

Analog Circuits for Self-Organizing Neural Networks Based on Mutual Information

Janusz Starzyk and Liang Jing

School of Electrical Engineering and Computer Science
Ohio University
Athens, OH 45701 USA

Abstract – A new self-organizing neural network concept based on mutual information is described in this paper. Comparing to conventional neural network structures, this organization greatly reduces the interconnections in neural network by using local interconnection based on statistical analysis, and eliminates the need to store large number of synaptic weights. The network is characterized by evolvable hardware structure and adjustable threshold values based on selection criteria, which use mutual information. Secondly, A mix-signal implementation scheme is proposed for this organization in order to achieve the best performance. The digital implementation is used for the evolvable structure of the network for the better ability to be reconfigured. Analog implementation is used for the entropy-based evaluator (EBE), which is used for statistical analysis and mutual information evaluation, in order to achieve smaller area and faster on-chip learning process. Either on-chip analog memory or off-chip digital memory can be used to store the threshold values of the neurons and organization of resulting interconnection of neurons. Finally, circuits used for the analog implementation of the EBE are presented, the simulation results of the circuits are shown and discussed.

1 INTRODUCTION

An artificial neural network (ANN) is a massively parallel distributed processor made up of simple processing units, which has a natural ability for storing experimental knowledge and making it available for use.[1] In general, there are many different classes of network architectures. All of them require extensive interconnections between source nodes and neurons, and/or between neurons. The knowledge acquired through a learning process is stored as synaptic weights for every connection. Since the number of interconnections and synaptic weights increase quickly as the number of input nodes and neurons increases, it is difficult to build a neural network with large number of neurons, even with the currently most advanced VLSI fabrication technology.

To overcome this difficulty, a new reconfigurable artificial neural network organization using learning algorithm based on maximum mutual information principle (MMIP) and statistical analysis was developed for data classification application.[2] This organization also uses multi-layer feedforward network, but the experimental knowledge is stored in the form of the evolved circuit structure instead of the synaptic weights like in conventional

neural networks. Only the threshold values to activate the neurons are stored for every neuron, the output signals of the preceding layer of neurons are processed through logical circuits before being feedforwarded to the next layer of neurons. Furthermore, by statistical analysis performed during learning process, only these features important for classification are selected and grouped as subsets to feed into the neural networks and only those outputs of the most useful neurons are forwarded to the next layer. After decisions are made for every subset of features, these results are combined together to get a more accurate final decision. The final organization of the network is reconstructed and fully arrived to through the learning process. The final network structure looks more like a multilayer combinational circuit rather than conventional neural networks, the interconnections between neurons are greatly reduced and the memory needed is much smaller, especially for those networks with large number of input nodes and neurons. As a result, the design area of the circuit, power consumption and the processing time can be greatly reduced.

The critical part of this new organization is the learning process, since it decides the final network structure. In [2], the algorithm based on MMIP and statistical analysis was used and verified by simulating the learning process with VHDL code and mapping the final hardware into Xilinx FPGA board. The approach presented in this paper focuses on the reconfigurable, mixed-signal hardware implementation of the learning process, hardware self-organization and its verification. To fully implement the whole learning system into hardware, we need to find an efficient implementation of the core unit of learning process, the entropy-based evaluator (EBE), either digital or analog. Since this unit needs to perform statistical analysis and entropy calculation, it must include a wide range counter, a logarithmic function circuit and a multiplication circuit. It would be very costly to build it as a digital circuit, therefore if implemented in digital VLSI technology the EBE must be shared among many neurons to reduce the final cost. This in term requires extensive wiring and multiplexing of the EBE. In addition, smaller number of EBEs limits the parallel operations during learning process making it longer, particularly for large databases. The analog approach is naturally suited for these functions and can attain much higher processing speed than the digital approach.[3] In this paper building blocks of a mixed signal EBE are presented. Since one of the fundamental functions of EBE is finding the maximum of the mutual information, operations needed to compare this maximum are implemented using analog

circuits. An analog counter inspired by charge injector is used to estimate probabilities. It consists of only five CMOS transistors and two capacitors, the count range can be easily adjusted by modifying any of the following parameters: reference voltage, pulse width of the input signal, the size ratio of the two current mirror transistors, or the size of the storage capacitor. A wide range of linear response with respect to the number of pulses is achieved.

Since it's difficult to build an analog circuit with response similar to the entropy function, different approximation methods are simulated by Matlab code to approach the mutual information function. Although calculation resolution do present an issue for practical analog implementation, generally speaking, in calculation for learning, smoothness, not absolute precision, is important.[3] The simulation results shows that even with piecewise linear approximation, similar smoothness is still

achieved, and with quadratic approximation, the results are even better. Thus, it is possible to implement the mutual information function with either linear circuits or a quadratic circuit. The paper presents hardware organization of analog circuits, which are used as building blocks of EBE, and compares the simulation results with results obtained by means of digital circuits.

2 SYSTEM DESCRIPTION

The system is shown in Fig. 1. Every feature of the signal to be classified is represented as one dimension data; the data should be normalized (and A/D converted if analog signal) then fed into the network. Before the data is fed into next layer of neurons, one M by N MUX (N<M, typically, N=2, M=4) will randomly select M inputs, to make sure every

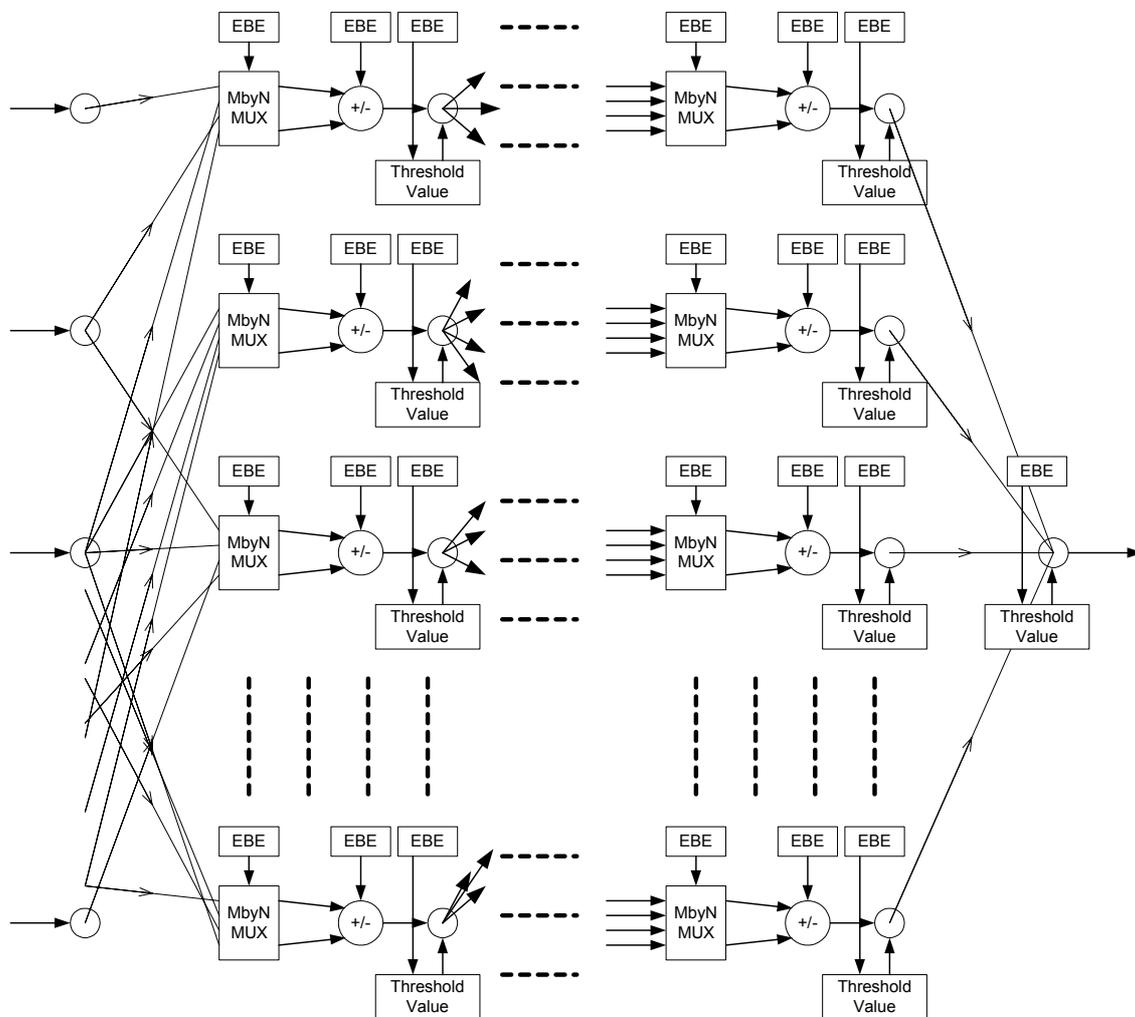


Fig. 1 System Description

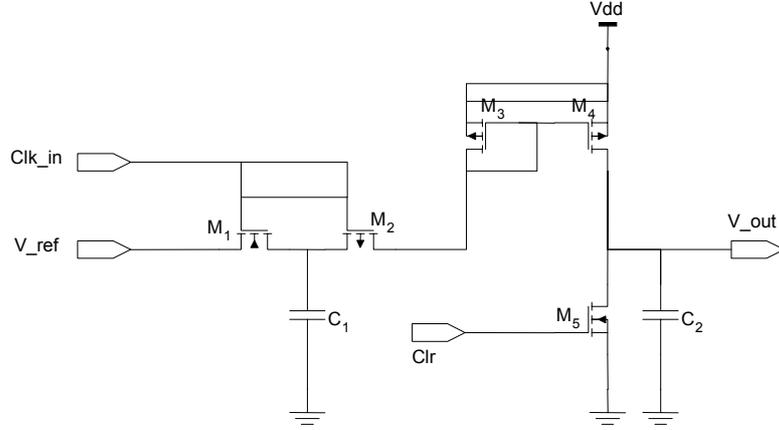


Fig. 2 The analog counter

feature has chance to be selected, from proceeding layers. Then, N inputs will be selected according to the result of EBE, to feed into logic circuits. Logic circuits will perform addition or subtraction decided also by EBE, thus a better feature is constructed and fed into one neuron. The neuron responds according to the threshold value set by EBE. Finally, the outputs of current layer of neurons are available to be selected by next layer of neurons, along with proceeding layers' outputs and original feature. The expansion of layers stops when no better feature can be generated, that is, there is no improvement of mutual information. Finally, the outputs of the last layer of neurons are fed into the output neuron, and the decision is made by this neuron.

3 ANALOG COUNTER

3.1 Circuit structure

The important issues in an analog counter design are the linearity, dynamic range, and strength against device dispersion; controllability of the step on a large dynamics, small area and power, high operation speed and small number of control lines are also required. To ensure these specifications, we come up with the analog counter circuit as shown in Fig. 2.

There are only five transistors and two capacitors used in this circuit. The capacitor C_2 is used as a memory to store the counter value in the form of analog charge, whereas the capacitor C_1 stores the increment charge for every incoming pulse. The transistors M_1 and M_2 make up of the increment circuit; the increment operation is performed whenever "Clk_in" is set to high. The current mirror circuit, which consists of transistors M_3 and M_4 , is used to eliminate the dependence of the injected charge quantity on the charge

stored in the memory capacitor C_2 . The transistor M_5 is used to discharge C_2 .

3.2 Analysis of the operation

By setting Clr to high and keeping it low thereafter, C_2 is discharged; the circuit output starts from "0". The voltage of C_1 is set to V_{ref} while the signal Clk_in is low. When the signal Clk_in rises to high, the voltage of capacitor C_1 is forced to $V_{dd}-V_{tp3}$ by the transistor M_3 and M_2 . The charging current is:

$$I_{C_1} = C_1 \frac{dV_{C_1}}{dt} \quad (3.1)$$

By the current mirror between M_3 and M_4 , the current flowing through M_4 is:

$$I_{M_4} = \frac{W_4/L_4}{W_3/L_3} I_{M_3} \quad (3.2)$$

Since $I_{M_3}=I_{C_1}$, the output current which charges C_2 is:

$$I_o = I_{M_4} = C_2 \frac{dV_{C_2}}{dt} \quad (3.3)$$

Integrating from 0 to infinity, we can get the voltage difference of the storage capacitor C_2 is:

$$\Delta V_{C_2} = \frac{W_4}{W_3} \frac{L_3}{L_4} \frac{C_1}{C_2} (V_{dd} - V_{ref} - V_{tp3}) \quad (3.4)$$

The equation (3.4) shows that the voltage difference of the storage capacitor depends linearly on the ratio of the transistors M_3 and M_4 , the ratio of the capacitors C_1 and C_2 , and the control voltage V_{ref} . Therefore, this analog counter achieves a small voltage difference through small parameter ratios with small area components. Moreover, this rational

design is robust with respect to the absolute dispersion of the devices.

3.3 Simulation

The circuit is simulated in AccuSim of Mentor Graphics. The CMOS FET SPICE models are provided by MOSIS website; they result from fabrication runs of 0.5 micron AMI C5N technology.

NMOS transistors with source, drain and base all connected to ground are used as capacitance. The transistor M_3 is set larger than the transistor M_4 , in order for the blocking of M_3 to result in the blocking of M_4 . Otherwise, the circuit will have a too large voltage corruption over C_2 to allow a long storage time.

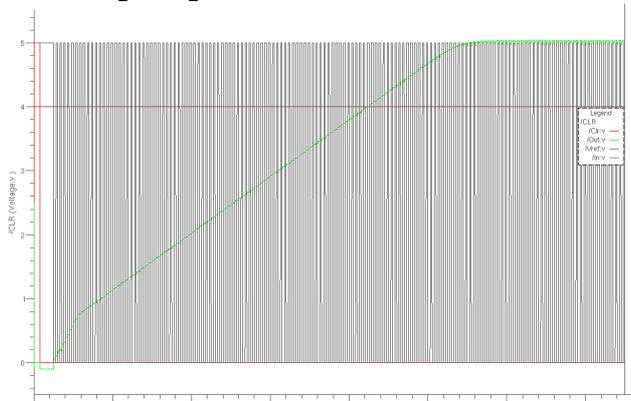


Fig. 3. Simulation result of analog counter's linearity and dynamic range.

The simulation result with the working frequency set to 10MHz is presented in Fig.3. It shows a highly linear characteristic over the dynamic range from 0.8V to 4.8V (in the simulation, the threshold voltage of NMOS transistors, which are used as capacitance, is about 0.8V).

4 MUTUAL INFORMATION PRINCIPLE

4.1 Mutual Information

The entropy based mutual information is defined as follows:

$$I = 1 - \frac{\Delta E}{E_{\max}} \quad (4.1)$$

where

$$\Delta E = -\sum_{a=0}^1 \sum_{c=1}^{n_c} p_{ac} \log(p_{ac}) + \sum_{a=0}^1 p_a \log(p_a) \quad (4.2)$$

and

$$E_{\max} = -\sum_{c=1}^{n_c} p_c \log(p_c) \quad (4.3)$$

a is the logic function of the signal column and is equal to 1 if the value exceeds the threshold and 0 otherwise, n_c is the number of classes in the training set, p_c , p_a , p_{ac} are probabilities of each class, attribute probability and joint probabilities respectively.

4.2 Approximations of Mutual Information Function

According to function (4.1~3), the entropy function $p \log p$ is the building block of mutual information calculation. The property of this function makes it very costly to implement in digital circuits and very hard to implement in analog circuits.

Since any classification problem can be decomposed into a number of classification problems of two classes, and for $n_c = 2$, (4.2) can be expressed as following:

$$\Delta E = -\sum_{a=0}^1 \frac{P_{a,c=1}}{P_a} \log\left(\frac{P_{a,c=1}}{P_a}\right) + \left(1 - \frac{P_{a,c=1}}{P_a}\right) \log\left(1 - \frac{P_{a,c=1}}{P_a}\right) \quad (4.4)$$

Similarly, (4.3) can be expressed as:

$$E_{\max} = -P_{c=0} \log P_{c=0} - (1 - P_{c=0}) \log(1 - P_{c=0}) \quad (4.5)$$

Thus, the following mutual information function of two classes is used as building block of mutual information calculation instead:

$$I(p) = -p \log p - (1 - p) \log(1 - p) \quad (4.6)$$

As shown in Fig. 4, this function is symmetric to the line $p = 0.5$ and has a maximum value of $\log 2$. Thus, different approximations can be used in order to implement it in analog circuits. Two different approximation are used here. One is the linear approximation:

$$I(p) = \log 2(1 - 2|p - 0.5|) \quad (4.7)$$

The other one is quadratic approximation:

$$I(p) = 4(\log 2)p(1 - p) \quad (4.8)$$

Their responses with respect to class probability p and its compare with original function is shown in Fig. 4.

4.3 Test Data Preparation

In order to compare the final decisions made by using the original mutual information function and its approximations, we randomly generated test data with normal distributions.

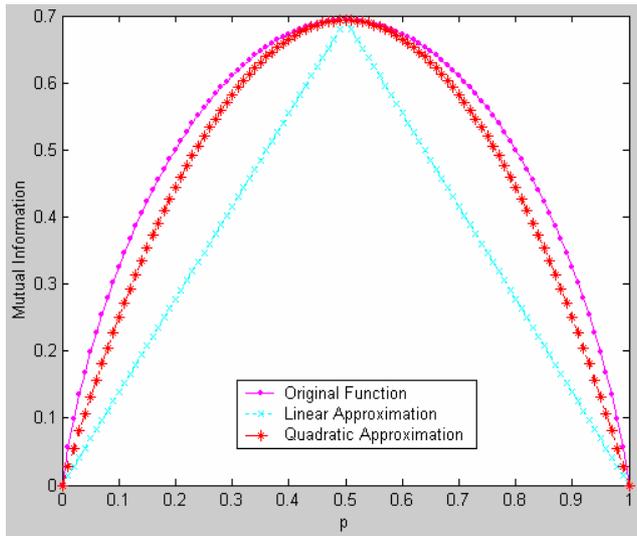


Fig. 4 Mutual information building function and its approximations

The mean value and covariance matrix are different for each class data set. To better shown the distribution of data, these data generated with two dimensions, although we only use one of the dimensions. Each class can have different ellipse shape with major axis in different directions.

4.4 Simulation Results

The simulation results are shown in Fig. 4(a), Fig. 4(b) and Fig. 4(c). They represent three situations: separated, overlapped, and messed up respectively. In all three situations, there are some deviations of mutual information index value between the original function and its approximations, but the shapes are similar. Furthermore, the location of maximum value, which is the most important information, are almost the same, especially for the completely separated situation.

As expected, the quadratic approximation is much better, but even for the linear approximation, the result is still acceptable.

5 CONCLUSION

We presented a new self-organizing neural network concept based on mutual information. This network is to be implemented in analog circuits in order to get better performance. To implement the most important circuit in this network, the EBE, an analog counter is presented as a building block to perform statistical calculation and simulated. Satisfying results are achieved.

Furthermore, different approximation approaches are used to implement the mutual information calculation; the simulation result shows that it is possible to implement it by linear circuits or quadratic circuits.

References

- [1] Simon Haykin, "Neural Networks – a comprehensive foundation", 2nd edition, Prentice Hall, New Jersey, pp. 1-44, 1999.
- [2] Janusz A. Starzyk, Jing Pang, "Evolvable binary artificial neural network for data classification.", The 2000 Int. Conf. on Parallel and Distributed Processing Techniques and Applications, (Las Vegas, NV, June 2000).
- [3] Takashi Morie, Kuniharu Uchimura, Yoshihito Amemiya, "Analog LSI implementation of self-learning neural networks", Computers and Electrical Engineering, 25 (1999), pp.339-355.

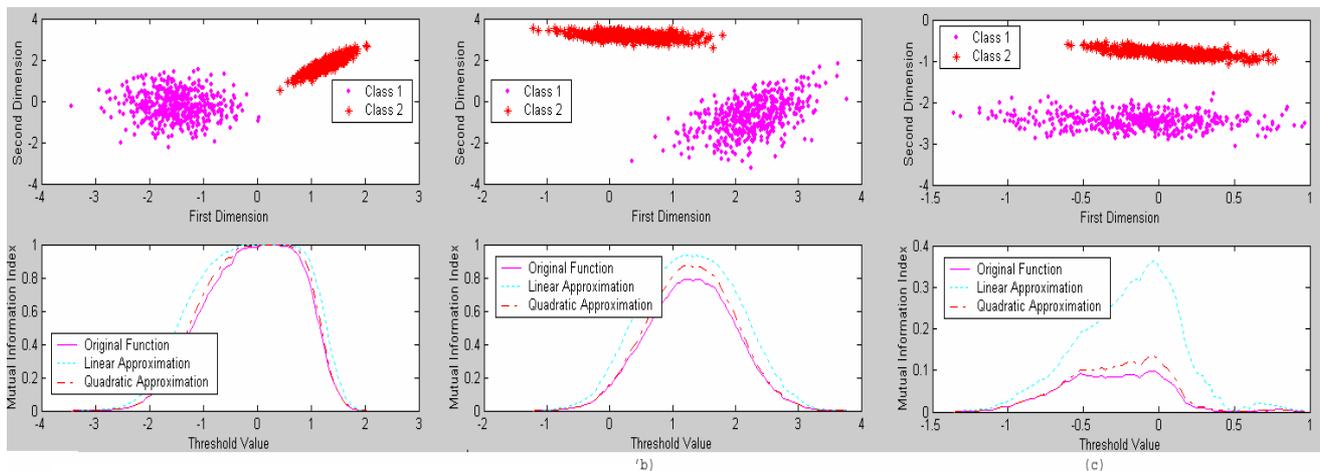


Fig. 5 Simulation results of different situations of two classes classification