In *Progress in Neural Networks*, vol. **1**, Ch. 5, pp. 105-126, O. Omidvar (ed), Ablex Publishing, 1990.

Adaptive Self-Organizing Concurrent Systems

Tony R. Martinez

Computer Science Dept.
Brigham Young University

1. Introduction

Current technology trends, demanding applications, and limitations of sequential computation, have combined to rejuvenate research into connectionist and neurally inspired computational mechanisms. A new class of multilayer connectionist systems, ASOCS (adaptive self-organizing concurrent systems), was created with the goal of fulfilling the desired functionality of connectionist methods, while overcoming some of the drawbacks of current models. A number of specific ASOCS architectures have been proposed [5,6] and VLSI implementation and testing are underway [2]. The goal of this chapter is to introduce the basic features and functional mechanisms of these new models.

ASOCS (adaptive self-organizing concurrent systems) is similar to most *decision-making* neural network models in that it attempts to learn an adaptive set of arbitrary vector mappings. However, it differs dramatically in its mechanisms. ASOCS is based on networks of adaptive digital elements which self-modify using local information. Function specification is entered incrementally by use of rules, rather than complete input-output vectors, such that a processing network is able to extract critical features from a large environment and give output in a parallel fashion. Learning also uses parallelism and self-organization such that a new rule is completely learned in time linear with the depth of the network. The model guarantees learning of any arbitrary mapping of boolean input-output vectors. The model is also stable in that learning does not erase any previously learned mappings except those explicitly contradicted.

The atomic knowledge input to the system is known as an *instance* and is made up of a variable length boolean input vector and an associated target output. The instance input vector typically contains only the critical features, (a small subset of the total environment input), sufficient to decide the current output. Thus, a single instance represents many possible states of the environment. Any propositional boolean function can be represented by a set of instances. Instances are input to and learned incrementally by the system. A new instance may contradict earlier instances, in which case only the contradicted portions of older instances optionally become invalid. A new instance is broadcast to a multilayer network of adaptive digital nodes. Using only local information (the internal state of the system or nodes is invisible to both the outside and inside), nodes adjust their functions and interconnectivity in a parallel and self-organizing fashion in order to maintain a network which consistently fulfills all instances received by the system. The network can build itself from scratch with new interconnections being dynamically made between nodes, together with the ability for nodes no longer aiding in correct network function to *self-delete*. During processing, the network acts as a parallel hardware circuit mapping input states to output states. Instances can be continually input over the life of the system.

The basic learning mechanisms of ASOCS departs appreciably from standard neural network models. Connectionist systems and applications differ in their needs and goals. Effective high level processing may require an integration of separate models, each processing in its domain of strength, in order to acquire an emergent multi-purpose mechanism. Although a simplistic and incomplete dichotomy, artificial neural systems can be classified into *categorization* systems and *decision* systems. Examples of categorization systems include adaptive resonance theory [1] and self-organizing feature maps [4]. These systems do unsupervised learning of categories and classifications of arbitrary inputs. They seek solutions in the important domain of building up reasonable categories from an unknown input environment, mapping similar things into similar classes.

A different application is that of taking mapped inputs, such as those from a categorization system, and further mapping them into outputs (motor actions) or inferences. In this case, many inputs which are very dissimilar may need to be mapped into the same output class, while similar inputs could be scattered into various output classes. This is the realm of decision systems. These systems are required to do more arbitrary mappings and require some type of supervised learning (teacher, environmental feedback, etc.) Attempts at this type of system include backpropagation techniques [10] and Boltzmann machines [3]. There is certainly overlap between categorization and decision systems, and classifying a model as one or the other is a matter of degree. ASOCS falls into the realm of decision making systems, and thus can be most accurately compared and measured against that class of models. Perhaps more general systems will include a categorization portion which does unsupervised classification of inputs into usable classes, and which feeds its outputs to a decision making system which controls motor or other output units (figure 1).

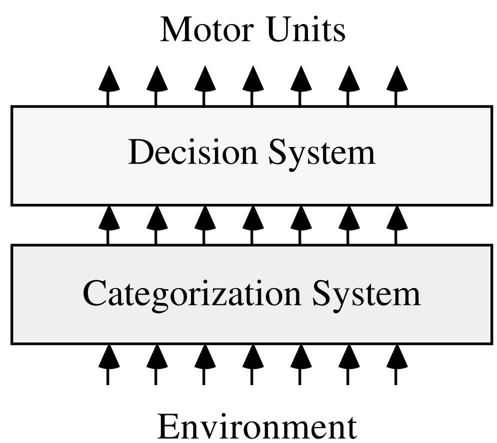

Figure 1 - Integrated categorization and decision systems

The goal of this chapter is to give an introduction to the features and mechanisms of the ASOCS system The following section overviews these features and contrasts them. The major portion of the chapter then overviews the different ASOCS learning algorithms currently proposed. Due to the introductory scope of this chapter, indepth presentation of the models is impossible. More detailed discussion of the learning algorithms can be found elsewhere [5-8].

2. Features of the Model

Although the goal functionality in ASOCS is similar to that of most decision making connectionist models, the method of operation and the features of the system are quite different. Much of this difference follows from the use of different types of atomic nodes. Most neural networks use a node which sums the products of inputs and weights, and then does some type of nonlinear function of that sum to attain an activation. Learning is accomplished through small changes in the weights. The ASOCS model uses as an atomic node, a two input boolean gate which can be set to any of the 16 possible boolean functions of two inputs. Rather than gradient seeking algorithms of small weight changes, ASOCS must seek methods of learning through changing the function and interconnection between nodes. Another cause of disparity is the use of rule-like instances which require only the critical variables of any pattern to be shown, rather than the complete input/output vectors of a typical training set. In the following we discuss the domain of boolean input and output vectors.

Although multilayer nonlinear networks can theoretically implement arbitrary mappings of input/output vectors, learning algorithms have not yet been established which can guarantee that a desired mapping will actually be learned. Since current attempts at multilayer learning use variants of gradient descent or stochastic mechanisms, local minima and finite time constraints keep these models from always attaining a desired mapping. The learning algorithms of ASOCS models, on the other hand, have been shown to produce guaranteed mappings of any boolean function with arbitrarily many inputs and outputs. It must be noted that layered models using exponential numbers of nodes, such as a RAM, can be trained in a guaranteed fashion to learn arbitrary mappings. In fact, certain mappings would appear to *require* an exponential amount of either nodes, links, or weight resolution. These type of problems, however, are not in the domain of applications for which neural networks are intended. Most current models are based on an a priori static topology, and even when a problem domain contains a portion of highly complex mappings, (which *is* a common situation), the model requires the full exponential size demands. ASOCS, on the other hand, through use of a reconfigurable topology, can account in a dynamically customized manner for complex mappings, while maintaining a parsimonious representation for the entire system.

A second feature of the ASOCS system is that the time necessary for the network to accept a new instance and reconfigure to a consistent network is both bounded and scales well as the network grows. Learning time is $O(d)$ where d is the depth of the network and $O(log(n))$ where n is the number of nodes in the network. This compares very favorably with gradient descent techniques which require large numbers of iterations for the weights to arrive at a proper value.

Along with a fast bounded learning time, ASOCS has the additional feature of *stability*. Learning of a new instance does not alter the systems ability to continue properly fulfilling previously learned instances. (At the same time, portions of previous instances explicitly contradicted by a new instance can be modified). This is the opposite of *unlearning*, where adjusting weights to correctly fulfill the current pattern can undo and invalidate earlier adjustments made to previous pattern values. This leads to oscillation and/or many repetitions of the same input/output pairs in order to converge to a proper network. ASOCS is able to maintain stability without sacrificing a distributed mechanism of processing and representation.

An important goal of most neural network models is that of *generalization*. This is the phenomenon that after learning, a network will with high probability respond correctly to input patterns which it has not previously seen. In weight summing algorithms, similar input vectors typically generate similar outputs. The generalization is then in terms of the similarity of inputs; (an example measure is that of hamming distance). However, this is but one type of generalization. Correct generalization is application specific. In the ASOCS scheme, generalization is attained through the incomplete nature of instances, which allow specification of only critical variables for a given pattern, rather than an entire input vector. One of the apparent powers of natural nervous systems is the ability to discriminate the current critical inputs from a massive barrage of total inputs. If one puts his hand to a hot stove, immediate retraction should take place regardless of the many other impinging input variables. Instances allow a natural mechanism to specify critical variables for specific situations, and generalization takes place in that many possible input vectors impinging during processing can match the single rule which has specified the critical variables. Thus a processing ASOCS system is able to extract the critical variables from a large input environment and make correct judgements although it has never before experienced the total current input environment before.

On the down side, the overall ASOCS system is not very plausible as a biological model. The ability to extract critical inputs from an extended environment *is* considered as a notion which demands more attention in neural modeling. Also, the basic ASOCS tenant of simple broadcast to many unobservable nodes, with concomitant local communication and modification between those nodes, is seen as a strong analog of the endocrine/nervous system interaction. Finally, the ability to create new connections and nodes, together with the self-deletion of obsolete nodes, seems closer to the reality of developmental neurophysiology than initially fixed topological structures. However, an immediate analogy is not seen for the purely boolean nature of the data and functions of the atomic ASOCS node.

3. Summary of ASOCS Computational Mechanisms

Before discussing particular mechanisms, an attempt is here made to give a feel for the overall adaptive process in an ASOCS system. Since a number of formal systems have been proposed, there is no single overall learning process. At any time while a system is processing, a new instance can be entered into the system. The instance is then broadcast to the network by one of a few defined mechanisms. The *instance set,* the total function described by all learned instances, need not be explicitly stored, but can be implicitly stored by the functions and interconnections within the network. If the network does not fulfill the newly introduced instance, nodes within the network self-select to be part of the modification process whereby the network reconfigures to consistently fulfill the newly changed instance set. Selected nodes then independently and concurrently combine to form the new network. At this point, nodes no longer aiding in the correct functioning of the network delete themselves after sending messages to neighbor nodes for them to also delete themselves. This recursive self-pruning allows the network to maintain a parsimonious representation of its current function.

ASOCS research was inspired by a revisitation of the seminal work of Rosenblatt [9] on perceptron systems. Verstraete [11] extended perceptron-like techniques into boolean logic devices, in contrast to the threshold unit. His and other [12] research sought centralized methods of programming boolean functions into passive fixed networks. The ASOCS notion seeks to support adaptive networks whose function can change over time, and where control of adaptation is distributed amongst active network nodes in a self-organizing fashion.

A discussion of knowledge input is first given, together with an overview of the generic ASOCS architecture. This is followed by summaries of the three formal ASOCS models proposed to date.

3.1 Knowledge Input

The atomic knowledge element of the system is the *instance*, which is a boolean rule defining what the system should output when confronted with a given input. Each instance is an if-then statement where the antecedent is a conjunction of boolean variables, and the consequent is a single boolean variable. Thus, the instance is a propositional production rule. Instances are universal in that any boolean function can be represented by a set of instances. Following are examples of instances.

$\overline{X1}\ \overline{X2} \Rightarrow Z1$

$X1\ \overline{X2}\ X3 \Rightarrow \overline{Z2}$

$\overline{X2}\ \overline{X3}\ Z2 \Rightarrow \overline{Z1}$

The $\Rightarrow$ is the implication operator that states if the antecedent is true, then the consequent must become true. ($\overline{}$ represents the complement of a variable). Note that output variables can be used in instance antecedents, in which case each instance can be viewed as a state transition. Output variables used in the antecedents are known as feedback variables, and in the current model they can be treated like any other input variable. Though sufficient to support any number of output variables, a single output is considered in this presentation for sake of simplicity.

Whenever the antecedent of an instance is not matched by the input environment, the instance says nothing about what the output of the system should be. Instances are incomplete by nature. The instance $A \Rightarrow Z$ states that Z should be 1 when A is 1 regardless of all other variables. Variables not appearing in an instance are don't care variables for that instance.

In the ASOCS model, instances are input to the system incrementally and the current collection of instances is called the *instance set*. The instance set is the rule base of the system and adaptation in ASOCS occurs by maintaining the instance set consistent and optionally minimal. By consistent it is meant that no instance in the instance set can contradict another one; i.e. no two instances can specify an opposite output for the same input. Newer instances are given precedence and the contradicted portion of an old instance is deleted from the set. Minimality means that the instance set is stored using a minimal representation. A minimal instance set representation however, is not crucial to the correct functioning of the system, and only near-minimality is typically supported by current models.

An input variable occurring in one instance, and occurring in its complemented form in another instance, is a *discriminant* variable with respect to the two instances. An instance implying a non-negated variable is a *positive* instance, while an instance implying a negated variable is a *negative* instance. This is known as the *polarity* of an instance. Two instances with the same polarity are *concordant*. They are *discordant* instances if they have opposite polarities.

We now consider the relationship between the instance set and a logic network for a single output Z. A logic network fulfills the instance set if, when any positive instance is matched the network output Z is 1, and when any negative instance is matched Z is 0. A positive and negative instance cannot be matched simultaneously in a consistent instance set. If the environment does not match any instance in the set, the network can arbitrarily output a value, or flag the output as unknown.

The relation between the number of input variables and output variables is arbitrary. Variables do not need to be predefined in a system. Rather, not until the first time an input or output variable is seen in an instance is a data line dynamically bound to the variable. Thus, new variables can be introduced at any point in the life of the model. Figure 2 shows a black box representation of the basic system.

Figure 2 - Black Box ASOCS Structure

A consistent set of instances is one in which no two instances contradict each other. Two instances contradict each other if and only if for some state of the environment, they are both matched, and they are discordant instances. For example, the two instances $A \Rightarrow Z$ and $B \Rightarrow \overline{Z}$ contradict each other when both A and B are 1. Contradictions occur when discordant instances exist without discriminant variables in their antecedents. Two instances with no discriminant variables, can always for some value of the variables both be matched. A corollary to the above is that all instances must contain at least one discriminant variable with respect to each discordant instance.

There are two basic ways by which the instance set is modified due to a contradiction. When the variables of a new instance (NI) are a subset or equal to those of an old instance (OI), the OI is deleted from the IS. For example:

$A\ B\ C \Rightarrow \overline{Z}$ (OI)

$A\ B \Rightarrow Z$ (NI)

By contrast, when there is no discriminant variable between the NI and the OI and the OI is not a subset of the NI, the OI is replaced by one or more modifications of the OI. The addition includes all of the OI which is not contradicted by the NI. Thus, the added instances contain the OI concatenated with one discriminant variable of the NI. For example,

$A\ B \Rightarrow Z$ (OI)

$A\ C\ D \Rightarrow \overline{Z}$ (NI)

will be replaced by

$A\ B\ \overline{C} \Rightarrow Z$

$A\ B\ \overline{D} \Rightarrow Z$

$A\ C\ D \Rightarrow \overline{Z}$

It should be noted that only the NI can cause that a network change be necessary. Instances added to the set in the manner of the last example are already fulfilled by the network since the root instance from which they are constructed was fulfilled.

Minimization of the instance set is an optional aspect of an ASOCS model. Minimality is not essential to correct functioning. It can provide savings in space and hardware, but only at the cost of increased adaptation time. ASOCS models using complete minimization cannot guarantee adaptation in linear time and its use is an optional decision for the implementor.

Although we have discussed the instance set as a model of the function of a network, the instance set need not be explicitly stored as such. The instance can be implicitly stored and maintained consistent in a distributed fashion by the function and interconnect of the network nodes.

3.2 ASOCS Architecture

The Adaptive Self-Organizing Concurrent System (ASOCS) is made up of boolean inputs and outputs and an input channel for instance input. There is no constraint on the number of outputs in relation to the number of inputs, although only 2^{2^n} unique outputs can exist for n inputs.

Figure 3 - ASOCS Structure

Figure 3 shows the internal structure of ASOCS. The two main components are the *adaption unit* (AU) and the *logic network*. During execution mode, only the logic network is active. The input data flows asynchronously through the network with only propagation delays. During adaptation mode, the other parts of the system become active. The AU and the logic network are connected by a broadcast bus. (The use of the word bus does not mean to imply any specific implementation. This and other names used in this chapter relate only to functionality and each can be implemented in a variety of ways.) The AU can broadcast to the entire logic network by placing a message on the broadcast bus, but it cannot address a specific node in the network. A node within the network can also place a message on the broadcast bus, which can be read in turn by the AU or any other node.

The AU can feed test data into the network on the *presentation path* and through the *input router*, which is controlled by the AU. During adaptation, the input router can get its input from the AU. During execution, the input router passes boolean inputs from the environment to the network. The AU can also monitor the outputs of the network through the test path.

It is also necessary to bind variables, in a flexible manner, to input and output lines. This is done for input and output variables by the *input binder* and *output binder*. Thus, when new (not previously used) input or output variables are used in a new instance, an input or output line will be allocated for the new variable. The feedback path allows output variables bound in the output binder to be fed back to the input binder.

The adaption unit (AU) guides the logic network through adaptation using global commands, which are messages placed on the broadcast bus. This allows all nodes within the network to work cooperatively by simultaneously executing procedures triggered by the command. Instances are input to the adaption unit. The global commands used are mentioned in the discussion of the individual algorithms.

3.3 Network Node

A single node within the logic network is represented in figure 4. Each node in the network has two basic parts: The control unit and a dyadic *dynamic programmable logic module* (DPLM). The DPLM is a logic gate which can be dynamically set to a subset of the 16 boolean functions of 2 inputs. During execution mode the DPLM is the main active element.

Figure 4 - Single Network Node

During adaptation the *control unit* can change the function of the DPLM. The control unit is a small finite state machine, with some random access memory, and the ability to communicate with the nodes to which it is directly connected. Each network node is identical, except for the contents of its memory, and the current function of its DPLM.

The control unit can send local commands on a bidirectional control path, which connects it to its immediate neighbors. The two nodes which give input to a node are known as the *child* nodes. The nodes to which a node sends its output are *parent* nodes. There may be more than one parent node for a given node, but there cannot be more than two children. The two nodes which input to a parent node are called *sibling* nodes with respect to each other.

ASOCS systems are amenable to fault tolerance by allowing testing and recovery mechanisms to be undertaken between the different nodes of the network. Testing and other activities can be continuously ongoing in the control units of the nodes, since the control units are not used during the normal processing activity of the network. These and other implementation issues are beyond the scope of this chapter.

4. Adaptive Algorithms

The next three sections give brief descriptions of three formally defined algorithms. These are called adaptive algorithms (AA) since they deal with how the system adapts to the input of new instances. The algorithms are labeled as AA1, AA2, and AA3. Each section first discusses any architectural variances from the basic architecture discussed in the last section. This is followed by an informal description of the working algorithm. The descriptions are regrettably short due to the scope of the chapter, but are intended to give a flavor of the type of methods used for the decentralized reconfigurations. More detailed discussion can be found in [5,6,7]. Although suited to any number of inputs, for sake of simplicity, the assumption of only one output variable continues.

4.1 Adaptive Algorithm 1

The main variant between the basic ASOCS architecture and that of adaptive algorithm 1 (AA1) is that 3-state DPLM's are used. The DPLM can be set to any of the 8 variations of the AND and OR functions with different inversions of the inputs.

In AA1, the instance set is maintained consistent and minimal in the adaption unit and is stored in a data structure called the *Instance Table* (IT) (figure 5).

P	N	D
$A\ \bar{B}\ D$	$\bar{A}\ \bar{B}\ \bar{E}$	
$B\ C\ D$	$B\ \bar{C}\ \bar{D}\ F$	
$\bar{A}\ E\ \bar{F}$	$A\ B\ \bar{D}$	
-	$A\ B\ \bar{C}$	
$A\ \bar{B}\ \bar{C}$	$B\ \bar{C}\ \bar{D}\ \bar{E}\ F$	
$A\ \bar{B}\ C$	$B\ \bar{C}\ \bar{E}\ G$	
$A\ \bar{B}\ D$	-	

Figure 5 - Instance Table

Positive instances are stored in the P column, negative instances in the N column, and the purpose of the D column is explained below. Cells which are empty are marked "-". The order of the stored instances is inconsequential.

When a new instance is input to the AU, it is stored in the first empty location in the instance table. For example, if the NI is a positive instance, then it is stored in the first empty cell in the positive column.

Each node control unit in the network contains a *node table* (NT). The function of the NT is to store the value of its DPLM output (1, 0, or ?, since it is a 3-state device) for each instance. For example, if a given node outputs 0 for all environment states matching an instance, a 0 is placed in the cell corresponding to that instance. The structure of the NT (i.e., its indexing), in terms of what each cell represents, is the same as that of the instance table. Thus, the third positive cell of every NT, corresponds to the instance in the third positive cell of the instance table. Consequently, a cell in the NT can be in one of four states represented by "1", "0", "?", and "-".

P	N	D
1	0	
?	0	
0	?	
-	1	
?	0	
0	1	
0	-	

Figure 6 - Node Table

Figure 6 shows an example of a possible NT which corresponds to the instance table shown in figure 5. When a new instance (NI) is input to the AU, it is *presented* to the network. A global presentation command is sent to the network giving the polarity of the NI. Presentation is done by setting the data inputs entering the network to the value of the NI. All inputs not corresponding to a variable found in the NI are set to "?". The values flow through the DPLM's and each control unit stores the value which its DPLM outputs in the first empty cell of the node table column corresponding to the NI. Since 3-state values are entered into the network, a node who's DPLM outputs a 1 (or 0) will output 1 (or 0) for all states of the environment that match the NI. If the top node of the network outputs the polarity of the NI, then the network already fulfills the NI and no changes need be made.

However, if the top node output does not match the NI polarity, then the network must be reconfigured to fulfill the instance set. This is done by selecting nodes in the network which can discriminate the NI from all old instances of the opposite polarity. Assume for a given node that the node table cell corresponding to the NI contains a 0. This node would then discriminate the NI from all discordant instances whose cell contains a 1 in its node table. Each node can sum up the number of discordant instances which it discriminates from the NI. This value is known as the *discriminant count*.

At this point the AU broadcasts the *selection* global command. This initiates a parallel heap sort, where each node sends its discriminant count to its parent, and the parent forwards the maximum count to its parent. Thus, the top node will have received the maximum discriminant count in the entire network in $O(log(n))$ time where n is the number of network nodes. A downward wave is then initiated, resulting in the selection of the single "best" node in the network. Ties are decided arbitrarily. The selected node sets itself as a *growth* node. Since a single node is selected, this node can place information on the broadcast bus without bus conflict, which can in turn be read by all other nodes in the network and by the AU. In this way the selected node can convey to the rest of the network the amount of discrimination it does and what still remains to be done. The message placed on the bus is a bit vector containing a 1 for each discordant instance that the node discriminated. Before the beginning of the selection wave, each node sets each cell in its D column to 0. After the bit vector is broadcast, each cell then replaces the contents of its D column with the logical OR of its D column and the broadcast vector. If the new vector still contains 0's a new selection wave is initiated. When each node sums its discriminant count, it only counts instances with a 0 in D cell of the same row, signifying that the instance has not yet been discriminated. The AU updates the D column of the instance table in the same fashion, and is thus able to correctly initiate selection waves. When the D column of the AU contains all 1's then sufficient nodes have been selected to discriminate the NI.

If the maximum discriminant count in the network is 0, or below some threshold, and the D column still contains 0's, then new nodes must be added to the network in order to discriminate the NI. The AU can guide this new node addition since it knows which instances remain un-discriminated. A minimal set of variables necessary to discriminate the NI can be calculated from the instance table and they become inputs to newly allocated nodes, which are also set as growth nodes.

Once sufficient growth nodes are selected, the combination phase ensues. Growth nodes combine, allocating a new parent to which their outputs enter. The DPLM function and the node table of the new parent can be set by the two children nodes without outside intervention. The new parent is then set as a growth node. Combinations can take place concurrently and independently throughout the network. After all growth nodes have combined, a new node which discriminates the NI from all discordant instances will have been created. This node can then be combined, again using only information of the children, with the old top node, which correctly discriminates all instances except the NI, to create a new top node which fulfills the total instance set.

After network modification, the final phase of processing is self-deletion. A node, by viewing the contents of its node table, can deduce to what extent it aids in the overall functioning of the network. If a node discovers that it is no longer necessary, for example if it does not discriminate at least one positive instance from a negative instance, then it will self-delete. Upon self-deletion, the node sends local commands to its children for them to also self-delete. Thus, large portions of the network can be pruned away, leaving behind a correct network which is at all times relatively close to an optimal network. If a node discovers that it fulfills the entire instance set (it has all

1's in its P column and all 0's in its N column) then all nodes above it can be deleted and it will become the new top node. There are a number of other ways by which self-deletion can take place.

To summarize, in AA1 there is always a single top node for each output variable which is a complete-discriminant node for that variable. When a new instance is added to the set, and if the top node does not already fulfill the new instance, a new node is created which discriminates the new instance from all old discordant instances. This node is created by combining nodes within the network which already discriminate a subset of the old discordant instances from the new instance. If the union of these selected nodes is not sufficient to discriminate the new instance from all discordant instances, then new nodes are allocated which discriminate the remaining instances. Once the new node (which discriminates the new instance from all discordant instances) is built, it can be combined with the old top node (which discriminates all old instances, but not the new instance) creating a new top node which is a complete discriminant node for the given output variable. After the network has reconfigured to fulfill the NI, which takes time linear with the depth of the network, the whole process can iterate and more instances be input to the AU. A detailed presentation of AA1 is given in [7].

4.2 Adaptive Algorithm 2

AA2 uses a method much different than that of AA1. Instead of having a single top node for an output variable, there can be a number of top nodes for an output. These are divided into *positive* and *negative* nodes. If any positive node outputs a 1, then the output of the network must be a 1. If any negative node outputs a 1, then the output of the network must be a 0. Outputs from top nodes are OR'ed together. In AA2 there is one top node matching each instance in the instance set. Thus the network is maintained as a consistent and minimal instance set. There is no need to maintain the instance set in the AU, since the network implicitly maintains it.

There are two major architectural differences between AA1 and AA2. First, the DPLM in AA2 is a 2-state gate which is always set to the AND function and need not be programmable. Rather than store a table of outputs for all instances, each node stores a list of the variables it conjuncts. A node whose variable list matches an instance is a D-node (discriminant node). The variable list of a D-node is built up by conjunctions of P-nodes (primitive nodes). Each D-node has a polarity which is that of the instance it matches.

The second major difference in AA2 is an added construct called the *or-plane*. Figure 7 shows a general representation of the structure of the logic network.

Figure 7 - AA2 Logic Network Structure

Inputs to D-nodes must be from either a P-node, or a literal variable (i.e. A, B, etc.). The D-nodes are the top nodes of the network. However, D-nodes can be found at any depth of the network. The inputs to the or-planes must be from D-nodes.

The outputs of all positive D-nodes enter into the positive or-plane. The outputs of all negative D-nodes enter into the negative or-plane. The or-planes need not be adaptive, The logical structure of the or-planes, which includes the positive and the negative planes, is shown in figure 8.

Figure 8 - Or-Planes

A 3-node structure handles the or-plane outputs. The middle node is the output node. It outputs 1 if the positive plane outputs a 1, and it outputs 0 if the negative plane outputs a 1. If both planes output a 0, then the don't know output is active. With this output, one can always know whether an instance is matched or whether the environment is in a don't know state. If both planes output a 1, then there is an error in the network.

As can be seen by viewing the AA2 structure, the algorithm implements a sum-of-products expression. The products are the instances represented by the D-nodes in the network and each D-node is equivalent to one instance in the instance set. The products represented by the D-nodes are summed by the or-planes.

We now discuss the operational sequence of the algorithm. When a new instance (NI) is input to the AU, there is no storage or processing necessary in the AU, since the instance is both stored and maintained implicitly in the network. Presentation is not used to introduce instances to the network as in AA1, thus there is no need for 3-state DPLM's. A NI is simply broadcast to the network as a list of variables (with some variables possibly negated), along with the polarity of the NI. Each node in the network can then directly compare the NI with their own variable list, and independently deduce a necessary action. There is no need for selection waves since nodes can select themselves immediately and independently after viewing the NI.

Nodes execute unique procedures depending on how they match with the NI. For example, if a D-node exactly matches the variables of the NI, but is of the opposite polarity, it simply changes its polarity and redirects its output to the opposite or-plane. This is equivalent to deleting the old instance and adding the NI to the set. If a node's variable list was a superset to the NI, it would self-delete regardless of its polarity. If no node matches the NI, then a new D-node is created to match the NI by combining P-nodes which are a subset of the NI. There are 15 ways in which a node can match a NI, and 8 basic actions nodes can take depending on how they match [5].

4.3 Adaptive Algorithm 3

Most of the constructs defined in the previous section carry over to AA3 and are not restated. AA3 is similar, although simpler than AA2. Rather than have one D-node for each instance in the instance set, it is only necessary to have sufficient D-nodes to properly discriminate the instance set. When a new instance is broadcast, only discordant D-nodes can be matched with the new instance, thus the only test is for contradiction. The main advantage to AA3, besides simplicity, is that fixed interconnections between nodes are possible, allowing for a potentially easier physical implementation.

The DPLM function of each node is always the AND function. There are positive D-nodes, negative D-nodes, and P-nodes. All nodes have one *variable* child and one *node* child. A variable child is a literal variable, possibly inverted, which inputs to the node's DPLM. The node child is a network node. Each P-node has two parent nodes. A D-node has no parent nodes. All nodes also have a *variable-list* and a *polarity-flag*. Outputs from D-nodes exit the network to the sides and join the or-planes. Since inputs enter at all levels of the network the variable inputs also enter the networks from the side. Figure 9 gives a representation of the AA3 network structure.

Figure 9 - AA3 Network Structure

The overall structure is that of a binary decision tree. The bottom node of the network is called the *root* node and it is unique from all other nodes in the network. It has no inputs, and its DPLM always outputs a "1". Otherwise it is identical to any other node. Figure 10 gives an illustration of a possible internal network structure of AA3. The connections from the D-nodes into the or-plane and from the input binder to the nodes are not shown.

Figure 10 - AA3 Internal Network

The flexible portion of the network is at the input and output binders. The perpendicular traffic (variable inputs and outputs) needs to be dynamically modifiable. It is by changing the variable inputs to the nodes, and the polarity of the nodes, that the network adapts.

As in the previous algorithm, there is no need to store the instance set in the AU, since it is implicitly stored in the network. However, in AA3 it is impossible to extract the exact instance set from the network, since there is not a unique node for each instance. The network still does, however, correctly fulfill the instance set. A NI is broadcast to the network and the nodes initiate a specific action depending on how their variable list and polarity match with the NI. In AA3 only discordant D-nodes can match a NI, and there are only two types of possible action.

If the node's variable list is a superset or equal to the NI, the node simply complements it's polarity flag and redirects its output to the opposite or-plane. If the node's variable list is not a superset of the NI and it is also not discriminated (i.e. there does not exist at least one discriminant variable between the NI and the variable list), then the operation of discriminant variable addition takes place. Briefly, this is accomplished by growing new nodes starting from the matched node using variables from the NI which do not occur in the node's variable list. The matched node becomes a P-node and the parents become D-nodes. As in AA2, all node actions take place concurrently and independently, allowing for high adaptation speed.

Self-deletion follows the network modification, and is defined in one simple operation. Any two sibling D-nodes which have the same polarity both self-delete and their node child becomes a D-node with the same polarity. This operation can continue recursively and in parallel on all D-nodes.

5. Comparison of Algorithms

In this section we compare the features of the three algorithms discussed above. Software simulation of all three algorithms was used to understand the efficiency of the algorithms for large instance sets. All metrics are dependent on the specific instance set and thus a large number of random instance sets were applied to each algorithm to obtain the average behavior of the different systems. One measure of the total size of the network is number of nodes relative to the number of instances in the current instance set. The number of nodes in an AA1 network is typically equal to the number of instances in the set. AA2 and AA3 usually have two times as many nodes as rules. However, the network depth in AA1 is larger than in the others. The maximum network depth, which is the main bound on speed in an ASOCS system, is equal to the number of bound input variables for both AA2 and AA3.

The main disadvantage of AA1 is the necessity to have memory at each node proportional to the size of the instance set (2 bits per instance). This memory must also grow in proportion to the number of output variables handled by the system. This problem is mitigated in both AA2 and AA3 by use of constant sized variable lists which can be no larger than the number of input variables in the system. This memory necessity can be mitigated completely through use of a different presentation technique [2]. Also, there is only negligible increase of memory per node for systems with multiple output and feedback variables. AA2 and AA3 have the advantage of using a fixed 2-state gate, and they do not require presentation and selection waves.

In AA1 it is necessary to both store and maintain the instance set in the adaption unit. In AA2 and AA3 the instance set is both stored and maintained implicitly within the network structure.

AA2 has the unique ability to signal whether the current environment input matches any instance in the set, a more powerful notion than fulfilling the set as previously defined. AA3 loses this ability by using a simpler scheme of only testing for contradiction, and not for minimization. AA3, on the other hand, can be implemented on a network without flexible connections between nodes.

6. Conclusion

This chapter has reviewed features of a new class of multilayered connectionist models called adaptive self-organizing concurrent systems (ASOCS). Features include guaranteed mapping of arbitrary boolean functions, bounded linear learning time, stability of previously learned patterns, and ability to extract critical features from a large environmental input. Initial fabrication and testing of VLSI chips is underway. Current research includes implementation mechanisms, expansion of knowledge input mechanisms to include higher level data and analog inputs, rules with variable bindings, coupling with memory mechanisms, and use of feedback and sequencing.

7. Bibliography

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

2. Chang, J. and J. J. Vidal, Inferencing in Hardware, *Proceedings of the MCC-University Research Symposium*, Austin, TX, (July 1987).

3. Hinton, G., Sejnowski, T, and D. Ackley, *Boltzmann Machines: Constraint Satisfaction Networks that Learn,"*Tech. Rep CMU-CS-84-119, CMU, Pittsburgh, PA, (1984).

4. Kohonen, T., *Self-organization and associative memory*, Springer Verlag, New York, (1984).

5. Martinez, T. R., *Adaptive Self-Organizing Logic Networks*, Ph.D. Dissertation, Technical Report - CSD 860093, University of California, Los Angeles, CA (May 1986).

6. Martinez T. R., Models of Parallel Adaptive Logic, *Proceedings of the 1987 IEEE Systems Man and Cybernetics Conference*, pp. 290-296, (October, 1987).

7. Martinez, T. R. and J. J. Vidal, Adaptive Parallel Logic Networks, *Journal of Parallel and Distributed Computing*, Vol. **5**, pp. 26-58, (1988).

8. Martinez, T.R., Digital Neural Networks, *Proceedings of the 1988 IEEE Systems Man and Cybernetics Conference*, pp. 681-684, (August, 1988).

9. Rosenblatt, F., Principles of Neurodynamics, Spartan Books, Washington, D.C. (1962).

10. Rumelhart, D. and McClelland, J., *Parallel Distributed Processing:*, Vol. I, pp. 318-362, MIT Press, (1986).

11. Verstraete, R.A., "Assignment of Functional Responsibility in Perceptrons," Ph.D. Dissertation, Computer Science Department, University of California, Los Angeles, CA, (June 1986).

12. Yau, S. S. and C. K. Tang, "Universal Logic Circuits and their Modular Realizations," AFIPS Conference Proceedings, vol. 32, pp. 297-305, (1968).