Pergamon

Neural Networks, Vol. 8, No. 2, pp. 203–213, 1995

0893-6080/95 $9.50 + .00

0893-6080(94)00073-5

CONTRIBUTED ARTICLE

Fuzzy ART Properties

JUXIN HUANG,[1] MICHAEL GEORGIOPOULOS,[1] AND GREGORY L. HEILEMAN[2]

[1]University of Central Florida and [2]University of New Mexico

(*Received* 10 *September* 1993; *revised and accepted* 1 *July* 1994)

Abstract—*This paper presents some important properties of the Fuzzy ART neural network algorithm introduced by Carpenter, Grossberg, and Rosen. The properties described in the paper are distinguished into a number of categories. These include template, access, and reset properties, as well as properties related to the number of list presentations needed for weight stabilization. These properties provide numerous insights as to how Fuzzy ART operates. Furthermore, the effects of the Fuzzy ART parameters* α *and* ρ *on the functionality of the algorithm are clearly illustrated.*

Keywords—Neural network, Pattern recognition, Clustering, Learning, Adaptive resonance theory, Fuzzy set theory, Fuzzy ART.

1. INTRODUCTION

A neural network model that can be used to cluster arbitrary binary or analog data was derived by Carpenter, Grossberg, and Rosen (1991b). This model is termed Fuzzy ART in reference to the adaptive resonance theory introduced by Grossberg (1976). One of the major reasons for the development of Fuzzy ART was to remedy the inability of ART1, as well as Predictive ART architectures based on ART1 modules, to classify analog data (see, for example, Carpenter, Grossberg, & Reynolds, 1991a). Although the learning properties of ART1 and Predictive ART architectures based on ART1 modules are well understood (see Carpenter and Grossberg, 1987; Georgiopoulos, Heileman, & Huang, 1991, 1992, 1994; Moore, 1989), the same cannot be said for the Fuzzy ART algorithm.

In this paper we present useful properties of the Fuzzy ART algorithm that facilitate the understanding of its operation. For clarity purposes we split the properties into four different categories: template properties (Section 3), access properties (Section 4), reset properties (Section 5), and properties related to the number of list presentations needed for the weight stabilization (Section 6). These properties are presented in the form of theorems, propositions, and corollaries. Some of the properties discussed in this paper involve the size/similarities of templates created in Fuzzy ART, as well as the number of list presentations required to learn an arbitrary list of binary input patterns repeatedly presented to Fuzzy ART. For most of the Fuzzy ART properties mentioned in this manuscript, the effects of parameters α and ρ are clearly illustrated.

2. PRELIMINARIES—NOTATION

The Fuzzy ART algorithm is described in detail by Carpenter et al. (1991b). In this section we only provide information that is necessary to understand the results developed here. The Fuzzy ART architecture consists of two layers of nodes, designated F_1 and F_2. Inputs are presented at the F_1 layer of Fuzzy ART. If $\mathbf{a} = (a_1, \ldots, a_M)$ denotes a vector, with each of its components in the interval $[0, 1]$, then the input to the F_1 layer of Fuzzy ART is a vector $\mathbf{I}$ such that

$$\mathbf{I} = (\mathbf{a}, \mathbf{a}^c) = (a_1, \ldots, a_M, a_1^c, \ldots, a_M^c) \tag{1}$$

where

$$a_i^c = 1 - a_i; \quad 1 \le i \le M. \tag{2}$$

This type of transformation, called *complement coding,* is necessary for the successful operation of Fuzzy ART, especially when the input vector $\mathbf{I}$ is of analog nature (for more details see Carpenter et al., (1991b). The F_2 layer in Fuzzy ART is usually referred to as the *cate-*

Acknowledgements: This research was supported in part by a grant from the Florida High Technology and Industry Council, in part by a grant from the Division of Sponsored Research at the University of Central Florida, and in part by a grant from Boeing Computer Services under contract W-300445.

Requests for reprints should be sent to Michael Georgiopoulos, Department of Electrical and Computer Engineering, University of Central Florida, Orlando, FL 32816.

gory representation layer because its nodes denote the categories to which the input patterns belong.

The F_1 layer has $2M$ nodes, and the F_2 layer has N nodes. We use the index i to designate nodes in the F_1 layer, and the index j to designate nodes in the F_2 layer. There are *bottom-up weight* connections emanating from the nodes in the F_1 layer and converging to the nodes in the F_2 layer. Similarly, there are *top-down weight* connections emanating from the nodes in the F_2 layer and converging to the nodes in the F_1 layer. The bottom-up weights converging to a node in the F_2 layer can be completely defined by the top-down weights emanating from this F_2 node. Hence, only the set of top-down weights need be defined. In particular, we let $\mathbf{W}_j = (W_{j1}, \ldots, W_{j(2M)})$ designate the vector of top-down weights emanating from node j in the F_2 layer. The initial value of all top-down weights is 1. When an input pattern $\mathbf{I}$ is applied at the F_1 layer, it produces an input $T_j(\mathbf{I})$ at node j in the F_2 layer. For notational simplicity, $T_j(\mathbf{I})$ is written as T_j in the rest of the paper. The input T_j is given by the equation

$$T_j = \frac{|\mathbf{I} \wedge \mathbf{W}_j|}{\alpha + |\mathbf{W}_j|}. \tag{3}$$

The function T_j is referred to as the *choice function.* In the above equation, α is a positive real-valued number called the *choice parameter,* $\mathbf{I} \wedge \mathbf{W}_j$ is a vector for which the ith component ($1 \le i \le 2M$) is equal to the minimum of $\mathbf{I}_i$ and $\mathbf{W}_{ji}$ (the operator $\wedge$ is referred to as the MIN operator and $\mathbf{I} \wedge \mathbf{W}_j$ is called the MIN of $\mathbf{I}$ and $\mathbf{W}_j$), and $|\cdot|$ designates the size of a vector, where the size of a vector is defined to be the sum of its components. The node in the F_2 layer that receives the largest input T_j will be chosen to represent the input pattern $\mathbf{I}$. Assume that node J in the F_2 layer receives the largest such input. The appropriateness of node J to represent the input pattern $\mathbf{I}$ is based on the *vigilance criterion.* This criterion is satisfied if

$$\frac{|\mathbf{I} \wedge \mathbf{W}_J|}{|\mathbf{I}|} \ge \rho \tag{4}$$

where ρ, the *vigilance parameter,* may take values in the interval $[0, 1]$. If node J fails the vigilance criterion, it is reset and a search for another node in the F_2 layer to represent the input pattern starts. The reset of node J is accomplished by the *orienting subsystem* in Fuzzy ART (for more details about the orienting subsystem see Carpenter and Grossberg (1987). If node J passes the vigilance criterion, learning starts and the top-down weight vector $\mathbf{W}_J$ is updated as

$$\mathbf{W}_J = (1 - \beta)\mathbf{W}_J + \beta(\mathbf{I} \wedge \mathbf{W}_J) \tag{5}$$

where β is a Fuzzy ART parameter, called *learning rate,* which may assume values in the interval $(0, 1]$. If $\beta = 1$ the learning is called *fast learning,* and if $0 < \beta < 1$ it is called *slow learning.* If a node has previously coded an input pattern, then it is said to be *committed;* otherwise, it is said to be *uncommitted.* We say that a node J in the F_2 layer has *coded* an input pattern $\mathbf{I}$ if, after the presentation of $\mathbf{I}$ at the F_1 layer, J is eventually chosen and not reset, and learning of pattern $\mathbf{I}$ by node J ensues. A special type of slow learning, called *fast-commit slow-recode,* is one in which fast learning occurs (i.e., $\beta = 1$) when the chosen node in the F_2 layer is uncommitted, and slow learning occurs (i.e., $0 < \beta < 1$) when the chosen node is committed.

The vector of top-down weights from a node in the F_2 layer is called a *template.* A template corresponding to a committed node is called a *committed template,* and a template corresponding to an uncommitted node is called *uncommitted template.* An uncommitted template has all of its components equal to 1. In this paper, when we refer to the word "template" we mean committed template. Consider now an input pattern $\mathbf{I}$ presented to the Fuzzy ART architecture, and an arbitrary template denoted by $\mathbf{W}_j$. A component of an input pattern $\mathbf{I}$ is indexed by i if it affects node i in the F_1 layer, and the corresponding component of template $\mathbf{W}_j$ is $\mathbf{W}_{ji}$. We can identify three types of templates with respect to an input pattern $\mathbf{I}$: *subset templates, superset templates,* and *mixed templates.* A template $\mathbf{W}_j$ is a subset template of an input pattern $\mathbf{I}$ if each one of the $\mathbf{W}_j$ components is smaller than or equal to its corresponding components in $\mathbf{I}$. A template $\mathbf{W}_j$ is a superset template of an input pattern $\mathbf{I}$ if each one of the $\mathbf{W}_j$ components is larger than its corresponding components in $\mathbf{I}$. A template is a mixed template if some of the $\mathbf{W}_j$ components are smaller than or equal to its corresponding components in $\mathbf{I}$, and the rest of the $\mathbf{W}_j$ components are larger than its corresponding components in $\mathbf{I}$. With reference to an input $\mathbf{I}$, we designate a committed node in the F_2 layer as subset, superset, or mixed depending on whether its corresponding template is a subset, superset, or mixed template with respect to $\mathbf{I}$. It is worth noting that in the case of fast-learning or fast-commit slow-recode learning we can only define subset, mixed, and uncommitted templates. Due to the complement coding nature of the input patterns, superset templates cannot be created in a Fuzzy ART architecture with fast-learning or fast-commit slow-recode learning.

The following assumptions will be used at various points in the remainder of the paper to guarantee the validity of specific results: (i) fast learning, (ii) fast-commit slow-recode learning, (iii) binary input patterns, (iv) repeated or cyclic presentations of an input list of patterns, and (v) a sufficient number of nodes in the F_2 layer. Assumptions (i) and (ii) have been discussed previously. Assumption (iii) implies that the input patterns presented to the Fuzzy ART architecture have binary (0 or 1) components. Most of the results in this paper are valid for analog or binary data. Only the properties presented in Propositions 6.1–6.4 of Section 6 require assumption (iii). Assumption (iv) corresponds to the case where we have a list of input patterns, designated $\mathbf{I}^1, \mathbf{I}^2, \ldots, \mathbf{I}^P$, which is presented either repeatedly or cyclically to Fuzzy ART. In *repeated presentations* of the list, the order of the pattern presentation within the list is of no consequence, but in *cyclic presentations* of the list the patterns are always presented in the same order within each list (e.g., $\mathbf{I}^1, \mathbf{I}^2, \ldots, \mathbf{I}^P, \mathbf{I}^1, \mathbf{I}^2, \ldots, \mathbf{I}^P$, and so on). Assumption (v) means that every time an input pattern $\mathbf{I}$ is presented to Fuzzy ART there is at least one uncommitted node available at the F_2 layer; this assumption is sufficient to guarantee that an appropriate node in the F_2 layer will always be found to represent the input pattern.

In the case where a list of input patterns is repeatedly presented to the Fuzzy ART architecture, it is reasonable to ask how many list presentations does Fuzzy ART need to learn the input list; or equivalently, how many list presentations are needed for the weights to stabilize. We say that the weights in the Fuzzy ART architecture are stabilized by the end of the nth list presentation, if in subsequent presentations of the list (i.e., list presentations $\geq n + 1$) weights cannot be modified. Under the aforementioned scenario (i.e., repeated presentations of a list of input patterns) when weights are stabilized we also say that *learning* (of the list) *is complete.* After stabilization of the weights occurs, each pattern from the input list will have *direct access* to a node in the F_2 layer (assuming there are enough nodes in the F_2 layer). We say that a pattern $\mathbf{I}$ has direct access to a node j in the F_2 layer if immediately after the presentation of $\mathbf{I}$ at the F_1 layer, j is chosen first and no reset of j occurs.

3. TEMPLATE PROPERTIES

In this section we discuss properties related to the templates created in a Fuzzy ART architecture. In particular, Theorem 3.1 states that the templates in Fuzzy ART are distinct, whereas Proposition 3.1 deals with the smallest possible size of the templates. Corollary 3.1 shows how the range of α is related to the smallest possible template size. Proposition 3.2 focuses on the similarity among the templates in Fuzzy ART. Finally, Corollary 3.2 shows how this similarity is affected by the ranges of the α and ρ parameters.

THEOREM 3.1. *In a Fuzzy ART architecture, all the templates are distinct.*

Proof. (By contradiction) Assume that Fuzzy ART creates two templates $\mathbf{W}_1$ and $\mathbf{W}_2$ that are equal (i.e., $\mathbf{W}_{1i} = \mathbf{W}_{2i}$ for all i). Assume also that $\mathbf{W}_1$ is created first, and template $\mathbf{W}_2$ is created by pattern $\mathbf{I}$ as it is coded by node 2 with template $\hat{\mathbf{W}}_2$. Then,

$$\mathbf{W}_2 = \beta(\mathbf{I} \wedge \hat{\mathbf{W}}_2) + (1 - \beta)\hat{\mathbf{W}}_2 = \mathbf{W}_1. \tag{6}$$

Equation (6) implies that each component of $\mathbf{W}_2$ is given by

$$\mathbf{W}_{2i} = \beta \min(\mathbf{I}_i, \hat{\mathbf{W}}_{2i}) + (1 - \beta)\hat{\mathbf{W}}_{2i}. \tag{7}$$

Based on eqn (7) it is not difficult to show [by distinguishing cases such as (i) $\hat{\mathbf{W}}_{2i} < \mathbf{I}_i$, (ii) $\hat{\mathbf{W}}_{2i} > \mathbf{I}_i$, and (iii) $\hat{\mathbf{W}}_{2i} = \mathbf{I}_i$] that

$$\mathbf{I} \wedge \mathbf{W}_1 = \mathbf{I} \wedge \hat{\mathbf{W}}_2. \tag{8}$$

Let us now identify cases under which pattern $\mathbf{I}$ will be coded by node 2 with template $\hat{\mathbf{W}}_2$ in the presence of node 1 with template $\mathbf{W}_1$.

Case 1. $\mathbf{I}$ chooses node 2 before it chooses node 1, and node 2 is not reset. One of the conditions for Case 1 to happen is

$$\frac{|\mathbf{I} \wedge \hat{\mathbf{W}}_2|}{\alpha + |\hat{\mathbf{W}}_2|} \geq \frac{|\mathbf{I} \wedge \mathbf{W}_1|}{\alpha + |\mathbf{W}_1|}. \tag{9}$$

Equations (8) and (9) imply that $|\mathbf{W}_1| \geq |\hat{\mathbf{W}}_2|$. This is a contradiction because eqn (6) implies $|\mathbf{W}_1| \leq |\hat{\mathbf{W}}_2|$, and the assumption that $\mathbf{W}_1 \neq \hat{\mathbf{W}}_2$ means that $|\mathbf{W}_1| < |\hat{\mathbf{W}}_2|$. Therefore, eqn (9) can never be true, or equivalently, Case 1 can never happen.

Case 2. $\mathbf{I}$ chooses node 1 before it chooses node 2, reset of node 1 occurs, $\mathbf{I}$ eventually chooses node 2, and node 2 is not reset. One of the conditions for Case 2 to happen is

$$\frac{|\mathbf{I} \wedge \mathbf{W}_1|}{|\mathbf{I}|} < \rho. \tag{10}$$

Combining eqns (8) and (10) we get

$$\frac{|\mathbf{I} \wedge \hat{\mathbf{W}}_2|}{|\mathbf{I}|} = \frac{|\mathbf{I} \wedge \mathbf{W}_1|}{|\mathbf{I}|} < \rho, \tag{11}$$

which implies that if node 1 is reset node 2 will also be reset, or equivalently, that Case 2 can never happen.

Because Cases 1 and 2 are the only two possible scenarios under which pattern $\mathbf{I}$ will create a template $\mathbf{W}_2$ equal to the already existing template $\mathbf{W}_1$, we conclude that templates that are equal can never be created in Fuzzy ART. ■

REMARKS. This theorem shows one of the good properties of Fuzzy ART: that templates can never be the same. It applies to binary or analog patterns, fast or slow learning, and for any values of the α and ρ parameters. In addition, because $|\mathbf{I}| = M$ is not used in our proof, the validity of Theorem 3.1 is not dependent on the complement-coded nature of the input patterns.

PROPOSITION 3.1. *In a Fuzzy ART architecture with a sufficient number of nodes in the F_2 layer, the size of a template is larger than $\alpha M/(\alpha + M)$. For the binary patterns and fast learning case, the size of a template is larger than or equal to $(\alpha + 1)M/(\alpha + M)$.*

Proof. Consider an input pattern $\mathbf{I}$ that creates a template of size $|\mathbf{W}|$ by destroying a template of size greater than $|\mathbf{W}|$. In particular, assume that $\mathbf{I}$ is coded by node j with template $\mathbf{W}_j^{\text{old}}$ of size $|\mathbf{W}| + \delta(\delta > 0)$, and creates a template $\mathbf{W}_j^{\text{new}}$ of size $|\mathbf{W}|$. Obviously,

$$\mathbf{W}_j^{\text{new}} = (1 - \beta)\mathbf{W}_j^{\text{old}} + \beta(\mathbf{I} \wedge \mathbf{W}_j^{\text{old}}). \tag{12}$$

One of the conditions for node j to code $\mathbf{I}$ is

$$\frac{|\mathbf{I} \wedge \mathbf{W}_j^{\text{old}}|}{\alpha + |\mathbf{W}_j^{\text{old}}|} \geq \frac{M}{\alpha + 2M}. \tag{13}$$

It is not difficult to show, using eqn (12), that $|\mathbf{I} \wedge \mathbf{W}_j^{\text{old}}| \leq |\mathbf{W}_j^{\text{new}}|$, or equivalently, that $|\mathbf{I} \wedge \mathbf{W}_j^{\text{old}}| = |\mathbf{W}| - \varepsilon(\varepsilon \geq 0)$. Consequently, another way of writing eqn (13) is

$$\frac{|\mathbf{W}| - \varepsilon}{\alpha + |\mathbf{W}| + \delta} \geq \frac{M}{\alpha + 2M}, \tag{14}$$

which gives us

$$|\mathbf{W}| \geq \frac{(\alpha + \delta + \varepsilon)M}{(\alpha + M)} + \varepsilon. \tag{15}$$

Knowing that δ and ε can be arbitrarily small and $\delta \neq 0$, we conclude

$$|\mathbf{W}| > \frac{\alpha M}{(\alpha + M)}. \tag{16}$$

The second part of the proposition is obvious if $|\mathbf{W}| = M$. For $|\mathbf{W}| \leq M - 1$ consider eqn (15) for fast learning ($\varepsilon = 0$) and binary inputs, which immediately implies that the smallest δ value is equal to 1. As a result, eqn (15) becomes

$$|\mathbf{W}| \geq (\alpha + 1)M/(\alpha + M). \tag{17}$$

This proves Proposition 3.1. ■

COROLLARY 3.1. *In a Fuzzy ART architecture with binary patterns, fast learning, and a sufficient number of nodes in the F_2 layer, if $\alpha > M(M - L - 1)/L$, then the smallest possible template size is equal to $M - L + 1$ and there are at most L different template sizes, where L is an integer in the interval $[1, M - 1]$.*

Proof. Corollary 3.1 is a direct consequence of Proposition 3.1. ■

REMARKS. Proposition 3.1 and Corollary 3.1 are valid independent of the value of the vigilance parameter ρ. The smallest possible template size increases as α increases. Furthermore, it is worth observing that under the Fuzzy ART conditions stated in Corollary 3.1, size-1 templates cannot be created because $1/(\alpha + 2) < M/(\alpha + 2M)$.

PROPOSITION 3.2. *In a Fuzzy ART architecture with either fast-commit slow-recode or fast learning, and a sufficient number of nodes in the F_2 layer, the size of the MIN of any two templates (the number of common 1s between any two templates in the binary input patterns and fast learning case) is smaller than*

$$\max\left\{\rho M, M\frac{\alpha + M}{\alpha + 2M}\right\}. \tag{18}$$

Proof. Let us consider the templates emanating from nodes 1 and 2 in the F_2 layer of Fuzzy ART. Assume that the F_2 nodes become committed in the order 1, 2, ... Assume also that node 2 becomes committed during the presentation of an input pattern $\mathbf{I}$, and at this time the template corresponding to node 1 is equal to $\mathbf{W}_1$. Because $\mathbf{I}$ is coded by node 2 in the presence of the template $\mathbf{W}_1$, one of the following two conditions must be valid:

$$\frac{|\mathbf{I} \wedge \mathbf{W}_1|}{\alpha + |\mathbf{W}_1|} < \frac{M}{\alpha + 2M} \tag{19}$$

or

$$\frac{|\mathbf{I} \wedge \mathbf{W}_1|}{M} < \rho. \tag{20}$$

The first condition gives

$$|\mathbf{I} \wedge \mathbf{W}_1| < \frac{M}{\alpha + 2M}(\alpha + |\mathbf{W}_1|) \leq \frac{M}{\alpha + 2M}(\alpha + M). \tag{21}$$

And the second condition gives

$$|\mathbf{I} \wedge \mathbf{W}_1| < \rho M. \tag{22}$$

Combining eqns (21) and (22) we have

$$|\mathbf{I} \wedge \mathbf{W}_1| < \max\left\{\rho M, M\frac{\alpha + M}{\alpha + 2M}\right\}. \tag{23}$$

At the time $\mathbf{I}$ is initially coded by node 2 $\mathbf{W}_2 = \mathbf{I}$, hence $|\mathbf{W}_1 \wedge \mathbf{W}_2| = |\mathbf{I} \wedge \mathbf{W}_1|$. As a result, initially $|\mathbf{W}_1 \wedge \mathbf{W}_2|$ is smaller than the right-hand side of eqn (23). Obviously, as learning progresses, $\mathbf{W}_1$ and $\mathbf{W}_2$ will either shrink or stay the same, and $|\mathbf{W}_1 \wedge \mathbf{W}_2|$ cannot increase. Consequently, the size of the MIN of the templates emanating from nodes 1 and 2 will always be smaller than the maximum of ρM and $M[(\alpha + M)/(\alpha + 2M)]$. ■

COROLLARY 3.2. *In a Fuzzy ART architecture with either fast-commit slow-recode or fast learning, and a sufficient number of nodes in the F_2 layer, if $\alpha \leq (M - 2L)M/L$ and $\rho \leq 1 - L/M$, then $|\mathbf{W}_1 \wedge \mathbf{W}_2| < M - L$, where $0 \leq L < M/2$.*

Proof. Corollary 3.2 is a direct result of Proposition 3.2. ■

The consequences of Corollaries 3.1 and 3.2 are depicted in Tables 1 and 2.

TABLE 1
Consequences of Corollary 3.1 for $M = 10$

Range of α	Max. Number of Template Sizes	Smallest Template Size
(0, 1.25]	9	2
(1.25, 20/7]	8	3
(20/7, 5]	7	4
(5, 8]	6	5
(8, 12.5]	5	6
(12.5, 20]	4	7
(20, 35]	3	8
(35, 80]	2	9
(80, ∞)	1	10

4. ACCESS PROPERTIES

In this section we present properties related to what type of nodes in the F_2 layer will be chosen during a pattern's presentation. In particular, Proposition 4.1 discusses the order of search among the nodes in the F_2 layer during a pattern's presentation. Theorem 4.1 states that with fast learning, uncommitted nodes in the F_2 layer will not be chosen after the first presentation of a list of input patterns. Theorem 4.2 states that a pattern will always directly access a node with a template equal to the pattern. Finally, Proposition 4.2 verifies that under certain conditions, after learning of an input list of patterns is complete, there may exist committed nodes in the F_2 layer that are not directly accessed by any pattern from the input list.

In Fuzzy ART the search order among the nodes in the F_2 layer depends on the choice parameter α. If α is small, a pattern tends to choose a node with the largest ratio $|\mathbf{I} \wedge \mathbf{W}_j|/|\mathbf{W}_j|$, regardless of the size of $|\mathbf{I} \wedge \mathbf{W}_j|$. In this case, subset nodes always have priority over other nodes. If α is large, the size $|\mathbf{I} \wedge \mathbf{W}_j|$ plays a more important role in the choice of a node in the F_2 layer. For any α, Fuzzy ART follows the rules stated in Proposition 4.1.

PROPOSITION 4.1. *In a Fuzzy ART architecture, when an input pattern* $\mathbf{I}$ *is presented at the* F_1 *layer, a node in the* F_2 *layer is chosen according to the following rules:*

(*a*) *A subset node (if there is one) will be chosen over an uncommitted node.*
(*b*) *Among all the subset nodes, the node with the largest template will be chosen first.*
(*c*) *If a mixed node* j *with template* $\mathbf{W}_j$ *is accessed prior to a subset node* J *with template* $\mathbf{W}_J$, *then* $|\mathbf{I} \wedge \mathbf{W}_j| > |\mathbf{W}_J|$ *must hold.*
(*d*) *If there are no subset nodes, and for every mixed node* j: $|\mathbf{I} \wedge \mathbf{W}_j|/|\mathbf{W}_j| \le 0.5$, *then an uncommitted node will be chosen over any mixed node.*

Proof. (a) Assume that $\mathbf{W}_J$ is a subset template for the input pattern $\mathbf{I}$. If pattern $\mathbf{I}$ accesses an uncommitted node over the subset node J, then

$$\frac{|\mathbf{W}_J|}{\alpha + |\mathbf{W}_J|} < \frac{M}{\alpha + 2M}, \tag{24}$$

which gives

$$|\mathbf{W}_J| < \frac{\alpha M}{\alpha + M}. \tag{25}$$

We have proved (in Proposition 3.1) that the smallest possible template size is larger than $[(\alpha M)/(\alpha + M)]$. Therefore, eqn (25) will never hold, and consequently pattern $\mathbf{I}$ will not access an uncommitted node if there is at least one subset node.

(b) This is a direct result of the way the choice function is defined in Fuzzy ART [see eqn (3)].

(c) This is a direct result of the fact that $T_j \ge T_J$.

(d) The sufficient condition for an uncommitted node to be chosen over a mixed node j is

$$\frac{|\mathbf{I} \wedge \mathbf{W}_j|}{\alpha + |\mathbf{W}_j|} < \frac{M}{\alpha + 2M}, \tag{26}$$

which can be written as

$$\alpha(M - |\mathbf{I} \wedge \mathbf{W}_j|) + M(|\mathbf{W}_j| - 2|\mathbf{I} \wedge \mathbf{W}_j|) > 0. \tag{27}$$

Because $\alpha > 0$ and $|\mathbf{I} \wedge \mathbf{W}_j| < M$, if $|\mathbf{I} \wedge \mathbf{W}_j|/|\mathbf{W}_j| \le 0.5$, eqn (27) will hold. Therefore, an uncommitted node will be chosen over the mixed node j. ■

THEOREM 4.1. *In a Fuzzy ART architecture with fast learning and repeated presentations of a list of input patterns, no uncommitted node will be chosen after the first list presentation. As a result, the total number of committed nodes (or templates) cannot exceed the total number of patterns in the input list.*

Proof. Consider a pattern $\mathbf{I}$ from the input list during list presentation x ($x \ge 2$). We know that after the first list presentation there is at least one subset node for the input pattern $\mathbf{I}$. In list presentation x ($x \ge 2$), according to Proposition 4.1, pattern $\mathbf{I}$ will either choose node J with the largest subset template $\mathbf{W}_J$ or it will choose a node j with a mixed template $\mathbf{W}_j$. Assume that node J is chosen first. Let us also assume that input pattern $\hat{\mathbf{I}}$ is the last pattern prior to $\mathbf{I}$'s presentation that modified the template of node J to its current form (i.e., $\mathbf{W}_J$). Obviously,

TABLE 2
Consequences of Corollary 3.2 for $M = 10$

Range of α	Range of ρ	Size of the MIN of Two Templates
(30, 80]	(0.8, 0.9]	<9
(40/3, 30]	(0.7, 0.8]	<8
(5, 40/3]	(0.6, 0.7]	<7
(0, 5]	(0, 0.6]	<6

$$\frac{|\mathbf{W}_J|}{|\hat{\mathbf{I}}|} = \frac{|\mathbf{W}_J|}{M} > \rho. \tag{28}$$

Knowing that $|\mathbf{I}| = |\hat{\mathbf{I}}| = M$ and $\mathbf{W}_J \subset \mathbf{I}$, we can conclude from eqn (28) that, if node J is chosen first by pattern $\mathbf{I}$, node J will not be reset. Assume now that node j is chosen first (before node J is chosen). Then, according to Proposition 4.1 (c),

$$|\mathbf{I} \wedge \mathbf{W}_j| > |\mathbf{W}_J|. \tag{29}$$

Equation (29) in conjunction with eqn (28), and the fact that $|\mathbf{I}| = |\hat{\mathbf{I}}| = M$, imply that if node j is chosen first by pattern $\mathbf{I}$, node j will not be reset. The above discussion proves that uncommitted nodes will not be chosen in list presentations ≥ 2. ■

REMARKS. This theorem provides an upper bound for the number of nodes needed in the F_2 layer so that Fuzzy ART will learn all the patterns in an input list, provided that fast learning is employed. In practice, the number of categories (i.e., the number of nodes needed in the F_2 layer) is usually much less than the number of patterns in the input list, and is an increasing function of the choice parameter α and the vigilance parameter ρ.

THEOREM 4.2 (Direct Access by Perfectly Learned Pattern). *In a Fuzzy ART architecture, if a node J in the F_2 layer has perfectly learned an input pattern $\mathbf{I}$ (i.e., $\mathbf{W}_J = \mathbf{I}$), then when $\mathbf{I}$ is presented it will directly access node J.*

Proof. When input pattern $\mathbf{I}$ is presented, the choice function for node J is equal to

$$T_J = \frac{|\mathbf{I} \wedge \mathbf{W}_J|}{\alpha + |\mathbf{W}_J|} = \frac{|\mathbf{I}|}{\alpha + |\mathbf{I}|}. \tag{30}$$

For any other node j, the choice function is equal to

$$T_j = \frac{|\mathbf{I} \wedge \mathbf{W}_j|}{\alpha + |\mathbf{W}_j|}. \tag{31}$$

Thus,

$$(T_J - T_j)(\alpha + |\mathbf{W}_j|)(\alpha + |\mathbf{I}|)$$
$$= \alpha(|\mathbf{I}| - |\mathbf{I} \wedge \mathbf{W}_j|) + |\mathbf{I}|(|\mathbf{W}_j| - |\mathbf{I} \wedge \mathbf{W}_j|). \tag{32}$$

Note that both $|\mathbf{I}| - |\mathbf{I} \wedge \mathbf{W}_j|$ and $|\mathbf{W}_j| - |\mathbf{I} \wedge \mathbf{W}_j|$ are greater than or equal to zero and they cannot be equal to zero at the same time for $\mathbf{W}_j \neq \mathbf{I}$. We therefore conclude from eqn (32) that $T_J > T_j$, which guarantees that pattern $\mathbf{I}$ will choose node J over all the other nodes. Node J will not be reset because $\mathbf{W}_J = \mathbf{I}$. Consequently, in the presence of a node J with template $\mathbf{W}_J = \mathbf{I}$, the input pattern $\mathbf{I}$ will directly access node J. ■

PROPOSITION 4.2. *In a Fuzzy ART architecture with repeated presentations of a list of input patterns, after learning is complete, there may exist committed nodes in the F_2 layer that are not directly accessed by any pattern in the input list.*

Proof. (By example) Suppose that the complement-coded patterns in the input list are as follows:

$\mathbf{I}^1 = (0.3\ 0.7 | 0.7\ 0.3)$
$\mathbf{I}^2 = (0.7\ 0.3 | 0.3\ 0.7)$
$\mathbf{I}^3 = (0.2\ 0.6 | 0.8\ 0.4)$
$\mathbf{I}^4 = (0.4\ 0.8 | 0.6\ 0.2)$
$\mathbf{I}^5 = (0.6\ 0.2 | 0.4\ 0.8)$
$\mathbf{I}^6 = (0.8\ 0.4 | 0.2\ 0.6)$

These patterns are presented repeatedly to Fuzzy ART in the order $\mathbf{I}^1\ \mathbf{I}^2\ \mathbf{I}^3\ \mathbf{I}^4\ \mathbf{I}^5\ \mathbf{I}^6$. Assume that $\rho = 0.59$, $\beta = 1.0$ (fast-learning) and α is small. In the first list presentation patterns $\mathbf{I}^1$ and $\mathbf{I}^2$ will choose node 1, patterns $\mathbf{I}^3$ and $\mathbf{I}^4$ will choose node 2, and patterns $\mathbf{I}^5$ and $\mathbf{I}^6$ will choose node 3. Learning will be complete at the end of the first list presentation. In the second list presentation patterns $\mathbf{I}^1$ and $\mathbf{I}^2$ will choose nodes 2 and 3, respectively, patterns $\mathbf{I}^3$ and $\mathbf{I}^4$ will choose node 2, and patterns $\mathbf{I}^5$ and $\mathbf{I}^6$ will choose node 3. Thus, after the completion of learning, node 1 will not be chosen by any pattern in the input list.

Similar results are obtained in the case of fast-commit, slow-recode learning (if $0.9 < \beta < 1$) or the regular slow learning (if $0.96 \leq \beta < 1$) for the given example. The only difference is that it will take more than one list presentation to complete the learning process. ■

5. RESET PROPERTIES

The properties discussed in this section are byproducts of the results mentioned earlier. They are important to report though because they provide a different perspective of viewing these results; this perspective involves the orienting subsystem in Fuzzy ART. For example, Corollary 5.1 states that under certain assumptions, no reset events are possible after the first presentation of a list of input patterns, whereas Corollaries 5.2 and 5.3 determine the effective range of the vigilance parameter, that is, the range of ρ values that will allow reset events to occur. Corollaries 5.2 and 5.3 are also useful in helping us to choose appropriate α and ρ values for Fuzzy ART simulations.

COROLLARY 5.1. *In a Fuzzy ART architecture with fast learning, and repeated presentations of a list of input patterns, no reset will occur after the first list presentation.*

Proof. This is an immediate byproduct of the proof of Theorem 4.1. ■

REMARKS. Corollary 5.1 tells us that with fast learning and repeated presentations of a list of input patterns, for list presentations ≥ 2, there is no need to check on the vigilance criterion. In terms of hardware, the orienting subsystem becomes inactive (automatically disengaged) after the first list presentation. In terms of a software simulation of Fuzzy ART, we can disregard the orienting subsystem after the first list presentation to speed up the learning.

COROLLARY 5.2. *In a Fuzzy ART architecture with a sufficient number of nodes in the F_2 layer, if $\rho \le (\alpha/\alpha + M)$ no resets will occur. In the case of binary patterns and fast learning, if $\rho \le [(\alpha + 1)/(\alpha + M)]$ no resets will occur.*

Proof. This is a direct result of Proposition 3.1 because the vigilance parameter should be larger than the smallest possible template size divided by M for it to have an effect on the operation of the network. ■

REMARKS. The above corollary demonstrates that if α is large, small vigilance cannot be effective. For example, if $\alpha = M$, the vigilance should be larger than 0.5, because if $\rho \le 0.5$, no reset will ever occur. On the other hand, given a vigilance parameter ρ, if $\alpha \ge \rho M/(1 - \rho)$, no reset will ever occur. Therefore, we can always eliminate the orienting subsystem and let larger α values take care of the vigilance. Corollary 5.2 illustrates that the Fuzzy ART parameters α and ρ should be carefully chosen if we want the vigilance parameter ρ to have an effect on the operation of the network.

COROLLARY 5.3. *In a Fuzzy ART architecture with binary patterns, fast learning, and a sufficient number of nodes in the F_2 layer, if $\alpha > M(M - L - 1)/L$ and $\rho \le 1 - (L - 1)/M$, then no resets will occur. The parameter L is an integer taking values in the interval* $[2, M - 1]$.

Proof. This corollary is an immediate consequence of Corollary 3.1. ■

6. NUMBER OF LIST PRESENTATIONS

In this section, we assume that a list of input patterns is repeatedly presented to the Fuzzy ART architecture, and we derive results related to the number of list presentations required by Fuzzy ART to learn this list. In particular, Theorem 6.1 states that if the choice parameter α is relatively small, then learning in Fuzzy ART will be completed in one list presentation. Furthermore, Propositions 6.1–6.4 constitute an effort to find upper bounds on the number of list presentations needed by Fuzzy ART to learn the input list when α is relatively large (i.e., when α is not necessarily as small as it is required to validate Theorem 6.1). Of course other assumptions, besides the range of the α parameter, are needed to guarantee the validity of Theorem 6.1 and Propositions 6.1–6.4. A common assumption for Theorem 6.1 and Propositions 6.1–6.4 is that the input patterns are binary. With a slight modification of the assumptions in Theorem 6.1, we can guarantee its validity for analog input patterns (see Remarks after the proof of Theorem 6.1).

It is worth mentioning that Theorem 6.1 in its analog version (i.e., without the assumption of binary input patterns) was first presented in Carpenter et al. (1991b). However, Theorem 6.1, by specializing to the case of binary input patterns, gives explicit conditions on the values of the parameter α that guarantee its validity. Furthermore, in the remarks, following the proof of Theorem 6.1, we stress the difficulty of imposing explicit conditions on the values of α to guarantee the validity of the theorem for the case of analog input patterns. In any case, despite the fact that Theorem 6.1 was first presented elsewhere (Carpenter et al., 1991b) its incorporation in this section makes the section more complete.

THEOREM 6.1. *In a Fuzzy ART architecture with binary patterns, fast learning, a sufficient number of nodes in the F_2 layer, and repeated presentations of a list of input patterns, if $\alpha \le \rho/(1 - \rho)$, then the weights will stabilize in one list presentation.*

Proof. With fast learning, after the first list presentation, every pattern has at least one subset node and no pattern will choose an uncommitted node (Proposition 4.1). The weights are stabilized in one list presentation if, in list presentations ≥ 2, every pattern in the input list chooses a subset node over any mixed node. This is due to the fact that no resets occur in list presentations ≥ 2 (Corollary 5.1). The sufficient condition required so that a pattern **I** from the input list chooses a subset node J prior to choosing a mixed node j is

$$\frac{|\mathbf{W}_J|}{\alpha + |\mathbf{W}_J|} > \frac{|\mathbf{I} \wedge \mathbf{W}_j|}{\alpha + |\mathbf{W}_j|}, \tag{33}$$

which implies that

$$|\mathbf{W}_J|(|\mathbf{W}_j| - |\mathbf{I} \wedge \mathbf{W}_j|) + \alpha(|\mathbf{W}_J| - |\mathbf{I} \wedge \mathbf{W}_j|) > 0. \tag{34}$$

Because any template satisfies

$$\rho M \le |\mathbf{W}_j| \le M \tag{35}$$

and in the case of binary input patterns and fast learning

$$|\mathbf{I} \wedge \mathbf{W}_j| \le |\mathbf{W}_j| - 1 \le M - 1, \tag{36}$$

the left-hand side of inequality (34) is larger than or equal to

$$\rho M + \alpha(\rho M - M + 1). \tag{37}$$

Consequently, eqn (34) is satisfied if $\alpha \le \rho/(1 - \rho)$. Concluding, in list presentations ≥ 2, if $\alpha \le \rho/(1 - \rho)$, every pattern in the input list will directly access its subset node with the largest template size. Hence, no weight changes will occur in list presentations ≥ 2 and equivalently the weights are stabilized in one list presentation. ■

REMARKS. (1) In the extreme case where $\rho = 1$, each pattern from the input list will choose a different node in the F_2 layer. In this case, for any value of α the weights will stabilize in one list presentation (see also Proposition 6.1 for stronger results). (2) By Corollary 3.1, for binary input patterns and fast learning, the smallest possible template size is greater than or equal to 2, and as a result the vigilance parameter ρ should be larger than $2/M$. Therefore, if $\rho \leq 2/M$ (including zero), Theorem 6.1 is valid for $\alpha \leq 2/(M - 2)$. (3) A sufficient condition on the α values that will guarantee the validity of Theorem 6.1, even when the input patterns are analog, is

$$\alpha < \frac{\rho(|\mathbf{W}_j| - |\mathbf{I} \wedge \mathbf{W}_j|)}{1 - \rho}. \tag{38}$$

Unfortunately, even if we know the exact ρ value, we cannot find a lower bound for the right-hand side of eqn (38) because $|\mathbf{W}_j| - |\mathbf{I} \wedge \mathbf{W}_j|$ can be arbitrarily small in the analog case. In conclusion, we can only state that Theorem 6.1 is valid for analog input patterns if α is chosen very small.

Before we present other results, let us state a lemma. The proof of the lemma is easy and therefore omitted.

LEMMA 6.1. *In a Fuzzy ART architecture with binary patterns, fast learning, and repeated presentations of a list of input patterns, if $|\mathbf{W}|$ is the minimum template size at the end of the first list presentation, then the following rules are valid in list presentations ≥ 2:*

(1) *No template of size $|\mathbf{W}|$ or smaller can be created.*
(2) *A template of size $|\mathbf{W}|$ or $|\mathbf{W}| + 1$ cannot be modified.*
(3) *A template of size $|\mathbf{W}| + L$ ($L \geq 2$) can be modified only by patterns for which the largest subset template is of size $H \leq |\mathbf{W}| + L - 2$. And the new template size should be greater than or equal to $H + 1$.*

PROPOSITION 6.1. *In a Fuzzy ART architecture with binary patterns, fast learning, a sufficient number of nodes in the F_2 layer, and repeated presentations of a list of input patterns, if $\alpha > \frac{1}{2}M(M - 3)$ or $\rho > 1 - 2/M$, then the weights will be stabilized in one list presentation.*

Proof. If $\alpha > \frac{1}{2}M(M - 3)$, by Corollary 3.1, the smallest possible template size is equal to $M - 1$. Similarly, if $\rho > 1 - 2/M$, the smallest template size is equal to $M - 1$. In either case, we have at most two different sizes of the templates: size M and size $M - 1$. By Lemma 6.1, no template can be changed after the first list presentation. Therefore, the weights are stabilized in one list presentation. ■

REMARKS. If $\alpha > M(M - 2)$, the smallest possible template size is M (Corollary 3.1). As a result, each distinct pattern will choose a different node in the F_2 layer during the first list presentation, and no reset will occur no matter what the value of ρ is. In this case, Fuzzy ART provides a fast way of distinguishing patterns.

PROPOSITION 6.2. *In a Fuzzy ART architecture with binary patterns, fast learning, a sufficient number of nodes in the F_2 layer, and repeated presentations of a list of input patterns, if $\frac{1}{3}M(M - 4) < \alpha \leq \frac{1}{2}M(M - 3)$ or $1 - 3/M < \rho \leq 1 - 2/M$, then the weights will stabilize in at most two list presentations.*

Proof. If $\frac{1}{3}M(M - 4) < \alpha \leq \frac{1}{2}M(M - 3)$, or $1 - 3/M < \rho \leq 1 - 2/M$, then there are at most three different sizes of templates: size M, size $M - 1$, and size $M - 2$. By Proposition 4.1, no new template of size M will be created after the first list presentation. By Lemma 6.1, the templates of size $M - 2$ or $M - 1$ cannot be changed after the first list presentation. These observations allow us to state that templates of size M cannot be modified after the second list presentation. In review, no templates are modified and no new templates are created after the second list presentation, or equivalently, the weights are stabilized within the first two list presentations. ■

PROPOSITION 6.3. *In a Fuzzy ART architecture with binary patterns, fast learning, a sufficient number of nodes in the F_2 layer, and repeated presentations of a list of input patterns, if $\frac{1}{4}M(M - 5) < \alpha \leq \frac{1}{3}M(M - 4)$ or $1 - 4/M < \rho \leq 1 - 3/M$, then the weights will be stabilized in at most three list presentations.*

Proof. If $\frac{1}{4}M(M - 5) < \alpha \leq \frac{1}{3}M(M - 4)$, or $1 - 4/M < \rho \leq 1 - 3/M$, then by Corollary 3.1, there are at most four different sizes of the templates, that is, sizes M, $M - 1$, $M - 2$, and $M - 3$. When the smallest template size is $M - 3$, both $\alpha \leq \frac{1}{3}M(M - 4)$ and $\rho \leq 1 - 3/M$ have to be true. By Proposition 3.2, for any two templates $\mathbf{W}_1$ and $\mathbf{W}_2$, $|\mathbf{W}_1 \wedge \mathbf{W}_2| \leq M - 3$ for $M \geq 4$. Based on Theorem 4.1 we can also claim that in list presentations ≥ 2 templates of size M cannot be created.

Proposition 6.3 will be proven in two steps. In Step 1 we will prove that in list presentations ≥ 3, templates of size M cannot be destroyed. In Step 2 we will prove that in list presentations ≥ 4, templates of size $M - 1$ cannot be destroyed. The combination of the two steps in conjunction with Lemma 6.1 prove Proposition 6.3. For the proof of Step 1 we distinguish cases.

Case 1. At the beginning of the second list presentation a pattern $\mathbf{I}$ has a subset template of size M, denoted by $\mathbf{W}_1$. This template, in list presentations ≥ 2, might be reduced in size to a template of size $M - 1$ or $M - 2$. Independently of what happens to $\mathbf{W}_1$ pattern $\mathbf{I}$ cannot, in list presentations ≥ 2, destroy another template $\mathbf{W}_2$ of size M. This is because $|\mathbf{W}_1 \wedge \mathbf{W}_2| \leq M - 3$ and consequently $|\mathbf{I} \wedge \mathbf{W}_2| = |\mathbf{W}_1 \wedge \mathbf{W}_2| \leq M - 3$.

Case 2. At the beginning of the second list presentation a pattern $\mathbf{I}$ has a subset template of size $M - 1$, denoted by $\mathbf{W}_1$. This template, in list presentations ≥ 2, might be reduced in size to a template of size $M - 2$. Independently of what happens to $\mathbf{W}_1$ pattern $\mathbf{I}$ cannot, in list presentations ≥ 2, destroy another template $\mathbf{W}_2$ of size M. This is because $|\mathbf{W}_1 \wedge \mathbf{W}_2| \leq M - 3$, which means that pattern $\mathbf{I}$ can have at most $M - 2$ common ones with $\mathbf{W}_2$.

Case 3(a). At the beginning of the second list presentation a pattern $\mathbf{I}$ has a subset template $\mathbf{W}_1$ of size $M - 2$. Furthermore, during $\mathbf{I}$'s presentation in the second list, $\mathbf{I}$ chooses node 1 with template $\mathbf{W}_1$ over all other nodes with templates of size M. It is obvious then that pattern $\mathbf{I}$ in list presentations ≥ 3 will always choose template $\mathbf{W}_1$ over all other templates of size M (note that in list presentations ≥ 2 template $\mathbf{W}_1$ cannot be destroyed and new templates of size M cannot be created).

Case 3(b). At the beginning of the second list presentation a pattern $\mathbf{I}$ has a subset template $\mathbf{W}_1$ of size $M - 2$. Furthermore, during $\mathbf{I}$'s presentation in the second list, pattern $\mathbf{I}$ destroys a template $\mathbf{W}_2$ of size M and thus it creates a template of size $M - 1$. Following similar arguments as the ones for Case 2, we can prove that pattern $\mathbf{I}$, in list presentations ≥ 3, cannot destroy another template of size M.

Case 4(a). At the beginning of the second list presentation a pattern $\mathbf{I}$ has a subset template $\mathbf{W}_1$ of size $M - 3$. Furthermore, during $\mathbf{I}$'s presentation in the second list, pattern $\mathbf{I}$ chooses node 1 with template $\mathbf{W}_1$ over all other nodes with templates of size M. For similar reasons as the ones mentioned in Case 3(a), pattern $\mathbf{I}$ in list presentations ≥ 3 cannot destroy templates of size M.

Case 4(b). At the beginning of the second list presentation a pattern $\mathbf{I}$ has a subset template $\mathbf{W}_1$ of size $M - 3$. Furthermore, during $\mathbf{I}$'s presentation in the second list pattern $\mathbf{I}$ destroys a template $\mathbf{W}_2$ of size M and thus creates (i) a template of size $M - 1$ or (ii) a template of size $M - 2$. Scenario (i) can be treated in a similar fashion as Case 2 to prove that in list presentations ≥ 3, pattern $\mathbf{I}$ cannot destroy another template $\mathbf{W}_3$ of size M. If scenario (ii) occurs, we know that in list presentations ≥ 3, pattern $\mathbf{I}$ can destroy template $\mathbf{W}_3$ of size M only if it were possible to create a template of size $M - 1$; but if this can happen in a list presentation ≥ 3, it should have also happened in the second list presentation. This is a contradiction because we are operating under scenario (ii). Hence, under scenario (ii), pattern $\mathbf{I}$ in list presentations ≥ 3 cannot destroy templates of size M.

Cases 1 through 4 cover all possible scenarios, and prove the validity of Step 1. Due to Step 1 we can claim that in list presentations ≥ 3, templates of size $M - 1$ cannot be created. Consequently, Step 2 can now be proved in the same manner that Proposition 6.2 was proved. The combination of Lemma 6.1, Step 1, and Step 2 guarantees that the weights will stabilize in at most three list presentations. ■

PROPOSITION 6.4. *In a Fuzzy ART architecture with binary patterns, fast learning, a sufficient number of nodes in the F_2 layer, cyclic presentations of a list of input patterns, and $M \geq 9$, if $\frac{1}{5}M(M - 6) < \alpha \leq \frac{1}{4}M(M - 5)$, or $1 - 5/M < \rho \leq 1 - 4/M$, then the weights will stabilize in at most four list presentations.*

Discussion of the proof. The proof of this proposition is complicated. Its complication arises from the fact that the number of cases that needs to be examined to

TABLE 3a
Summary of the Results in Theorem 6.1 and Propositions 6.1–6.4 for an Arbitrary M

Range of α		Range of ρ	Number of Template Sizes	Number of Lists Needed
$\alpha \in \left(0, \max\left(\frac{\rho}{1-\rho}, \frac{2}{M-2}\right)\right]$	and	$\rho \in (0, 1]$	$\leq M - 1$	1
$\alpha \in (\frac{1}{2}M(M - 3), \infty)$	or	$\rho \in \left(1 - \frac{2}{M}, 1\right]$	≤ 2	1
$\alpha \in (\frac{1}{3}M(M - 4), \frac{1}{2}M(M - 3)]$	or	$\rho \in \left(1 - \frac{3}{M}, 1 - \frac{2}{M}\right]$	≤ 3	≤ 2
$\alpha \in (\frac{1}{4}M(M - 5), \frac{1}{3}M(M - 4)]$	or	$\rho \in \left(1 - \frac{4}{M}, 1 - \frac{3}{M}\right]$	≤ 4	≤ 3
$\alpha \in (\frac{1}{5}M(M - 6), \frac{1}{4}M(M - 5)]$	or	$\rho \in \left(1 - \frac{5}{M}, 1 - \frac{4}{M}\right]$	≤ 5	≤ 4

TABLE 3b
Summary of the Results in Theorem 6.1 and Propositions 6.1–6.4 for the Special Case of $M = 10$

Range of α		Range of ρ	Number of Template Sizes	Number of Lists Needed
$\alpha \in \left(0, \max\left(\frac{\rho}{1-\rho}, 0.25\right)\right]$	and	$\rho \in (0, 1]$	≤ 9	1
$\alpha \in (35, \infty)$	or	$\rho \in (0.8\ 1]$	≤ 2	1
$\alpha \in (20, 35]$	or	$\rho \in (0.7, 0.8]$	≤ 3	≤ 2
$\alpha \in (12.5, 20]$	or	$\rho \in (0.6, 0.7]$	≤ 4	≤ 3
$\alpha \in (8, 12.5]$	or	$\rho \in (0.5, 0.6]$	≤ 5	≤ 4

guarantee its validity is much larger than the number of cases investigated in the proof of Proposition 6.3. Here we only provide a sketch of the proof.

If $\frac{1}{5}M(M-6) < \alpha \leq \frac{1}{4}M(M-5)$, or $1 - 5/M < \rho \leq 1 - 4/M$, then there are at most five different sizes of the templates: M, $M-1$, $M-2$, $M-3$, and $M-4$. When the smallest template size is $M-4$, both $\alpha \leq \frac{1}{4}M(M-5)$ and $\rho \leq 1 - 4/M$ have to be true. By Proposition 3.2, for any two templates $\mathbf{W}_1$ and $\mathbf{W}_2$, $|\mathbf{W}_1 \wedge \mathbf{W}_2| \leq M - 4$ for $M \geq 9$. We can prove Proposition 6.4 in two steps. In Step 1, we prove that no templates of size $M-1$ can be created in list presentations ≥ 3, and if a template of size M is modified in the third list, then the new template size must be $M-2$. In Step 2 we prove that no templates of size $M-1$ or M can be modified in list presentations ≥ 4, and as a result, no templates of size $M-2$ can be created in list presentations ≥ 4.

Steps 1 and 2, in conjunction with Theorem 4.1, prove that new templates cannot be created and old templates cannot be modified in list presentations ≥ 5. That is, the weights are stabilized in four list presentations.

REMARKS. Proposition 6.4 covers only the cases where $M \geq 9$, which ensures $|\mathbf{W}_1 \wedge \mathbf{W}_2| \leq M - 4$ (obtained from Proposition 3.2). If $M \leq 8$ and $\rho > 0.5$, the weights will be stabilized in at most three list presentations (see Propositions 6.1–6.3). Hence, Propositions 6.1–6.3 cover most of the practical cases when $M \leq 8$.

In Table 3 we present a summary of our findings as they are predicted by Theorem 6.1 and Propositions 6.2–6.4. We depict the results for arbitrary M and for $M = 10$. It is easy to see from Table 3 that for small M values, Theorem 6.1 and Propositions 6.1–6.4 cover all the practical cases of interest regarding the number of list presentations needed by Fuzzy ART to learn a list of binary patterns that is repeatedly presented to it. Furthermore, by looking at the results of Table 3, we are encouraged to believe that there is a pattern relating the range of α or ρ values, and the number of list presentations needed to learn an arbitrary binary list repeatedly presented to Fuzzy ART. Hence, we were tempted to formulate a conjecture that extends the results of Table 3 over the entire range of α and ρ values. We decided not to do so because of the additional assumptions needed to verify the validity of Proposition 6.4 (i.e., cyclic presentations of the input list, and $M \geq 9$). Nevertheless, it is worth mentioning that out of hundreds of simulations performed with random input patterns, we found that the maximum number of list presentations needed for weight stabilization in Fuzzy ART was three for two of the simulations, and two for the rest of the simulations.

7. SUMMARY

We have examined the Fuzzy ART algorithm carefully from a number of different perspectives. For example, in Section 3 we demonstrated that Fuzzy ART templates are distinct, we calculated a lower bound on the template size, and we found an upper bound on the similarity of the templates created. Furthermore, in Section 4 we focused on access properties, investigating the order of search of the F_2 layer nodes, finding an upper bound on the number of nodes needed in the F_2 layer of Fuzzy ART to learn an arbitrary list of input patterns, and proving the direct access property of patterns to perfectly learned templates. Also, in Section 5 we concentrated on the orienting subsystem and reset events, and elaborated on the interrelationship between the α and ρ parameter values needed in Fuzzy ART simulations. Finally, in Section 6 we shifted our attention to the number of list presentations required by Fuzzy ART to learn an arbitrary list of patterns repeatedly presented to it. Most of the results presented in Section 6 were valid for binary input patterns and fast learning. The strongest result proven in Section 6 stated that for small α values [i.e., $\alpha \leq \rho/(1-\rho)$], learning will be complete in one list presentation. Weaker results were also presented in Section 6, where, to come up with a definite upper bound on the number of list presentations needed, we restricted either the α range or the ρ range.

REFERENCES

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

Carpenter, G. A., Grossberg, S., & Reynolds, J. H. (1991a). ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks,* **4**(5), 565–588.

Carpenter, G. A., Grossberg, S., & Rosen, D. B. (1991b). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks,* **4**(6), 759–771.

Georgiopoulos, M., Heileman, G. L., & Huang, J. (1990). Convergence properties of learning in ART1. *Neural Computation,* **2**(4), 502–509.

Georgiopoulos, M., Heileman, G. L., & Huang, J. (1991). Properties of learning related to pattern diversity in ART1. *Neural Networks,* **4**(6), 751–757.

Georgiopoulos, M., Heileman, G. L., & Huang, J. (1992). The *N-N-N* conjecture in ART1. *Neural Networks,* **5**(5), 745–753.

Georgiopoulos, M., Huang, J., & Heileman, G. L. (1994). Properties of learning in ARTMAP. *Neural Networks,* **7**(3), 495–506.

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

Moore, B. (1989). ART1 and pattern clustering. In Touretzky, D. S., Hinton, G., & Sejnowski, T. (Eds.), *Proceedings of the* 1988 *Connectionist Summer School* (pp. 174–185). San Mateo, CA: Morgan Kaufmann.