Hardware Implementation of ART1 Memories Using A Mixed Analog/Digital Approach

C. S. Ho, J. J. Liou, and M. Georgiopoulos
Department of Electrical and Computer Engineering
University of Central Florida, Orlando, FL 32816

Abstract

This paper presents a VLSI circuit implementation for both the short-term memory (STM) and long-term memory (LTM) of the adaptive resonance theory neural network (ART1-NN). The circuit is implemented based on the transconductance-mode approach and mixed analog/digital components, in which analog circuits are used to fully incorporate the parallel mechanism of the neural network, whereas digital circuits provide a reduced circuit size as well as a more precise multiplication operation. A simple analog-to-digital (A/D) converter is also included to realize binary STM activities and characterize the quenching threshold. The PSpice simulation results of the implemented circuits are in good agreement with the exact solutions of the coupled nonlinear differential equations.

1. Introduction

The adaptive resonance theory neural network (ART1-NN), introduced by G. Carpenter and S. Grossberg [1,2], is a self-contained neural network which has the ability to classify an arbitrary set of binary input patterns into different clusters very quickly and stably. This network can be used in a variety of applications including group technology and pattern recognition.

In this paper, a mixed analog/digital VLSI implementation for a generalized ART1 system including both the LTM and STM is presented based on the transconductance-mode (T-mode) approach [3]. The transconductance multipliers [4] are utilized to construct the summation of synaptic weights. Furthermore, in order to reduce the circuit size and thus the chip area, 2-input digital multiplexers are used to perform certain multiplication operations in which the output signal can be selected as one of the two input signals. A simple analog-to-digital (A/D) converter is also included to realize binary STM activities and characterize the quenching threshold [1]. In addition, we have designed a generalized circuit which is suitable for both the LTM and STM implementations with compatible input/output voltage ranges.

2. STM and LTM Activities of ART1 Neural Network

The activities of a node v_i (or neuron v_i) in the F_1 (bottom) field and a node v_j (or neuron v_j) in the F_2 (top) field are described by the following nonlinear first-order differential equations:

$$\epsilon_1 \frac{dx_i}{dt} = -x_i + (1-A_1x_i)J_i^+ - (B_1+C_1x_i)J_i^- , \quad \epsilon_2 \frac{dx_j}{dt} = -x_j + (1-A_2x_j)J_j^+ - (B_2+C_2x_j)J_j^- \tag{1}$$

where x_i and x_j are the activities of nodes v_i and v_j, ϵ_1 and ϵ_2 are the learning rates of these nodes, J_i^+ and J_i^- represent the total excitatory and inhibitory inputs to the node v_i, respectively, J_j^+ and J_j^- represent the total excitatory and inhibitory inputs to the node v_j, respectively, and A_1, A_2, B_1, B_2, C_1, and C_2 are positive constants. In particular, J_i^+, J_i^-, J_j^+, and J_j^- are given by the following equations:

$$J_i^+ = I_i + D_1 \sum_{j=M+1}^{N} f_2(x_j)z_{ji} , \quad J_i^- = \sum_{j=M+1}^{N} f_2(x_j) \tag{2}$$

$$J_j^+ = f_2(x_j) + D_2 \sum_{i=1}^{M} f_1(x_i)z_{ij} , \quad J_j^- = \sum_{k=M+1,k\neq j}^{N} f_2(x_k) \tag{3}$$

In the above quations, D_1 and D_2 are positive constants, I_i is the component of the input pattern I (binary array) that is received by node v_i, and $f_1(x_i)$ and $f_2(x_j)$ are the output activities denervated by node v_i with activity x_i and node v_j with activity x_j, respectively. Here, $f_1(x_i)$ and $f_2(x_j)$ are characterized by the following equations:

0-7803-1901-X/94 $4.00 ©1994 IEEE

$$f_1(x_i) = \begin{cases} 1, & \text{if } x_i > \delta_1 \\ 0, & \text{otherwise} \end{cases}, \quad f_2(x_j) = \begin{cases} 1, & \text{if } x_j > \delta_2 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where δ_1 and δ_2 are the quenching thresholds of every node v_i in the F_1 field and every node v_j in the F_2 field, respectively.

The values of the bottom-up LTM trace, z_{ij}, and top-down LTM trace, z_{ji}, can be determined from the following equations:

$$\epsilon_z \frac{dz_{ij}}{dt} = -[(L-1)f_1(x_i) + \sum_{k=1}^{M} f_1(x_k)]\, f_2(x_j) z_{ij} + L f_1(x_i) f_2(x_j) \tag{5}$$

$$\epsilon_z \frac{dz_{ji}}{dt} = -f_2(x_j) Z_{ji} + f_1(x_i) f_2(x_j) \tag{6}$$

Here ε_z is the learning rate of the LTM traces and L is a positive constant larger than unity.

3. Circuit Implementation Of Short-Term Memory

Figure 1 shows a circuit implemented for (1) based on the transconductance-mode approach, in which the time-dependent voltage across the capacitor C represents the STM activity. The synaptic elements are built using the wide-range transconductance multipliers (see Fig. 2(a)) and 2-input multiplexers (see Fig. 2(b)). In addition, an analog-to-digital (A/D) converter is employed (see Fig. 2(c)) to characterize the quenching threshold and generate a digital output signal (either 0 or 5 V) to be used as the input of the multiplexer (see Fig. 2(b)).

3.1 The Wide-Range Transconductance Multiplier

The tranconductance multiplier is shown in Fig. 2(a). To prevent the current-voltage characteristics of the differential amplifiers from being saturated, $V_{I1} << \sqrt{4I_1/\sqrt{\beta_P}}$ and $V_{I2} << \sqrt{I_b/\sqrt{\beta_N}}$ (I_b is shown in Fig. 2(a)) are needed, where $V_{I1} = V_3 - V_4$, $V_{I2} = V_1 - V_2$, and β_N and β_P are the device parameters of the NMOS and PMOS differential pairs, respectively. Within these voltage limits, I_{out} can be expressed as

$$I_{out} \approx \sqrt{\frac{\beta_N \beta_P}{2}} V_{I1} V_{I2} = a\, V_{I1} V_{I2} \tag{7}$$

Note that a can be changed by scaling the W/L ratios of the transistors MP7, MP8, MP9, MP10, MN3, and MN4, which yields different I_{out}, to fit the positive constants required in (1).

3.2 The 2-input Multiplexer

To simplify the ART-1 circuit, the multiplexer (see Fig. 2(b)) is used to perform certain multiplication functions in which the output response can be selected as one of the two input signals. In addition to the smaller circuit size, this design can yield a more accurate response than an analog multiplier. In Fig. 2(a), the control signal $F_1(x_i)$ is a digital signal (either 0 or 5 V) which is generated by the A/D converter of the STM in F_1 fields (to be discussed in the next section). Here, in order to perform the multiplication function using the multiplexer, we define a digital function for a neuron v_i in the F_1 field such that

$$F_1(x_i) = \begin{cases} 5V, & \text{if } f_1(x_i)=1 \quad (x_i > \delta_1) \\ 0V, & \text{if } f_1(x_i)=0 \quad (x_i \le \delta_1) \end{cases} \tag{8}$$

Based on this property, an output voltage of $V_c f_1(x_i)$ (see Fig. 2(b)) can be generated by this multiplexer. Similarly, we also define $F_2(x_j) = 5$ V if $f_2(x_j) = 1$ and $F_2(x_j) = 0$ V if $f_2(x_j) = 0$ for the neuron v_j in the F_2 field, and $F_1(I_i) = 5$ V if the input pattern $I_i = 1$ and $F_1(I_i) = 0$ V if $I_i = 0$.

3.3 The Analog-to-Digital Converter

The A/D converter shown in Fig. 2(c) comprises of three circuits: an operational transconductance amplifier (OTA) and two CMOS inverters. Utilizing the OTA circuit, x_i is increased to a higher voltage level V_O, thus allowing the CMOS inverter with threshold voltage V_{th} to generate a digital output signal $F_1(x_i)$ (5 V or 0).

In the OTA circuit, the bias voltage V_{bi} (shown in Fig. 2(c)) is selected such that transistors M_1 and M_2 are operated in the saturation region. If the gate voltage V_s of M3 is chosen as $(V_{bi} + V_{TN})/2$, then the current $I_O = \beta (V_{bi} - V_{TN}) (x_i - V_s)$, where x_i is the output voltage (activity) generated by the STM circuit and V_{TN} is the threshold voltage for NMOS devices in the OTA circuit. And then, we get $V_O = V_{ref} + RI_O$.

Next, according to the characteristics of a CMOS inverter, its threshold voltage V_{th} can be determined by

$$V_{th} = \frac{V_{TN} + K(V_{DD} + V_{TP})}{1 + K}, \quad Here \quad K = \sqrt{\frac{\beta_p}{\beta_n}} \tag{9}$$

where β_n and β_p are the device parameters for NMOS and PMOS transistors, and V_{TN} and V_{TP} are the device threshold voltages for NMOS and PMOS transistors. Thus, once V_{th} is determined, the output voltage $F_1(x_i)$ can be characterized by a digital signal such that

$$F_1(x_i) = \begin{cases} 5V, & if\ V_O > V_{th} \\ 0V, & otherwise \end{cases} \tag{10}$$

In the circuit, we choose $V_{th} = V_{ref} + \delta_1$. Therefore, when $F_1(x_i) = 5$ V, neuron v_i is activated ($x_i > \delta_1$). This result is consistent with $F_1(x_i)$ defined in (8).

A PSpice simulation is carried out for a neuron v_1 in the F_1 field connected with two neurons in the F_2 field (nodes v_3 and v_4). In the simulation, parameters $\delta_1 = 0.01$, $\varepsilon_1 = 1$, $A_1 = 2$, $B_1 = 10$, and $C_1 = 10$ are used. The simulation results are shown in Fig 3. The line with empty squares represents the response of the implemented STM circuit, and the line with solid squares is the exact solution calculated numerically from the coupled differential equations. Excellent agreement is found between the two results.

4. Circuit Implementation Of Long-Term Memory

The behavior of the bottom-up LTM trace described in (5) can be re-written as

$$\varepsilon_z \frac{dz_{ij}}{dt} = -Lf_1(x_i)f_2(x_j)z_{ij} + Lf_1(x_i)f_2(x_j) - [\sum_{k=1,k\neq i}^{M} f_1(x_k)]f_2(x_j)z_{ij} \tag{11}$$

It can be seen from (11) that the bottom-up LTM trace z_{ij} can change its value only when node v_j is activated ($f_2(x_j) = 1$). Also z_{ij} approaches a positive value when $f_1(x_i) = 1$ and decays to zero when $f_1(x_i) = 0$. The bottom-up LTM can be implemented using the T-mode approach discussed above, provided that the range of the LTM trace is compatible with that of the STM (0 to 1 V). Fig. 4 shows a general-purpose bottom-up LTM circuit. The first two items on the right-hand side of (11) are implemented in Block A. The sigmoidal item in (11) is implemented in Block B. Notice that k is not equal to i in this bottom-up LTM implementation.

PSpice simulation is also carried out for a bottom-up LTM z_{15} having one node (node v_5) in the F_2 field and four nodes (nodes v_1, v_2, v_3, v_4) in the F_1 field. Parameter values $\varepsilon_z = 10$, $L = 1.01$ are chosen in the simulation. The simulation results are shown in Fig. 5. The simulation results (empty squares) agree well with the exact solutions (solid squares) calculated directly from (5), as shown in Fig. 5.

Since the top-down LTM trace z_{ji} (described in (6)) has the same form as (11) but without the sigmoidal item, z_{ji} can be implemented utilizing the circuit shown in Block A of Fig. 4, provided that $L = 1$ is used and one of the two $F_1(x_i)$ is maintained at 5 V, as required in (6).

5. Conclusion

In this paper, a generalized ART1 system including both the long- and short-term memories has been implemented using the T-mode approach and with both digital and analog VLSI components. The analog circuit is used to fully incorporate the parallel mechanism of the ART1-NN, whereas the digital circuit provides a smaller circuit size as well as a more precise multiplication operation. An analog-to-digital converter is also implemented

in the circuit to couple the analog and digital signals. Results from the PSpice simulation show that the implemented circuits perform the same functionalities as that expected from the ART1-NN algorithm. The circuit implemented is general for all hardware utilizing ART1 memories and thus can be readily used as a building block to construct a large-scale neural network.

Acknowledgement--This work was supported in part by the Florida High Technology Council and Harris Semiconductor, Melbourne, FL. The donation of PSpice circuit simulator by MicroSim Corp., Irvine, CA is gratefully acknowledged.

References

[1] G. Carpenter and S. Grossberg, "A massively parallel architecture for self-organizing neural pattern recognition machine," Computer Vision, Graphics, and Image Process., vol. 37, pp. 54-115, 1987.

[2] S. Grossberg, "Nonlinear neural networks: Principles, mechanisms, and architecture," Neural Networks, vol. 1, pp. 17-61, 1988.

[3] B. Linares-Barranco, E. Sánchez-Sinencio, A. Rodríguez-Vázquez, and J. L. Huertas, "A modular T-mode design approach for analog neural network hardware implementations," IEEE J. Solid-State Cir., pp. 701-713, May 1992.

[4] C. A. Mead, *Analog VLSI and Neural Systems*, Chapter 6, Addison Wesley Publishing Co., 1989.

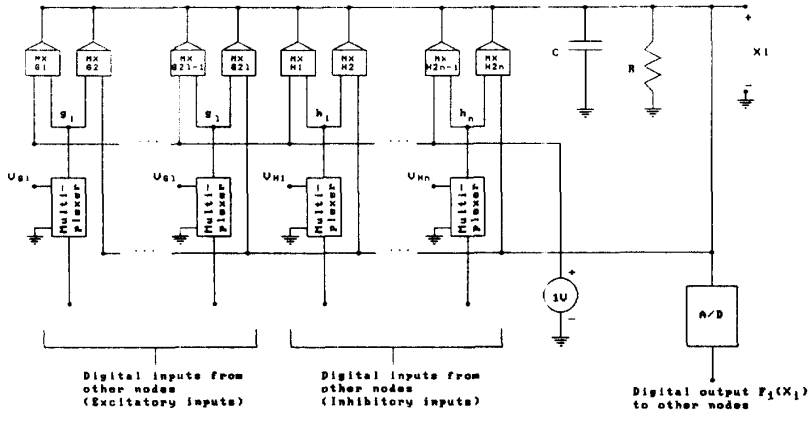
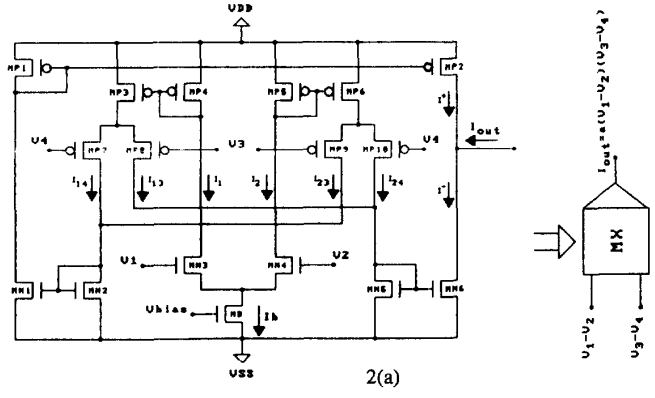


Fig. 1: The circuit block diagram for the STM activity of the ART1-NN.



2(a)

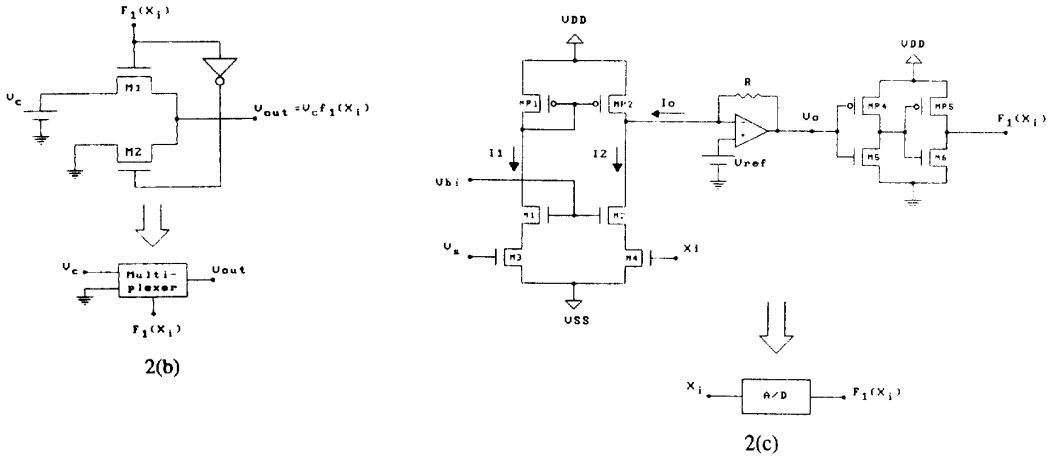


Fig. 2: Circuits implemented for (a) the wide-range transconductance multiplier, (b) the 2-input digital multiplexer, (c) the A/D converter.

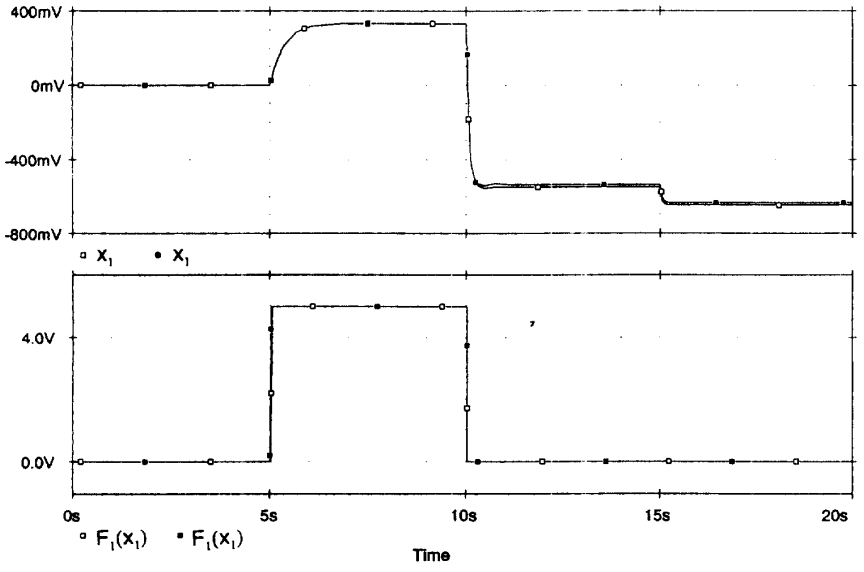


Fig. 3: Activities of the STM simulated from PSpice (empty squares) and calculated from the coupled differential equations (solid squares).

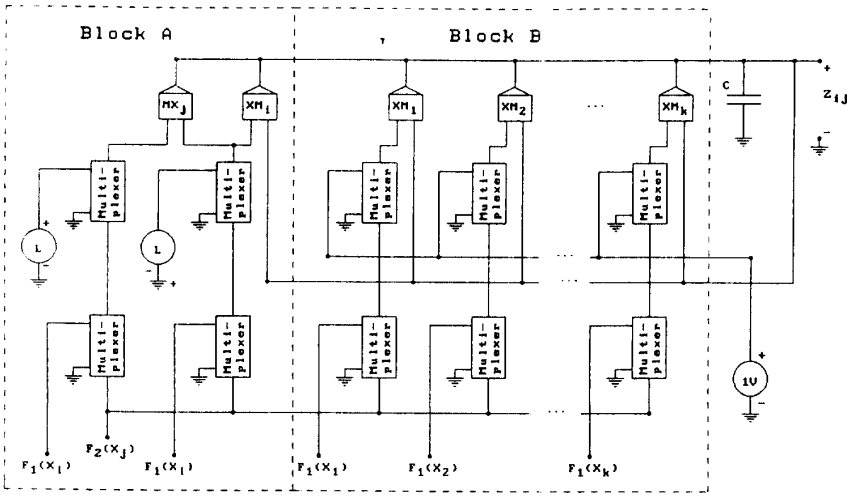


Fig. 4: Circuit block diagram for the bottom-up LTM trace in the ART1-NN.

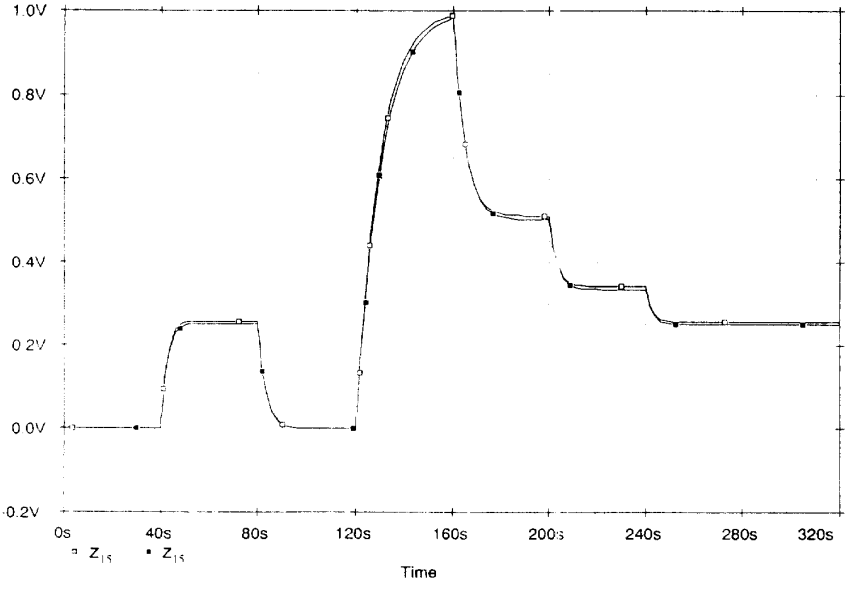


Fig. 5: Activities of a bottom-up LTM with one neuron v_5 in the F_2 field and four neurons v_1, v_2, v_3, and v_4 in the F_1 field; the line with empty squares is the simulation result from Pspice and the line with solid squares is the exact solution of the differential equations.