

N 9 3 - 2 5 9 7 4

## DATA FUSION WITH ARTIFICIAL NEURAL NETWORKS (ANN) FOR CLASSIFICATION OF EARTH SURFACE FROM MICROWAVE SATELLITE MEASUREMENTS

Y. M. Fleming Lure<sup>+</sup>, Norman C. Grody<sup>\*</sup>, Y. S. Peter Chiou<sup>+</sup>, and H. Y. Michael Yeh<sup>+</sup>

<sup>+</sup> Caelum Research Corporation, Silver Spring, MD 20901 (Tel) (301) 593-1748, (FAX) (301) 593-3951

<sup>\*</sup> National Oceanic and Atmospheric Administration (NOAA), Washington, D. C. 20233

### ABSTRACT

A data fusion system with artificial neural networks (ANN) is used for fast and accurate classification of five earth surface conditions and surface changes, based on seven SSMI multi-channel microwave satellite measurements. The measurements include brightness temperatures at 19, 22, 37, and 85 GHz at both H and V polarizations (only V at 22 GHz). The seven channel measurements are processed through a convolution computation such that all measurements are located at same grid. Five surface classes including non-scattering surface, precipitation over land, over ocean, snow, and desert are identified from ground-truth observations. The system processes sensory data in three consecutive phases : (1) pre-processing to extract feature vectors and enhance separability among detected classes; (2) preliminary classification of Earth surface patterns using two separate and parallelly acting classifiers: back-propagation neural network and binary decision tree classifiers; and (3) data fusion of results from preliminary classifiers to obtain the optimal performance in overall classification. Both the binary decision tree classifier and the fusion processing centers are implemented by neural network architectures. The fusion system configuration is a hierarchical neural network architecture, in which each functional neural net will handle different processing phases in a pipelined fashion. There is a total of around 13,500 samples for this analysis, of which 4% are used as the training set and 96% as the testing set. After training, this classification system is able to bring up the detection accuracy to 94% compared with 88% for back-propagation artificial neural networks and 80% for binary decision tree classifiers. The neural network data fusion classification is currently under progress to be integrated in an image processing system at NOAA and to be implemented in a prototype of a massively parallel and dynamically reconfigurable Modular Neural Ring (MNR).

# 1. INTRODUCTION

Artificial neural networks (ANN) have demonstrated capabilities for robust pattern classification in the presence of noise and object-to-background sensory uncertainty, and have found applications in environmental monitoring including land cover determination, vegetable mapping, soil survey, etc., or multichannel satellite imagery. This paper presents a data fusion system with artificial neural networks which will utilize multichannel SSMI satellite imagery, to combine supervised trainable and self-organized neural network architectures with specific knowledge-based classification techniques, with reference to fast and accurate classification of the earth surface. This neural approach is intended to compensate for different classification techniques by using the data fusion method and to reduce the lengthy training time required in a supervised learning network. The overall neural network data fusion system, which will be described in more detail, can also be seen as a four-layered supervised network which is composed of several modular and hierarchical networks. In this paper, we will start with a background discussion of the measurement used in this study. The data fusion classification system will be presented. Hardware implementation of each component in a Modular Parallel Ring (MPR) will also be discussed. Some experimental results will be presented and a summary will be given.

# 2. BACKGROUND

The SSMI instrument, flown on board the Defense Meteorological Satellite Program (DMSP) polar orbiting satellites, is a seven-channel conically-scanning microwave radiometer, measuring brightness temperatures at 19, 22, 37, and 85 GHz. All measurements are obtained with dual polarizations (H and V) except for 22 GHz channel. The 19 and 22 GHz channels are mainly responsive to variations in temperature and water vapor at large spatial scale. The 37 and 85 GHz channels, due to the scattering effects at high frequencies, respond to precipitation at smaller scale. Polarization measurements have been used to infer the wind speed, precipitation, and snow cover over the land and ocean. The spatial resolution (field of view) of the different channels decreases in proportion to the wavelength (inverse with frequency). It provides unique signatures for identifying surface features and obtaining the temperature and condition of the Earth's atmosphere. In comparing the measurements at different frequencies, effects due to different spatial resolutions are minimized by convolving all measurements to the 55-km resolution of the lowest-frequency channel (Grody, 1991). This enables one to investigate the spectral variations without having to consider the effects of spatial inhomogeneity on the different channel measurements. The measurements (brightness temperature, sometimes called antenna temperatures) used in this study were made between November 1988 and January 1989 and covers the entire northern hemisphere. The data was identified and confirmed by "ground truth" as five different data sets corresponding to five different surface classes: non-scattering medium (Non-Sm), precipitation over the ocean (R-Ocean), snow cover land (Snow), precipitation over the land (R-Land), and the desert (Desert). Each class has different samples ranging from 445 to 5535 and there is a total of over 13,034 samples. Table 1 illustrates some SSMI measurement classification characteristics including SSMI measurements, surface features and their corresponding samples. The brightness temperatures are normalized within the range of (-1, +1), denoted as  $X_i$ , and the desired output classes are represented by mutually orthogonal vectors, denoted as  $C_j$ .

SSMI	Channel frequencies and polarizations						
	19 H $T_H(19)$	19 V $T_V(19)$	22 V $T_V(22)$	37 H $T_H(37)$	37 V $T_V(37)$	85 H $T_H(85)$	85 V $T_V(85)$
Surface features:	Non-Sm		R-Ocean	Snow	R-Land	Desert	
Number of samples:	4294		505	5535	2255	445	

### 3. DATA FUSION CLASSIFICATION SYSTEM

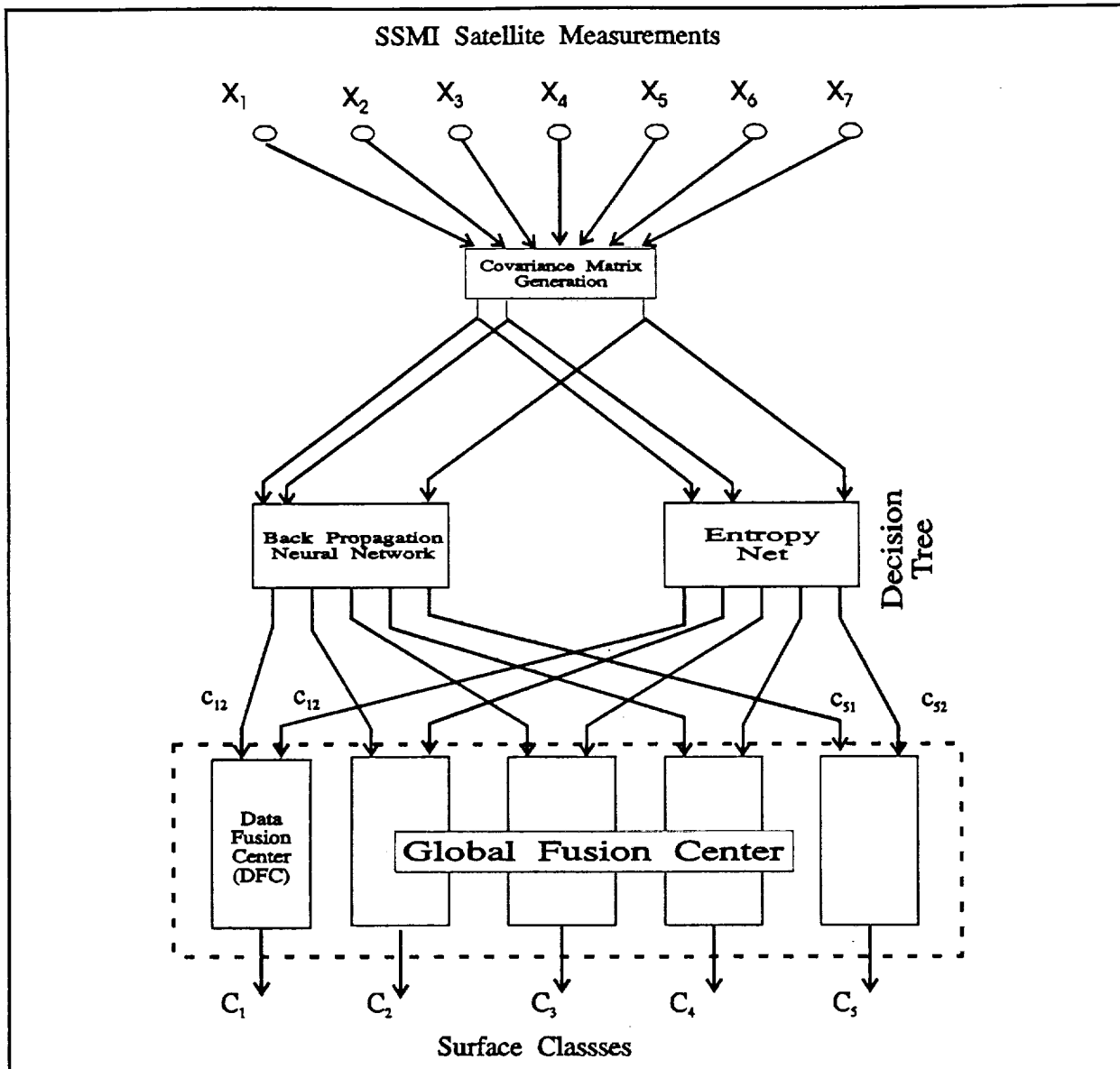


Figure 1. Data Fusion System with Artificial Neural Networks for SSMI Measurements

Although existing neural network paradigms have demonstrated excellent capabilities in learning and generalization, efficient training and determination of internal topology (such as number of hidden neurons) still remain challenging tasks. This data fusion classification system implemented with ANNs provides an alternative approach to attack these problems and can be easily implemented in hardware. Basically, this system treats each classifier as a different sensor and fuses each classification result to obtain the optimal or better results. The term "optimal" is used such that the probability of error is minimized in the likelihood ratio test. The sizes and connections of intermediate layers (or hidden layers) can be determined based upon the desired data flow.

This fusion classification system will process sensory data in three consecutive phases, as follows: (1) pre-processing, aimed at extracting feature vectors and at enhancing separability among detected classes; (2) preliminary classification of Earth surface patterns at two separate and parallel acting classifiers: back-propagation ANN (BP ANN) and a binary decision tree (BDT); and (3) fusion of classification results performed at global fusion center (GFC) from different classifiers and imagery to obtain the optimal decision. The configuration is a hierarchical neural network architecture, in which each functional neural net will handle different processing phases in a pipelined fashion.

### **3.1 Pre-processing**

Pre-processing for SSMI imagery includes mainly the generation of (7 x 7) covariance matrices from measured brightness temperatures at each pixel. Information about pixel brightness temperatures, covariance matrix elements, and desired surface class definitions is collected in a feature vector for the supervised training of a neural network classifier. It has been demonstrated that increasing the elements of the feature vector by adding more relevant parameters, derived nonlinearly from original features, can reduce the number and size of hidden layers, and can also reduce the training time (Marks, et al., 1988). Since the covariance matrix evaluation involves the manipulation of two matrices, the operations involved are suitable to neural network implementation by feed-forward topologies, by merely assigning two manipulated matrices to the weights and input vectors of the back propagation neural architecture, as has been investigated.

### **3.2 Preliminary Classification**

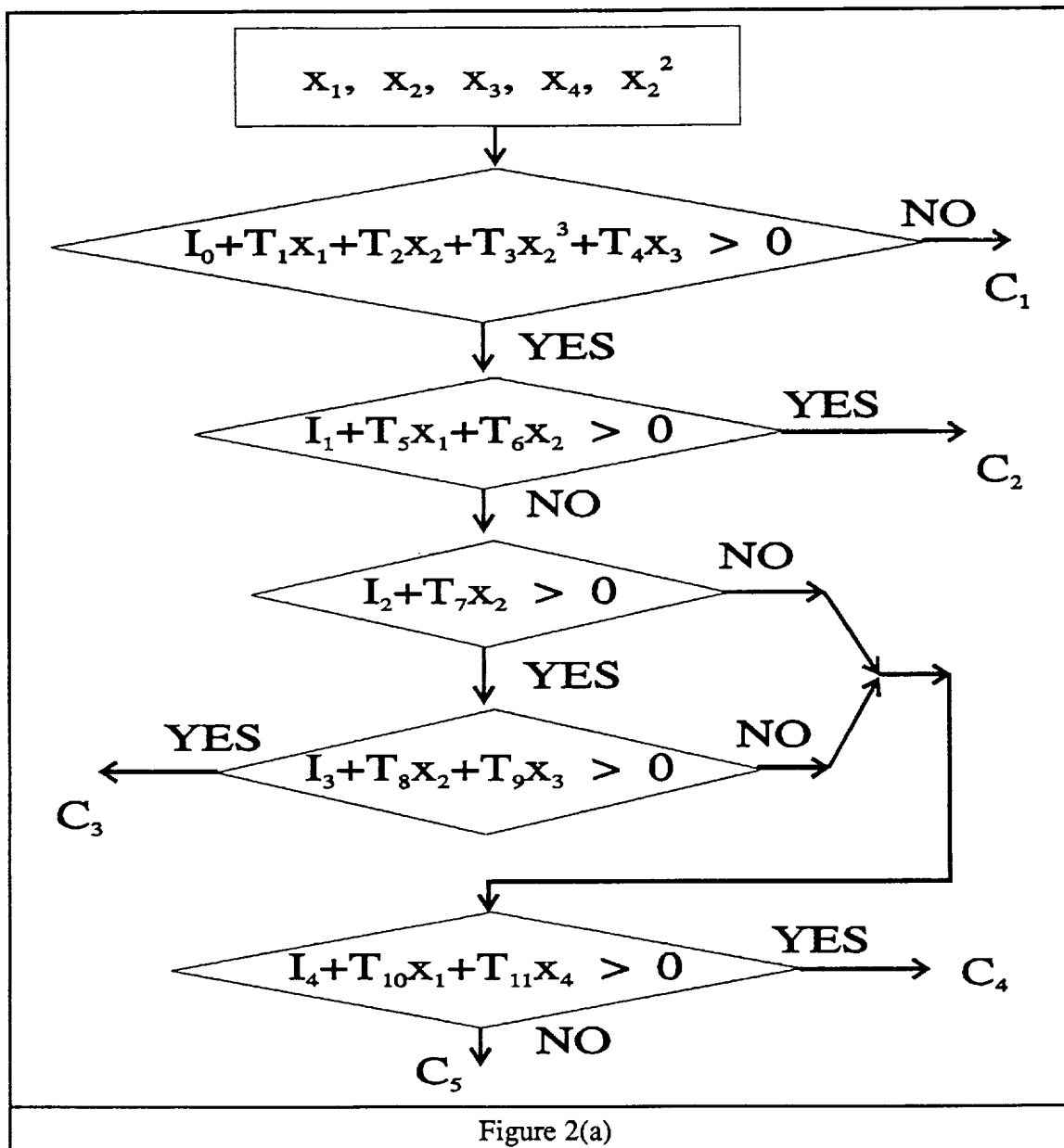
#### **3.2.1 BP ANN Classifier**

A three-layer (one hidden layer) supervised back propagation (BP ANN) algorithm is used to train the network to become a feed forward pattern recognition engine (Rumelhart and McClelland, 1991) to learn the input feature vectors corresponding to different output classes. There are 14 input neurons corresponding to SSMI measurements as well as to their covariance matrix, 60 hidden neurons, and 5 output neurons representing 5 surface conditions. It takes around 40 and 160 epoches to train the BP ANN classifier to learn up to 90% and 100% accuracy of the training data set, respectively. With a fully-trained BP ANN, the classification accuracy can reach up to 88% (Lure, et al., 1992a, 1992b). For the data fusion classification system, the BP ANN is only trained to a "satisfactory" accuracy (e.g., 75%). Such a "partially" trained ANN only takes around 50% of the training time required in fully-trained nets. A single fully-trained network can only reach a certain detection accuracy limit whereas a combination of several networks such as this one can reach even higher precision since the fusion processor will make an optimal decision based on the statistics of preliminary classification accuracy.

#### **3.2.2 BDT Classifier**

The BDT classifier is constructed to implement Grody's global classification algorithms as in Figure 2 (Grody, 1991). They are designed to analyze global coverage of satellite data sets and to classify based on the physical characteristics of measurements and on surface types. This technique performs a hierarchical tree-structured decision procedure through the evaluation of polynomial functions of input feature elements and through thresholding. The special topology of BDT classifiers used for surface condition classification based on SSMI measurements is drawn from the so-called Entropy Net architecture (Sethi, 1990). This architecture includes a two-layered topology, of which the lower layer performs arbitrary mapping of thresholding operations, while the upper layer performs logical operations (e.g. AND, OR) which allow us to convert the hierarchical decision procedure into a fully parallel process. The weight vectors between the layers are determined from

the coefficients of polynomial functions of the decision tree functions. The logical operations, such as AND, OR, NOR, and NAND, are implemented by using a simple BP ANN architecture with sigmoid transfer functions (Lippmann, 1987). A striking advantage of the neural implementation architecture is that it allows us to specify the number of neurons needed in each layer, along with the desired output. This, in turn, leads to an accelerated progressive training procedure that also allows each layer to be trained separately.



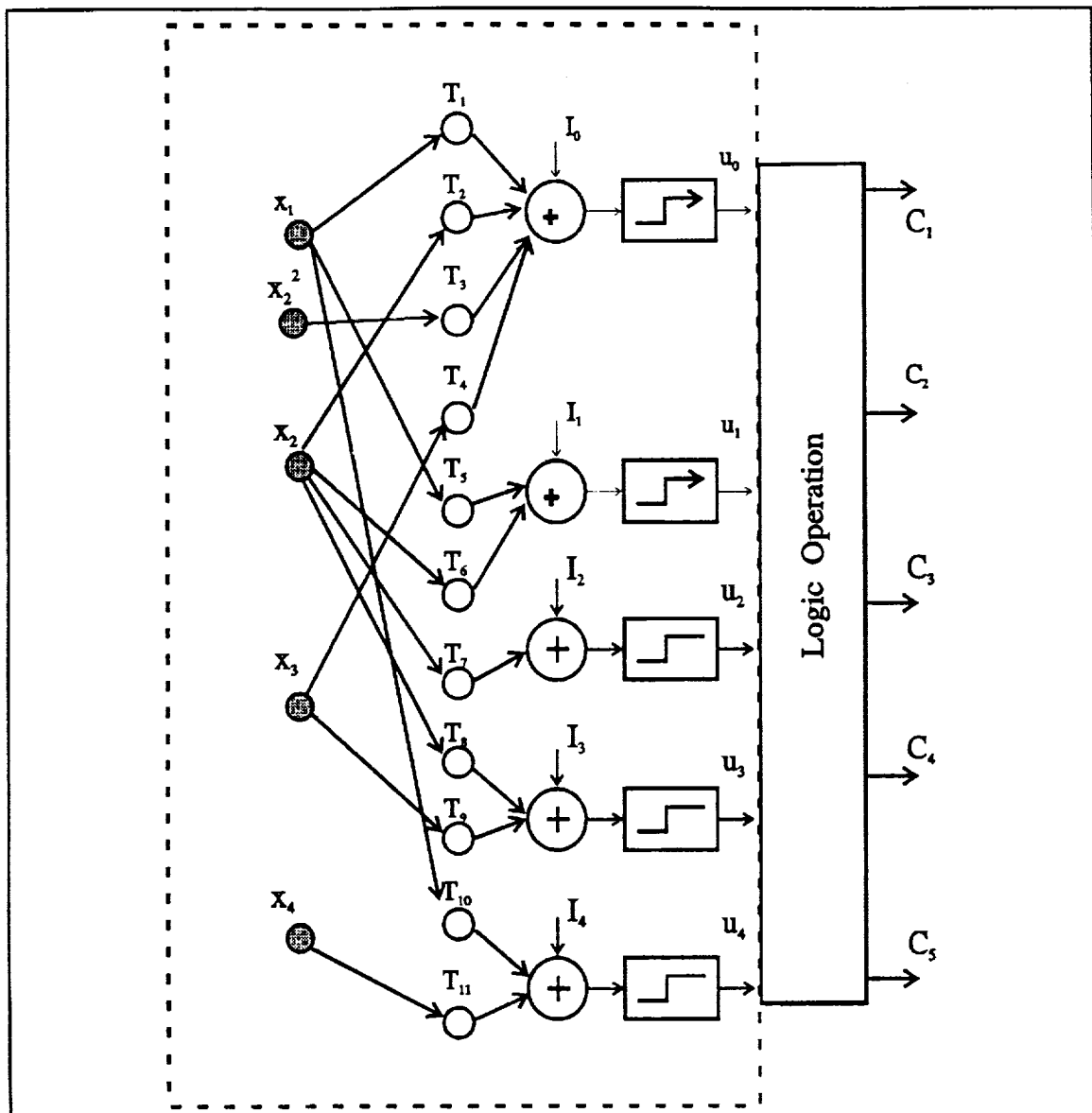
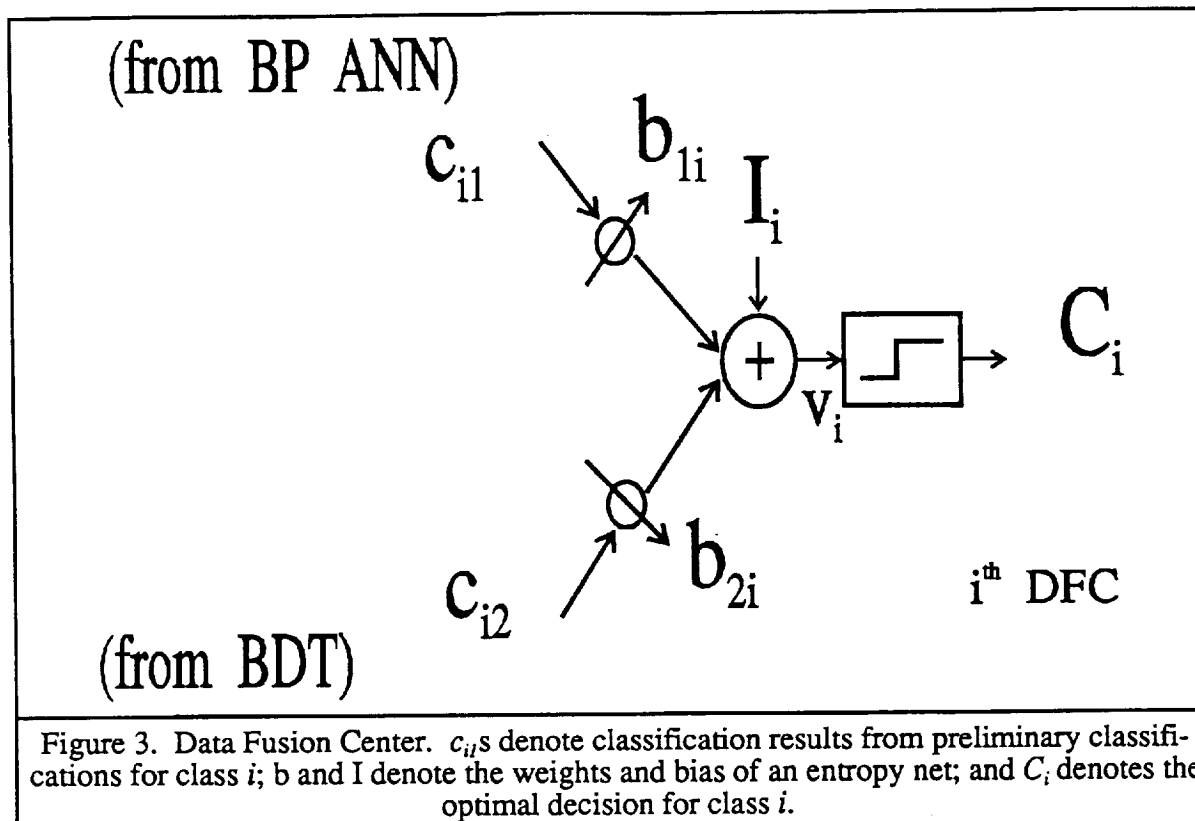


Figure 2(b)

Figure 2. (a) BDT Classifier and (b) its Neural Implementation.  $X_i$ 's denote the SSMI measurements;  $T_i$ 's denote the higher order polynomial coefficients in (a) and weights in (b); and  $I_i$ 's denote constants in (a) and biases in (b), respectively.

There are 5 neurons corresponding to 4 selected SSMI measurements and to one element of the covariance matrix ( $X_1$ ,  $X_2$ ,  $X_3$ ,  $X_4$ , and  $X_2^2$ ), and 5 output neurons for each surface class. The individual decision from both BP ANN and BDT modules are sent to the global fusion center (GFC) for the final decision. The two-trainable-layered BP neural net for logical operation is trained based upon the data derived from known logic relationships from the decision tree. As for other neural networks for logic operations, it only takes a few epoches for them to learn the desired patterns.

### 3.3 Fusion Processing



The fusion processing involves global fusion center (GFC) operations, which integrate results from both BP ANN and BDT classifiers. The GFC is composed of several different data fusion centers (DFC), each of which corresponds to different types of output classes as in Figure 3. A self-adjusted or self-trained learning algorithm is used in each DFC to set the optimal decision rules such that the total probability of detection is maximized. This data fusion scheme, also called distributed-detection scheme, corresponds to a two-layered network of nonlinear threshold elements, e.g., binary or sigmoidal functions (Tenney, 1981). The decision operation, weights and bias of these elements are obtained as

$$v_i = I_i + \sum_{j=1}^n b_{ji} u_{ij}$$

$$b_i = \log\left(\frac{(1 - P_{M_i})(1 - P_{F_i})}{P_{F_i} P_{M_i}}\right)$$

and

$$I = -2 \log\left(\frac{P(H_0)}{P(H_1)}\right) + \log\left(\prod_{i=1}^n \frac{(1 - P_{M_i})}{P_{F_i}}\right) + \log\left(\prod_{i=1}^n \frac{P_{M_i}}{(1 - P_{F_i})}\right)$$

where  $n$  denotes the number of classifiers ( $n = 2$ ),  $P_{M_i}$  represents missed detection in the  $i$ th classifier,  $P_{F_i}$  represents a false alarm in the  $i$ th classifier,  $P(H_1)$  denotes the probability that the desired class is present, and  $P(H_0)$  denotes the probability that the desired class is absent.

The probability functions  $P_s$  are obtained during training by comparing individual classification results with the desired class. The fusion networks are trained by self-adapting off-line stochastic information to form the detection system. The stochastic information including a priori probability, the probability of false alarm, and of missed detection is obtained by comparing classification results from individual classifiers with ground-truth data. The approximation rules are obtained from the nonlinear combination of the statistics of previous classification results from individual classifiers.

#### 4. HARDWARE IMPLEMENTATION

The neural network data fusion system for real time processing is implemented in a prototype of a massively parallel and dynamically reconfigurable Modular Neural Ring (MNR) architecture (Ligomenides, et al., 1991), which is capable of maintaining a high performance for digital and neural applications. The MNR architecture is composed of multiple primitive processing rings (pRing) embedded in a global communication structure and is interfaced to a host workstation as in Figure 4. It is a multiple-SIMD (single instruction multiple data) architecture. Each of the pRings consists of 40 processing elements (PE) that are capable of mapping any number of neurons. It has been shown that the MNR provides very highly efficient hardware utilization and very low communication delay overhead. The achieved speed/capacity performance is increased linearly with the number of processing elements, without upper limit.

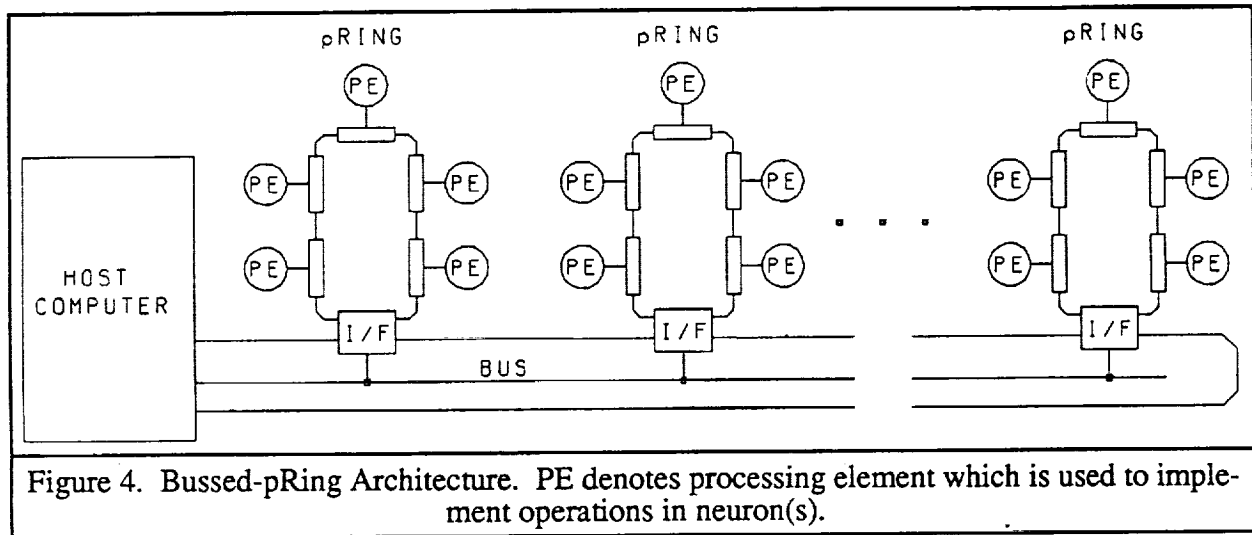


Figure 4. Bussed-pRing Architecture. PE denotes processing element which is used to implement operations in neuron(s).

Covariance matrix evaluation, involving the manipulation of two matrices, is performed by merely assigning two manipulated matrices to the weights and input vectors of the feed forward neural architecture. Two pRings are used to implement the BP ANN module: one for handling the  $16 \times 64$  weight matrix of input-hidden connection and one for the  $64 \times 16$  weight matrix of hidden-output connections. The third pRing is used for the parallel implementation of the BDT, which handles a  $16 \times 16$  weight matrix. Since some weights are not utilized (for example, the input-hidden connection in BP ANN only requires a  $14 \times 61$  weight matrix), they are filled with zero weights to satisfy hardware implementation requirements. The operation and performance of the hardware-based networks remain almost unchanged. Once the training is finished, the weights and bias are then stored in the memory of each PE for future processing. Both BP ANN and DBT operations are performed at the MNR architecture simultaneously. The individual decision from each operation is then fed to the data fusion center (DFC) for final optimum decision performed at the host computer.



## 5. CLASSIFICATION RESULTS

There are a total of 13,034 samples of data used in this study. Each of five different classes contains from 400 to 5,000 different samples. We used 500 samples of data as training sets which represent 3.8% of the total samples. Each training set, obtained randomly from the total data set, consists of an equal number of samples from five different classes. The rest of the samples (over 96%) are used for testing the network and the classification results are shown in Table 2. Once the BP ANN is trained either fully or partially, it is used to perform the classification. The classification accuracies, using the fully-trained BP ANN classifier (i.e., all training patterns are recognized by this BP ANN), are 82%, 98%, 97%, 78%, and 79% for non-scattering medium, precipitation over ocean, snow, desert, and precipitation over land, respectively (Lure, et al., 1992). The classification accuracies are 99%, 56%, 81%, 57%, and 70% for each surface class. Note that the class of non-scattering medium represents the surface which can not easily be specifically identified as any of the other four surfaces. The overall accuracy for BP ANN approach is around 88% whereas it is around 80% for BDT classifier. The preliminary results show that the neural network data fusion system improves the classification accuracy for all classes by around 4% from BP ANN's results. The overall accuracy of neural network data fusion is improved to 94%. Even without fully-trained being (e.g., 75% of training set are learned correctly by BP ANN) the overall classification accuracy can still achieve similar classification accuracies. From the coefficients of the data fusion center, it is also found that the BP ANN plays a more important role in classifying the non-scattering medium, snow, and desert; whereas the BDT is more dominant in classifying the other two surfaces. The significance of each SSMI measurement to classification of each of five surface types can also be obtained through the linearization procedure of the weights described in the previous study.

ALGORITHM	Non-sm	R-Ocean	Snow	Desert	R-Land	Overall
BDT	99%	56%	81%	57%	70%	80%
BP ANN	82%	98%	97%	78%	79%	88%
ANN FUSION	86%	98%	97%	84%	83%	94%

## 6. SUMMARY

In this research effort, a data fusion system with artificial neural networks is presented to classify surface types based on the SSMI measurements. Both back propagation ANN (BP ANN) and binary decision tree (BDT) classifiers are used for this study. Seven SSMI measurements (brightness temperature at 19, 22, 37, and 85 GHz for H and V polarizations, except V for 37 GHz) at each image pixel are extracted as an input feature vector. Five surface types including non-scattering medium, precipitation over the ocean, snow cover land, precipitation over the land, and the desert are used as target patterns. After training by using less than 4% of the samples, both BP ANN and BDT are able to perform the classification over 13,000 samples. The training for this data fusion system is performed progressively. The BP ANN, first module of entropy net, and logical operation net, are trained separately. Once these are trained, each data fusion center network is trained separately. The overall accuracy for the BP ANN and the BDT approaches 88% and 80%, respectively. The neural network data fusion system which fused the individual decision from the BP ANN and the BDT improved the overall accuracy to 94%. The significance of the contribution from either approach is determined based on the coefficients of the data fusion center. The fusion system is currently implemented in a massively parallel and dynamically reconfigurable hardware neural network (Modular Neural Ring) for real time parallel processing and integrated in an image processing system at NOAA/NESDIS. The data fusion classification system not only preserves

the advantages of both BP ANN and BDT classifiers (for example, the capability of physical interpretation of input feature space from the BDT classifier and robust classification from the BP ANN), but also reduce the pitfall of individual classifiers (for example, brute-force training of the BP ANN module and sensitivity to noise of the BDT module).

## ACKNOWLEDGEMENTS

This work is partially supported by NOAA/NESDIS and Air Force Rome Development Center under SBIR Phase II contract: F30602-89-C-0130. The SSMI data set was provided by Hughes Aircraft Co. through NOAA/NESDIS. The authors are grateful for the discussion with P. A. Ligomenides and L. Jump of Computer Systems and Architecture Group at Caelum Research Corporation. The authors are also thankful for the comments by the reviewers. This paper is submitted to IEEE Transaction of Neural Network. Due to page limitation, a more detailed information can be found in the submitted paper.

## REFERENCES:

- Grody, N.C., 1991, "Classification of Snow Cover and Precipitation Using the Special Sensor Microwave Imager", J. Geophys. Res., Vol. 96, No. D5, 7423-7435.
- Ligomenides, P. A., L. B. Jump, and Y-S. Chiou, 1991: "A Reconfigurable Ring Architecture for Large Scale Neural Networks", Proc. of Conf. Fuzzy and Neural Syst. and Vehi. Appli. 1991, Tokyo, Japan.
- Lippmann, R. P., 1987, "An Introduction to Computing with Neural Nets". IEEE Acous. Spec. Signal Proc. Magazine, 4-22.
- Marks II, R. J., L. E. Atlas, and S. Oh, 1988: "Generalization in Layered Classification Neural Networks", ISCAS'88, IEEE, 503-506.
- Lure, Y.M.F., N.C. Grody, Y.S.P. Chiou, and H.Y.M. Yeh, 1992a: "Classification of Earth Surface from Special Sensor Microwave Imager (SSMI) Using Artificial Neural Network (ANN) Data Fusion," IEEE Proc. IGARSS '92, May 26-29, Houston, TX.
- Lure, Y.M.F., Y.S.P. Chiou, H.Y.M. Yeh, and N.C. Grody, 1992b: "Hardware-based Neural Network Data Fusion for Classification of Earth Surface Conditions", Invited Paper, 26th Annual Asilomer Conf. on Signal. System. and Computers, Oct. 26-29, 1992, CA, IEEE Compu. Society.
- Sethi, I.K., 1990: "Entropy Nets: From Decision Trees to Neural Networks." Proc. of the IEEE, Vol. 78, No. 10, October 1990.
- Tenney, R. R., and N. R. Sandell, 1981: "Detection with distributed sensors", IEEE Trans. Aeros. Elec. Sys., pp 501-509.
- Rumelhart, D. E., J. L. McClelland, 1986, Parallel Distributed Processing: Explorations in the Microstructure of Cognition, MIT Press, Cambridge, MA.