

# Equivalence in Knowledge Representation: Automata, Recurrent Neural Networks, and Dynamical Fuzzy Systems

C. LEE GILES, FELLOW, IEEE, CHRISTIAN W. OMLIN, AND KARVEL K. THORNBER

## Invited Paper

*Neurofuzzy systems—the combination of artificial neural networks with fuzzy logic—have become useful in many application domains. However, conventional neurofuzzy models usually need enhanced representational power for applications that require context and state (e.g., speech, time series prediction, control). Some of these applications can be readily modeled as finite state automata. Previously, it was proved that deterministic finite state automata (DFA) can be synthesized by or mapped into recurrent neural networks by directly programming the DFA structure into the weights of the neural network. Based on those results, a synthesis method is proposed for mapping fuzzy finite state automata (FFA) into recurrent neural networks. Furthermore, this mapping is suitable for direct implementation in very large scale integration (VLSI), i.e., the encoding of FFA as a generalization of the encoding of DFA in VLSI systems. The synthesis method requires FFA to undergo a transformation prior to being mapped into recurrent networks. The neurons are provided with an enriched functionality in order to accommodate a fuzzy representation of FFA states. This enriched neuron functionality also permits fuzzy parameters of FFA to be directly represented as parameters of the neural network. We also prove the stability of fuzzy finite state dynamics of the constructed neural networks for finite values of network weight and, through simulations, give empirical validation of the proofs. Hence, we prove various knowledge equivalence representations between neural and fuzzy systems and models of automata.*

**Keywords**— Dynamic systems, finite automata, fuzzy systems, recurrent neural networks.

## I. INTRODUCTION

### A. Motivation

As our applications for intelligent systems become more ambitious, our processing models become more powerful.

Manuscript received December 2, 1998; revised April 13, 1999.

C. L. Giles is with NEC Research Institute, Princeton, NJ 08540 USA and with UMIACS, University of Maryland, College Park, MD 20742 USA.

C. W. Omlin is with the Department of Computer Science, University of Stellenbosch, 7600 Stellenbosch, South Africa.

K. K. Thornber is with NEC Research Institute, Princeton, NJ 08540 USA.

Publisher Item Identifier S 0018-9219(99)06915-7.

One approach to increasing this power is through hybrid systems—systems that include several different models intelligent processing [1]. There has also been an increased interest in hybrid systems as more applications with hybrid models emerge [2]. However, there are many definitions of hybrid systems [3]–[5].

One example of hybrid systems is in combining artificial neural networks and fuzzy systems (see [6]–[9]). Fuzzy logic [10] provides a mathematical foundation for approximate reasoning; fuzzy logic has proven very successful in a variety of applications [11]–[20]. The parameters of adaptive fuzzy systems have clear physical meanings that facilitate the choice of their initial values. Furthermore, rule-based information can be incorporated into fuzzy systems in a systematic way. Artificial neural networks propose to simulate on a small scale the information processing mechanisms found in biological systems that are based on the cooperation and computation of artificial neurons that perform simple operations, and on their ability to learn from examples. Artificial neural networks have become valuable computational tools in their own right for tasks such as pattern recognition, control, and forecasting (for more information on neural networks, please see [21]–[23]). Recurrent neural networks (RNN's) are dynamical systems with temporal state representations; they are computationally quite powerful [24], [25] and can be used in many different temporal processing models and applications [26].

Fuzzy finite state automata (FFA), fuzzy generalizations of deterministic finite state automata,<sup>1</sup> have a long history [32], [33]. The fundamentals of FFA have been discussed in [34] without presenting a systematic machine synthesis method. Their potential as design tools for modeling a variety of systems is beginning to be exploited in various

<sup>1</sup>Finite state automata also have a long history as theoretical [27] and practical [28] models of computation and were some of the earliest implementations of neural networks [29], [30]. Besides automata, other symbolic computational structures can be used with neural networks [31], [26].

applications [35], [36]. Such systems have two major characteristics: 1) the current state of the system depends on past states and current inputs and 2) the knowledge about the system's current state is vague or uncertain.

Finally, the proofs of representational properties of artificial intelligence, machine learning, and computational intelligence models are important for a number of reasons. Many users of a model want guarantees about what it can do theoretically, i.e., its performance and capabilities; others need this for justification of use and acceptance of the approach. The capability of representing a model, say FFA, in an intelligent system can be viewed as a foundation for the problem of learning that model from examples (if a system cannot represent FFA, then it certainly will have difficulty learning FFA).

Since recurrent neural networks are nonlinear dynamical systems, the proof of their capability to represent FFA amounts to proving that a neural network representation of fuzzy states and transitions remains stable for input sequences of arbitrary length and is robust to noise. Neural networks that have been trained to behave like FFA do not necessarily share this property, i.e., their internal representation of states and transitions may become unstable for sufficiently long input sequences [37]. Finally, with the extraction of knowledge from trained neural networks, the methods discussed here could potentially be applied to incorporating and refining [38] fuzzy knowledge previously encoded into recurrent neural networks.

## B. Background

A variety of implementations of FFA have been proposed, some in digital systems [39], [40]. However, here we give a proof that such implementations in sigmoid activation RNN's are stable, i.e., guaranteed to converge to the correct prespecified membership. This proof is based on previous work of stably mapping deterministic finite state automata (DFA) in recurrent neural networks reported in [41]. In contrast to DFA, a set of FFA states can be occupied to varying degrees at any point in time; this fuzzification of states generally reduces the size of the model, and the dynamics of the system being modeled is often more accessible to a direct interpretation.

From a control perspective, fuzzy finite state automata have been shown to be useful for modeling fuzzy dynamical systems, often in conjunction with recurrent neural networks [35], [42]–[45]. There has been much work on the learning, synthesis, and extraction of finite state automata in recurrent neural networks (see, for example, [46]–[53]). A variety of neural network implementations of FFA have been proposed [39], [40], [54], [55]. We have previously shown how fuzzy finite state automata can be mapped into recurrent neural networks with second-order weights using a crisp representation<sup>2</sup> of FFA states [56]. That encoding required a transformation of a fuzzy finite state automaton into a deterministic finite state automaton that computes the membership functions for strings; it is only applicable

<sup>2</sup>A crisp mapping is one from a fuzzy to a nonfuzzy variable.

to a restricted class of FFA that has final states. The transformation of a fuzzy automaton into an equivalent deterministic acceptor generally increases the size of the automaton and thus the network size. Furthermore, the fuzzy transition memberships of the original FFA undergo modifications in the transformation of the original FFA into an equivalent DFA that is suitable for implementation in a second-order recurrent neural network. Thus, the direct correspondence between system and network parameters is lost which may obscure the natural fuzzy description of systems being modeled.

The existence of a crisp recurrent network encoding for all FFA raises the question of whether recurrent networks can also be trained to compute the fuzzy membership function, and how they represent FFA states internally. Based on our theoretical analysis, we know that they have the ability to represent FFA in the form of equivalent deterministic acceptors. Recent work reported in [57] addresses these issues. Instead of augmenting a second-order network with a linear output layer for computing the fuzzy string membership as suggested in [56], they chose to assign a distinct output neuron to each fuzzy string memberships  $\mu_i$  occurring in the training set. Thus, the number of output neurons became equal to the number of distinct membership values  $\mu_i$ . The fuzzy membership of an input string was then determined by identifying the output neuron whose activation was highest after the entire string had been processed by a network. Thus, they transformed the fuzzy inference problem into a classification problem with multiple classes or classifications. This approach lessens the burden on the training and improves the accuracy and robustness of string membership computation.

Apart from the use of multiple classes, training networks to compute the fuzzy string membership is identical to training networks to behave like DFA. This was verified empirically [57] through information extraction methods [46], [37] where recurrent networks trained on fuzzy strings develop a crisp internal representation of FFA, i.e., that they represent FFA in the form of equivalent deterministic acceptors.<sup>3</sup> Thus, our theoretical analysis correctly predicted the knowledge representation for such trained networks.

## C. Overview of Paper

In this paper, we present a method for encoding FFA using a fuzzy representation of states.<sup>4</sup> The objectives of the FFA encoding algorithm are: 1) ease of encoding FFA into recurrent networks; 2) the direct representation of “fuzziness,” i.e., the fuzzy memberships of individual transitions in FFA are also parameters in the recurrent networks; and 3) achieving a fuzzy representation by making only minimal changes to the underlying architecture used for encoding DFA (and crisp FFA representations).

<sup>3</sup>The equivalence of FFA and deterministic acceptors was first discussed in [58] and first used for encoding FFA in [56].

<sup>4</sup>For reasons of completeness, we have included the main results from [41] which lay the foundations for this and other papers [59], [56]. Thus, by necessity, there is some overlap.

Representation of FFA in recurrent networks requires that the internal representation of FFA states and state transitions be stable for indefinite periods of time. We will demonstrate how the stability analysis for neural DFA encodings carries over to and generalizes the analysis of stable neural FFA representations.

In high-level very large scale integration (VLSI) design a DFA (actually finite state machines) is often used as the first implementation of a design and is mapped into sequential machines and logic [28]. Previous work has shown how a DFA can be readily implemented in RNN's and neural networks have been directly implemented in VLSI chips [60]–[62]. Thus, with this approach FFA could be readily mapped into electronics and could be useful for applications, such as real-time control (see e.g., [13])<sup>5</sup> and could potentially be applied to incorporate *a priori* knowledge into recurrent neural networks for knowledge refinement [63].

The remainder of this paper is organized as follows. FFA are introduced in Section II. The fuzzy representation of FFA states and transitions in recurrent networks are discussed in Section III. The mapping “fuzzy automata → recurrent network” proposed in this paper requires that FFA be transformed into a special form before they can be encoded in a recurrent network. The transformation algorithm can be applied to arbitrary FFA; it is described in Section IV. The recurrent network architecture for representing FFA is described in Section V. The stability of the encoding is derived in Section VI. A discussion of simulation results in Section VII and a summary of the results and possible directions for future research in Section VIII conclude this paper.

## II. FFA

In this section, we give a formal definition of FFA [64] and illustrate the definition with an example.

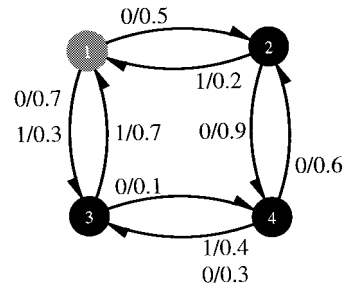
*Definition 2.1:* A fuzzy finite state automaton  $M$  is a 6-tuple  $M = \langle \Sigma, Q, R, Z, \delta, \omega \rangle$ , where  $\Sigma = \{a_1, \dots, a_m\}$  is the alphabet,  $Q = \{q_1, \dots, q_n\}$  is a set of states,  $R \in Q$  is the automaton's fuzzy start state,<sup>6</sup>  $Z$  is a finite output alphabet,  $\delta: \Sigma \times Q \times [0, 1] \rightarrow Q$  is the fuzzy transition map, and  $\omega: Q \rightarrow Z$  is the output map.<sup>7</sup>

Weights  $\theta_{ijk} \in [0, 1]$  define the “fuzziness” of state transitions, i.e., FFA can simultaneously be in different states with a different degree of certainty. The particular

<sup>5</sup>Alternative implementations of FFA have been proposed (see e.g., [54]). The method proposed uses recurrent neurons with sigmoidal discriminant functions and a fuzzy internal representation of FFA states.

<sup>6</sup>In general, the start state of a FFA is fuzzy, i.e., it consists of a set of states that are occupied with varying memberships. It has been shown that a restricted class of FFA whose initial state is a single crisp state is equivalent with the class of FFA described in Definition 2.1 [64]. The distinction between the two classes of FFA is irrelevant in the context of this paper.

<sup>7</sup>This is in contrast to stochastic finite state automata where there exists no ambiguity about which is an automaton's current state. The automaton can only be in exactly one state at any given time and the choice of a successor state is determined by some probability distribution. For a discussion of the relationship between probability and fuzziness, see, for instance, [65].



**Fig. 1.** Example of an FFA. A fuzzy finite state automaton is shown with weighted state transitions. State 1 is the automaton's start state. A transition from state  $q_j$  to  $q_i$  on input symbol  $a_k$  with weight  $\theta$  is represented as a directed arc from  $q_j$  to  $q_i$  labeled  $a_k/\theta$ . Note that transitions from states 1 and 4 on input symbols “0” are fuzzy ( $\delta(1, 0, \cdot) = \{2, 3\}$  and  $\delta(4, 0, \cdot) = \{2, 3\}$ ).

output mapping depends on the nature of the an application. Since our goal is to construct a fuzzy representation of FFA states and their stability over time, we will ignore the output mapping  $\omega$  for the remainder of this discussion and not concern ourselves with the language  $L(M)$  defined by  $M$ . For a possible definition, see [64]. An example of a FFA over the input alphabet  $\{0, 1\}$  is shown in Fig. 1.

## III. REPRESENTATION OF FUZZY STATES

### A. Preliminaries

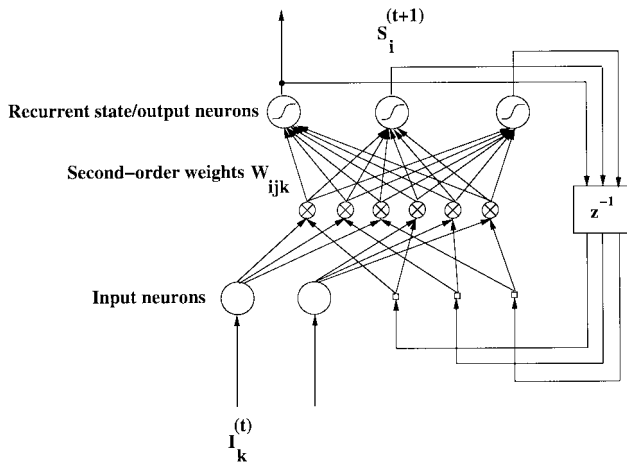
The current fuzzy state of a fuzzy finite state automaton  $M$  is a collection of states  $\{q_i\}$  of  $M$  that are occupied with different degrees of fuzzy membership. A fuzzy representation of the states in fuzzy finite state automaton thus requires knowledge about the membership of each state  $q_i$ . This requirement then dictates the representation of the current fuzzy state in a recurrent neural network. Since the method for encoding FFA in recurrent neural networks is a generalization of the method for encoding DFA, we will briefly discuss the DFA encoding algorithm.

### B. DFA Encoding Algorithm

We make use of an algorithm used for encoding DFA [41], [59]. For encoding DFA, we used discrete-time, second-order RNN's with sigmoidal discriminant functions that update their current state according to the following equations:

$$S_i^{(t+1)} = g(\alpha_i(t)) = \frac{1}{1 + e^{-\alpha_i(t)}} \\ \alpha_i(t) = b_i + \sum_{j,k} W_{ijk} S_j^{(t)} I_k^{(t)} \quad (1)$$

where  $b_i$  is the bias associated with hidden recurrent state neurons  $S_i$ ,  $W_{ijk}$  is a second-order weight, and  $I_k$  denotes the input neuron for symbol  $a_k$ . The indexes  $i, j$ , and  $k$  run over all state and input neurons, respectively. The product  $S_j^{(t)} I_k^{(t)}$  corresponds directly to the state transition  $\delta(q_j, a_k) = q_i$ . The architecture is illustrated in Fig. 2.



**Fig. 2.** Recurrent network architecture for deterministic finite state automata. The recurrent state neurons are connected and implement the stable finite state dynamics. One of the recurrent neurons also is the dedicated network output neuron (i.e., the neuron which with its output value classifies whether or not a given string is a member of a regular language).

DFA can be encoded in discrete-time, second-order RNN's with sigmoidal discriminant functions such that the DFA and constructed network accept the same regular language [41]. The desired finite state dynamics are encoded into a network by programming a small subset of all available weights to values  $+H$  and  $-H$ ; this leads to a nearly orthonormal internal DFA state representation for sufficiently large values of  $H$ , i.e., a one-to-one correspondence between current DFA states and recurrent neurons with a high output. Since the magnitude of all weights in a constructed network is equal to  $H$ , the equation governing the dynamics of a constructed network is of the special form

$$S_i^{(t+1)} = g(x, H) = \frac{1}{1 + e^{H(1-2x)/2}} \quad (2)$$

where  $x$  is the input to neuron  $S_i$ .

The objective of mapping DFA into recurrent networks is to assign DFA states to neurons and to program the weights such that the assignment remains stable for input sequence of arbitrary length, i.e., exactly one neuron corresponding to the current DFA state has a high output at any given time. Such stability is trivial for recurrent networks whose neurons have hard-limiting (or "step function") discriminant functions. However, this is not obvious for networks with continuous, sigmoidal discriminant functions. The nonlinear dynamical nature of recurrent networks makes it possible for intended internal DFA state representations to become unstable, i.e., the requirement of a one-to-one correspondence between DFA states and recurrent neurons may be violated for sufficiently long input sequences. We have previously demonstrated that it is possible to achieve a stable internal DFA state representation that is independent of the string length: in constructed networks, the recurrent state neurons always operate near their saturation regions for sufficiently large values of  $H$ ; as a consequence, the internal

DFA state representation remains stable indefinitely. The internal representation of fuzzy states proposed in this paper is a generalization of the method used to encode DFA states since FFA may be in several states at the same time. We will apply the same tools and techniques to prove stability of the internal representation of fuzzy states in recurrent neural networks.

### C. Recurrent State Neurons with Variable Output Range

We extend the functionality of recurrent state neurons in order to represent fuzzy states as illustrated in Fig. 3. The main difference between the neuron discriminant function for DFA and FFA is that the neuron now receives as inputs the weight strength  $H$ , the signal  $x$  that represents the collective input from all other neurons, and the transition weight  $\theta_{ijk}$ , where  $\delta(q_j, a_k, \theta_{ijk}) = q_i$ ; we will denote this triple with  $(x, H, \theta_{ijk})$ . The value of  $\theta_{ijk}$  is different for each of the states that collectively make up the current fuzzy network state. This is consistent with the definition of FFA.

The following generalized form of the sigmoidal discriminant function  $g(\cdot)$  will be useful for representing FFA states:

$$S_i^{(t+1)} = \tilde{g}(x, H, \theta_{ijk}) = \frac{\theta_{ijk}}{1 + e^{H(\theta_{ijk}-2x)/2\theta_{ijk}}} \quad (3)$$

Compared to the discriminant function  $g(\cdot)$  for the encoding of DFA, the weight  $H$  that programs the network state transitions is strengthened by a factor  $1/\theta_{ijk}$  ( $0 < \theta_{ijk} \leq 1$ ); the range of the function  $\tilde{g}(\cdot)$  is squashed to the interval  $[0, \theta_{ijk}]$ , and it has been shifted toward the origin. Setting  $\theta_{ijk} = 1$  reduces the function (3) to the sigmoidal discriminant function (2) used for DFA encoding.

More formally, the function  $\tilde{g}(x, H, \theta)$  has the following important invariant property that will later simplify the analysis:

$$\text{Lemma 3.1: } \tilde{g}(\theta x, H, \theta) = \theta \tilde{g}(x, H, 1) = \theta g(x, H).$$

*Proof:*

$$\begin{aligned} \tilde{g}(\theta x, H, \theta) &= \frac{\theta}{1 + e^{H(\theta-2\theta x)/2\theta}} \\ &= \frac{\theta}{1 + e^{H(1-2x)/2}} = \theta \tilde{g}(x, H, 1) = \theta g(x, H). \end{aligned}$$

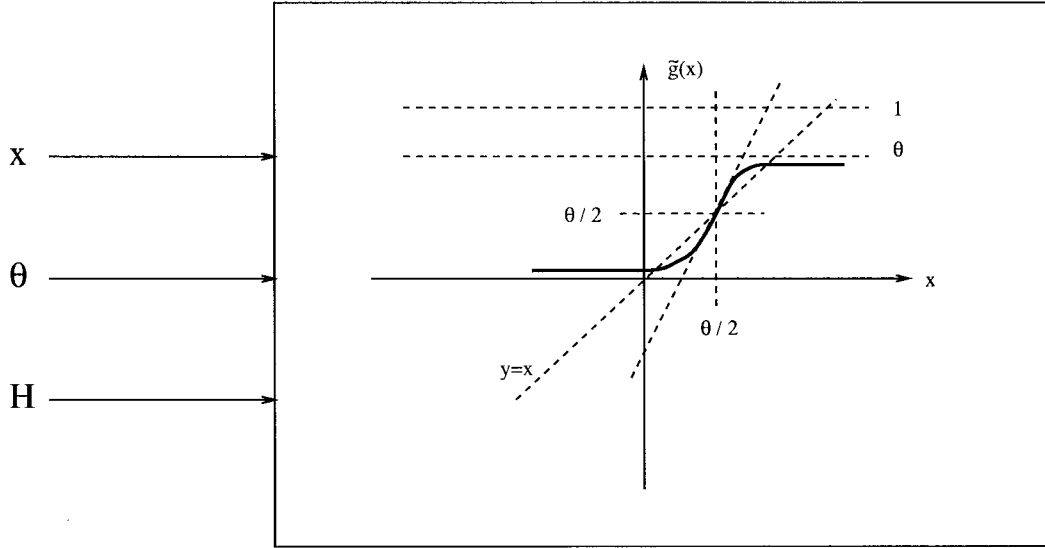
□

Thus,  $g(x, H)$  can be obtained by scaling  $\tilde{g}(x, H, 1)$  uniformly in the  $x$  and  $y$  directions by a factor  $\theta$ .

The above property of  $\tilde{g}$  allows a stability analysis of the internal FFA state representation similar to the analysis of the stability of the internal DFA state representation.

### D. Programming Fuzzy State Transitions

Consider state  $q_j$  of FFA  $M$  and the fuzzy state transition  $\delta(q_j, a_k, \{\theta_{ijk}\} = \{q_{i_1}, \dots, q_{i_r}\})$ . We assign recurrent state neuron  $S_j$  to FFA state  $q_j$  and neurons  $S_{i_1}, \dots, S_{i_r}$  to FFA states  $q_{i_1}, \dots, q_{i_r}$ . The basic idea is as follows. The activation of recurrent state neuron  $S_i$  represents the certainty



**Fig. 3.** Neuron discriminant function for fuzzy states. A neuron is represented figuratively by the box and receives as input the collective signal  $x$  from all other neurons, the weight strength  $H$ , and the fuzzy transition membership  $\theta$  to compute the function  $\bar{g}(x, H, \theta) = \theta / (1 + e^{H(\theta - 2x)/2\theta})$ . Thus, the sigmoidal discriminant function that represents FFA states has a variable output range.

$\theta_{ijk}$  with which some state transition  $\delta(q_j, a_k, \theta_{ijk}) = q_i$  is carried out, i.e.,  $S_i^{t+1} \simeq \theta_{ijk}$ . If  $q_i$  is not reached at time  $t + 1$ , then we have  $S_i^{t+1} \simeq 0$ .

We program the second-order weights  $W_{ijk}$  as follows:

$$W_{ijk} = \begin{cases} +H, & \text{if } q_i \in \delta(q_j, a_k, \theta_{ijk}) \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

$$W_{jjk} = \begin{cases} +H, & \text{if } q_j \in \delta(q_j, a_k, \theta_{jjk}) \\ -H, & \text{otherwise} \end{cases} \quad (5)$$

$$b_i = -H/2 \quad \text{if } q_i \in M. \quad (6)$$

Setting  $W_{ijk}$  to a large positive value will ensure that  $S_i^{t+1}$  will be arbitrarily close to  $\theta_{ijk}$  and setting  $W_{jjk}$  to a large negative value will guarantee that the output  $S_j^{t+1}$  will be arbitrarily close to zero. This is the same technique using for programming DFA state transitions in recurrent networks [41] and for encoding partial prior knowledge of a DFA for rule refinement [66].

#### IV. AUTOMATA TRANSFORMATION

##### A. Preliminaries

The above encoding algorithm leaves open the possibility for ambiguities when FFA are encoded in a recurrent network as follows. Consider two FFA states  $q_j$  and  $q_l$  with transitions  $\delta(q_j, a_k, \theta_{ijk}) = \delta(q_l, a_k, \theta_{ilk}) = q_i$  where  $q_i$  is one of all successor states reached from  $q_j$  and  $q_l$ , respectively, on input symbol  $a_k$ . Further assume that  $q_j$  and  $q_l$  are members of the set of current FFA states (i.e., these states are occupied with some fuzzy membership). Then, the state transition  $\delta(q_j, a_k, \theta_{ijk}) = q_i$  requires that recurrent state neuron  $S_i$  have dynamic range  $[0, \theta_{ijk}]$  while state transition  $\delta(q_l, a_k, \theta_{ilk}) = q_i$  requires that state neuron  $S_i$  asymptotically approach  $\theta_{ilk}$ . For  $\theta_{ijk} \neq \theta_{ilk}$ , we have ambiguity for the output range of neuron  $S_i$ .

*Definition 4.1:* We say an ambiguity occurs at state  $q_i$  if there exist two states  $q_j$  and  $q_l$  with  $\delta(q_j, a_k, \theta_{ijk}) = \delta(q_l, a_k, \theta_{ilk}) = q_i$  and  $\theta_{ijk} \neq \theta_{ilk}$ . A fuzzy finite state automaton  $M$  is called ambiguous if an ambiguity occurs for any state  $q_i \in M$ .

##### B. Transformation Algorithm

That ambiguity could be resolved by testing all possible paths through the FFA and identifying those states for which the above described ambiguity can occur. However, such an endeavor is computationally expensive. Instead, we propose to resolve that ambiguity by transforming any FFA  $M$ .

Before we state the transformation theorem, and give the algorithm, it will be useful to define the concept of equivalent FFA.

*Definition 4.2:* Consider a FFA  $M$  that is processing some string  $s = \sigma_1\sigma_2, \dots, \sigma_L$  with  $\sigma_i \in \Sigma$ . As  $M$  reads each symbol  $\sigma_i$ , it makes simultaneous weighted state transitions  $\Sigma \times Q \times [0, 1]$  according to the fuzzy transition map  $\delta(q_j, a_k, \theta_{ijk}) = q_i$ . The set of distinct weights  $\{\theta_{ijk}\}$  of the fuzzy transition map at time  $t$  is called the active weight set.

Note that the active weight set can change with each symbol  $\sigma_i$  processed by  $M$ . We will define what it means for two FFA to be equivalent:

*Definition 4.3:* Two FFA  $M$  and  $M'$  with alphabet  $\Sigma$  are called equivalent if their active weight sets are at all times identical for any string  $s \in \Sigma^*$ .

We will prove the following theorem.

*Theorem 4.1:* Any fuzzy finite state automaton  $M$  can be transformed into an equivalent, unambiguous fuzzy finite state automaton  $M'$ .

The tradeoff for making the resolution of ambiguities computationally feasible is an increase in the number of FFA states. The algorithm that transforms a FFA  $M$  into

Input: FFA  $M = \langle \Sigma, Q, R, Z, \delta, \omega \rangle$  with  $\Sigma = \{a_1, \dots, a_M\}$  and  $Q = \{q_1, \dots, q_N\}$ .  
Output: FFA  $M' = \langle \Sigma, Q', R', Z, \delta', \omega \rangle$  with  $\Sigma = \{a_1, \dots, a_M\}$  and  $Q' = \{q_1, \dots, q_N, q_{N+1}, \dots, q_X\}$   
with the properties

- (1)  $M \equiv M'$  and
- (2) there exist no two states  $q_j$  and  $q_l$  in  $M'$  with  $\delta(q_j, a_k, \theta_{ijk}) = \delta(q_l, a_k, \theta_{ilk}) = q_i$  with  $\theta_{ijk} \neq \theta_{ilk}$ .

Algorithm:

```

1.   $X \leftarrow N$ ;  $list \leftarrow Q$ ;
    while  $list \neq \emptyset$  do
2.       $list \leftarrow list \setminus \{q_i\}$ ;
        for  $k = 1 \dots M$  do
3.           $visit \leftarrow \emptyset$ ;
            for  $j = 1 \dots N$  do
4.                if  $\delta(q_j, a_k, \cdot) = q_i$  then  $visit \leftarrow visit \cup \{q_j\}$ ;
            end
5.           $class \leftarrow \{q_l \in visit \mid \delta(q_l, a_k, \theta_{ilk}) = q_i \text{ with } \theta_{ilk} = \theta_{ik}\}$ ;
6.           $visit \leftarrow visit \setminus \{class\}$ ;
            while  $class \neq \emptyset$  do
7.                 $class \leftarrow \{q_l \in visit \mid \delta(q_l, a_k, \theta_{ilk}) = q_i \text{ with } \theta_{ilk} = \theta_{ik}\}$ ;
8.                 $visit \leftarrow visit \setminus \{class\}$ ;
9.                 $X \leftarrow X + 1$ ;
10.                $Q \leftarrow Q \cup \{q_X\}$ ; /* create new FFA state  $q_X$  */
                for each  $q_j$  in  $class$  do
11.                    $\delta(q_j, a_k, \theta_{ijk}) \leftarrow q_X$ ; /* change transition */
                   for  $l = 1 \dots N$  do
12.                       for  $k = 1 \dots M$  do
                            $\delta(q_X, a_k, \theta_{lXk}) \leftarrow \delta(q_i, a_k, \theta_{ilk})$ ; /* implies  $\theta_{lXk} \leftarrow \theta_{ilk}$  */
                       end
                   end
                end
            end
        end
    end

```

Fig. 4. Algorithm for the FFA transformation.

a FFA  $M'$  such that  $L(M) = L(M')$  is shown in Fig. 4. Before we prove the above theorem, we will discuss an example of FFA transformation.

### C. Example

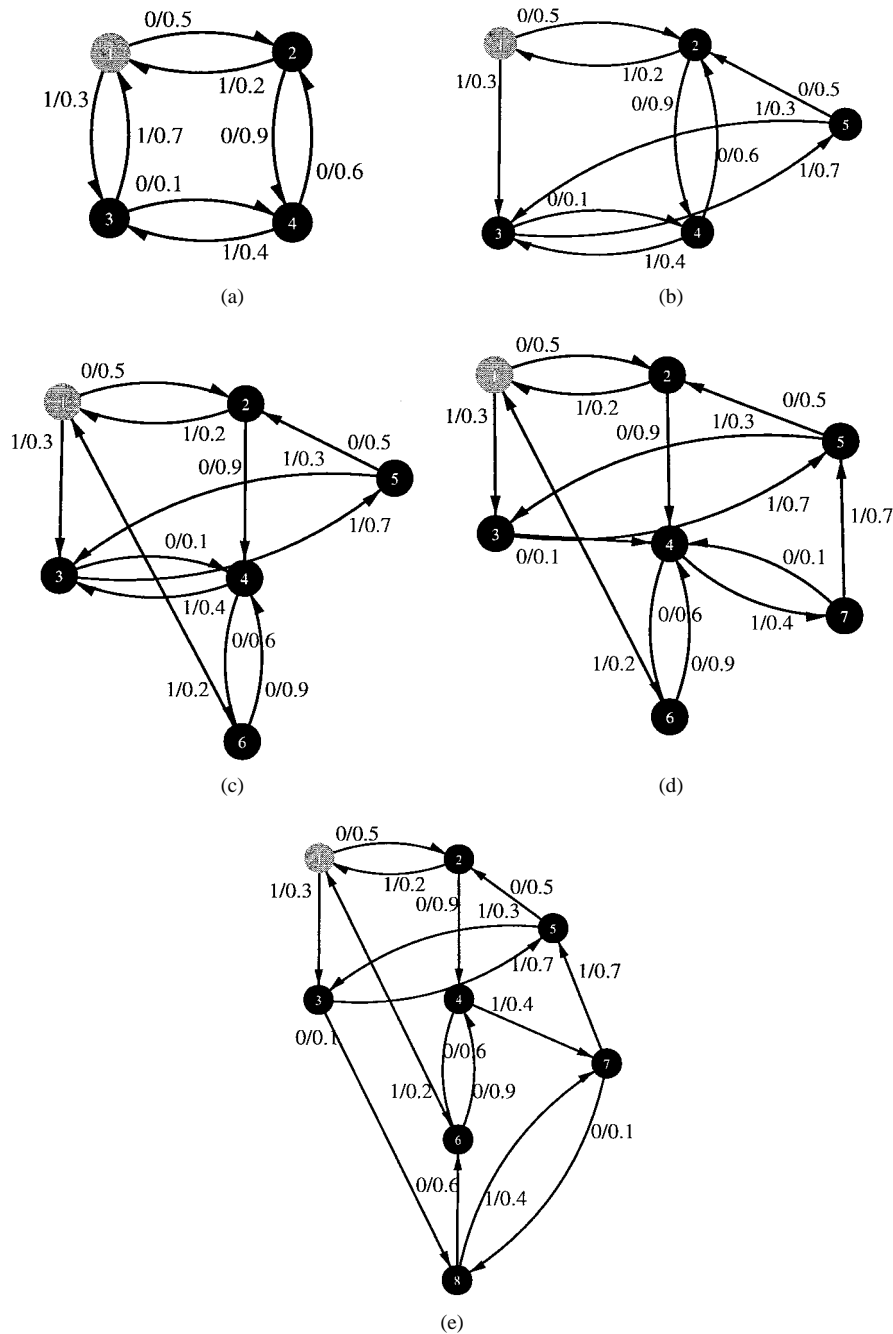
Consider the FFA shown in Fig. 5(a) with four states and input alphabet  $\Sigma = \{0, 1\}$ ; state  $q_1$  is the start state.<sup>8</sup> The algorithm initializes the variable “list” with all FFA states, i.e.,  $list = \{q_1, q_2, q_3, q_4\}$ . First, we notice that no ambiguity exists for input symbol ‘0’ at state  $q_1$  since there are no state transitions  $\delta(\cdot, 0, \cdot) = q_1$ . There exist two state transitions that have state  $q_1$  as its target, i.e.,  $\delta(q_2, 1, 0.2) = \delta(q_3, 1, 0.7) = q_1$ . Thus, we set the variable  $visit = \{q_2, q_3\}$ . According to Definition 4.1, an ambiguity exists since  $\theta_{121} \neq \theta_{131}$ . We resolve that ambiguity by introducing a new state  $q_5$  and setting  $\delta(q_3, 1, 0.7) = q_5$ . Since  $\delta(q_3, 1, 0.7)$  no longer leads to state  $q_1$ , we need to introduce new state transitions leading from state  $q_5$  to the target states  $\{q\}$  of all possible state transitions:  $\delta(q_1, \cdot, \cdot) = \{q_2, q_3\}$ . Thus, we set  $\delta(q_5, 0, \theta_{250}) = q_2$  and  $\delta(q_5, 1, \theta_{351}) = q_3$  with  $\theta_{250} = \theta_{210}$  and  $\theta_{351} = \theta_{311}$ .

<sup>8</sup>The FFA shown in Fig. 5(a) is a special case in that it does not contain any fuzzy transitions. Since the objective of the transformation algorithm is to resolve ambiguities for states  $q_i$  with  $\delta(\{q_{j_1}, \dots, q_{j_r}\}, a_k, \{\theta_{ij_1k}, \dots, \theta_{ij_rk}\}) = q_i$ , fuzziness is of no relevance; therefore, we omitted it for reasons of simplicity.

One iteration through the outer loop thus results in the FFA shown in Fig. 5(b), another in Fig. 5(c). Consider Fig. 5(d) which shows the FFA after three iterations. State  $q_4$  is the only state left that has incoming transitions  $\delta(\cdot, a_k, \theta_{4.k}) = q_4$  where not all values  $\theta_{4.k}$  are identical. We have  $\delta(q_2, 0, 0.9) = \delta(q_6, 0, 0.9) = q_4$ ; since these two state transition do not cause an ambiguity for input symbol ‘0,’ we leave these state transitions as they are. However, we also have  $\delta(q_2, 0, \theta_{420}) = \delta(q_3, 0, \theta_{430}) = \delta(q_7, 0, \theta_{470}) = q_4$  with  $\theta_{430} = \theta_{470} \neq \theta_{420} = 0.9$ . Instead of creating new states for both state transitions  $\delta(q_3, 0, \theta_{430})$  and  $\delta(q_7, 0, \theta_{470})$ , it suffices to create one new state  $q_8$  and to set  $\delta(q_3, 0, 0.1) = \delta(q_7, 0, 0.1) = q_8$ . States  $q_6$  and  $q_7$  are the only possible successor states on input symbols ‘0’ and ‘1,’ respectively. Thus, we set  $\delta(q_8, 0, 0.6) = q_6$  and  $\delta(q_8, 1, 0.4) = q_7$ . There exist no more ambiguities and the algorithm terminates [Fig. 5(e)].

### D. Properties of the Transformation Algorithm

We have shown with an example how the algorithm transforms any FFA  $M$  into a FFA  $M'$  without ambiguities. We now need to show that the algorithm correctly transforms  $M$  into  $M'$ , i.e., we need to show that  $M$  and  $M'$  are equivalent. In addition, we also need to demonstrate that the algorithm terminates for any input  $M$ .



**Fig. 5.** Example of FFA transformation. Transition weight ambiguities are resolved in a sequence of steps: (a) the original FFA—there exist ambiguities for all four states; (b) the ambiguity of transition from state 3 to state 1 on input symbol 1 is removed by adding a new state 5; (c) the ambiguity of transition from state 4 to state 2 on input symbol 0 is removed by adding a new state 6; (d) the ambiguity of transition from state 4 to state 3 on input symbol 1 is removed by adding a new state 7; (e) the ambiguity of transition from states 3 and 7—both transition have the same fuzzy membership—to state 4 is removed by adding a new state 8.

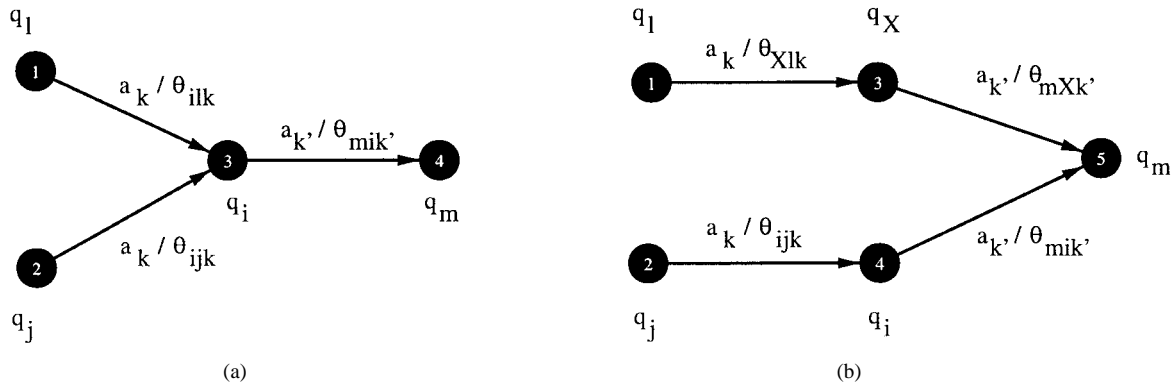
First, we prove the following property of the transformation algorithm:

*Lemma 4.1:* Resolution of an ambiguity does not result in a new ambiguity.

*Proof:* Consider the situation illustrated in Fig. 6(a). Let  $q_i, q_j, q_l$ , and  $q_m$  be four FFA states and let there be an ambiguity at state  $q_i$  on input symbol  $a_k$ , i.e.,  $\delta(q_j, a_k, \theta_{ijk}) = \delta(q_l, a_k, \theta_{ilk}) = q_i$  with  $\theta_{ijk} \neq \theta_{ilk}$ . Furthermore, let  $\delta(q_i, a_{k'}, \theta_{mik'}) = q_m$ . The ambiguity is resolved by creating a new state  $q_X$ . We arbitrarily

choose the state transition  $\delta(q_l, a_k, \theta_{ilk}) = q_i$  and set  $\delta(q_l, a_k, \theta_{Xlk}) = q_X$  with  $\theta_{Xlk} = \theta_{ilk}$ . This removes the ambiguity at state  $q_i$ . We now need to introduce a new state transition  $\delta(q_X, a_{k'}, \theta_{mXk'}) = q_m$ . By observing that  $\theta_{mXk'} = \theta_{mik'}$  we conclude that no new ambiguity has been created at state  $q_m$  following the resolution of the ambiguity at state  $q_i$ .  $\square$

We observe that  $M'$  is not unique, i.e., the order in which states are visited and the order in which state transition ambiguities are resolved determine the final FFA



**Fig. 6.** Resolution of ambiguities. The transition ambiguity from states  $q_l$  and  $q_j$  to state  $q_i$  on input symbol  $a_k$  is resolved by adding a new state  $q_X$  and adjusting the transition as shown.

$M'$ . Consider the FFA in Fig. 5(a). In our example, if we had chosen to change transition  $\delta(q_2, 1, 0.2) = q_1$  instead of state transition  $\delta(q_3, 1, 0.7) = q_1$ , then the resulting FFA  $M'$  would have been different. However, all possible transformations  $M'$  share a common invariant property.

*Lemma 4.2:* The number of states in  $M'$  is constant regardless of the order in which states are visited and state transition ambiguities are resolved.

*Proof:* To see that the lemma's claim holds true, we observe that resolving an ambiguity consists of creating a new state for each set of states  $\{q_j\}$  with  $\delta(q_j, a_k, \theta_{ijk}) = q_i$  with  $\forall j \neq j': \theta_{ijk} \neq \theta_{i'jk}$ . Since resolving the ambiguity for any state  $q_i$  does not introduce new ambiguities (see Lemma 4.1), the number of newly created states only depends on the number FFA states with ambiguities.  $\square$

The following definitions will be convenient.

*Definition 4.4:* The outdegree  $d_{out}(q_i)$  of a state  $q_i$  in FFA  $M$  is the maximum number of states  $q_j$  for which we have  $\delta(q_i, a_k, \theta_{ijk}) = \{q_j\}$  for fixed  $a_k$  with  $\theta_{ijk} > 0$  where the maximum is taken over all symbols  $a_k$ . The maximum outdegree  $D_{out}(M)$  of some FFA  $M$  is the maximum over all  $d_{out}(q_i)$  with  $q_i \in M$ .

*Definition 4.5:* The indegree  $d_{in}(q_i)$  of a state  $q_i$  in FFA  $M$  is the maximum number of states  $q_j$  for which we have  $\delta(\{q_j\}, a_k, \theta_{ijk}) = q_i$  for fixed  $a_k$  with  $\theta_{ijk} > 0$  where the maximum is taken over all symbols  $a_k$ . The maximum indegree  $D_{in}(M)$  of some FFA  $M$  is the maximum over all  $d_{in}(q_i)$  with  $q_i \in M$ .

We can give a very loose upper bound for the number of states in  $M'$  as follows.

*Lemma 4.3:* For a fuzzy finite state automaton  $M$  with  $N$  states and  $K$  input symbols, the transformed FFA has at most  $D_{in}KN(N-1)$  states.

*Proof:* Consider some arbitrary state  $q_i$  of  $M$ . It can have at most  $D_{in}N$  incoming transitions for input symbol  $a_k$ . The resolution of ambiguity for state  $q_i$  requires that all but one transition  $\delta(\cdot, a_k, \theta_{i\cdot k})$  lead to a new state. In the case where the fuzzy transition memberships  $\theta_{i\cdot k}$  are all different,  $N-1$  new states are created per ambiguous state. Thus, for  $K$  input symbols, at the most  $D_{in}KN(N-1)$  new states are created.  $\square$

The results in Table 1 show the size of randomly generated FFA  $M$  with input alphabet  $\{0, 1\}$ , the maximum outdegree  $D_{out}(M)$ , the upper bound on the size of transformed FFA  $M'$ , and average and standard deviation of actual sizes for transformed FFA  $M'$  taken over 100 experiments. The random FFA  $M$  were generated by connecting each state of  $M$  to at most  $D_{out}$  other states for given input symbol. We observe that the average actual size of transformed FFA depends on the maximum outdegree  $D_{out}(M)$  and appears to be linear in  $N$  and  $D_{out}$ . Lemma 4.3 has the following corollary:

*Corollary 4.1:* The FFA transformation algorithm terminates for all possible FFA.

*Proof:* The size of the set list in the algorithm decreases monotonically with each iteration. Thus, the outer while loop terminates when list =  $\emptyset$ . Likewise, the inner while loop terminates since there is only a finite number of states  $q_i$  in the set "class" and the size of that set monotonically decreases with each iteration. Thus, the algorithm terminates.  $\square$

We now return to the proof of Theorem 4.1. We have already proven that the applying the FFA transformation algorithm results in a FFA where no ambiguities exist. It is easy to see that the a transformed FFA  $M'$  is equivalent with the original FFA  $M$  since no new fuzzy transition memberships have been added, and the algorithm leaves unchanged the order in which FFA transitions are executed. This completes the proof of Theorem 4.1.  $\square$

The above transformation algorithm removes all ambiguities for incoming transitions. However, a minor adjustment for the neural FFA encoding is needed. Given the FFA state  $q_i$  with  $\delta(q_j, a_k, \theta_{ijk}) = q_i$  and  $\delta(q_j, a_k, \cdot) \neq q_i$ , the corresponding weight  $W_{ijk}$  is set to  $-H$ . We also need to specify an implicit value  $\theta_{iik} > 0$  for the neural FFA encoding even though we have  $\theta_{iik} = 0$  in the FFA. In order to be consistent with regard to neurons with variable output range, we set  $\theta_{iik} = \theta_{ijk}$ .

## V. NETWORK ARCHITECTURE

The architecture for representing FFA is shown in Fig. 7. A layer of sparsely connected recurrent neurons implements



**Table 1**

Scaling of a Transformed FFA. The Results Show the Increase of the Size of  $M$  FFA Due to Its Transformation into an Fuzzy Finite State Automaton  $M'$  Without Ambiguities as a Function of the Size of  $M$  and the Maximum Outdegree  $D_{out}(M)$ ; the FFA Were Randomly Generated and the Average Was Computed over 100 Transformations; The Average Size of Transformed FFA Appears to Be Linear in  $N$  and  $D_{out}$

size of $M$	$D_{out}(M)$	upper limit on size of $M'$	average size of $M'$	standard deviation
10	1	180	12	2
	2	360	16	5
	3	540	19	15
	4	720	25	29
	5	900	28	84
20	1	760	25	6
	2	1520	32	19
	3	2280	40	40
	4	3040	50	191
	5	3800	68	278
30	1	1740	38	7
	2	3480	49	27
	3	5220	61	64
	4	6960	84	266
	5	8700	111	578
40	1	2400	51	6
	2	4800	65	29
	3	7200	85	104
	4	9600	117	342
	5	12000	154	1057
50	1	4900	65	14
	2	9800	84	41
	3	14700	107	217
	4	19600	154	704
	5	24500	198	1478
100	1	19800	129	26
	2	39600	161	64
	3	59400	215	285
	4	78800	309	1845
	5	98600	401	3916

the finite state dynamics. Each neuron  $S_i$  of the state transition module has a dynamical output range  $[0, \theta_{ijk}]$  where  $\theta_{ijk}$  is the rule weight in the FFA state transition  $\delta(q_j, a_k, \theta_{ijk}) = q_i$ . Notice that each neuron  $S_i$  is only connected to pairs  $(S_i, I_k)$  for which  $\theta_{ijk} = \theta_{i'j'k}$  since we assume that  $M$  is transformed into an equivalent, unambiguous FFA  $M'$  prior to the network construction. The weights  $W_{ijk}$  are programmed as described in Section III-D. Each recurrent state neuron receives as inputs the value  $S_j^t$  and an output range value  $\theta_{ijk}$ ; it computes its output according to (3).

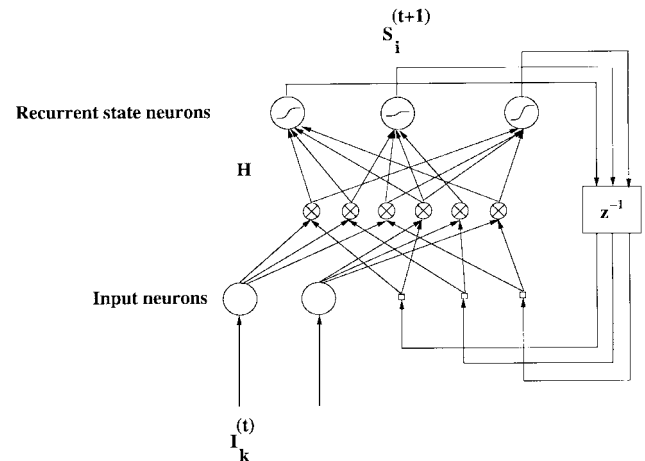
## VI. NETWORK STABILITY ANALYSIS

### A. Preliminaries

In order to demonstrate how the FFA encoding algorithm achieves stability of the internal FFA state representation for indefinite periods of time, we need to understand the dynamics of signals in a constructed RNN.

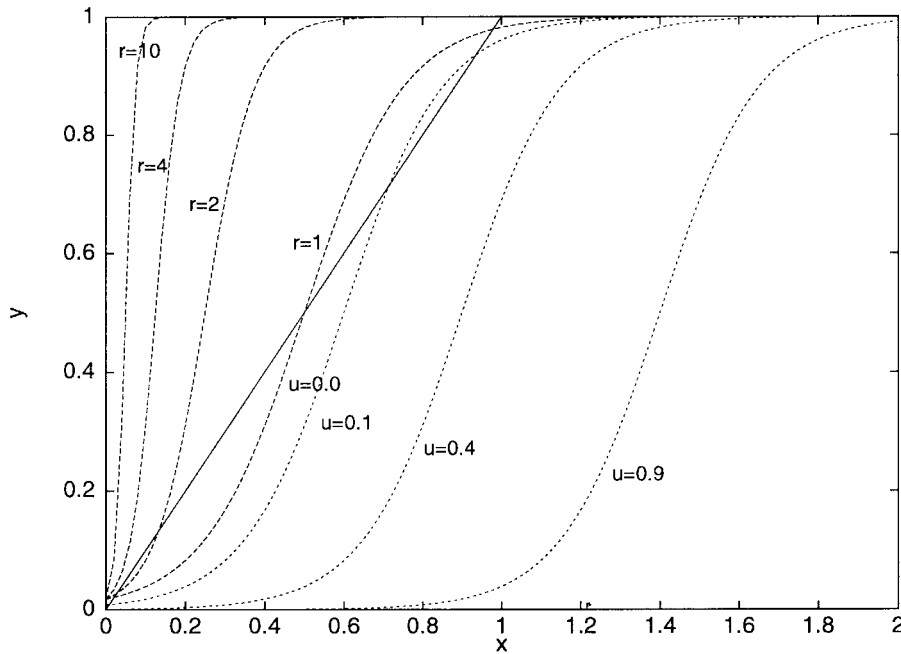
We define stability of an internal FFA state representation as follows.

*Definition 6.1:* A fuzzy encoding of FFA states with transition weights  $\{\theta_{ijk}\}$  in a second-order recurrent neural



**Fig. 7.** Network architecture for FFA representation. The architecture for representing FFA differs from that for DFA in that: 1) the recurrent state neurons have variable output range; 2) the resolution of ambiguities causes a sparser interconnection topology; and 3) there is no dedicated output neuron.

network is called stable if only state neurons corresponding to the set of current FFA states have an output greater than  $\theta_{ijk}/2$  where  $\theta_{ijk}$  is the dynamic range of recurrent



**Fig. 8.** Fixed points of the sigmoidal discriminant function. Shown are the graphs of the function  $f(x, H, 1, r) = 1/(1 + e^{H(1-2r^x)/2})$  (dashed graphs) for  $H = 8$  and  $r = \{1, 2, 4, 10\}$  and the function  $p(x, u) = 1/(1 + e^{H(1-2(x-u))/2})$  (dotted graphs) for  $H = 8$  and  $u = \{0.0, 0.1, 0.4, 0.9\}$ . Their intersection with the function  $y = x$  shows the existence and location of fixed points. In this example,  $f(x, r)$  has three fixed points for  $r = \{1, 2\}$ , but only one fixed point for  $r = \{4, 10\}$ , and  $p(x, u)$  has three fixed points for  $u = \{0.0, 0.1\}$ , but only one fixed point for  $u = \{0.4, 0.9\}$ .

state neurons, and all remaining recurrent neurons have low output signals less than  $\theta_{ijk}/2$  for all possible input sequences.

It follows from that definition that there exists an upper bound  $0 < \phi^- < \theta_{ijk}/2$  for low signals and a lower bound  $\theta_{ijk}/2 < \phi^+ < \theta_{ijk}$  for high signals in networks that represent stable FFA encodings. The ideal values for low and high signals are zero and  $\theta_{ijk}$ , respectively.

A detailed analysis of all possible network state changes in [41] revealed that, for the purpose of demonstrating stability of internal finite state representations, it is sufficient to consider the following two worst cases: 1) a neuron that does not correspond to a current fuzzy automaton state receives the same residual low input from all other neurons that it is connected to, and that value is identical for all neurons, and 2) a neuron that changes its output from low to high at the next time step receives input only from one other neuron (i.e., the neuron which corresponds to the current fuzzy automaton state), and it may inhibit itself. In the case of FFA, a neuron  $S_i$  undergoing a state change from  $S_i^t \approx 0$  to  $S_i^{t+1} \approx \theta_{ijk}$  may receive principal inputs from more than one other neuron. However, any additional input only serves to strengthen high signals. Thus, the case of a neuron receiving principal input from exactly one other neuron represents a worst case.

### B. Fixed Point Analysis for Sigmoidal Discriminant Function

Here, we summarize without proofs some of the results that we used to demonstrate stability of neural DFA encodings; details of the proofs can be found in [41].

In order to guarantee low signals to remain low, we have to give a tight upper bound for low signals that remains valid for an arbitrary number of time steps.

*Lemma 6.1:* The low signals are bounded from above by the fixed point  $[\phi_f^-]_\theta$  of the function  $f$

$$\begin{aligned} f^0 &= 0 \\ f^{t+1} &= \tilde{g}(r \cdot f^t) \end{aligned} \quad (7)$$

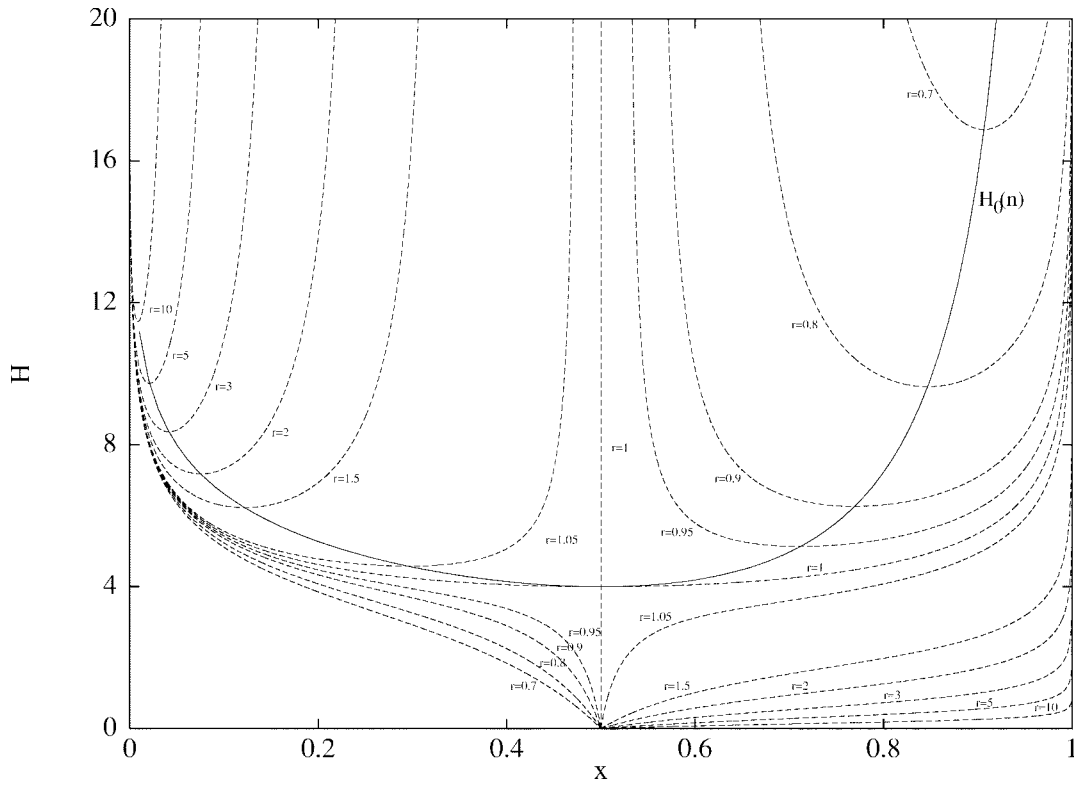
where  $[\phi_f^-]_\theta$  represents the fixed point of the discriminant function  $\tilde{g}(\cdot)$  with variable output range  $\theta$ , and  $r$  denotes the maximum number of neurons that contribute to a neuron's input. For reasons of simplicity, we will write  $\phi_f^-$  for  $[\phi_f^-]_\theta$  with the implicit understanding that the location of fixed points depends on the particular choice of  $\theta$ . This lemma can easily be proven by induction on  $t$ .

It is easy to see that the function to be iterated in (7) is  $f(x, H, \theta, r) = \theta/(1 + e^{H(\theta-2rx)/2\theta})$  and fixed points are shown in Fig. 8. The graphs of the function for  $\theta = 1.0$  are shown in Fig. 9 for different values of the parameter  $r$ . It is obvious that the location of fixed points depends on the particular values of  $\theta$ . We will show later in this section that the conditions that guarantee the existence of one or three fixed points are independent of the parameter  $\theta$ .

The function  $f(x, H, \theta, r)$  has some desirable properties.

*Lemma 6.2:* For any  $H > 0$ , the function  $f(x, H, \theta, r)$  has at least one fixed point  $\phi_f^0$ .

*Lemma 6.3:* There exists a value  $H_0^-(r)$  such that for any  $H > H_0^-(r)$ ,  $f(x, H, \theta, r)$  has three fixed points  $0 < \phi_f^- < \phi_f^0 < \phi_f^+ < \theta$ .



**Fig. 9.** Contour plot of  $f(x, H, \theta, r) = x$ . The contour plots (dotted graphs) show the relationship between  $H$  and  $x$  for various values of  $r$  and fixed value  $\theta = 1$ . If  $H$  is chosen such that  $H > \max(H_0^-(r), H_0^+(r))$  (solid graphs), then a line parallel to the  $x$  axis intersects the surface satisfying  $f(x, H, \theta, r) = x$  in three points which are the fixed points of  $f(x, \theta, r)$ .

*Lemma 6.4:* If  $f(x, H, \theta, r)$  has three fixed points  $\phi_f^-, \phi_f^0$ , and  $\phi_f^+$ , then

$$\lim_{t \rightarrow \infty} f^t = \begin{cases} \phi_f^-, & x_0 < \phi_f^0 \\ \phi_f^0, & x_0 = \phi_f^0 \\ \phi_f^+, & x_0 > \phi_f^0 \end{cases} \quad (8)$$

where  $x_0$  is an initial value for the iteration of  $f(\cdot)$ .

The above lemma can be proven by defining an appropriate Lyapunov function  $P$  and showing that  $P$  has minima at  $\phi_f^-$  and  $\phi_f^+$ .<sup>9</sup>

The basic idea behind the network stability analysis is to show that neuron outputs never exceed or fall below some fixed points  $\phi^-$  and  $\phi^+$ , respectively. The fixed points  $\phi_f^-$  and  $\phi_f^+$  are only valid upper and lower bounds on low and high signals, respectively, if convergence toward these fixed points is monotone. The following corollary establishes monotone convergence of  $f^t$  toward fixed points:

*Corollary 6.1:* Let  $f^0, f^1, f^2, \dots$  denote the finite sequence computed by successive iteration of the function  $f$ . Then we have  $f^0 < f^1 < \dots < \phi_f$  where  $\phi_f$  is one of the stable fixed points of  $f(x, H, \theta, r)$ .

<sup>9</sup>Lyapunov functions can be used to prove the stability of dynamical systems [67]. For a given dynamical system  $S$ , if there exists a function  $P$  (we can think of  $P$  as an energy function) such that  $P$  has at least one minimum, then  $S$  has a stable state. Here, we can choose  $P(x_i) = (x_i - \phi_f)^2$  where  $x_i$  is the value of  $f(\cdot)$  after  $i$  iterations and  $\phi$  is one of the fixed points. It can be shown algebraically that, for  $x_0 \neq \phi_f^0$ ,  $P(x_i)$  decreases with every step of the iteration of  $f(\cdot)$  until a stable fixed point is reached.

With these properties, we can quantify the value  $H_0^-(r)$  such that for any  $H > H_0^-(r)$ ,  $f(x, H, \theta, r)$  has three fixed points. The low and high fixed points  $\phi_f^-$  and  $\phi_f^+$  are the bounds for low and high signals, respectively. The larger  $r$ , the larger  $H$  must be chosen in order to guarantee the existence of three fixed points. If  $H$  is not chosen sufficiently large, then  $f^t$  converges to a unique fixed point  $\theta/2 < \phi_f < \theta$ . The following lemma expresses a quantitative condition that guarantees the existence of three fixed points.

*Lemma 6.5:* The function  $f(x, H, \theta, r) = (\theta/1 + e^{H(\theta-2rx)/2\theta})$  has three fixed points  $0 < \phi_f^- < \phi_f^0 < \phi_f^+ < \theta$  if  $H$  is chosen such that

$$H > H_0^-(r) = \frac{2 \left( \theta + (\theta - x) \log \left( \frac{\theta - x}{x} \right) \right)}{\theta - x}$$

where  $x$  satisfies the equation

$$r = \frac{\theta^2}{2x \left( \theta + (\theta - x) \log \left( \frac{\theta - x}{x} \right) \right)}$$

*Proof:* We only present a sketch of the proof; for a complete proof, see [41]. Fixed points of the function  $f(x, H, \theta, r)$  satisfy the equation  $(\theta/1 + e^{H(\theta-2rx)/2\theta}) = x$ . Given the parameter  $r$ , we must find a minimum value

$H_0^-(r)$  such that  $f(x, H, \theta, r)$  has three fixed points. We can think of  $x, r$ , and  $H$  as coordinates in a three-dimensional Euclidean space. Then the locus of points  $(x, r, H)$  satisfying relation the above equation is a curved surface. What we are interested in is the number of points where a line parallel to the  $x$  axis intersects this surface.

Unfortunately, the fixed point equation cannot be solved explicitly for  $x$  as a function of  $r$  and  $H$ . However, it can be solved for either of the other parameters, giving the intersections with lines parallel to the  $r$  axis or the  $H$  axis

$$r = r(x, \theta, H) = \frac{\theta}{2x} - \frac{\theta \log\left(\frac{\theta - x}{x}\right)}{Hx} \quad (9)$$

$$H = H(r, \theta, x) = \frac{2\theta \log\left(\frac{\theta - x}{x}\right)}{\theta - 2rx}. \quad (10)$$

The contours of these functions show the relationship between  $H$  and  $x$  when  $r$  is fixed (Fig. 9). We need to find the point on each contour where the tangent is parallel to the  $x$  axis, which will indicate where the transition occurs between one and three solutions for  $f(x, H, \theta, r) = x$ . Solving  $(\partial r(x, \theta, H)/\partial x) = 0$ , we obtain the conditions of the lemma.  $\square$

Even though the location of fixed points of the function  $f$  depends on  $H, r$ , and  $\theta$ , we will use  $[\phi_f]_\theta$  as a generic name for any fixed point of a function  $f$ .

Similarly, we can quantify high signals in a constructed network.

*Lemma 6.6:* The high signals are bounded from below by the fixed point  $[\phi_h^+]_\theta$  of the function

$$\begin{aligned} h^0 &= 1 \\ h^{t+1} &= \tilde{g}(h^t - f^t). \end{aligned} \quad (11)$$

Notice that the above recurrence relation couples  $f^t$  and  $h^t$  which makes it difficult if not impossible to find a function  $h(x, \theta, r)$  which when iterated gives the same values as  $h^t$ . However, we can bound the sequence  $h^0, h^1, h^2, \dots$  from below by a recursively defined function  $p^t$ , i.e.,  $\forall t: p^t \leq h^t$ , which decouples  $h^t$  from  $f^t$ .

*Lemma 6.7:* Let  $[\phi_f]_\theta$  denote the fixed point of the recursive function  $f$ , i.e.,  $\lim_{t \rightarrow \infty} f^t = [\phi_f]_\theta$ . Then the recursively defined function  $p$

$$\begin{aligned} p^0 &= 1 \\ p^{t+1} &= \tilde{g}(g^t - [\phi_f]_\theta) \end{aligned} \quad (12)$$

has the property that  $\forall t: p^t \leq h^t$ .

Then, we have the following sufficient condition for the existence of two stable fixed point of the function defined in (11).

*Lemma 6.8:* Let the iterative function  $p^t$  have two stable fixed points and  $\forall t: p^t \leq h^t$ . Then the function  $h^t$  has also two stable fixed points.

The above lemma has the following corollary.

*Corollary 6.2:* A constructed network's high signals remain stable if the sequence  $p^0, p^1, p^2, \dots$  converges toward the fixed point  $\theta/2 < [\phi_p^+]_\theta < \theta$ .

Since we have decoupled the iterated function  $h^t$  from the iterated function  $f^t$  by introducing the iterated function  $p^t$ , we can apply the same technique to  $p^t$  for finding conditions for the existence of fixed points as in the case of  $f^t$ . In fact, the function that when iterated generates the sequence  $p^0, p^1, p^2, \dots$  is defined by

$$\begin{aligned} p(r, \theta, x) &= \frac{\theta}{1 + e^{H(\theta - 2(x - [\phi_f^-]_\theta))/2\theta}} \\ &= \frac{\theta}{1 + e^{H'(\theta - 2r'x)/2\theta}} \end{aligned} \quad (13)$$

with

$$H' = H(1 + 2[\phi_f^-]_\theta), r' = \frac{1}{1 + 2[\phi_f^-]_\theta}. \quad (14)$$

We can iteratively compute the value of  $[\phi_p]_\theta$  for given parameters  $H$  and  $r$ . Thus, we can repeat the original argument with  $H'$  and  $r'$  in place of  $H$  and  $r$  to find the conditions under which  $p(r, x)$  and thus  $g(r, x)$  have three fixed points.

*Lemma 6.9:* The function  $p(x, [\phi_f^-]_\theta) = (1/1 + e^{H(\theta - 2(x - [\phi_f^-]_\theta))/2\theta})$  has three fixed points  $0 < [\phi_p^-]_\theta < [\phi_p^0]_\theta < [\phi_p^+]_\theta < 1$  if  $H$  is chosen such that

$$H > H_0^+(r) = \frac{2\left(\theta + (\theta - x) \log\left(\frac{\theta - x}{x}\right)\right)}{(1 + 2[\phi_f^-]_\theta)(\theta - x)}$$

where  $x$  satisfies the equation

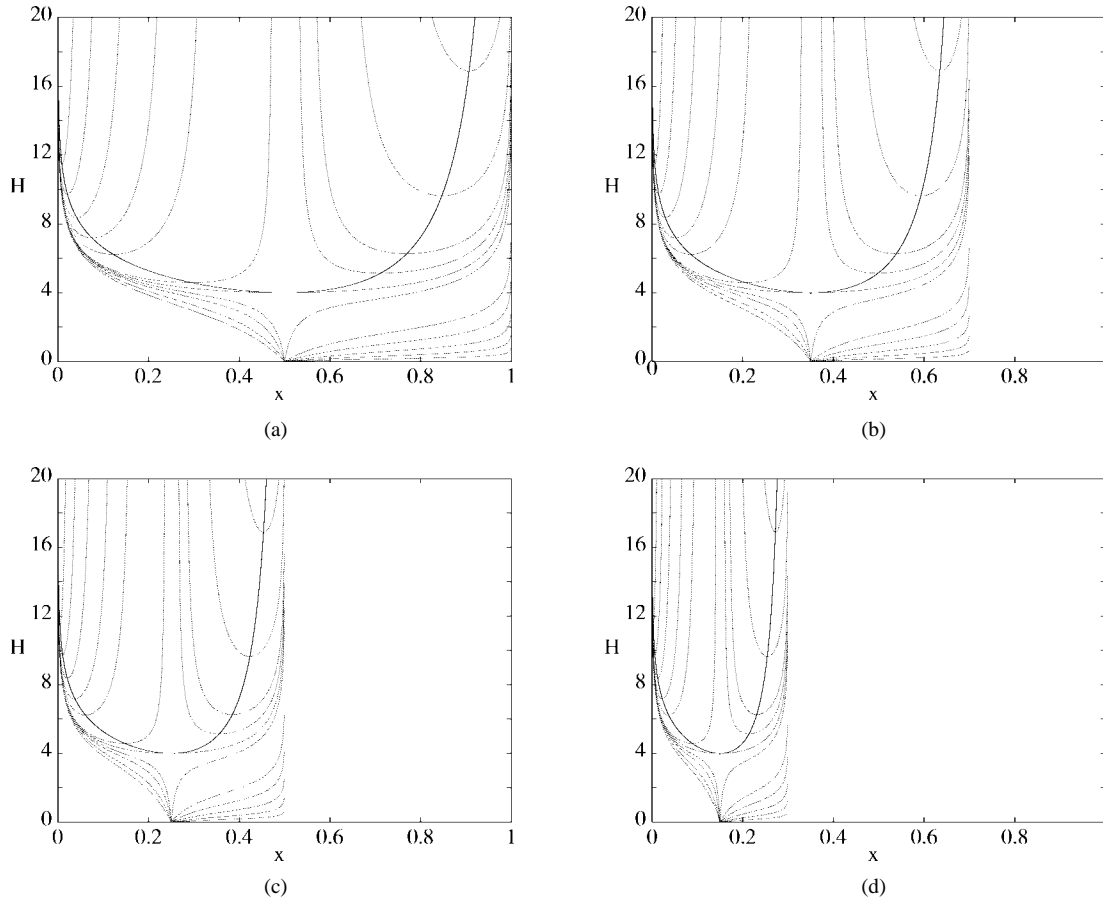
$$\frac{1}{1 + 2[\phi_f^-]_\theta} = \frac{\theta^2}{2x\left(\theta + (\theta - x) \log\left(\frac{\theta - x}{x}\right)\right)}.$$

$\square$

Since there is a collection of fuzzy transition memberships  $\theta_{ijk}$  involved in the algorithm for constructing fuzzy representations of FFA, we need to determine whether the condition of Lemmas 6.5 and 6.9 hold true for all rule weights  $\theta_{ijk}$ . The following corollary establishes a useful invariant property of the function  $H_0(x, r, \theta)$ .

*Corollary 6.3:* The value of the minima  $H(x, r, \theta)$  only depends on the value of  $r$  and is independent of the particular values of  $\theta$

$$\inf H(x, r, \theta) = \inf \frac{2\theta \log\left(\frac{\theta - x}{x}\right)}{\theta - 2rx} = H_0(r). \quad (15)$$



**Fig. 10.** Invariant fixed points. The contour plots illustrating the existence and location of fixed points of the function  $\bar{g}(x, H, \theta, r) = \theta / (1 + e^{H(\theta-2rx)/2\theta})$  are shown for: (a)  $\theta = 1.0$ ; (b)  $\theta = 0.7$ ; (c)  $\theta = 0.5$ ; and (d)  $\theta = 0.3$ . The location of fixed points depends on the value of  $\theta$ . The condition on  $H$  and  $r$  for the existence of one versus two stable fixed points is independent of  $\theta$ . The scaling of the graphs illustrates that invariant property.

*Proof:* The term  $\log(\theta - x/x)$  scales the function  $H(x, r, \theta)$  along the  $x$  axis. We introduce a scaling factor  $\epsilon$  and set  $\theta' = \epsilon\theta$  and  $x' = \epsilon x$ . Then, (10) becomes

$$\begin{aligned}
 H_\epsilon(x', r, \theta') &= \frac{2\epsilon\theta \log\left(\frac{\epsilon\theta - \epsilon x}{\epsilon x}\right)}{\epsilon\theta - 2r\epsilon x} = \frac{2\epsilon\theta \log\left(\frac{\theta - x}{x}\right)}{\epsilon(\theta - 2rx)} \\
 &= \frac{2\theta \log\left(\frac{\theta - x}{x}\right)}{\theta - 2rx} = H(x, r, \theta) \quad (16)
 \end{aligned}$$

for fixed  $r$ . Thus the values of  $H(x, r, \theta)$  are identical for fixed values of  $r$ , and their local minima have the same values independent of  $\theta$ .

The relevance of the above corollary is that there is no need to test conditions for all possible values of  $\theta$  in order to guarantee the existence of fixed points. The graphs in Fig. 10 illustrate that invariant property of the sigmoidal discriminant function.

We can now proceed to prove stability of low and high signals, and thus stability of the fuzzy representation of FFA states, in a constructed recurrent neural network.

### C. Network Stability

The existence of two stable fixed points of the discriminant function is only a necessary condition for network stability. We also need to establish conditions under which these fixed points are upper and lower bounds of stable low and high signals, respectively.

Before we define and derive the conditions for network stability, it is convenient to apply the result of Lemma 3.1 to the fixed points of the sigmoidal discriminant function (Section III-C).

*Corollary 6.4:* For any value  $\theta$  with  $0 < \theta \leq 1$ , the fixed points  $[\phi]_\theta$  of the discriminant function

$$\frac{\theta}{1 + e^{H(\theta-2rx)/2\theta}}$$

have the following invariant relationship:

$$[\phi]_\theta = \theta[\phi]_1.$$

*Proof:* By definition, fixed points  $\phi$  of  $\tilde{g}(\cdot)$  have the property that  $[\phi]_\theta = \tilde{g}([\phi]_\theta)$ . According to Lemma 3.1, we also have

$$\begin{aligned}
 [\phi]_\theta &= \tilde{g}([\phi]_\theta) = \tilde{g}([\phi]_\theta, H, \theta) \\
 &= \theta \tilde{g}(\theta[\phi]_1, H, 1) = \theta \tilde{g}([\phi]_1) = \theta[\phi]_1
 \end{aligned}$$

because the invariant scaling property applies to all points of the function  $\tilde{g}$ , including its fixed points. Thus, we do not have to consider the conditions separately for all values of  $\theta$  that occur in a given fuzzy finite state automaton.  $\square$

We now redefine stability of recurrent networks constructed from DFA in terms of fixed points.

*Definition 6.2:* An encoding of DFA states in a second-order recurrent neural network is called stable if all the low signals are less than  $[\phi_f^0]_{\theta_i}$ , and all the high signals are greater than  $[\phi_h^0]_{\theta_i}$  for all  $\theta_i$  of all state neurons  $S_i$ .

We have simplified  $\theta_i \dots$  to  $\theta_i$  because the output of each neuron  $S_i$  has a fixed upper limit  $\theta$  for a given input symbol, regardless which neurons  $S_j$  contribute residual inputs. We note that this new definition is stricter than that we gave in Definition 6.1. In order for the low signal to remain stable, the following condition has to be satisfied:

$$-\frac{H}{2} + Hr[\phi_f^-]_{\theta_j} < [\phi_f^0]_{\theta_j}. \quad (17)$$

Similarly, the following inequality must be satisfied for stable high signals:

$$-\frac{H}{2} + H[\phi_h^+]_{\theta_j} - H[\phi_f^-]_{\theta_i} > [\phi_h^0]_{\theta_i}. \quad (18)$$

The above two inequalities must be satisfied for all neurons at all times. Instead of testing for all values  $\theta_{ijk}$  separately, we can simplify the set of inequalities as follows.

*Lemma 6.10:* Let  $\theta_{\min}$  and  $\theta_{\max}$  denote the minimum and maximum, respectively, of all fuzzy transition memberships  $\theta_{ijk}$  of a given FFA  $M$ . Then, inequalities (17) and (18) are satisfied for all transition weights  $\theta_{ijk}$  if the inequalities

$$-\frac{H}{2} + Hr[\phi_f^-]_{\theta_{\max}} < [\phi_f^0]_{\theta_{\min}} \quad (19)$$

$$-\frac{H}{2} + H[\phi_h^+]_{\theta_{\min}} - H[\phi_f^-]_{\theta_{\max}} > [\phi_h^0]_{\theta_{\max}} \quad (20)$$

are satisfied.

*Proof:* Consider the two fixed points  $[\phi_f^-]_{\theta_{\min}}$  and  $[\phi_h^+]_{\theta_{\max}}$ . According to Corollary 6.4, we have

$$\begin{aligned} [\phi_f^-]_{\theta_{\min}} &= \theta_{\min}[\phi_f^-]_1 < \theta_{ijk}[\phi_f^-]_1 < \theta_{\max}[\phi_f^-]_1 \\ &= [\phi_f^-]_{\theta_{\max}}. \end{aligned}$$

Thus, if inequalities (19) and (20) are not violated for  $[\phi_f^-]_{\theta_{\min}}$  and  $[\phi_h^+]_{\theta_{\max}}$ , then they will not be violated for  $\theta_{\min} \leq \theta_{ijk} \leq \theta_{\max}$  either.  $\square$

We can rewrite inequalities (19) and (20) as

$$-\frac{H}{2} + Hr\theta_{\max}[\phi_f^-]_1 < \theta_{\min}[\phi_f^0]_1 \quad (21)$$

and

$$-\frac{H}{2} + H\theta_{\min}[\phi_h^+]_1 - H\theta_{\max}[\phi_f^-]_1 > \theta_{\max}[\phi_h^0]_1. \quad (22)$$

Solving inequalities (21) and (22) for  $[\phi_f^-]_1$  and  $[\phi_h^+]_1$ , respectively, we obtain conditions under which a constructed recurrent network implements a given FFA. These conditions are expressed in the following theorem.

*Theorem 6.1:* For some given unambiguous FFA  $M$  with  $n$  states and  $m$  input symbols, let  $r$  denote the maximum number of transitions to any state over all input symbols of  $M$ . Furthermore, let  $\theta_{\min}$  and  $\theta_{\max}$  denote the minimum and maximum, respectively, of all transitions weights  $\theta_{ijk}$  in  $M$ . Then, a sparse recurrent neural network with  $n$  state and  $m$  input neurons can be constructed from  $M$  such that the internal state representation remains stable if

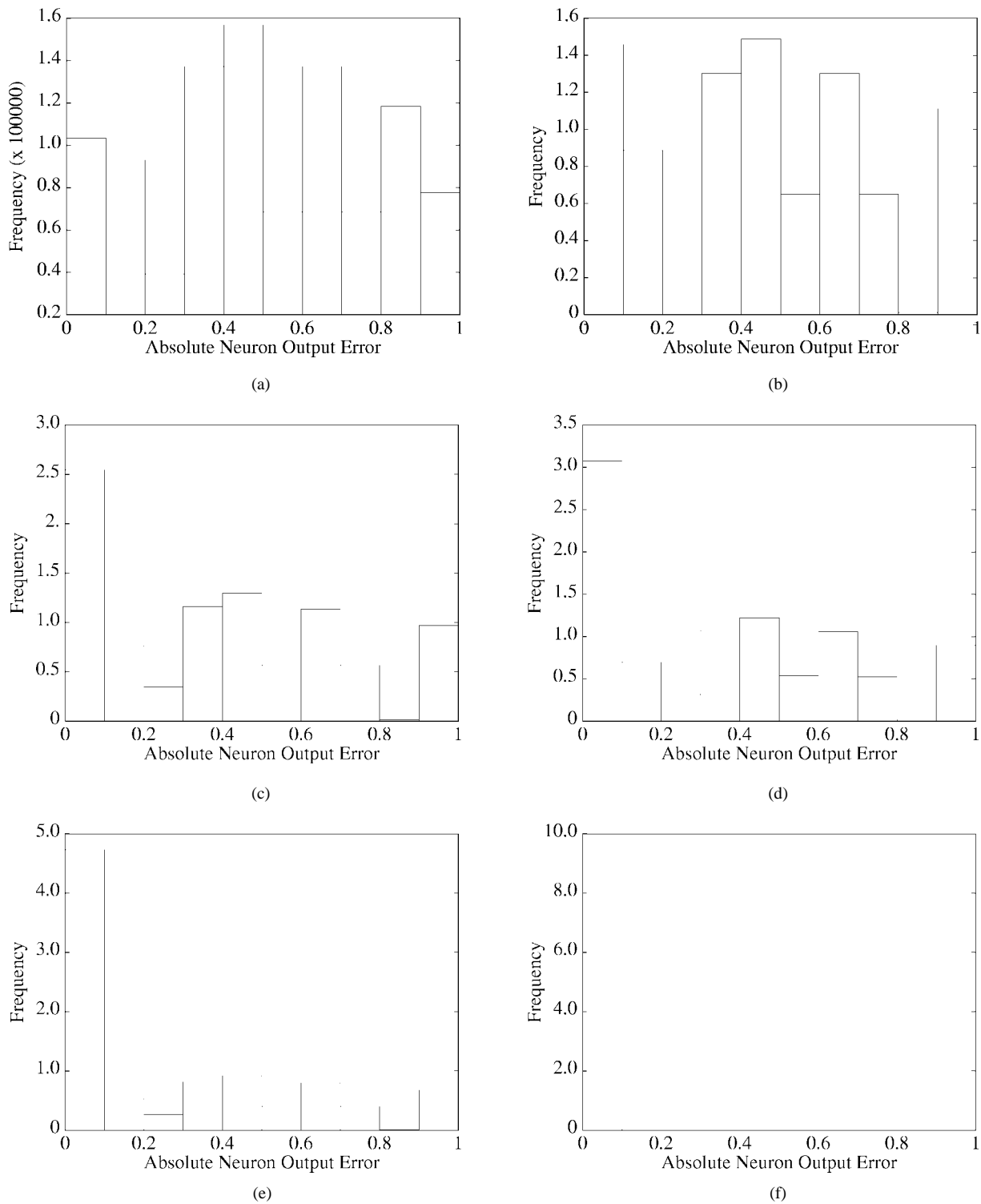
- 1)  $[\phi_f^-]_1 < \frac{1}{r\theta_{\max}} \left( \frac{1}{2} + \theta_{\min} \frac{[\phi_f^0]_1}{H} \right)$
- 2)  $[\phi_h^+]_1 > \frac{1}{\theta_{\min}} \left( \frac{1}{2} + \theta_{\max} [\phi_f^-]_1 + \frac{[\phi_h^0]_1}{H} \right)$
- 3)  $H > \max(H_0^-(r), H_0^+(r)).$

Furthermore, the constructed network has at most  $3mn$  second-order weights with alphabet  $\Sigma_w = \{-H, 0, +H\}$ ,  $n+1$  biases with alphabet  $\Sigma_b = \{-H/2\}$ , and maximum fan out  $3m$ .  $\square$

For  $\theta_{\min} = \theta_{\max} = 1$ , conditions 1)–3) of the above theorem reduce to those found for stable DFA encodings [41]. This is consistent with a crisp representation of DFA states.

## VII. SIMULATIONS

In order to test our theory, we constructed a fuzzy encoding of a randomly generated FFA with 100 states (after the execution of the FFA transformation algorithm) over the input alphabet  $\{0, 1\}$ . We randomly assigned weights in the range  $[0, 1]$  to all transitions in increments of 0.1. The maximum indegree was  $D_{\text{in}}(M) = r = 5$ . We then tested the stability of the fuzzy internal state representation on 100 randomly generated strings of length 100 by comparing, at each time step, the output signal of each recurrent state neuron with its ideal output signal (since each recurrent state neuron  $S_i$  corresponds to a FFA state  $q_i$ , we know the degree to which  $q_i$  is occupied after input symbol  $a_k$  has been read: either zero or  $\theta_{ijk}$ ). A histogram of the differences between the ideal and the observed signal of state neurons for selected values of the weight strength  $H$  over all state neurons and all tested strings is shown in Fig. 11. As expected, the error decreases for increasing values of  $H$ . We observe that the number of discrepancies between the desired and the actual neuron output decreases “smoothly” for the shown values of  $H$  (almost no change can be observed for values up to  $H = 6$ ). The most significant change can be observed by comparing the histograms for  $H = 9.7$  and  $H = 9.75$ . The existence of significant neuron output errors for  $H = 9.7$  suggests that the internal FFA representation is highly unstable. For  $H \geq 9.75$ , the internal FFA state representation becomes stable. This discontinuous change can be explained by observing that there exists a critical value  $H_0(r)$  such that the number of stable fixed points also changes discontinuously from one to two for  $H < H_0(r)$  and  $H > H_0(r)$ , respectively (see Fig. 11). The “smooth” transition from large output



**Fig. 11.** Stability of an FFA state encoding. The histogram shows the frequency ( $\times 10^5$ ) of absolute neuron output errors of a network with 100 neurons that implements a randomly generated fuzzy finite state automaton and reads 100 randomly generated strings of length 100 for different values of the weight strength  $H$ . The error was computed by comparing, at each time step, the actual with the desired output of each state neuron. The distribution of neuron output signal errors are for weight strengths: (a)  $H = 6.0$ ; (b)  $H = 9.0$ ; (c)  $H = 9.60$ ; (d)  $H = 9.65$ ; and (e)  $H = 9.70$ ; and (f)  $H = 9.75$ .

errors to very small errors for most recurrent state neurons [Fig. 11(a)–(e)] can be explained by observing that not all recurrent state neurons receive the same number of residual inputs; some neurons may not receive any residual input

for some given input symbol  $a_k$  at time step  $t$ ; in that case, the low signals of those neurons are strengthened to  $\tilde{g}(0, H, \theta_{i,k}) \simeq 0$  (note that strong low signals imply strong high signals by Lemma 6.7).

## VIII. CONCLUSIONS

Theoretical work that proves representational relationships between different computational paradigms is important because it establishes the equivalences of those models. Previously it has been shown that it is possible to deterministically encode FFA in recurrent neural networks by transforming any given FFA into a deterministic acceptor which assign string membership [56]. In such a deterministic encoding, only the network's classification of strings is fuzzy, whereas the representation of states is crisp. The correspondence between FFA and network parameters—i.e., fuzzy transition memberships and network weights, respectively—is lost in the transformation.

Here, we have demonstrated analytically and empirically that it is possible to encode FFA in recurrent networks without transforming them into deterministic acceptors. The constructed network directly represents FFA states with the desired fuzziness. That representation requires: 1) a slightly increased functionality of sigmoidal discriminant functions (it only requires the discriminants to accommodate variable output range), and 2) a transformation of a given FFA into an equivalent FFA with a larger number of states. (We have found empirically that the increase in automaton size is roughly proportional  $N * K$  where  $N$  and  $K$  are the automaton and alphabet size, respectively.) In the proposed mapping  $FFA \rightarrow$  recurrent network, the correspondence between FFA and network parameters remains intact; this can be significant if the physical properties of some unknown dynamic, nonlinear system are to be derived from a trained network modeling that system. Furthermore, the analysis tools and methods used to demonstrate the stability of the crisp internal representation of DFA carried over and generalized to show stability of the internal FFA representation.

We speculate that other encoding methods are possible and that it is an open question as to which encoding methods are better. One could argue that, from an engineering point of view, it may seem more natural to use radial basis functions to represent fuzzy state membership (they are often used along with triangular and trapezoidal membership functions in the design of fuzzy systems) instead of sigmoidal discriminant functions (DFA can be mapped into recurrent neural networks with radial basis functions [49]). It is an open question how mappings of FFA into RNN's with radial basis discriminant functions would be implemented and how such mappings would compare to the encoding algorithm described in this work.

The usefulness of training RNN's with fuzzy state representation from examples to behave like a FFA—the variable output range  $\theta$  can be treated as a variable parameter and an update rule similar to that for networks weights can be derived—and whether useful information can be extracted from trained networks has yet to be determined. In particular, it would be interesting to compare training and knowledge representation of networks whose discriminant functions have fixed and variable output ranges, respectively. Discriminant functions with variable neuron output range may open the door to novel methods for the extraction of symbolic knowledge from recurrent neural networks.

## ACKNOWLEDGMENT

The authors would like to acknowledge useful discussions with K. Bollacker, D. Handscomb, and B. G. Horne, as well as suggestions from the referees.

## REFERENCES

- [1] C. L. Giles, R. sun, and J. M. Zurada, *IEEE Trans. Neural Networks (Special Issue on Neural Networks and Hybrid Intelligent Models: Foundations, Theory, and Applications)*, vol. 9, pp. 721–723, Sept. 1998.
- [2] L. A. Bookman and R. Sun, *Connection Sci. (Special Issue on Architectures for Integrating Symbolic and Neural Processes)*, vol. 5, nos. 3 and 4, 1993.
- [3] J. Hendler, "Developing hybrid symbolic/connectionist models," in *Advances in Connectionist and Neural Computation Theory*, J. Barnden and J. Pollack, Eds. Ablex, 1991.
- [4] V. Honavar and L. Uhr, Eds., *Artificial Intelligence and Neural Networks: Steps Toward Principled Integration*. San Diego, CA: Academic, 1994.
- [5] R. Sun, "Learning, action, and consciousness: A hybrid approach toward modeling consciousness," *Neural Networks*, vol. 10, no. 7, pp. 1317–1332, 1997.
- [6] J. Bezdek, *IEEE Trans. Neural Networks (Special Issue on Fuzzy Logic and Neural Networks)*, vol. 3, 1992.
- [7] C. S. Herrmann, "A hybrid fuzzy-neural expert system for diagnosis," in *Proc. 14th Int. Joint Conf. Artificial Intelligence*, vol. 1, 1995, pp. 494–502.
- [8] M. Palaniswami, Y. Attikiouzel, R. J. Marks, and D. Fogel, Eds., *Computational Intelligence: A Dynamic System Perspective*. Piscataway, NJ: IEEE Press, 1995.
- [9] N. Kasabov, *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering*. Cambridge, MA: MIT Press, 1996.
- [10] L. Zadeh, "Fuzzy sets," *Inform. Control*, vol. 8, no. 3, pp. 338–353, 1965.
- [11] H. R. Berenji and P. Khedkar, "Learning and fine tuning fuzzy logic controllers through reinforcement," *IEEE Trans. Neural Networks*, vol. 3, pp. 724–740, Sept. 1992.
- [12] P. P. Bonissone, V. Badami, K. H. Chiang, P. S. Khedkar, K. W. Marcelle, and M. J. Schutten, "Industrial applications of fuzzy logic at General Electric," *Proc. IEEE*, vol. 83, pp. 450–465, Mar. 1995.
- [13] S. Chiu, S. Chand, D. Moore, and A. Chaudhary, "Fuzzy logic for control of roll and moment for a flexible wing aircraft," *IEEE Control Syst. Mag.*, vol. 11, pp. 42–48, Apr. 1991.
- [14] J. Corbin, "A fuzzy logic-based financial transaction system," *Embedded Syst. Programming*, vol. 7, no. 12, p. 24, 1994.
- [15] L. G. Franquelo and J. Chavez, "Fasy: A fuzzy-logic based tool for analog synthesis," *IEEE Trans. Computer-Aided Design*, vol. 15, p. 705, July 1996.
- [16] T. L. Hardy, "Multi-objective decision-making under uncertainty fuzzy logic methods," NASA, Washington, DC, Tech. Rep. TM 106796, 1994.
- [17] W. J. M. Kickert and H. R. van Nauta Lemke, "Application of a fuzzy controller in a warm water plant," *Automatica*, vol. 12, no. 4, pp. 301–308, 1976.
- [18] C. C. Lee, "Fuzzy logic in control systems: Fuzzy logic controller," *IEEE Trans. Syst., Man, Cybern.*, vol. 20, pp. 404–435, Feb. 1990.
- [19] C. P. Pappis and E. H. Mamdani, "A fuzzy logic controller for a traffic junction," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-7, pp. 707–717, Oct. 1977.
- [20] X. M. Yang and G. J. Kalambur, "Design for machining using expert system and fuzzy logic approach," *J. Mater. Eng. Performance*, vol. 4, no. 5, p. 599, 1995.
- [21] C. M. Bishop, *Neural Networks for Pattern Recognition*. New York: Oxford, 1995.
- [22] A. Cichocki and R. Unbehauen, Eds., *Neural Networks for Optimization and Signal Processing*. New York: Wiley, 1993.
- [23] S. Haykin, *Neural Networks, A Comprehensive Foundation*. New York: Prentice-Hall, 1998.
- [24] H. T. Siegelmann and E. D. Sontag, "On the computational power of neural nets," *J. Comput. Syst. Sci.*, vol. 50, no. 1, pp. 132–150, 1995.
- [25] H. T. Siegelmann, *Neural Networks and Analog Computation: Beyond the Turing Limit*. Boston, MA: Birkhauser, 1999.



- [26] C. L. Giles and M. Gori, Eds., *Adaptive Processing of Sequences and Data Structures* (Lecture Notes in Artificial Intelligence). New York: Springer Verlag, 1998.
- [27] J. E. Hopcroft and J. D. Ullman, *Introduction to Automata Theory, Languages, and Computation*. Reading, MA: Addison-Wesley, 1979.
- [28] P. Ashar, S. Devadas, and A. R. Newton, *Sequential Logic Synthesis*. Norwell, MA: Kluwer, 1992.
- [29] S. C. Kleene, "Representation of events in nerve nets and finite automata," in *Automata Studies*, C. E. Shannon and J. McCarthy, Eds. Princeton, NJ: Princeton Univ. Press, 1956, pp. 3–42.
- [30] M. Minsky, *Computation: Finite and Infinite Machines*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1967, ch. 3, pp. 32–66.
- [31] L.-M. Fu, *Neural Networks in Computer Intelligence*. New York: McGraw-Hill, 1994.
- [32] E. S. Santos, "Maximin automata," *Inform. Control*, vol. 12, pp. 363–377, 1968.
- [33] L. Zadeh, "Fuzzy languages and their relation to human and machine intelligence," Electron. Res. Lab., Univ. California, Berkeley, Tech. Rep. ERL-M302, 1971.
- [34] B. Gaines and L. Kohout, "The logic of automata," *Int. J. General Syst.*, vol. 2, pp. 191–208, 1976.
- [35] E. B. Kosmatopoulos and M. A. Christodoulou, "Neural networks for identification of fuzzy dynamical systems: An application to identification of vehicle highway systems," in *Proc. 4th IEEE Mediterranean Symp. New Directions in Control and Automation*, 1996, pp. 23–38.
- [36] S. I. Mensch and H. M. Lipp, "Fuzzy specification of finite state machines," in *Proc. Europ. Design Automation Conf.*, 1990, pp. 622–626.
- [37] C. W. Omlin and C. L. Giles, "Extraction of rules from discrete-time recurrent neural networks," *Neural Networks*, vol. 9, no. 1, pp. 41–52, 1996.
- [38] R. Maclin and J. W. Shavlik, "Using knowledge-based neural networks to improve algorithms: Refining the chou-fasman algorithm for protein folding," *Machine Learning*, vol. 11, nos. 2–3, pp. 195–215, 1993.
- [39] J. Grantner and M. J. Patyra, "Synthesis and analysis of fuzzy logic finite state machine models," in *Proc. 3rd IEEE Conf. Fuzzy Syst.*, vol. 1, 1994, pp. 205–210.
- [40] E. Khan and F. Unal, "Recurrent fuzzy logic using neural networks," in *Advances in Fuzzy Logic, Neural Networks, and Genetic Algorithms (Lecture Notes in Artificial Intelligence)*, T. Furuhashi, Ed. New York: Springer Verlag, 1995.
- [41] C. W. Omlin and C. L. Giles, "Constructing deterministic finite-state automata in recurrent neural networks," *J. ACM*, vol. 43, no. 6, pp. 937–972, 1996.
- [42] F. E. Cellier and Y. D. Pan, "Fuzzy adaptive recurrent counter-propagation neural networks: A tool for efficient implementation of qualitative models of dynamic processes," *J. Syst. Eng.*, vol. 5, no. 4, pp. 207–222, 1995.
- [43] E. B. Kosmatopoulos and M. A. Christodoulou, "Structural properties of gradient recurrent high-order neural networks," *IEEE Trans. Circuits Syst.*, vol. 42, pp. 592–603, Sept. 1995.
- [44] E. B. Kosmatopoulos, M. M. Polycarpou, M. A. Christodoulou, and P. A. Ioannou, "High-order neural networks for identification of dynamical systems," *IEEE Trans. Neural Networks*, vol. 6, pp. 422–431, Mar. 1995.
- [45] E. B. Kosmatopoulos and M. A. Christodoulou, "Recurrent neural networks for approximation of fuzzy dynamical systems," *Int. J. Intell. Control Syst.*, vol. 1, no. 2, pp. 223–233, 1996.
- [46] M. P. Casey, "The dynamics of discrete-time computation, with application to recurrent neural networks and finite state machine extraction," *Neural Computation*, vol. 8, no. 6, pp. 1135–1178, 1996.
- [47] A. Cleeremans, D. Servan-Schreiber, and J. McClelland, "Finite state automata and simple recurrent neural networks," *Neural Computation*, vol. 1, no. 3, pp. 372–381, 1989.
- [48] J. L. Elman, "Finding structure in time," *Cognitive Sci.*, vol. 15, no. 2, pp. 179–211, 1990.
- [49] P. Frasconi, M. Gori, M. Maggini, and G. Soda, "Representation of finite state automata in recurrent radial basis function networks," *Machine Learning*, vol. 23, no. 1, pp. 5–32, 1996.
- [50] C. L. Giles, C. B. Miller, D. Chen, H. H. Chen, G. Z. Sun, and Y. C. Lee, "Learning and extracting finite state automata with second-order recurrent neural networks," *Neural Computation*, vol. 4, no. 3, pp. 393–405, 1992.
- [51] J. B. Pollack, "The induction of dynamical recognizers," *Machine Learning*, vol. 7, nos. 2/3, pp. 227–252, 1991.
- [52] R. L. Watrous and G. M. Kuhn, "Induction of finite-state languages using second-order recurrent networks," *Neural Computation*, vol. 4, no. 3, pp. 406–411, 1992.
- [53] Z. Zeng, R. M. Goodman, and P. Smyth, "Learning finite state machines with self-clustering recurrent networks," *Neural Computation*, vol. 5, no. 6, pp. 976–990, 1993.
- [54] J. Grantner and M. J. Patyra, "VLSI implementations of fuzzy logic finite state machines," in *Proc. 5th IFSA Congr.*, 1993, pp. 781–784.
- [55] F. A. Unal and E. Khan, "A fuzzy finite state machine implementation based on a neural fuzzy system," in *Proc. 3rd Int. Conf. Fuzzy Syst.*, vol. 3, 1994, pp. 1749–1754.
- [56] C. W. Omlin, K. K. Thornber, and C. L. Giles, "Fuzzy finite-state automata can be deterministically encoded into recurrent neural networks," *IEEE Trans. Fuzzy Syst.*, vol. 6, pp. 76–89, Feb. 1998.
- [57] A. Blanco, M. Delgado, and M. C. Pegalajar, "Fuzzy grammar inference using neural networks," Dep. Comput. Sci. Artificial Intell., Univ. Granada, Spain, Tech. Rep., 1997.
- [58] M. G. Thomason and P. N. Marinos, "Deterministic acceptors of regular fuzzy languages," *IEEE Trans. Syst., Man, Cybern.*, vol. 4, pp. 228–230, Mar. 1974.
- [59] C. W. Omlin and C. L. Giles, "Stable encoding of large finite-state automata in recurrent neural networks with sigmoid discriminants," *Neural Computation*, vol. 8, no. 7, pp. 675–696, 1996.
- [60] L. A. Akers, D. K. Ferry, and R. O. Grondin, "Synthetic neural systems in VLSI," in *An Introduction to Neural and Electronic Systems*. San Diego, CA: Academic, 1990, pp. 317–336.
- [61] B. J. Sheu, *Neural Information Processing and VLSI*. Boston, MA: Kluwer, 1995.
- [62] C. Mead, *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley, 1989.
- [63] C. L. Giles and C. W. Omlin, "Extraction, insertion and refinement of symbolic rules in dynamically driven recurrent neural networks," *Connection Sci.*, vol. 5, nos. 3 & 4, pp. 307–337, 1993.
- [64] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications (Mathematics in Science and Engineering)*, vol. 144. New York: Academic, 1980, pp. 220–226.
- [65] S. F. Thomas, *Fuzziness and Probability*. Wichita, KS: ACG Press, 1995.
- [66] C. W. Omlin and C. L. Giles, "Rule revision with recurrent neural networks," *IEEE Trans. Knowledge and Data Eng.*, vol. 8, pp. 183–188, Jan. 1996.
- [67] H. K. Khalil, *Nonlinear Systems*. New York: Macmillan, 1992.



**C. Lee Giles** (Fellow, IEEE) received the B.S. degree from the University of Tennessee, the B.A. degree from Rhodes College, the M.S. degree from the University of Michigan, and the Ph.D. in optical sciences from the University of Arizona.

He is a Senior Research Scientist in Computer Science at NEC Research Institute, Princeton, NJ, an Adjunct Faculty Member at the Institute for Advanced Computer Studies at the University of Maryland, College Park, and an Adjunct

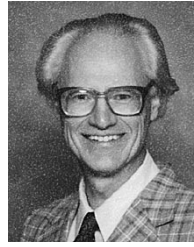
Professor in Computer and Information Science at the University of Pennsylvania, Philadelphia. Previously, he was a Program Manager at the Air Force Office of Scientific Research in Washington, DC, where he initiated and managed research programs in neural networks and artificial intelligence, and in optics in computing and processing. Before that he was a Research Scientist at the Naval Research Laboratory, Washington, DC, and an Assistant Professor of Electrical and Computer Engineering at Clarkson University. During part of his graduate education he was a research engineer at Ford Motor Scientific Research Laboratory. His research interests are in novel applications and tools in web computing and intelligent information systems and in the foundations of intelligent systems.

Dr. Giles is a member of AAAI, ACM, INNS, OSA, AAAS, and the Center for Discrete Mathematics and Theoretical Computer Science, Rutgers University.

**Christian W. Omlin** received the M.S. degree from the Swiss Federal Institute of Technology, Zurich, in 1987 and the Ph.D. degree from Rensselaer Polytechnic Institute, Troy, NY, in 1995.

He is a Senior Lecturer in the Computer Science Department at the University of Stellenbosch, South Africa. Prior to his appointment at Stellenbosch University, he was with Adaptive Computing Technologies, Troy, NY, where he specialized in real-world applications of intelligent information processing methods. He was also a graduate assistant and instructor at Rensselaer Polytechnic Institute from 1987 to 1991. From 1991 to 1996, he was with NEC Research Institute, Princeton, NJ. His expertise and research interests include foundations and applications of artificial intelligence, neural networks, fuzzy systems, hybrid systems, intelligent autonomous agents, and applications of machine learning methods to information retrieval on the world wide web. He has published over 30 papers in conferences, journals, and book chapters. He is the coauthor of the upcoming book *Knowledge Representation and Acquisition in Recurrent Neural Networks: Foundations, Algorithms, and Applications*.

Dr. Omlin is a member of ACM and INNS.



**Karvel K. Thornber** received the Ph.D. degree from the California Institute of Technology, Pasadena, in 1966.

After spending two years at Stanford University, one year at Bristol, and 20 years at Bell Labs, he joined NEC Research Institute, Princeton, NJ, in 1989 as a Senior Research Scientist. Having contributed earlier to theoretical and device physics, he is currently working on inference and vision.