ORIGINAL CONTRIBUTION

Fuzzy ART: Fast Stable Learning and Categorization of Analog Patterns by an Adaptive Resonance System

GAIL A. CARPENTER[1], STEPHEN GROSSBERG[2], AND DAVID B. ROSEN[3]

Boston University

Abstract—*A Fuzzy Adaptive Resonance Theory* (ART) *model capable of rapid stable learning of recognition categories in response to arbitrary sequences of analog or binary input patterns is described. Fuzzy ART incorporates computations from fuzzy set theory into the* ART 1 *neural network, which learns to categorize only binary input patterns. The generalization to learning both analog and binary input patterns is achieved by replacing appearances of the intersection operator* (∩) *in* ART 1 *by the* MIN *operator* (∧) *of fuzzy set theory. The* MIN *operator reduces to the intersection operator in the binary case. Category proliferation is prevented by normalizing input vectors at a preprocessing stage. A normalization procedure called complement coding leads to a symmetric theory in which the* MIN *operator* (∧) *and the* MAX *operator* (∨) *of fuzzy set theory play complementary roles. Complement coding uses on-cells and of-cells to represent the input pattern, and preserves individual feature amplitudes while normalizing the total on-cell/off-cell vector. Learning 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. With fast learning 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.*

Keywords—Fuzzy set theory, Adaptive resonance theory, Neural network, Pattern recognition, Learning, ART 1, Categorization, Memory search.

1. INTRODUCTION: A CONNECTION BETWEEN ART SYSTEMS AND FUZZY LOGIC

Adaptive Resonance Theory, or ART, was introduced as a theory of human cognitive information processing (Grossberg, 1976, 1980). The theory has since led to an evolving series of real-time neural network models for unsupervised category learning and pattern recognition. These models are capable of learning stable recognition categories in response to arbitrary input sequences with either fast or slow learning. Model families include ART 1 (Carpenter & Grossberg, 1987a), which can stably learn to categorize binary input patterns presented in an arbitrary order; ART 2 (Carpenter & Grossberg, 1987b), which can stably learn to categorize either analog or binary input patterns presented in an arbitrary order; and ART 3 (Carpenter & Grossberg, 1990a, 1990b), which can carry out parallel search, or hypothesis testing, of distributed recognition codes in a multilevel network hierarchy. Variations of these models adapted to the demands of individual applications have been developed by a number of authors (Baloch & Waxman, 1991; Baxter, 1991; Carpenter, Grossberg, & Rosen, 1991a; Galindo & Michaux, 1990; Gjerdingen, 1990; Gochin, 1990; Harvey et al., 1990; Hecht-Nielsen, 1990; Johnson, 1990; Kosko, 1987a, 1987b, 1987c; Kumar & Guez, 1989, 1991; Levine & Penz, 1990; Li & Wee, 1990; Liao, Liang, & Lin, 1990; Mekkaoui & Jespers, 1990; Michalson & Heldt, 1990; Moore, 1989; Nigrin, 1990; Rajapakse, Jaku-

bowicz, & Acharya, 1990; Ryan, 1988; Seibert & Waxman, 1990a, 1990b; Simpson, 1990; Weingard, 1990; Wilson, Wilkinson, & Ganis, 1990; Winter, 1989; Winter, Ryan, & Turner, 1987).

Recently the ART 1 model has been used to design a hierarchical network architecture, called ART-MAP, that can rapidly self-organize stable categorical mappings between m-dimensional input vectors and n-dimensional output vectors (Carpenter, Grossberg, & Reynolds, 1991a,b). Under supervised learning conditions, ARTMAP's internal control mechanisms create stable recognition categories of optimal size by maximizing predictive generalization while minimizing predictive error in an on-line setting. ARTMAP was originally used to learn mappings between binary input and binary output vectors. The Fuzzy ART model (Carpenter, Grossberg, & Rosen, 1991b) developed herein generalizes ART 1 to be capable of learning stable recognition categories in response to both analog and binary input patterns. This Fuzzy ART model has been incorporated into a Fuzzy ARTMAP architecture (Carpenter, Grossberg, Markuzon, Reynolds, & Rosen, 1991) that can rapidly learn stable categorical mappings between analog or binary input and output vectors. For example, Fuzzy ARTMAP learns in five training epochs a benchmark that requires twenty thousand epochs for back propagation to learn (Lang & Witbrock, 1989). The Fuzzy ART system is summarized in Section 3.

Fuzzy ART incorporates the basic features of all ART systems, notably, pattern matching between bottom-up input and top-down learned prototype vectors. This matching process leads either to a resonant state that focuses attention and triggers stable prototype learning or to a self-regulating parallel memory search. If the search ends by selecting an established category, then the category's prototype may be refined to incorporate new information in the input pattern. If the search ends by selecting a previously untrained node, then learning of a new category takes place.

Figure 1 illustrates a typical example chosen from the family of ART 1 models, and Figure 2 illustrates a typical ART search cycle. As shown in Figure 2a, 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 long-term memory (LTM) traces, to generate a net input vector T to level F_2. The internal competitive dynamics of F_2 contrast-enhance vector T. 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
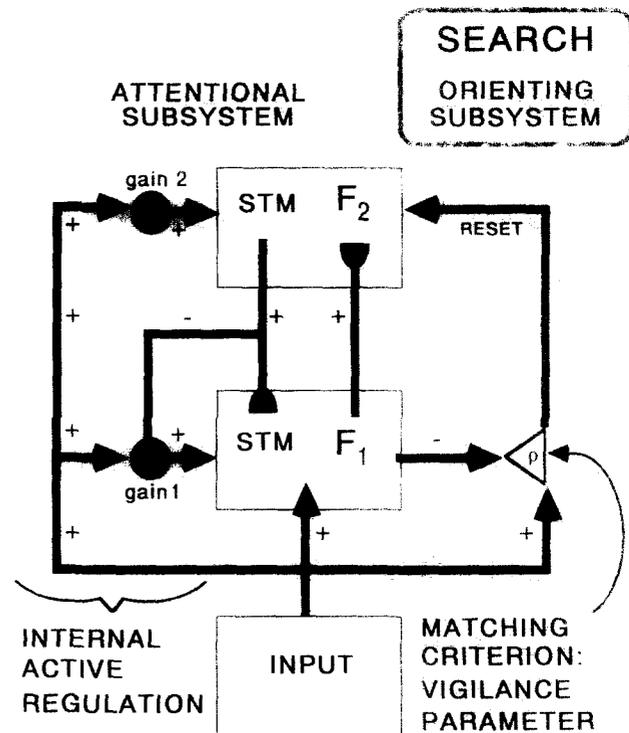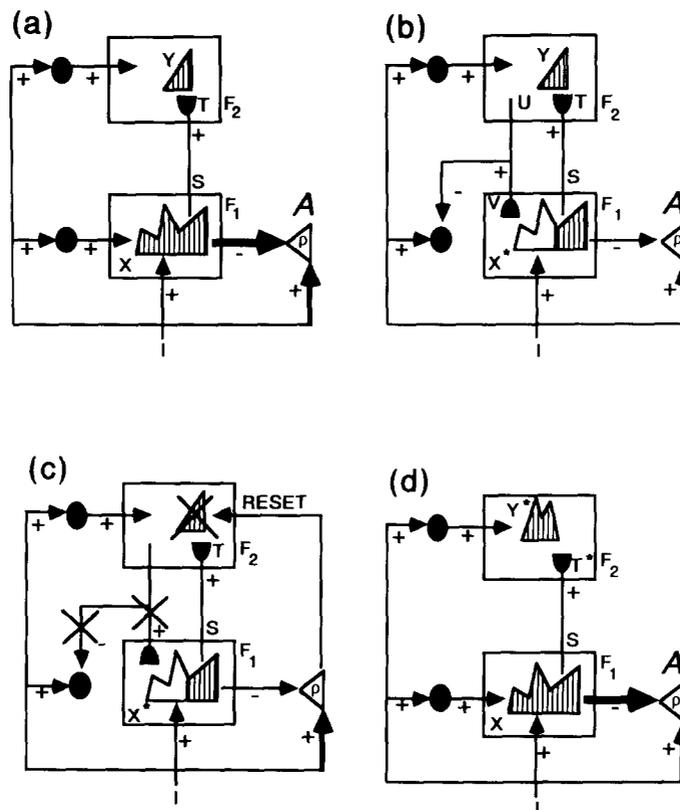

FIGURE 1. Typical ART 1 neural network (Carpenter & Grossberg, 1987a).

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.

Activation of an F_2 node may be interpreted as "making a hypothesis" about an input I. When Y is activated, it generates a signal 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 2b). 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.

This 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 V. The resultant pattern 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 occurs as the attentional focus takes hold. The resonant state persists long enough for learning to occur; hence the term adaptive resonance theory. ART 1 learns prototypes, rather than exemplars, because the attended feature vector X^*, rather than the input I itself, is learned.

FIGURE 2. ART search for an F_2 code: (a) The input pattern I generates the specific STM activity pattern X at F_1 as it nonspecifically activates the orienting subsystem A. Pattern X both inhibits A and generates the output signal pattern S. Signal pattern S is transformed into the input pattern T, which activates the STM pattern Y across F_2; (b) Pattern Y generates the top-down signal pattern U, which is transformed into the prototype pattern V. If V mismatches I at F_1, then a new STM activity pattern X* is generated at F_1. The reduction in total STM activity which occurs when X is transformed into X* causes a decrease in the total inhibition from F_1 to A; (c) If the matching criterion fails to be met, 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 X to activate a different STM pattern Y* at F_2. If the top-down prototype due to Y* also mismatches I at F_1, then the search for an appropriate F_2 code continues.

The criterion of an acceptable match is defined by a dimensionless parameter called *vigilance*. Vigilance weighs how close the input exemplar I must be to the top-down prototype V for resonance to occur. Because vigilance can vary across learning trials, recognition categories capable of encoding widely differing degrees of generalization, or morphological variability, can be learned by a single ART system. Low vigilance leads to broad generalization and abstract prototypes. High vigilance leads to narrow generalization and to prototypes that represent fewer input examplars. In the limit of very high vigilance, prototype learning reduces to exemplar learning. Thus a single ART system may be used, say, to recognize abstract categories of faces and dogs, as well as individual faces and dogs.

If the top-down expectation V and the bottom-up input I are too novel, or unexpected, to satisfy the vigilance criterion, then a bout of hypothesis testing, or memory search, is triggered. Search leads to selection of a better recognition code, symbol, cat-

egory, or hypothesis to represent input I at level F_2. An *orienting subsystem* A mediates the search process (Figure 1). The orienting subsystem interacts with the attentional subsystem, as in Figures 2c and 2d, to enable the attentional subsystem to learn about novel inputs without risking unselective forgetting of its previous knowledge.

The search process prevents associations from forming between Y and X* if X* is too different from I to satisfy the vigilance criterion. The search process resets Y before such an association can form. A familiar category may be selected by the search if its prototype is similar enough to the input I to satisfy the vigilance criterion. The prototype may then be refined in light of new information carried by I. If I is too different from any of the previously learned prototypes, then an uncommitted F_2 node is selected and learning of a new category is initiated.

A network parameter controls how deeply the search proceeds before an uncommitted node is chosen. As learning of a particular category self-stabi-

lizes, all inputs coded by that category access it directly in a one-pass fashion, and search is automatically disengaged. The category selected is then the one whose prototype provides the globally best match to the input pattern. Learning can proceed on-line, and in a stable fashion, with familiar inputs directly activating their categories, while novel inputs continue to trigger adaptive searches for better categories, until the network's memory capacity is fully utilized.

The read-out of the top-down expectation **V** may be interpreted as a type of hypothesis-driven query. The matching process at F_1 and the hypothesis testing process at F_2 may be interpreted as query-driven symbolic substitutions. From this perspective, ART systems provide examples of new types of self-organizing production systems (Laird, Newell, & Rosenbloom, 1987). By incorporating predictive feedback into their control of the hypothesis testing cycle, ARTMAP systems embody self-organizing production systems that are also goal-oriented. ARTMAP systems are thus a new type of self-organizing expert system, which is capable of stable autonomous fast learning about nonstationary environments that may contain a great deal of morphological variability. The fact that fuzzy logic may also be usefully incorporated into ARTMAP systems blurs even further the traditional boundaries between artificial intelligence and neural networks.

The new Fuzzy ART model incorporates the design features of other ART models due to the close formal homolog between ART 1 and Fuzzy ART operations. Figure 3 summarizes how the ART 1 operations of category choice, matching, search, and

ART 1
(BINARY)

FUZZY ART
(ANALOG)

CATEGORY CHOICE

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

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

MATCH CRITERION

$$\frac{|I \cap w|}{|I|} \geq \rho$$

$$\frac{|I \wedge w|}{|I|} \geq \rho$$

FAST LEARNING

$$w_j^{(new)} = I \cap w_j^{(old)}$$

$$w_j^{(new)} = I \wedge w_j^{(old)}$$

$\cap$ = logical AND intersection

$\wedge$ = fuzzy AND minimum

FIGURE 3. Analogy between ART 1 and Fuzzy ART. The notation w_j in ART 1 denotes the index set of top-down LTM traces of the j-th category that exceed a prescribed positive threshold value. See Carpenter and Grossberg (1987a) for details.

learning translate into Fuzzy ART operations by replacing the set theory intersection operator ($\cap$) of ART 1 by the fuzzy set theory conjunction, or MIN operator ($\wedge$). Despite this close formal homology, Fuzzy ART is described as an algorithm, rather than a locally defined neural model. A neural network realization of Fuzzy ART is described elsewhere (Carpenter, Grossberg, & Rosen, 1991c). For the special case of binary inputs and fast learning, the computations of Fuzzy ART are identical to those of the ART 1 neural network. The Fuzzy ART algorithm also includes two optional features, one concerning learning and the other input preprocessing, as described in Section 2.

2. FAST-LEARN SLOW-RECODE AND COMPLEMENT CODING

Many applications of ART 1 use fast learning, whereby adaptive weights fully converge to new equilibrium values in response to each input pattern. Fast learning enables a system to adapt quickly to inputs that may occur only rarely and that may require immediate accurate performance. The ability of humans to remember many details of an exciting movie is a typical example of fast learning. It has been mathematically proved that ART 1 can carry out fast learning of stable recognition categories in an on-line setting in response to arbitrary lists of binary input patterns (Carpenter & Grossberg, 1987a). In contrast, error-based learning models like backpropagation become unstable in this type of learning environment. This is because back propagation learning is driven by the difference between the actual output and a target output. Fast learning would zero this error signal on each input trial and would thus force unselective forgetting of past learning. This feature of back propagation restricts its domain to off-line learning applications carried out with a slow learning rate. Off-line learning is needed because real-time presentation of inputs with variable durations has a similar effect on learning as presenting the same inputs with a fixed duration but variable learning rates. In particular, longer duration inputs reduce the error signal more on each input trial and thus have an effect similar to fast learning. In addition, lacking the key feature of competition, a back propagation system tends to average rare events with similar frequent events that may have different consequences.

For some applications, it is useful to combine fast initial learning with a slower rate of forgetting. We call this the *fast-commit slow-recode* option. This combination of properties retains the benefit of fast learning; namely, an adequate response to inputs that may occur only rarely and in response to which accurate performance may be quickly demanded. The slow-recode operation also prevents features that have already been incorporated into a category's proto-

type from being erroneously deleted in response to noisy or partial inputs. With slow recoding, only a statistically persistent change in a feature's relevance to a category can delete it from the prototype of the category. The fast-commit slow-recode option in Fuzzy ART corresponds to ART 2 learning at intermediate learning rates (Carpenter, Grossberg, & Rosen, 1991a).

The input preprocessing option concerns normalization of input patterns. It is shown below that input normalization prevents a problem of category proliferation that could otherwise occur (Moore, 1989). A normalization procedure called *complement coding* is of particular interest from three vantage points. From a neurobiological perspective, complement coding uses both on-cells and off-cells to represent an input pattern, and preserves individual feature amplitudes while normalizing the total on-cell/off-cell vector. From a functional perspective, the on-cell portion of the prototype encodes features that are critically present in category exemplars, while the off-cell portion encodes features that are critically absent. Features that are occasionally present in a category's input exemplars lead to low weights in both the on-cell and the off-cell portions of the prototype. Finally, from a set theoretic perspective, complement coding leads to a more symmetric theory in which both the MIN operator ($\wedge$) and the MAX operator ($\vee$) of fuzzy set theory play a role (Figure 4). Using both the MIN and the MAX operations, a geometrical interpretation of Fuzzy ART learning is given in terms of box-shaped recognition categories whose corners are iteratively defined in terms of the $\wedge$ and $\vee$ operators. Complement coding hereby establishes a connection between on-cell/off-cell representations and fuzzy set theory operations. This linkage further develops a theme concerning the

relationship between ART on-cell/off-cell representations, hypothesis testing, and probabilistic logic that was outlined at the theory's inception and used to explain various perceptual and cognitive data (Grossberg, 1980, Sections 7–9; Grossberg, 1982, Section 47).

Section 4 discusses Fuzzy ART systems in a parameter range called the conservative limit. In this limit, an input always selects a category whose weight vector is a fuzzy subset of the input, if such a category exists. Given such a choice, no weight change occurs during learning; hence the name conservative limit, since learned weights are conserved wherever possible. Section 5 describes Fuzzy ART coding of two-dimensional analog vectors that are preprocessed into complement coding form before being presented to the Fuzzy ART system. The geometric interpretation of Fuzzy ART dynamics is introduced here and illustrative computer simulations are summarized. The geometric formulation allows comparison between Fuzzy ART and aspects of the NGE (Nested Generalized Exemplars) algorithms of Salzberg (1990). Section 6 further develops the geometric interpretation and provides a simulation of Fuzzy ART without complement coding to show how category proliferation can occur. Section 7 compares the stability of Fuzzy ART to that of related clustering algorithms that were discussed by Moore (1989). The Fuzzy ART computations of choice, search, learning, and complement coding endow the system with stability properties that overcome limitations of the algorithms described by Moore.

$\wedge$ 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 4. Fuzzy operations.

3. SUMMARY OF THE FUZZY ART ALGORITHM

Input vector: Each input $\mathbf{I}$ is an M-dimensional vector $(I_1, \ldots, I_M)$, where each component I_i is in the interval $[0, 1]$.

Weight vector: Each category (j) corresponds to a vector $\mathbf{w}_j = (w_{j1}, \ldots, w_{jM})$ of adaptive weights, or LTM traces. The number of potential categories N ($j = 1, \ldots, N$) is arbitrary. Initially
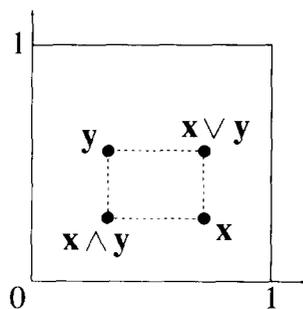
$$w_{j1} = \ldots = w_{jM} = 1, \tag{1}$$

and each category is said to be *uncommitted*. Alternatively, initial weights w_{ji} may be taken greater than 1. Larger weights bias the system against selection of uncommitted nodes, leading to deeper searches of previously coded categories.

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 category j, the choice function T_j is defined by

$$T_j(\mathbf{I}) = \frac{|\mathbf{I} \wedge \mathbf{w}_j|}{\alpha + |\mathbf{w}_j|}, \tag{2}$$

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

$$(\mathbf{x} \wedge \mathbf{y})_i \equiv \min(x_i, y_i), \tag{3}$$

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

$$|\mathbf{x}| \equiv \sum_{i=1}^{M} |x_i|. \tag{4}$$

For notational simplicity, $T_j(\mathbf{I})$ in eqn (2) is often written as T_j when the input $\mathbf{I}$ is fixed. The category choice is indexed by J, where

$$T_J = \max\{T_j : j = 1 \ldots N\}. \tag{5}$$

If more than one T_j is maximal, the category j with the smallest index is chosen. In particular, nodes become committed in order $j = 1, 2, 3, \ldots$.

Resonance or reset: *Resonance* occurs if the match function of the chosen category meets the vigilance criterion; that is, if

$$\frac{|\mathbf{I} \wedge \mathbf{w}_J|}{|\mathbf{I}|} \geq \rho. \tag{6}$$

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

$$\frac{|\mathbf{I} \wedge \mathbf{w}_J|}{|\mathbf{I}|} < \rho. \tag{7}$$

Then the value of the choice function T_J is reset to -1 for the duration of the input presentation to prevent its persistent selection during search. A new index J is chosen, by eqn (5). The search process continues until the chosen J satisfies eqn (6).

Learning: The weight vector $\mathbf{w}_J$ is updated according to the equation

$$\mathbf{w}_J^{(new)} = \beta(\mathbf{I} \wedge \mathbf{w}_J^{(old)}) + (1 - \beta)\mathbf{w}_J^{(old)}. \tag{8}$$

Fast learning corresponds to setting $\beta = 1$ (Figure 3). The learning law (8) is the same as one used by Moore (1989) and Salzberg (1990).

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 $\mathbf{w}_J^{(new)} = \mathbf{I}$ the first time category J becomes active.

Input normalization option: Moore (1989) 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. Proliferation of categories is avoided in Fuzzy ART if inputs are normalized; that is, for some $\gamma > 0$,

$$|\mathbf{I}| \equiv \gamma, \tag{9}$$

for all inputs $\mathbf{I}$. Normalization can be achieved by preprocessing each incoming vector $\mathbf{a}$, for example setting

$$\mathbf{I} = \frac{\mathbf{a}}{|\mathbf{a}|}. \tag{10}$$

An alternative normalization rule, called *complement coding*, achieves normalization while preserving amplitude information. Complement coding represents both the on-response and the off-response to $\mathbf{a}$ (Figure 5). To define this operation in its simplest form, let $\mathbf{a}$ itself represent the on-response. The complement of $\mathbf{a}$, denoted by $\mathbf{a}^c$, represents the off-response, where

$$a_i^c \equiv 1 - a_i. \tag{11}$$

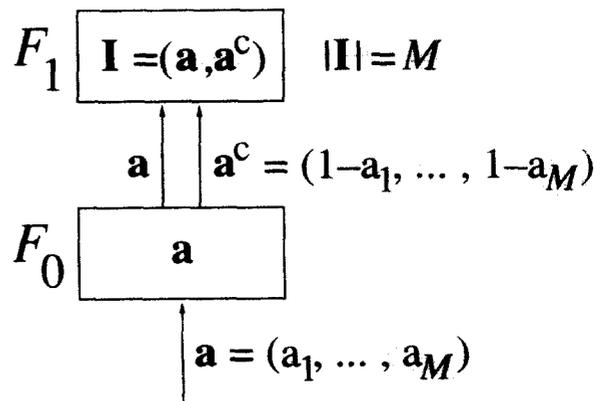The complement coded input $\mathbf{I}$ to the recognition system is the $2M$-dimensional vector

$$\mathbf{I} = (\mathbf{a}, \mathbf{a}^c) \equiv (a_1, \ldots, a_M, a_1^c, \ldots, a_M^c). \tag{12}$$

Note that

$$\begin{aligned}|\mathbf{I}| &= |(\mathbf{a}, \mathbf{a}^c)| \\ &= \sum_{i=1}^{M} a_i + \left(M - \sum_{i=1}^{M} a_i\right) \\ &= M,\end{aligned} \tag{13}$$

so inputs preprocessed into complement coding form are automatically normalized. Where complement coding is used, the initial condition (1) is replaced by

$$w_{j1} = \ldots = w_{j,2M} = 1. \tag{14}$$

FIGURE 5. Complement coding uses on-cell and off-cell pairs to normalize input vectors.

4. FUZZY SUBSET CHOICE AND ONE-SHOT FAST LEARNING IN THE CONSERVATIVE LIMIT

In fast-learn ART 1, if the choice parameter α in (2) is chosen close to 0 (see Figure 3), then the first category chosen by (5) will always be the category whose weight vector $\mathbf{w}_j$ is the largest coded subset of the input vector $\mathbf{I}$, if such a category exists (Carpenter & Grossberg, 1987a). In other words, $\mathbf{w}_j$ is chosen if it has the maximal number of 1's ($w_{ji} = 1$) at indices i where $I_i = 1$, and 0's elsewhere, among all weight vectors $\mathbf{w}_j$. Moreover, when $\mathbf{w}_j$ is a subset of $\mathbf{I}$ during resonance, $\mathbf{w}_j$ is unchanged, or conserved, during learning. More generally, $\mathbf{w}_j$ encodes the attentional focus induced by $\mathbf{I}$, not $\mathbf{I}$ itself. The limit $\alpha \to 0$ is called the *conservative limit* because small values of α tend to minimize recoding during learning.

For analog vectors, the degree to which $\mathbf{y}$ is a fuzzy subset of $\mathbf{x}$ is given by the term

$$\frac{|\mathbf{x} \wedge \mathbf{y}|}{|\mathbf{y}|} \quad (15)$$

(Kosko, 1986). In the conservative limit of Fuzzy ART, the choice function T_j in eqn (2) reflects the degree to which the weight vector $\mathbf{w}_j$ is a fuzzy subset of the input vector $\mathbf{I}$. If

$$\frac{|\mathbf{I} \wedge \mathbf{w}_j|}{|\mathbf{w}_j|} = 1, \quad (16)$$

then $\mathbf{w}_j$ is a fuzzy subset of $\mathbf{I}$ (Zadeh, 1965), and category j is said to be a *fuzzy subset choice* for input $\mathbf{I}$. In this case, by eqn (8), no recoding occurs if j is selected since $\mathbf{I} \wedge \mathbf{w}_j = \mathbf{w}_j$.

Resonance depends on the degree to which $\mathbf{I}$ is a fuzzy subset of $\mathbf{w}_j$, by eqns (6) and (7). In particular, if category j is a fuzzy subset choice, then the match function value is given by

$$\frac{|\mathbf{I} \wedge \mathbf{w}_j|}{|\mathbf{I}|} = \frac{|\mathbf{w}_j|}{|\mathbf{I}|}. \quad (17)$$

Thus, choosing J to maximize $|\mathbf{w}_j|$ among fuzzy subset choices also maximizes the opportunity for resonance in eqn (6). If reset occurs for the node that maximizes $|\mathbf{w}_j|$, reset will also occur for all other subset choices.

Consider a Fuzzy ART system in the conservative limit with fast learning and normalized inputs. Then $\alpha \cong 0$ in eqn (2), $\beta = 1$ in eqn (8), and eqn (9) holds. Under these conditions, one-shot stable learning occurs; that is, no weight change or search occurs after each item of an input set is presented *just once*, although some inputs may select different categories on future trials. To see this, note by eqns (6), (8), and (9) that when $\mathbf{I}$ is presented for the first time, $\mathbf{w}_j^{(new)} \to \mathbf{I} \wedge \mathbf{w}_j^{(old)}$ for some category node $J = j$ such that $|\mathbf{I} \wedge \mathbf{w}_j^{(old)}| \geq \rho|\mathbf{I}| = \rho\gamma$. Thereafter, category j is

a fuzzy subset choice of $\mathbf{I}$, by eqn (16). If $\mathbf{I}$ is presented again, it will either choose $J = j$ or make another fuzzy subset choice, maximizing $|\mathbf{w}_j|$, because fuzzy subset choices (16) maximize the category choice function (2) in the conservative limit. In either case, $\mathbf{w}_j^{(new)} = \mathbf{I} \wedge \mathbf{w}_j^{(old)} = \mathbf{w}_j^{(old)}$, which implies that neither reset nor additional learning occurs.

5. FUZZY ART WITH COMPLEMENT CODING

A geometric interpretation of Fuzzy ART with complement coding will now be developed. For definiteness, let the input set consist of two-dimensional vectors $\mathbf{a}$ preprocessed into the four-dimensional complement coding form. Thus

$$\mathbf{I} = (\mathbf{a}, \mathbf{a}^c) = (a_1, a_2, 1 - a_1, 1 - a_2). \quad (18)$$

In this case, each category j has a geometric representation as a rectangle R_j, as follows. Following eqn (18), the weight vector $\mathbf{w}_j$ can be written in complement coding form:

$$\mathbf{w}_j = (\mathbf{u}_j, \mathbf{v}_j^c), \quad (19)$$

where $\mathbf{u}_j$ and $\mathbf{v}_j$ are two-dimensional vectors. Let vector $\mathbf{u}_j$ define one corner of a rectangle R_j and let $\mathbf{v}_j$ define another corner of R_j (Figure 6a). The size of R_j is defined to be

$$|R_j| \equiv |\mathbf{v}_j - \mathbf{u}_j|, \quad (20)$$

which is equal to the height plus the width of R_j in Figure 6a.

In a fast-learn Fuzzy ART system, with $\beta = 1$ in eqn (8), $\mathbf{w}_j^{(new)} = \mathbf{I} = (\mathbf{a}, \mathbf{a}^c)$ when J is an uncommitted node. The corners of $R_j^{(new)}$ are then given by $\mathbf{a}$ and $(\mathbf{a}^c)^c = \mathbf{a}$. Hence, $R_j^{(new)}$ is just the point $\mathbf{a}$. Learning increases the size of each R_j. In fact the size of R_j grows as the size of $\mathbf{w}_j$ shrinks during learning, and the maximum size of R_j is determined by the vigilance parameter ρ, as shown below. During each fast-learning trial, R_J expands to $R_J \oplus \mathbf{a}$, the minimum rectangle containing R_J and $\mathbf{a}$ (Figure 6b). The corners of $R_J \oplus \mathbf{a}$ are given by $\mathbf{a} \wedge \mathbf{u}_J$ and $\mathbf{a} \vee \mathbf{v}_J$, where

$$(\mathbf{x} \vee \mathbf{y})_i \equiv \max(x_i, y_i) \quad (21)$$

(Zadeh, 1965). Hence, by eqn (20), the size of $R_J \oplus \mathbf{a}$ is given by

$$|R_J \oplus \mathbf{a}| = |(\mathbf{a} \vee \mathbf{v}_J) - (\mathbf{a} \wedge \mathbf{u}_J)|. \quad (22)$$

However, reset leads to a new category choice if $|R_J \oplus \mathbf{a}|$ is too large. These properties will now be proved.

Suppose that $\mathbf{I} = (\mathbf{a}, \mathbf{a}^c)$ chooses category J, by eqn (5). The weight vector $\mathbf{w}_J$ is updated according
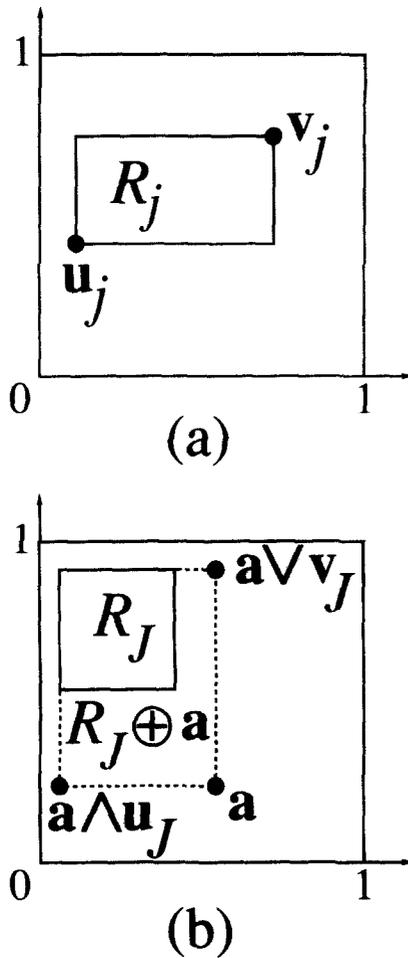
(a)

(b)

FIGURE 6. (a) In complement coding form with $M = 2$, each weight vector w_j has a geometric interpretation as a rectangle R_j with corners (u_j, v_j); (b) During fast learning, R_J expands to $R_J \oplus a$, the smallest rectangle that includes R_J and a, provided that $|R_J \oplus a| \leq 2(1 - \rho)$.

to the fast-learn equation

$$\mathbf{w}_J^{(new)} = \mathbf{I} \wedge \mathbf{w}_J^{(old)} \tag{23}$$

only if the match criterion eqn (6) is satisfied. Because of complement coding, $|\mathbf{I}| = M$, by eqn (13). Thus, when $M = 2$, the match criterion eqn (6) is satisfied iff

$$|\mathbf{I} \wedge \mathbf{w}_J| \geq 2\rho. \tag{24}$$

However,

$$
\begin{aligned}
|\mathbf{I} \wedge \mathbf{w}_J| &= |(\mathbf{a}, \mathbf{a}^c) \wedge (\mathbf{u}_J, \mathbf{v}_J^c)| \\
&= |(\mathbf{a} \wedge \mathbf{u}_J), (\mathbf{a}^c \wedge \mathbf{v}_J^c)| \\
&= |(\mathbf{a} \wedge \mathbf{u}_J), (\mathbf{a} \vee \mathbf{v}_J)^c| \\
&= |\mathbf{a} \wedge \mathbf{u}_J| + 2 - |\mathbf{a} \vee \mathbf{v}_J| \\
&= 2 - |R_J \oplus \mathbf{a}|, \tag{25}
\end{aligned}
$$

by eqn (22). Thus by eqns (24) and (25), the match criterion is met iff the expanded rectangle $R_J \oplus \mathbf{a}$

satisfies

$$|R_J \oplus \mathbf{a}| \leq 2(1 - \rho). \tag{26}$$

By eqn (26), if vigilance ρ is close to 1, then all R_j's are small. If ρ is close to 0, then some R_j's may grow to fill most of the unit square $[0, 1] \times [0, 1]$.

Suppose now that the match criterion is satisfied. By eqn (23)

$$
\begin{aligned}
\mathbf{w}_J^{(new)} &= \mathbf{I} \wedge \mathbf{w}_J^{(old)} \\
&= (\mathbf{a}, \mathbf{a}^c) \wedge (\mathbf{u}_J^{(old)}, (\mathbf{v}_J^{(old)})^c) \\
&= (\mathbf{a} \wedge \mathbf{u}_J^{(old)}, \mathbf{a}^c \wedge (\mathbf{v}_J^{(old)})^c) \\
&= (\mathbf{a} \wedge \mathbf{u}_J^{(old)}, (\mathbf{a} \vee \mathbf{v}_J^{(old)})^c) \\
&= (\mathbf{u}_J^{(new)}, (\mathbf{v}_J^{(new)})^c). \tag{27}
\end{aligned}
$$

Thus

$$R_J^{(new)} = R_J^{(old)} \oplus \mathbf{a} \tag{28}$$

In particular, no weight changes occur if $\mathbf{a} \in R_J^{(old)}$. In summary, with fast learning, each R_j equals the smallest rectangle that encloses all vectors $\mathbf{a}$ that have chosen category j, under the constraint that $|R_j| \leq 2(1 - \rho)$.

In general, if $\mathbf{a}$ has dimension M, the hyper-rectangle R_j includes the two vertices $\wedge_j\mathbf{a}$ and $\vee_j\mathbf{a}$, where the i-th component of each vector is

$$(\wedge_j\mathbf{a})_i = \min\{a_i : \mathbf{a} \text{ has been coded by category } j\} \tag{29}$$

and

$$(\vee_j\mathbf{a})_i = \max\{a_i : \mathbf{a} \text{ has been coded by category } j\} \tag{30}$$

(Figure 7). The size of R_j is given by

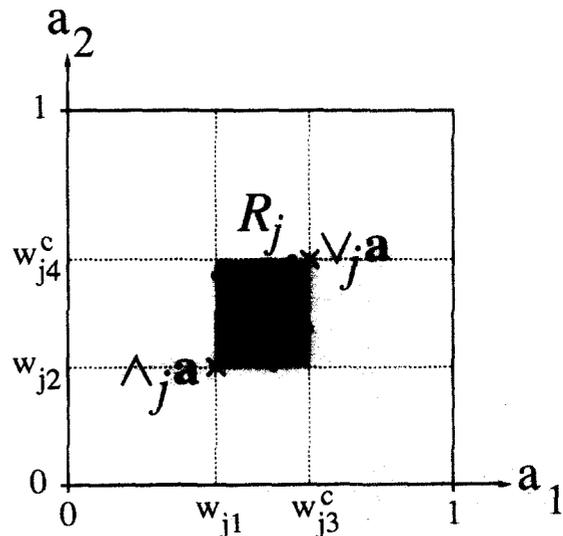$$|R_j| = |\vee_j\mathbf{a} - \wedge_j\mathbf{a}| \tag{31}$$

FIGURE 7. With fast learning and complement coding, the j-th category rectangle R_j includes all those vectors $\mathbf{a}$ in the unit square which have activated category j without reset.

As in eqn (27),

$$\mathbf{w}_j = (\wedge_i \mathbf{a}, (\vee_i \mathbf{a})^c) \tag{32}$$

so

$$|\mathbf{w}_j| = \sum_i (\wedge_i \mathbf{a})_i + \sum_i [1 - (\vee_i \mathbf{a})_i]$$

$$= M - |\vee_i \mathbf{a} - \wedge_i \mathbf{a}|. \tag{33}$$

The size of the hyper-rectangle R_j is thus

$$|R_j| = M - |\mathbf{w}_j|. \tag{34}$$

By eqns (6), (8), and (13),

$$|\mathbf{w}_j| \geq M\rho. \tag{35}$$

By eqns (34) and (35),

$$|R_j| \leq M(1 - \rho). \tag{36}$$

Thus, in the M-dimensional case, high vigilance ($\rho \cong 1$) again leads to small R_j while low vigilance ($\rho \cong 0$) permits large R_j. If j is an uncommitted node, $|\mathbf{w}_j| = 2M$, by eqn (14), and so, $|R_j| \cong -M$, by eqn (34). These observations may be combined into the following theorem.
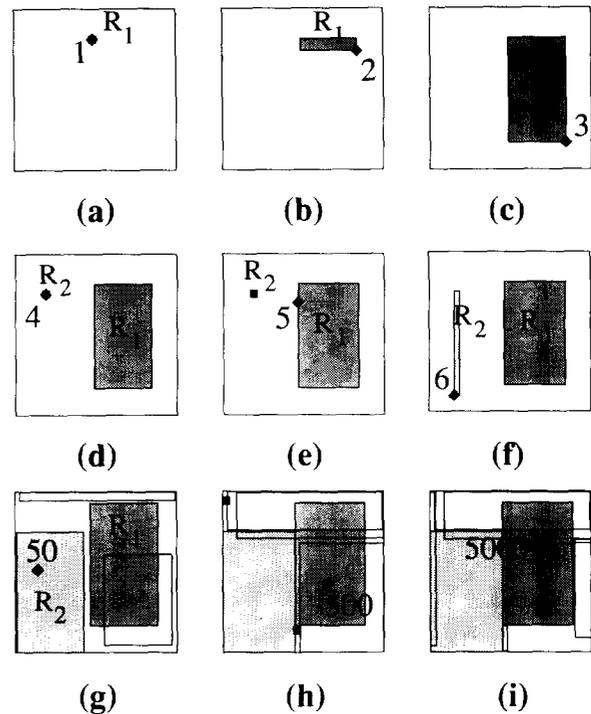
THEOREM (Stable Category Learning):

In response to an arbitrary sequence of analog or binary input vectors, a Fuzzy ART system with complement coding and fast learning forms stable hyper-rectangular categories R_j, which grow during learning to a maximum size $|R_j| \leq M(1 - \rho)$ as $|\mathbf{w}_j|$ monotonically decreases. In the conservative limit, one-pass learning obtains such that no reset or additional learning occurs on subsequent presentations of any input.

Similar properties hold for the fast-learn slow-recode case, except that repeated presentations of an input may be needed before stabilization occurs.

The geometry of the hyper-rectangles R_j resembles part of the Nested Generalized Exemplar (NGE) algorithm (Salzberg, 1990). Both NGE and Fuzzy ARTMAP (Carpenter, Grossberg, Reynolds, & Rosen, 1991) construct hyper-rectangles that represent category weights in a supervised learning paradigm. Both algorithms use the learning law eqn (8) to update weights when an input correctly predicts the output. The two algorithms differ significantly, however, in their response to an incorrect prediction. In particular, NGE has no analogue of the ART vigilance parameter ρ, and its rules for search differ from those of Fuzzy ART. In addition, NGE allows hyper-rectangles to shrink as well as to grow, so the Fuzzy ART stability properties do not obtain.

In the computer simulation summarized in Figure 8, $M = 2$ and vectors $\mathbf{a}^{(1)}$, $\mathbf{a}^{(2)}$, . . . are selected at random from the unit square. Each frame shows the vector $\mathbf{a}^{(t)}$ and the set of rectangles R_j present after learning occurs. The system is run in the fast-learn,

(a)   **(b)**   **(c)**

(d)   **(e)**   **(f)**

(g)   **(h)**   **(i)**

FIGURE 8. Fuzzy ART complement coding simulation with $\alpha \cong 0$, $\beta = 1$, $\rho = .4$, and input vectors $\mathbf{a}^{(t)}$ selected at random from the unit square. Rectangles R_j grow during learning, and new categories are established, until the entire square is covered by 8 rectangles. Categories do not proliferate. A new point rectangle, R_2, is established at $t = 4$, since $R_1 \oplus \mathbf{a}^{(4)}$ is too large to satisfy the match criterion eqn (26).

conservative limit, and $\rho = .4$. When the first category is established, R_1 is just the point $\mathbf{a}^{(1)}$. If $\mathbf{a}^{(t)}$ lies within one or more established R_j, the rectangle chosen is the one that has the smallest size $|R_j|$. In this case, neither reset nor weight change occurs. Each new input that activates category j, but does not lie within its previously established boundaries, expands R_j unless (as in (d)) such an expansion would cause the size of R_j to exceed $2(1 - \rho) = 1.2$. As more and more inputs sample the square, all points of the square are eventually covered by a set of eight rectangles R_j, as illustrated by (g)–(i).

6. FUZZY ART WITHOUT COMPLEMENT CODING

The advantages of complement coding are highlighted by consideration of Fuzzy ART without this preprocessing component. Consider again a fast-learn Fuzzy ART system with $M = 2$. Let the input set consist of two-dimensional vectors $\mathbf{I} = \mathbf{a}$. During learning,

$$\mathbf{w}_j^{(new)} = \mathbf{a} \wedge \mathbf{w}_j^{(old)}. \tag{37}$$

since $\beta = 1$ in eqn (8). Without complement coding,

the monotone decrease in $|\mathbf{w}_J|$ that is implied by eqn (37) can lead to a proliferation of categories, as follows.

Geometry of Fuzzy ART: The choice, match, and learning computations of Fuzzy ART (Figure 3) will now be described geometrically. Without complement coding, these computations can be described in terms of polygonal regions. With complement coding, the analogous regions would be four-dimensional sets.

For a given input $\mathbf{a}$, Fuzzy ART choice is characterized by a nested set of polygons. These polygons are defined by the choice function

$$T_j(\mathbf{a}) = \frac{|\mathbf{a} \wedge \mathbf{w}_j|}{\alpha + |\mathbf{w}_j|}. \tag{38}$$

and thus are called *choice polygons* (Figure 9). As in eqn (5), the category choice is indexed by J, where

$$T_J(\mathbf{a}) = \max\{T_j(\mathbf{a}): j = 1, 2, \ldots, N\}. \tag{39}$$

The choice J can be geometrically interpreted as follows. For each input vector $\mathbf{a}$ and number $T \in [0,$

1], let

$$P_T(\mathbf{a}) = \left\{ \mathbf{w}: \frac{|\mathbf{a} \wedge \mathbf{w}|}{\alpha + |\mathbf{w}|} \geq T \right\}. \tag{40}$$

where $\mathbf{w}$ is a vector in the unit square. For each T, region $P_T(\mathbf{a})$ is a choice polygon with boundary

$$\partial P_T(\mathbf{a}) = \left\{ \mathbf{w}: \frac{|\mathbf{a} \wedge \mathbf{w}|}{\alpha + |\mathbf{w}|} = T \right\}. \tag{41}$$

Progressively smaller values of T induce a family of progressively larger polygons that are nested within each other. In the conservative limit ($\alpha \cong 0$), $P_1(\mathbf{a}) \cong \mathbf{a}$ and all $P_T(\mathbf{a})$ include $\mathbf{a}$ for $0 \leq T \leq 1$.
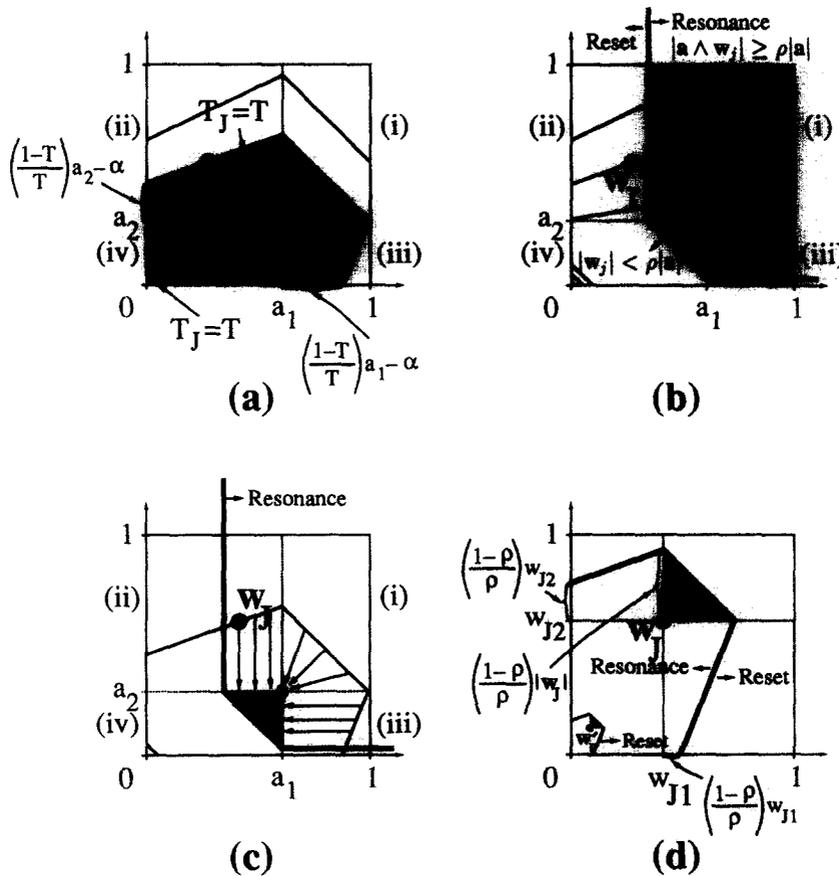
By eqns (38) and (40), an LTM vector

$$\mathbf{w}_j \in P_T(\mathbf{a}) \tag{42}$$

iff

$$T_j(\mathbf{a}) \geq T. \tag{43}$$

Thus, to find a category choice J that maximizes $T_j(\mathbf{a})$, the largest T must be found such that

$$\mathbf{a} \in P_T(\mathbf{a}), \tag{44}$$

FIGURE 9. Geometry of Fuzzy ART without complement coding. Input a divides the unit square into four rectangular regions (i), . . . , (iv). (a) Choice polygons $P_T(a)$ for $0 < T < 1$ and $0 < \alpha \ll 1$. If input a chooses a category J with $T_J = T$, then w_J is a point on $\partial P_T(a)$, but no weight vector w_J lies inside $P_T(a)$; (b) Resonance and reset regions for input a; (c) During learning, w_J is projected onto the shaded triangle; (d) For a given category J, the size of the set of inputs a that will resonate upon choosing w_J shrinks to 0 as $|w_J| \to 0$. If a is in the shaded triangle, w_J is a fuzzy subset of a. There, if a chooses category J, no weight change occurs during learning.

$$w_j \in \partial P_T(\mathbf{a}),\qquad(45)$$

but

$$w_j \notin P_S(\mathbf{a})\qquad(46)$$

for any $j = 1, 2, \ldots, N$ and $S > T$. To accomplish this, T is decreased and the corresponding polygon $P_T(\mathbf{a})$ expands until its boundary $\partial P_T(\mathbf{a})$ intersects the first LTM vector w_j, $j = 1, 2, \ldots, N$, at some value T. Then set $J = j$ and $T_j(\mathbf{a}) = T$.

Resonance and reset regions: Once chosen, node J remains active only if the vigilance criterion is met; namely, by eqn (6),

$$|\mathbf{a} \wedge w_j| \geq \rho|\mathbf{a}|.\qquad(47)$$

The resonance and reset regions can be visualized in terms of the four rectangular regions into which $\mathbf{a}$ divides the unit square (Figure 9b). If w_j is in region (i),

$$|\mathbf{a} \wedge w_j| = |\mathbf{a}|.\qquad(48)$$

Thus, eqn (47) is satisfied given any vigilance value $\rho \in [0, 1]$. On the other hand, if w_j is in region (iv), then

$$|\mathbf{a} \wedge w_j| = |w_j|,\qquad(49)$$
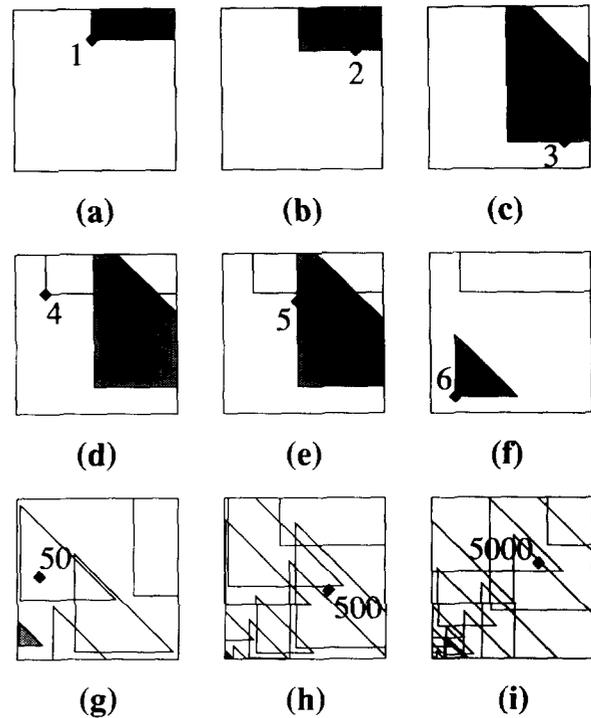
so node J will be reset if

$$|w_j| < \rho|\mathbf{a}|.\qquad(50)$$

The boundary of the reset region in (iv) is thus defined by the straight line $\{\mathbf{w}:|\mathbf{w}| = \rho|\mathbf{a}|\}$, which approaches $\mathbf{a}$ as ρ approaches 1. The fact that the reset boundary is a vertical line in region (ii) and a horizontal line in region (iii) is checked by evaluating $|\mathbf{a} \wedge w_j|$ in these regions. Figure 9b pieces together these reset regions and depicts the complementary resonance region in gray at a vigilance value $\rho < 1$.

Learning: After search, some node J is chosen with w_j in the resonance region. During learning, $\mathbf{a} \wedge w_j$ becomes the new weight vector for category J, by eqn (37). That is, w_j is projected to region (iv); specifically, to the shaded triangle in Figure 9c. Thus, unless w_j already lies in region (iv), w_j is drawn toward the origin during learning. However, as w_j approaches the origin, it leaves the resonance region of most inputs $\mathbf{a}$ (Figure 9b). To satisfy the resonance criterion, future inputs are forced to drag other weight vectors toward the origin, or to choose uncommitted nodes, even though the choice value of these nodes is small. Figure 9d illustrates, for two different weight vectors w_j and w_j', the sets of points $\mathbf{a}$ where resonance will occur if category J or J' is chosen. As shown, this set shrinks to zero as $|w_j|$ approaches 0.

Category proliferation: Figure 10 shows how the properties of Fuzzy ART described above can lead to category proliferation. In the simulation illustrated, the same randomly chosen sequence of inputs

FIGURE 10. Fuzzy ART simulation with $\alpha \cong 0$, $\beta = 1$, and $\rho = .4$. The sequence of input vectors $\mathbf{a}_{(t)}$ is the same as in Figure 8. The resonance triangle is shown for each category, with the triangle for category 1 shaded. Categories proliferate as $|w_j| \to 0$.

$\mathbf{a}^{(t)}$ as in Figure 8 were presented to a Fuzzy ART system. Each frame shows the vector $\mathbf{a}^{(t)}$ and the triangular subset of the resonance region (Figure 9d) for all established categories. As shown, proliferating categories cluster near the origin, where they are rarely chosen for resonance, while new categories are continually created. This problem is solved by complement coding of the input vectors, as was illustrated in Section 5.

7. STABILITY OF CLUSTERING ALGORITHMS

Moore (1989) described a variety of clustering algorithms, some of them classical and others, based on ART 1, that are similar to Fuzzy ART. All use, however, a choice function that includes a dot product or Euclidean distance measure that differs from the choice function T_j in eqn (2). In addition, complement coding is not used. For example, the *Cluster Euclidean* algorithm (Moore, 1989, p. 176) chooses the coded category J whose weight vector w_j is the minimal Euclidean distance $d(w_j, \mathbf{I})$ to $\mathbf{I}$, and which satisfies

$$d(w_j, \mathbf{I}) \leq \theta\qquad(51)$$

If such a J exists, $\mathbf{w}_J$ is updated by

$$\mathbf{w}_J^{(new)} = \beta\mathbf{I} + (1 - \beta)\mathbf{w}_J^{(old)}. \qquad (52)$$

If no such category exists, an uncommitted node J is chosen and $\mathbf{w}_J^{(new)} = \mathbf{I}$, as in the fast commitment option. The *Cluster Unidirectional algorithm* (Moore, 1989, p. 177) is similar except that weights are updated according to eqn (8). Moore pointed out that the *Cluster Euclidean* algorithm is unstable in the sense that weight vectors and category boundaries may cycle endlessly. Moore also showed that the unidirectional weight update rule eqn (8) avoids this type of instability, but introduces the category proliferation problem described in Section 6.

As noted in the Stable Category Learning Theorem, normalization of inputs using complement coding allows Fuzzy ART to overcome the category proliferation problem while retaining the stable coding properties of the weight update rule eqn (8). The strong stability and rapid convergence properties of Fuzzy ART models are due to the direct relationship between the choice function eqn (2), the reset rule eqn (7), and the weight update rule eqn (8). Choice, search, and learning are made computationally consistent by the common use of the vector $\mathbf{I} \wedge \mathbf{w}_j$. This direct relationship enables Fuzzy ART models to be embedded in multilevel Fuzzy ARTMAP systems for supervised learning of categorical maps between m-dimensional and n-dimensional analog vector pairs (Carpenter, Grossberg, Reynolds, & Rosen, 1991).

REFERENCES

Baloch, A. A., & Waxman, A. M. (1991). Visual learning, adaptive expectations, and behavioral conditioning of the mobile robot MAVIN. *Neural Networks*, **4**, 271–302.

Baxter, R. (1991). Error propagation and supervised learning in adaptive resonance networks. *International Joint Conference on Neural Networks (Seattle)*II (pp. 423–429). Piscataway, NJ: IEEE Service Center.

Carpenter, G. A., & Grossberg, S. (1987a). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, **37**, 54–115. Reprinted in G. A. Carpenter & S. Grossberg (Eds.) (1991). *Pattern recognition by self-organizing neural networks* (Chap. 10). Cambridge, MA: MIT Press.

Carpenter, G. A., & Grossberg, S. (1987b). ART 2: Stable self-organization of pattern recognition codes for analog input patterns. *Applied Optics*, **26**, 4919–4930. Reprinted in G. A. Carpenter & S. Grossberg (Eds.) (1991). *Pattern recognition by self-organizing neural networks* (Chap. 12) Cambridge, MA: MIT Press.

Carpenter, G. A., & Grossberg, S. (1990a). ART 3 hierarchical search: Chemical transmitters in self-organizing pattern recognition architectures. *International Joint Conference on Neural Networks (Washington, DC)* II (pp. 30–33). Hillsdale, NJ: Erlbaum Associates.

Carpenter, G. A., & Grossberg, S. (1990b). ART 3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural Networks*, **3**, 129–152. Reprinted in G. A. Carpenter & S. Grossberg (Eds.) (1991).

Pattern recognition by self-organizing neural networks (Chap. 14). Cambridge, MA: MIT Press.

Carpenter, G. A., Grossberg, S., & Reynolds, J. H. (1991a). ARTMAP: A self-organizing neural network architecture for fast supervised learning and pattern recognition. *International Joint Conference on Neural Networks (Seattle)*I (pp. 863–868). Piscataway, NJ: IEEE Service Center.

Carpenter, G. A., Grossberg, S., & Reynolds, J. H. (1991b). ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*, **4**, 565–588. Reprinted in G. A. Carpenter & S. Grossberg (Eds.) (1991). *Pattern recognition by self-organizing neural networks* (Chap. 15). Cambridge, MA: MIT Press.

Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H., & Rosen, D. B. (1991). *Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps*. Technical Report CAS/CNS-91-016. Boston, MA: Boston University.

Carpenter, G. A., Grossberg, S., & Rosen D. B. (1991a). ART 2-A: An adaptive resonance algorithm for rapid category learning and recognition. *Neural Networks*, **4**, 493–504.

Carpenter, G. A., Grossberg, S., & Rosen, D. B. (1991b). Fuzzy ART: An adaptive resonance algorithm for rapid, stable classification of analog patterns (Technical Report CAS/CNS-91-006). *International Joint Conference on Neural Networks (Seattle)*II (pp. 411–416). Piscataway, NJ: IEEE Service Center.

Carpenter, G. A., Grossberg, S., & Rosen, D. B. (1991c). *A neural network realization of Fuzzy ART*. Technical Report CAS/CNS-91-021. Boston, MA: Boston University.

Galindo, P. L., & Michaux, T. (1990). An improved competitive learning algorithm applied to high level speech processing. *International Joint Conference on Neural Networks (Washington, DC)*. I (pp. 142–146). Hillsdale, NJ: Erlbaum Associates.

Gjerdingen, R. O. (1990). Categorization of musical patterns by self-organizing neuronlike networks. *Music Perception*, **7**, 339–370.

Gochin, P. (1990). Pattern recognition in primate temporal cortex: But is it ART? *International Joint Conference on Neural Networks (Washington, DC)*. I (pp. 77–80). Hillsdale, NJ: Erlbaum Associates.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics*, **23**, 187–202. Reprinted in G. A. Carpenter & S. Grossberg (Eds.) (1991). *Pattern recognition by self-organizing neural networks* (Chap. 9). Cambridge, MA: MIT Press.

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, **1**, 1–51. Reprinted in S. Grossberg (Ed.) (1982). *Studies of mind and brain: Neural principles of learning, perception, development, cognition, and motor control* (Chap. 1). Boston, MA: Reidel Press.

Grossberg, S. (1982). Processing of expected and unexpected events during conditioning and attention: A psychophysiological theory. *Psychological Review*, **89**, 529–572. Reprinted in S. Grossberg (Ed.) (1987). *The adaptive brain I* (Chap. 3). Amsterdam: North-Holland.

Harvey, R. L., DiCaprio, P. N., Heinemann, K. G., Silverman, M. L., & Dugan, J. M. (1990). A neural architecture for visual recognition of general stationary objects by machines. *International Joint Conference on Neural Networks (San Diego)*, II (pp. 937–942). Piscdataway, NJ: IEEE Service Center.

Hecht-Nielsen, R. (1990). Spatiotemporal pattern segmentation by expectation feedback. *International Joint Conference on Neural Networks (Washington, DC)*, II (p. 40). Hillsdale, NJ: Erlbaum Associates.

Johnson, K. (1990). Characteristics of neural population codes in hierarchical, self-organizing vision machines. *International Joint*

Conference on Neural Networks (*Washington, DC*), **II** (pp. 35–39). Hillsdale, NJ: Erlbaum Associates.

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

Kosko, B. (1987a). Adaptive bidirectional associative memories. *Applied Optics*, **26**, 4947–4960. Reprinted in G. A. Carpenter & S. Grossberg (Eds.) (1991). *Pattern recognition by self-organizing neural networks* (Chap. 13). Cambridge, MA: MIT Press.

Kosko, B. (1987b). Bidirectional associative memories. *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-17**.

Kosko, B. (1987c). Competitive bidirectional associative memories. In M. Caudill & C. Butler (Eds.), *IEEE International Conference on Neural Networks*. Piscataway, NJ: IEEE Service Center, **2**, 759–766.

Kumar, S. S., & Guez, A. (1989). A neural network approach to target recognition. *International Joint Conference on Neural Networks* (*Washington, DC*), **II** (p. 573). San Diego, CA: IEEE.

Kumar, S. S., & Guez, A. (1991). ART based adaptive pole placement for neurocontrollers. *Neural Networks*, **4**, 319–335.

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

Lang, K. J., & Witbrock, M. J. (1989). Learning to tell two spirals apart. In D. Touretzky, G. Hinton, & T. Sejnowski (Eds.), *Proceedings of the 1988 Connectionist Models Summer School*. (pp. 52–59). San Mateo, CA: Morgan Kaufmann Publishers.

Levine, D. S., & Penz, P. A. (1990). ART 1.5—A simplified adaptive resonance network for classifying low-dimensional analog data. *International Joint Conference on Neural Networks* (*Washington, DC*), **II** (pp. 639–642). Hillsdale, NJ: Erlbaum Associates.

Li, D., & Wee, W. G. (1990). A new neocognitron structure modified by ART and back propagation. *International Joint Conference on Neural Networks* (*Washington, DC*), **I** (pp. 420–423). Hillsdale, NJ: Erlbaum Associates.

Liao, F.-Y., Liang, C.-C., & Lin, W.-C. (1990). ART 1 neural network for reducing search space in 3-D object recognition using multiple views. *International Joint Conference on Neural Networks* (*San Diego*), **II** (pp. 931–936). Piscataway, NJ: IEEE Service Center.

Mekkaoui, A., & Jespers, P. (1990). An optimal self-organizing pattern classifier. *International Joint Conference on Neural Networks* (*Washington, DC*), **I** (pp. 447–450). Hillsdale, NJ: Erlbaum Associates.

Michalson, W. R., & Heldt, P. (1990). A hybrid architecture for the ART 2 neural model. *International Joint Conference on*

Neural Networks (*Washington, DC*), **II**, pp. 167–170. Hillsdale, NJ: Erlbaum Associates.

Moore, B. (1989). ART 1 and pattern clustering. In D. Touretzky, G. Hinton, & T. Sejnowski (Eds.), *Proceedings of the 1988 Connectionist Models Summer School* (pp. 174–1985). San Mateo, CA: Morgan Kaufmann Publishers.

Nigrin, A. L. (1990). Real-time classification of temporal sequences with an adaptive resonance circuit. *International Joint Conference on Neural Networks* (*Washington, DC*), **I**, (pp. 525–528). Hillsdale, NJ: Erlbaum Associates.

Rajapakse, J. C., Jakubowicz, O. G., & Acharya, R. S. (1990). A real time ART-1 based vision system for distortion invariant recognition and autoassociation. *International Joint Conference on Neural Networks* (*Washington, DC*), **II** (pp. 298–301). Hillsdale, NJ: Erlbaum Associates.

Ryan, T. W. (1988). The resonance correlation network. *IEEE International Conference on Neural Networks*. Piscataway, NJ: IEEE Service Center, **1**, 673–680.

Salzberg, S. L. (1990). *Learning with nested generalized exemplars*. Boston: Kluwer Academic Publishers.

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

Seibert, M., & Waxman, A. M. (1990b). Learning aspect graph representations of 3D objects in a neural network. *International Joint Conference on Neural Networks* (*Washington, DC*), **II** (pp. 233–236). Hillsdale, NJ: Erlbaum Associates.

Simpson, P. K. (1990). Neural networks for sonar signal processing. In A. J. Maren, C. T. Harston, & R. M. Pap (Eds.), *Handbook of neural computing applications* (pp. 319–335). San Diego: Academic Press.

Weingard, F. S. (1990). Self-organizing analog fields (SOAF). *International Joint Conference on Neural Networks* (*Washington, DC*), **II**, (p. 34). Hillsdale, NJ: Erlbaum Associates.

Wilson, C. L., Wilkinson, R. A., & Ganis, M.D. (1990). Self-organizing neural network character recognition on a massively parallel computer. *International Joint Conference on Neural Networks* (*San Diego*), **II** (pp. 325–329). Piscataway, NJ: IEEE Service Center.

Winter, C. L. (1989). An adaptive network that learns sequences of transitions. In D. S. Touretzky (Ed.). *Advances in neural information processing systems 1* (pp. 653–660). San Mateo, CA: Morgan Kaufmann.

Winter, C. L., Ryan, T. W., & Turner, C. J. (1987). TIN: A trainable inference network. In M. Caudill & C. Butler (Eds.), *IEEE International Conference on Neural Networks*, **2**, 777–785. Piscataway, NJ: IEEE Service Center.

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