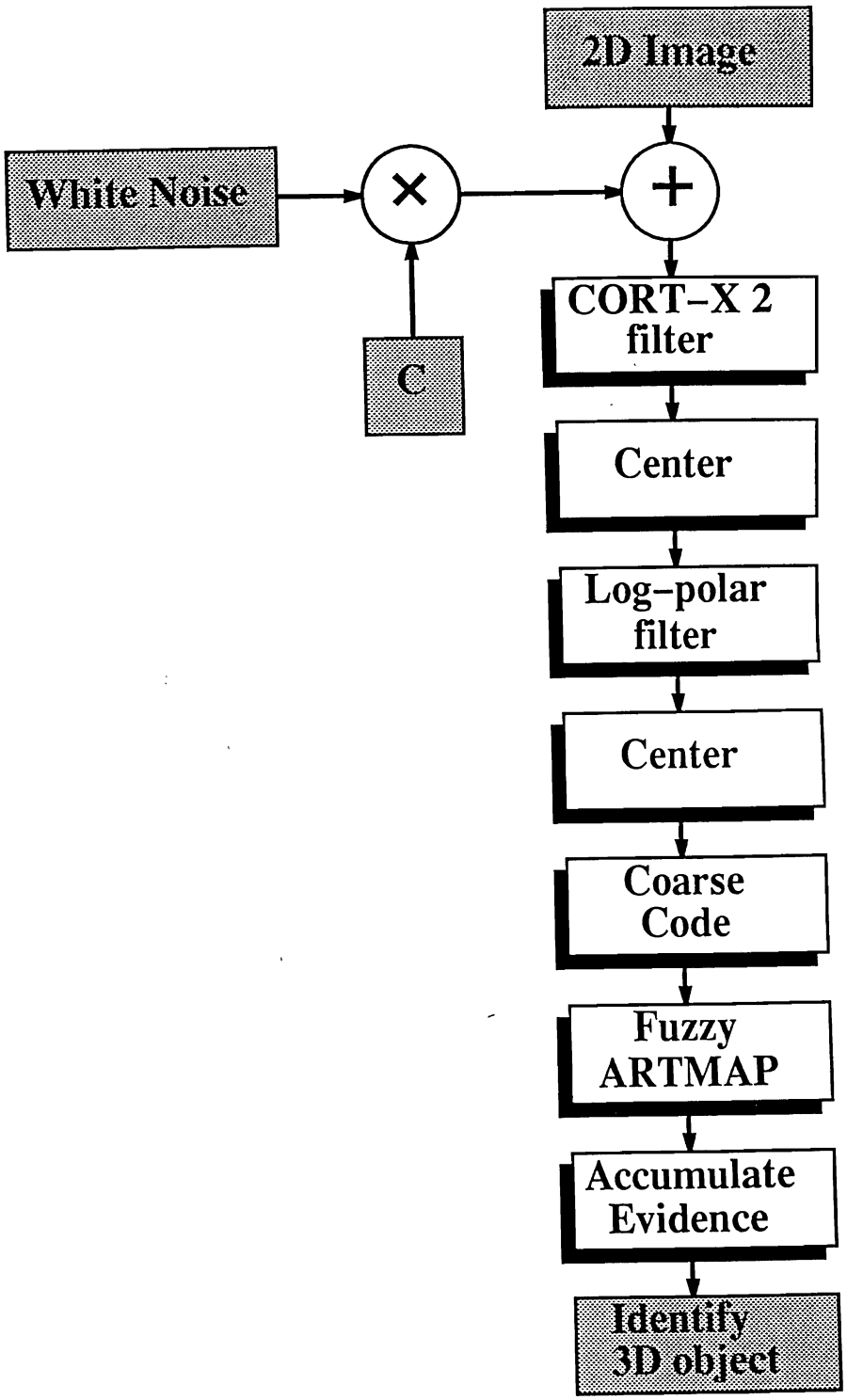
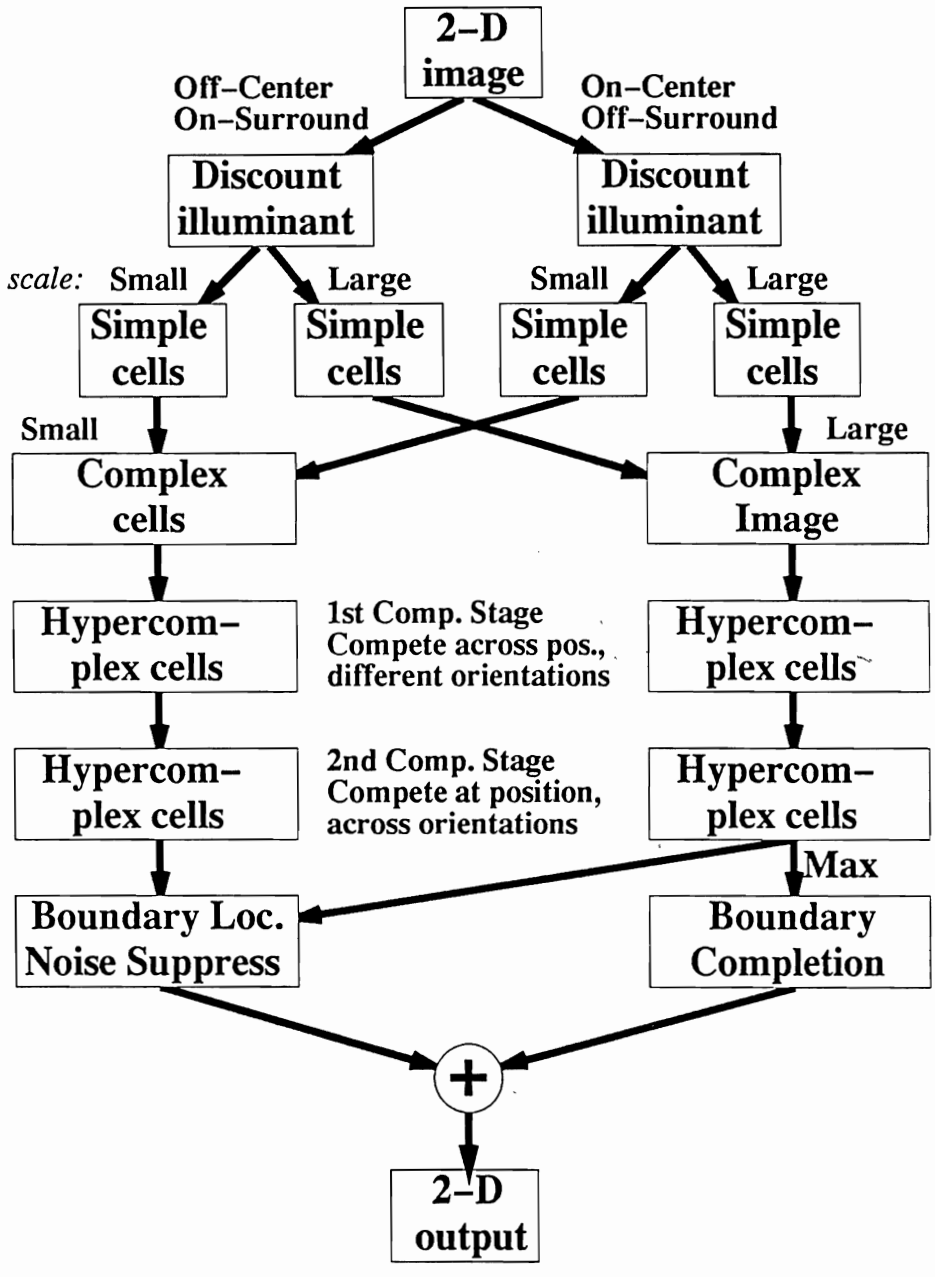

2D Image
White Noise
×
+
C
CORT-X 2 filter
Center
Log-polar filter
Center
Coarse Code
Fuzzy ARTMAP
Accumulate Evidence
Identify 3D object

Figure 12

CORT-X 2 Flow Chart



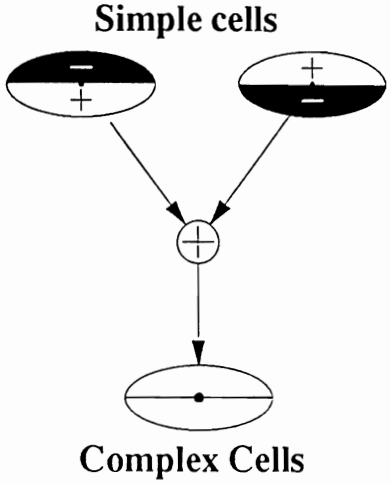


Figure 13

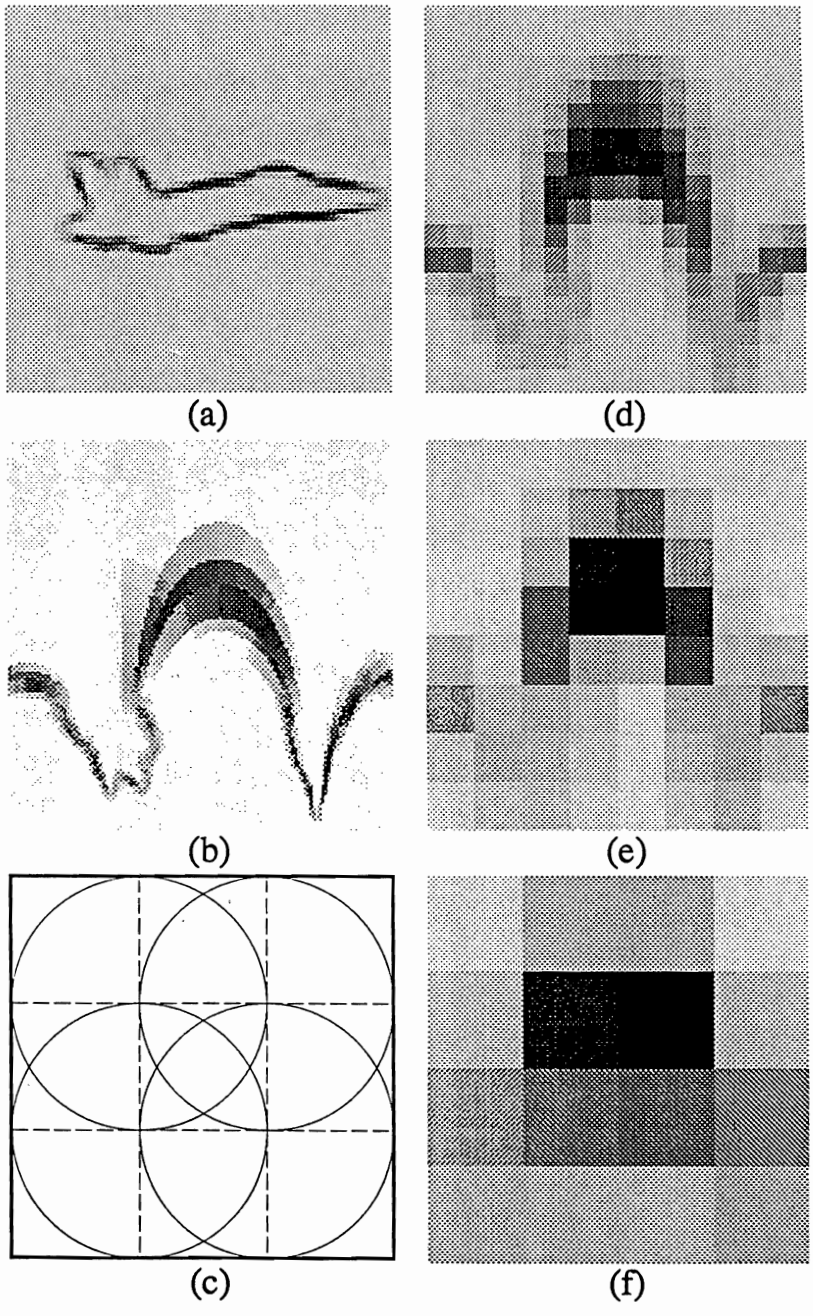
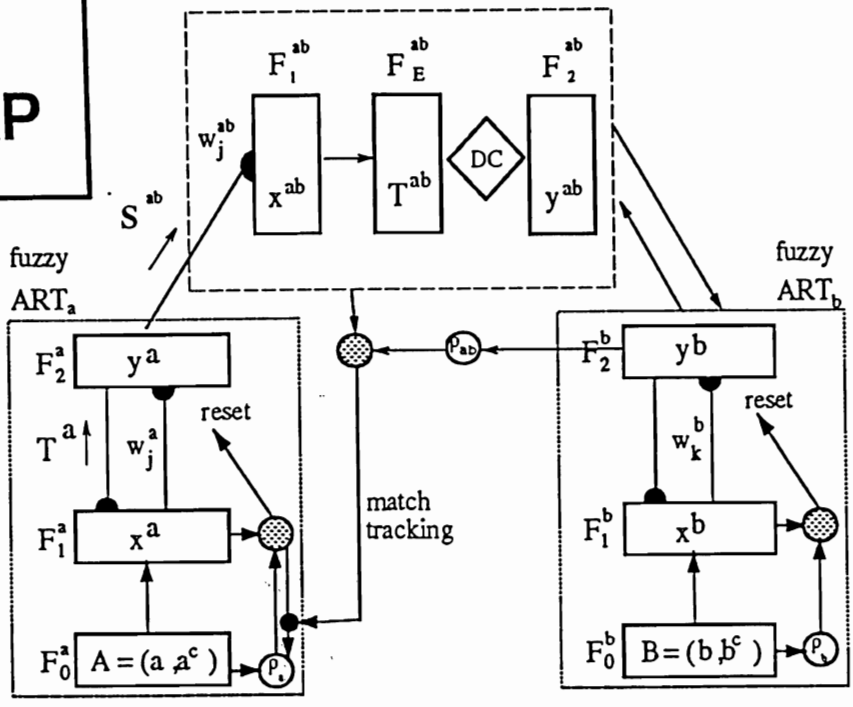


Figure 14

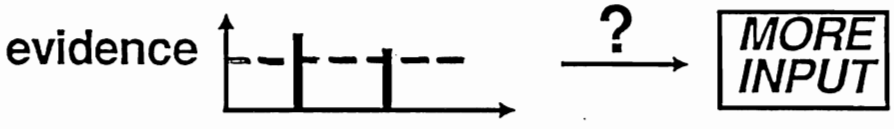
Train:
Fuzzy
ARTMAP

ART-EMAP

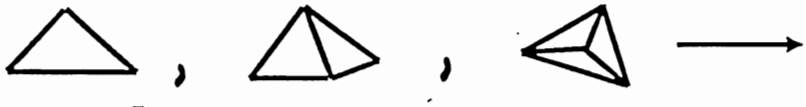


Stage 1: Spatial evidence accumulation

Stage 2: Predictive decision criterion



Stage 3: Spatio-temporal evidence accumulation



Stage 4: Unsupervised rehearsal learning

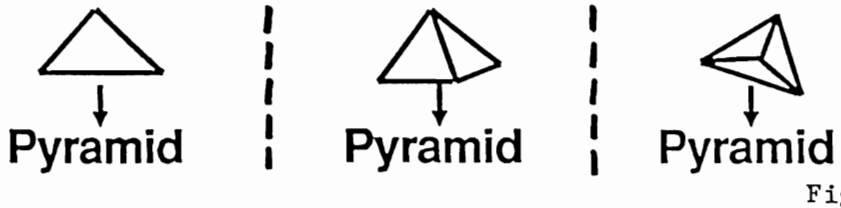


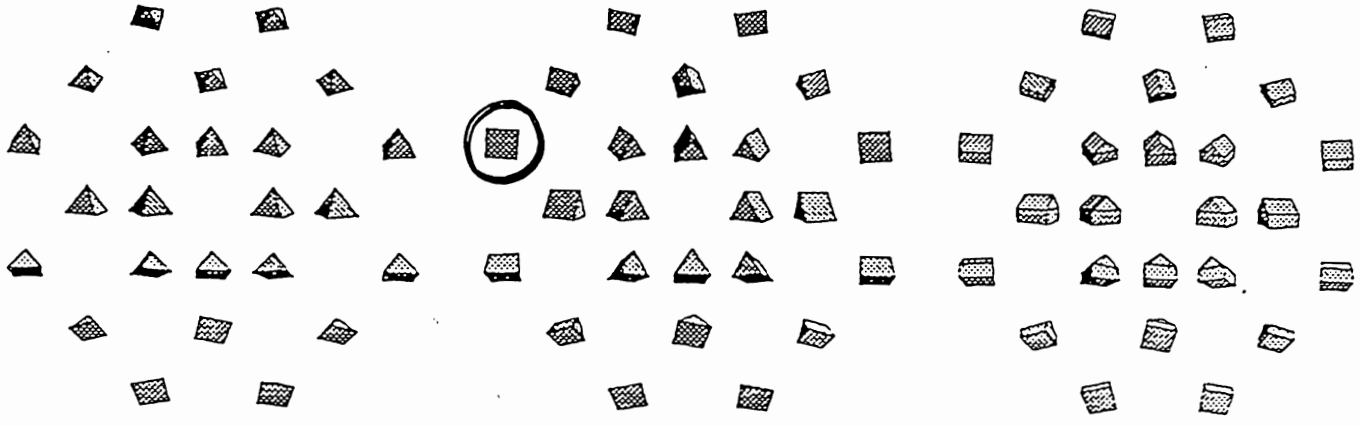
Figure 15

ART-EMAP SIMULATIONS

3-D object recognition from 2-D views (noise-free and noisy)

pyramid prism house

Train (sparse views)



Test (intermediate views)

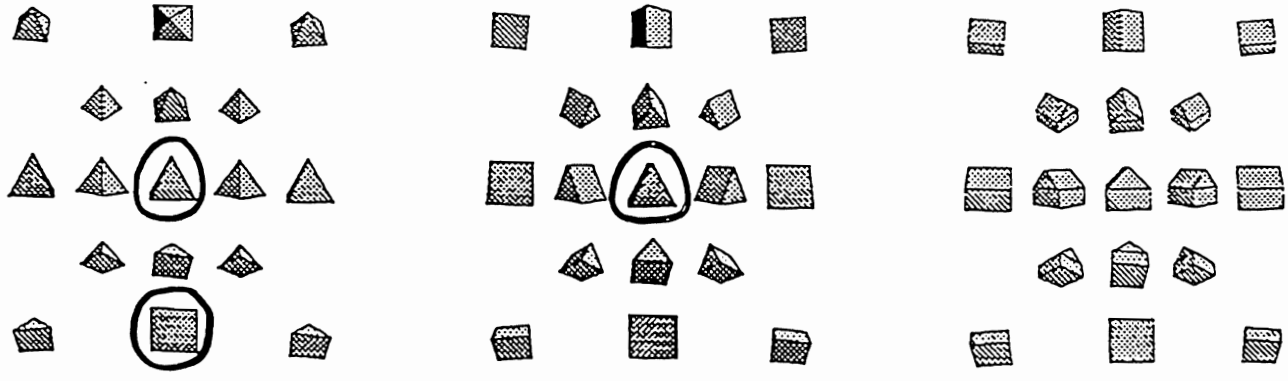
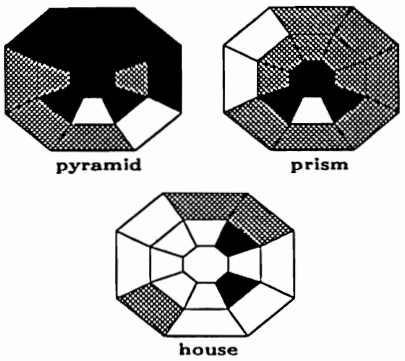


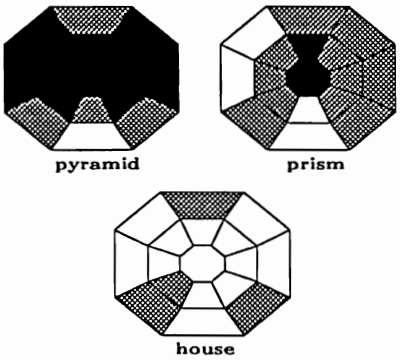
Figure 16

black - pyramid
(a)



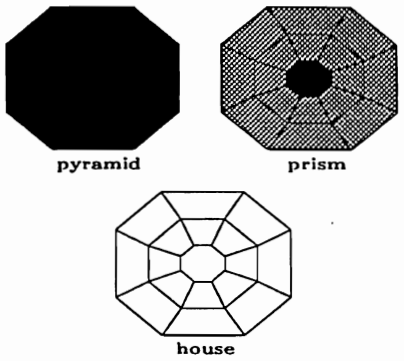
Fuzzy ARTMAP
64.7% on test set

gray - prism
(b)



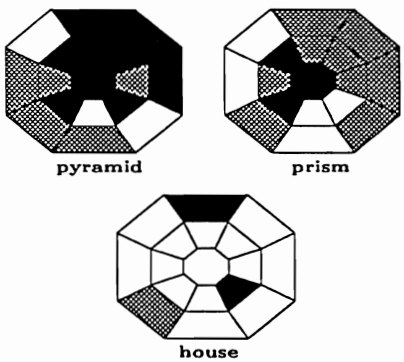
Stage 1 ART-EMAP
Distributed F_2^a activity
p=24 , DC=1.0
70.6% on test set

white - house
(c)



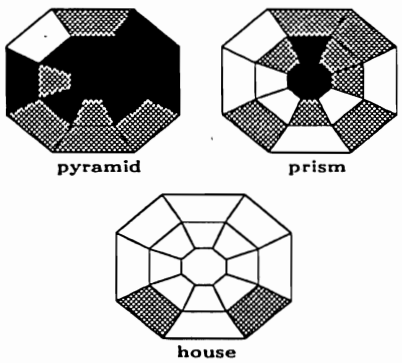
Stage 3 ART-EMAP
Multiple views
p=24 , decreasing DC
98.0% on test set

(d)



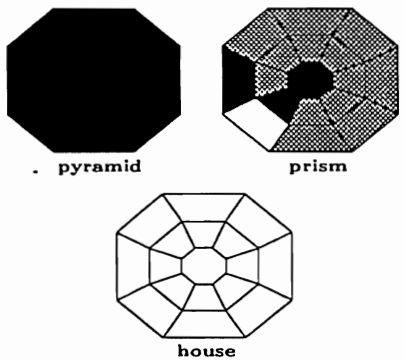
Fuzzy ARTMAP
60.8% on test set

(e)



Stage 1 ART-EMAP
Distributed F_2^a activity
p=24 , DC=1.0
64.7% on test set

(f)



Stage 3 ART-EMAP
Multiple views
p=24 , decreasing DC
92.2% on test set

Figure 17

ARTMAP BENCHMARK STUDIES

1. Medical database - mortality following coronary bypass grafting (CABG) surgery

   FUZZY ARTMAP significantly outperforms:

   LOGISTIC REGRESSION

   ADDITIVE MODEL

   BAYESIAN ASSIGNMENT

   CLUSTER ANALYSIS

   CLASSIFICATION AND REGRESSION TREES

   EXPERT PANEL-DERIVED SICKNESS SCORES

   PRINCIPAL COMPONENT ANALYSIS

2. Mushroom database

   DECISION TREES (90-95 % correct)

   ARTMAP (100% correct)

   Training set an order of magnitude smaller

3. Letter recognition database

   GENETIC ALGORITHM (82% correct)

   FUZZY ARTMAP (96% correct)

4. Circle-in-the-Square task

   BACK PROPAGATION (90% correct)

   FUZZY ARTMAP (99.5% correct)

5. Two-Spiral task

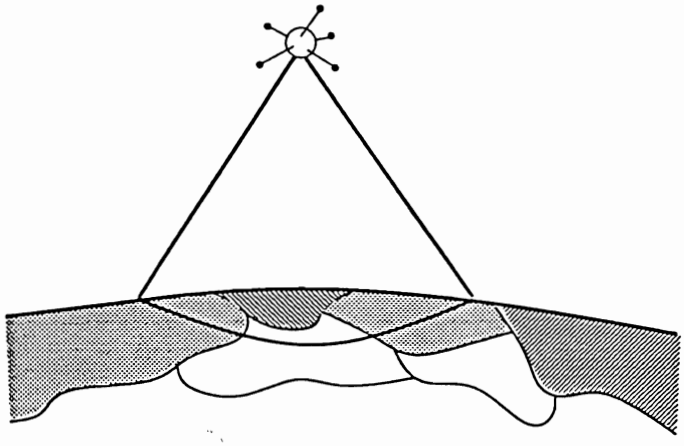
   BACK PROPAGATION (10,000 - 20,000 training epochs)

   FUZZY ARTMAP (1-5 training epochs)

Table 1

Application:
Landsat Satellite Image Classification

(Feng, King, Sutherland, Muggleton, & Henery, 1993)



Algorithm	*Accuracy(%)*
k-N-N	91
fuzzy ARTMAP	89
RBF	88
Alloc80	87
INDCART	86
CART	86
Backprop	86
C4.5	85

Algorithm	*Accuracy(%)*
NewID	85
CN2	85
Quadra	85
SMART	84
LogReg	83
Discrim	83
CASTLE	81

	Test (%)	Compression
k-NN	91%	1:1
fuzzy ARTMAP	89%	6:1

Table 2

Supervised Learning

Training	Test
576	192

(1) ADAP (Smith, Everhart, Dickson, Knowler, and Johannes, 1988)

- **100,000** rules
- **76%** correct on test set
- slow learning

(2) Fuzzy ARTMAP (Carpenter et al., 1992)

- **50-80** rules
- **76%** correct on test set
- fast learning (**6-15** epochs)

Table 3

CORT-X 2 filter set	*Data presentation*	*Coarse code using spatial avg / Gaussian avg*		
		4x4	*8x8*	*16x16*
Small	Ordered	79.9/83.1	84.0/85.6	84.7/89.9
Small	Unordered	78.8/83.3	83.2/85.7	84.9/89.1
Large	Ordered	76.3/78.2	78.5/81.5	77.0/78.8
Large	Unordered	77.4/80.2	79.6/80.41	75.8/79.2

Table 4