Hardware Implementation of An Adaptive Resonance Theory (ART) Neural Network Using Compensated Operational Amplifiers

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

ABSTRACT

This paper presents an analog circuit design and implementation for an adaptive resonance theory neural network architecture called the augmented ART1 neural network (AART1-NN). Practical monolithic operational amplifiers (Op-Amps) LM741 and LM318 are selected to implement the circuit, and a simple compensation scheme is developed to adjust the Op-Amp electrical characteristics to meet the design requirement. A 7-node prototype circuit has been designed and verified using the Pspice circuit simulator run on a Sun workstation. Results simulated from the AART1-NN circuit using the LM741, LM318, and ideal Op-Amps are presented and compared.

1. INTRODUCTION

Neural networks are basically a means of correlating data using many simple processing elements. These elements work together to form a transfer function that maps input data to output data. Neural networks "learn" by reading known input/output samples and adapting themselves to map one to the other. After enough "training", the network is ready to respond to new data by producing a best estimated output. For example, neural networks have exhibited the capability of recovering signals embedded in a noisy environment and to adjust quickly to changing signal/noise scenaria. This ability has led to many applications in signal processing and feedback control. One of popular neural network models are the ones based on the the adaptive resonance theory developed by Grossberg[1]. Among them, the adaptive resonance theory-1 neural network (ART1-NN) was the first complete network to be introduced by Carpenter and Grossberg[2]. The ART1-NN has the ability to classify an arbitrary set of binary input patterns into different clusters. Recently, a modification of the ART1-NN was designed by Heileman et al [3]. The major difference between the AART1-NN in [3] and the ART1-NN is that the AART1-NN is completely described by a set of nonlinear differential equations, while the ART1-NN incorporates algorithmic components in its description.

In this paper an analog, prototype AART1-NN circuit is designed and implemented using practical operational amplifiers (Op-Amps) with compensated circuits. Operational amplifiers are relatively inexpensive and have been used in various circuits [4], including neural networks [5]. Ideal Op-Amps were used in [6] to demonstrate the validity of the approach without extensive circuit design and simulation. This paper follows the same approach proposed in [6], but two practical Op-Amps will be considered in designing the AART1-NN. A simple compensation scheme is also implemented to offset the non-ideal Op-Amp characteristics. Simulation results obtained from Pspice circuit simulator run on a Sun workstation are also presented.

It should be mentioned that our design approach differs from the VLSI approach in which CMOS technology is normally utilized and all electronic components are fabricated on a silicon wafer using advanced processing technology. Our "off-the-shelf" approach utilizes the off-shelf components like Op-Amps, resistors, and capacitors, thus allowing one to build a prototype neural network in a flexible manner and without requiring advanced equipments and technology. The drawback of our approach is of course the much larger circuit size.

2. BASIC STRUCTURE OF THE AART1-NN MODEL

0-8194-1547-2/94/$6.00

A schematic illustration of the AART1-NN is shown in Figure 1. It consists of two subsystems--the attentional subsystem and the orienting subsystem. The attentional subsystem consists of two fields of nodes designated as F_1 and F_2 fields. The F_1 field is often referred to as the input field because the input patterns are presented to it. The nodes in the F_1 and F_2 fields are used to encode patterns of short term memory (STM) activity. Each node in the F_1 field and in the first layer of F_2 field is connected via a bottom-up weighted connection, called the long term memory (LTM) trace. The pathway from F_1 field to F_2 field is called the bottom-up LTM trace (denoted as z_{ij}) and, likewise, the pathway from F_2 field to F_1 field is called top-down LTM trace (denoted z_{ji}). The bottom-up and top-down LTM traces are adaptive, and they are used to store the knowledge acquired by the neural network as the training progresses. The first layer of the F_2 field is often referred to as the category representation layer, because it is the layer that indicates the category to which the input pattern belongs to. The objective of the second layer of nodes in the F_2 field is to deactivate the erroneous category representation in the first layer of the F_2 field, whenever such an erroneous representation occurs, and to keep this erroneous category deactivated for as long as the same input pattern persists at the F_1 field. Furthermore, the orienting subsystem in the AART1-NN architecture consists of a single node designated as v_r. The primary purpose of the orienting subsystem is to generate a reset wave to the F_2 field, whenever the category representation in the first layer of the F_2 field is not a good match for the input pattern presented at the F_1 field.

We denote nodes in the F_1 field by v_i (i=1,2,...,M), nodes in the first layer of the F_2 field by v_j (j=M+1,M+2,...,N), and nodes in the second layer of the F_2 field by v_j (j=M+1,M+2,...,N). The STM activities of the nodes v_i, v_j, v_j, and v_r are denoted by x_i, x_j, x_j and x_r, respectively. In addition, as we mentioned before, the bottom-up LTM traces are designated as z_{ij} and the top-down LTM traces are designated as z_{ji}. The detailed differential equations governing the AART1-NN have been reported in [6].

3. OP-AMP COMPENSATION AND CIRCUIT IMPLEMENTATION

Two practical Op-Amps (LM741 and LM318) are considered for the AART1-NN implementation. The popularity of these Op-Amps arise from the fact that they have large current gain, have good common-mode and differential-mode input voltage range, and have relatively simple circuits that can be fit on a die less than 40 mils on a side [7]. The LM741 is an improved industry standard, medium-performance, general-purpose Op-Amp. The LM318, on the other hand, is a precision, high-speed Op-Amp. Compared to the LM741 Op-Amp, the LM318 Op-Amp offers a higher slew rate as well as a factor of ten increase in speed without sacrificing dc performance [8]. Both Op-Amps are pin compatible.

3.1 Op-Amp Compensation

An ideal Op-Amp offers the following characteristics: high input resistance, low input bias current, small input offset voltage and current, high common-mode rejection ratio, high common-mode input range, high voltage gain, large output voltage swing, and high slew rate. The behavior of a practical Op-Amp, however, deviates considerably from its ideal counterpart. For the circuit under study, which is used in transient operations, only the deviations in the input bias current, input offset voltage, and slew rate are of concern.

The slew rate is the maximum rate of change of the output voltage for a step change at the input. When the input is driven by a sudden changing input, the output is forced to ramp, or slew, at some limited rate determined by the internal currents and capacitances. Using an Op-Amp with a high slew rate can of course avoid such a limitation.

Next, we try to minimize the output voltage change caused by the input bias currents. There are two bias currents, I_B^+ and I_B^-, shown in Figure 2(a). Normally, the bias currents are at the same order of magnitude but are not exactly equal. If an output voltage change is generated due to the input bias currents, we can place a

compensation resistor (R_C) between the noninverting input and ground to offset the change. In the figure, the bias current I_B^+ flowing through R_C would then produce a negative voltage on the noninverting input, which, when multiplied by the noninverting gain, would cancel the output voltage shift due to I_B^- on the inverting input. According to this concept, R_C is related to other circuit components (Fig. 2(a)) as follows

$$I_B^- R_F = I_B^+ R_C[(R_F + R_I)/R_I] \quad (1)$$

If I_B^+ and I_B^- are presumed to be equal, then the compensation resistor R_C can be obtained by $R_C = R_F R_I/(R_F + R_I)$, which is the value of the parallel combination of R_I and R_F.

Furthermore, the undesired output-voltage shift which results from the input offset voltage (current) should also be compensated. The simplest method for such compensation is the attachment of a potentiometer P_{ot} on the two "offset null" terminals of the Op-Amp, as shown in Fig. 2(b). In the figure, the value of R_O is selected to be 1000 times of R_I. From the inverting circuit, we can derive the output voltage

$$V_O = -(R_F/R_I)V_{IN} - (R_F/R_O)V_N \quad (2)$$

where V_N is the 'null' voltage. The first item on the right hand side of (2) is the uncompensated output voltage and the second item is the voltage introduced to compensate the voltage shift.

3.2 Circuit Implementation

Circuits have been designed: node v_j (Fig. 3), node v_i (Fig. 4), node v_j (Fig. 5), the reset node v_r (Fig. 6), and the bottom-up and top-down LTM traces (Figs. 7(a) and 7(b)) [6]. In Figure 3, Blocks A, B, C, and D denote the inverting summer cell, integrator cell, multiplier cell, and comparator cell. Here the Op-Amps are not compensated.

Before proceeding to circuit simulation, we need to compensate the LM741 and LM318 Op-Amps based on the approaches discussed in the previous section. Fig. 8 shows the Pspice circuit simulation results for the activity of an arbitrary node v_j in the first layer of the F_2 field using the ideal, uncompensated LM318, and uncompensated LM741 Op-Amps. From results presented in Fig. 8, both the LM318 and LM741 give rise to a notable voltage shift compared to the ideal Op-Amp. After the LM741 and LM318 are compensated (with voltage compensations of 0.032 mV and 0.08 mV, respectively), such voltage shifts are eliminated, and an excellent agreement is found among the three Op-Amps. We also conclude that these shifts are caused mainly by the input offset voltage and input bias current, not the slew rate of the Op-Amp, since the LM318 yields a larger voltage shift than the LM741 (see Fig. 8). Pspice simulations given in the next section will strengthen this assessment.

4. CIRCUIT SIMULATION

In this section we carry out Pspice circuit simulation for the AART1-NN in a Sun SPARC-2 workstation. The prototype circuit consists of 7 nodes; two nodes in the F_1 field (nodes v_1 and v_2), a reset node v_r, and four nodes in the F_2 field (nodes v_3 and v_4 in the first layer and nodes v_3 and v_4 in the second layer). The various choices of the parameter values for the differential equations were defined in [3], and the values chosen for our simulation are depicted in Table 1.

Pspice simulation results for the transient response of the prototype circuit are shown in Figures 9, 10 and 11, and the trends agree closely with those discussed qualitatively above. Binary pattern I^1 (10) is presented in the interval [0, 700] (seconds), binary pattern I^2 (00) is presented in the interval [700, 850] (seconds), and binary pattern I^3 (11) is presented in the interval [850, 1550] (seconds).

For the first pattern, all the ideal, compensated LM318, and compensated LM714 Op-Amps are used, and the two practical Op-Amps yield similar results. In Figure 9(b), the results demonstrate that node v_3 is the one chosen to represent the input pattern I^1. Also, the activity of v_4, which is the node receiving the next largest input, never manages to reach a level above the node threshold δ_2. After the activation of node v_3, node v_1 in the F_1 field receives both bottom-up input (from the input pattern I^1) and top-down input (from node v_3). This causes the activity of v_1 to decrease and eventually reach a limiting value (0.125), which is above the node threshold δ_1 (see Figure 9(a)). Note that the activity of field F_1 throughout I^1's presentation is large enough so that the reset node cannot generate a reset wave. Consequently, node v_3 is deemed by the architecture as the right node in the F_2 field to represent the input pattern I^1, and the bottom-up and top-down LTM traces corresponding to node v_3 will learn pattern I^1. In particular, the LTM traces z_{31} and z_{13} converge to large values ($z_{31} \rightarrow 1$, $z_{13} \rightarrow 1$) while the remaining traces corresponding to node v_3 converge to small values (i.e., values close to zero). Since the LM714 and LM318 Op-Amps are shown to behave similarly during the presentation of the first pattern, only the LM318 Op-Amp will be used in the simulation for the next two patterns.

In Figure 10, we illustrate what happens during the presentation of the zero pattern at the F_1 field of the AART1-NN. As is shown in Figure 10, all STM node activities converge to their resting value of zero within the first 120 units of time after the appearance of I^2. It is worth pointing out that all the LTM traces remain at their values achieved at the end of the I^1 pattern except the top-down LTM trace z_{31} during the presentation of pattern I^2. During that time, the value of z_{31} decreases to 0.41 because of the smaller learn rates (ε_z of Table 1) chosen for the bottom-up and top-down traces. Normally, ε_z is chosen much larger than ε_1 and ε_2 (eg. $\varepsilon_z > 100\varepsilon_1$, $\varepsilon_z > 100\varepsilon_2$). The reason we chose ε_z=100 is because we wanted to speed up the simulation time.

The behavior of the network after the presentation of I^3 is shown in Figures 11(a) and 11(b). In Figure 11(a), we can see that after the presentation of I^3 node v_3 becomes active first because it is node in the first layer of the F_2 field that receives the largest bottom-up input. Once node v_3 becomes active the activities of nodes v_1 and v_2 begin to decrease. The activity of v_1 remains above the threshold (δ_1=0.01 V), while the activity of v_2 decreases to a level below the threshold. This a consequence of the fact that node v_1 receives strong top-down input, while node v_2 receives weak top-down input both nodes receive bottom-up input. When v_2 becomes non-active, the activity of the reset node starts increasing due to the mismatch between the bottom-up and top-down inputs that is now occurring at the F_1 field. When v_r becomes active (i.e., its activity exceeds the threshold δ_r), it generates a reset wave that deactivates v_3 almost instantaneously (see Figure 11(a)). After v_3 is deactivated, v_1 and v_2 receive only bottom-up input, and their activities increase towards the limiting value of 0.5 (see Figure 11(a)). Now that v_3 is deactivated, v_4 becomes active since it is the node in the first layer of the F_2 field that receives the next largest bottom-up input from the F_1 field. The activation of v_4 is shown in Figures 11(a) and 11(b). When v_4 becomes active the activities of v_1 and v_2 begin to decrease from the value of 0.5; but they both remain above the threshold (see Figure 11(b)). This is a consequence of both v_1 and v_2 receiving bottom-up input and strong top-down input.

It is worth mentioning that the circuit of AART1-NN presented in this paper can be implemented with other Op-Amps having similar characteristics as that of the LM741 or LM318 (eg., LM118, LM108, and LF411 Op-Amps), provided the Op-Amps are properly compensated. The CPU time required for a complete Pspice simulation (i.e., presentation of I^1, I^2, and I^3 with a duration of 1550 seconds) is about 2 hours for the LM318 circuit and about 18 hours for the LM741 circuit. We believe this is caused by that the circuit model for the LM318 is simpler than that for the LM741.

5. CONCLUSION

The AART1-NN, like its predecessor the ART1-NN, can cluster in a parallel manner an arbitrary collection of binary input patterns. This capability makes the AART1-NN very attractive for high speed signal processing

applications. In this paper, a prototype 7-node AART1-NN circuit has been successfully designed and implemented with practical LM318 and LM741 Op-Amps with compensation circuits. Pspice circuit simulation was carried out to verify the circuit subject to three different input patterns. It has been shown that both LM318 and LM741 can be used to implement the ART1-NN with satisfactory performance, provided that they are properly compensated to eliminate the voltage drift caused mainly by the nonideal input offset volatge and input bias current.

6. ACKNOWLEDGMENT

This work was supported in part by the Florida High Technology Council and by the Division of Sponsored Research at UCF.

7: REFERENCES

[1] S. Grossberg, "Adaptive Pattern Recognition and Universal Recoding II: Feedback Exception, Olfaction and Illusions", Biological Cybernetics, Vol. 23, pp. 187-202, 1976.
[2] G. A. Carpenter and S. Grossberg, "A Massively Parallel Architecture for a Self-Organizing Neural Pattern Recognition Machine," Computer Vision, Graphics and Image Proce., Vol. 37, pp. 54-115, 1987.
[3] G. L. Heileman, M. Georgiopoulos and C. Abdallah, "A Dynamical Adaptive Resonance Architecture," accepted to the IEEE Transactions on Neural networks, September 1992; also partially published in the Proceedings of the IJCNN, Vol. 3, Singapore, November 1991, pp. 2658-2663.
[4] P. R. Gray and R. G. Meyer,"Recent Advances in Monolithic Operational Amplifiers Design," IEEE Transactions on Circuits and Systems, pp. 317-327, May 1974.
[5] R. D. Read and R. L. Geiger, "A Multi-Input OTA Circuit for Neutral Networks," IEEE Transaction on Circuits and Systems, vol. 36, pp. 767, 1989.
[6] C. S. Ho, J. J. Liou, M. Georgiopoulos, G. L. Heileman, and C. Christodoulou, "Design and Simulation of An Adaptive Resonance Theory Neural Network," Intl. J. Electronics, accepted September 1993; also partially published in Symposium on SMS, Taipei, March 1993, pp. 57-58.
[7] R. G. Irvine, Operational Amplifier Characteristics and Applications, 2nd Edition, Prentice-Hall New Jersey, 1987, Chap 3, pp. 69-85.
[8] Linear 1 Databook, National Semiconductor Corporation, 1988 Edition.

Table 1: Parameter values and initial conditions used for the PSpice circuit simulation.

$A_1=1$	$B_1=1.5$	$C_1=1$	$D_1=1$	$\delta_1=0.01$
$\varepsilon_1=1$	$A_2=0.3$	$B_2=100$	$C_2=100$	$D_2=1$
$\delta_2=0.01$	$\varepsilon_2=10$	$A_r=2$	P=0.99	Q=1
$\delta_r=0.02$	$\varepsilon_r=1$	L=1.01	$\delta_2=10^{-4}$	$\varepsilon_z=100$
$z_{13}(0)=0.24$	$z_{23}(0)=0.24$	$z_{14}(0)=0.22$	$z_{24}(0)=0.22$	$z_{31}(0)=1$
$z_{32}(0)=1$	$z_{41}(0)=1$	$z_{42}(0)=1$	$x_1(0)=0$	$x_2(0)=0$
$x_3(0)=0$	$x_4(0)=0$	$x_r(0)=0$	$x_3(0)=0$	$x_4(0)=0$

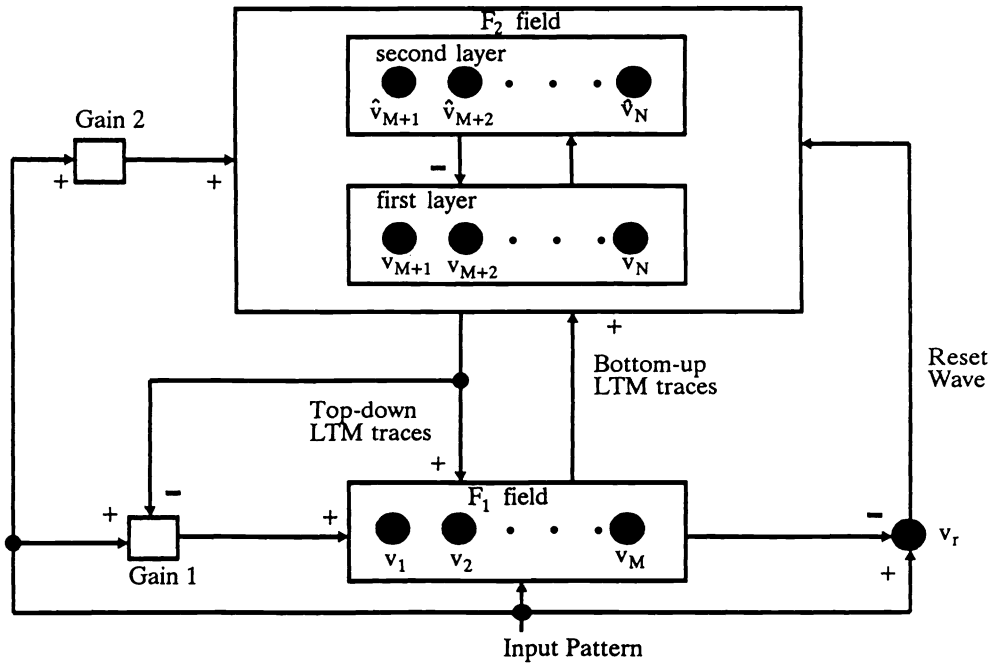


Figure 1: The structure of the augmented ART1 neural network.

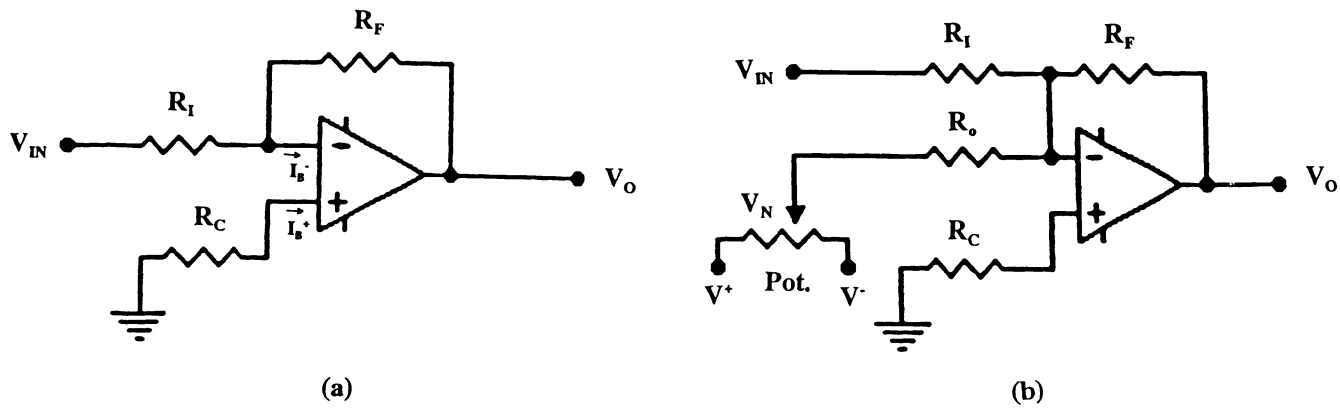


Figure 2: Op-Amp offset compensation; (a) inverting amplifier with bias-current compensation resistor, (b) inverting amplifier with voltage-offset compensation.

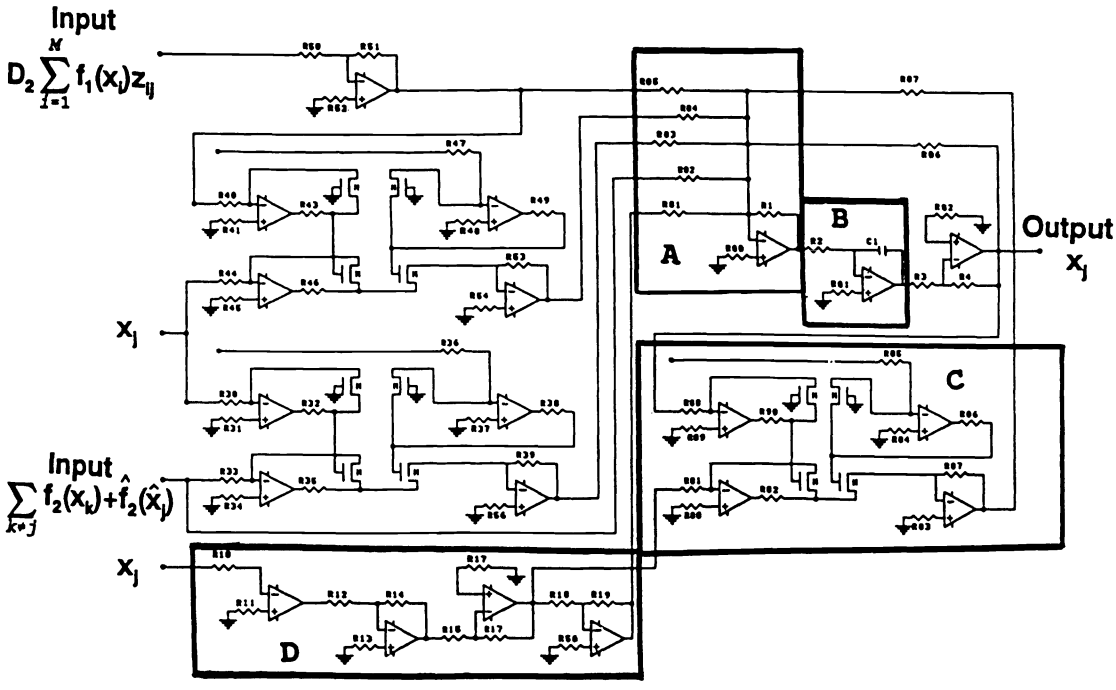


Figure 3: Circuit diagram for an arbitrary node v_j in the first layer of F_2 field.

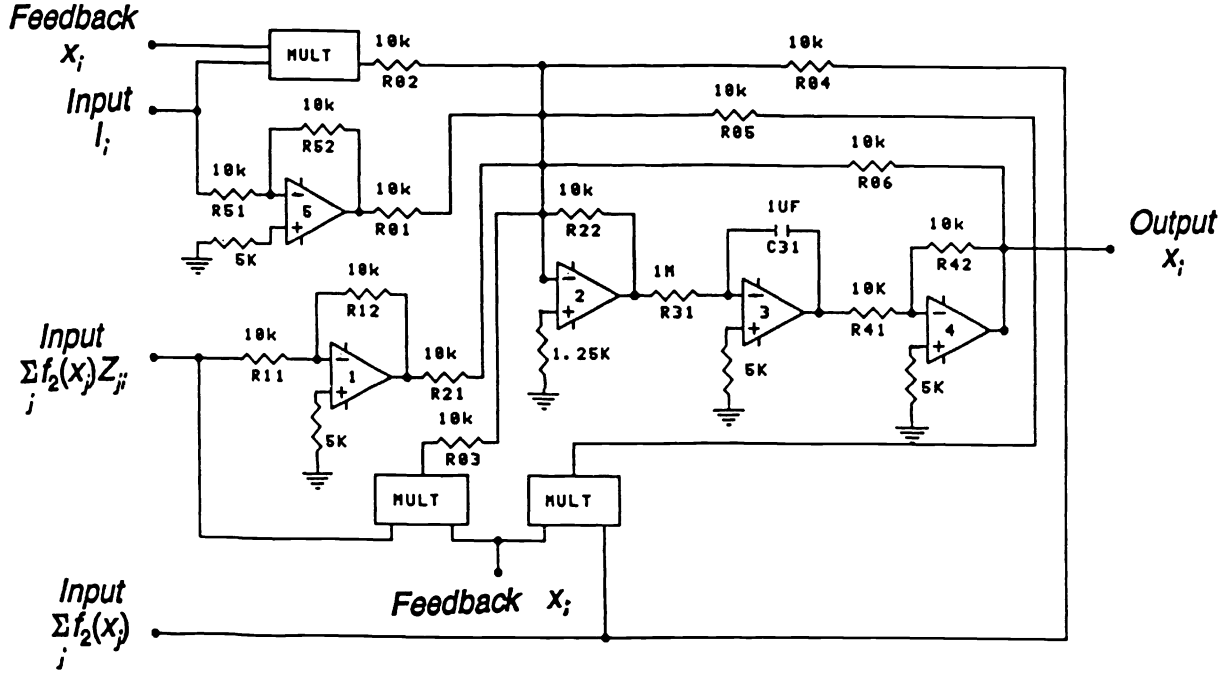


Figure 4: Circuit diagram for an arbitrary node v_i in the F_1 field.

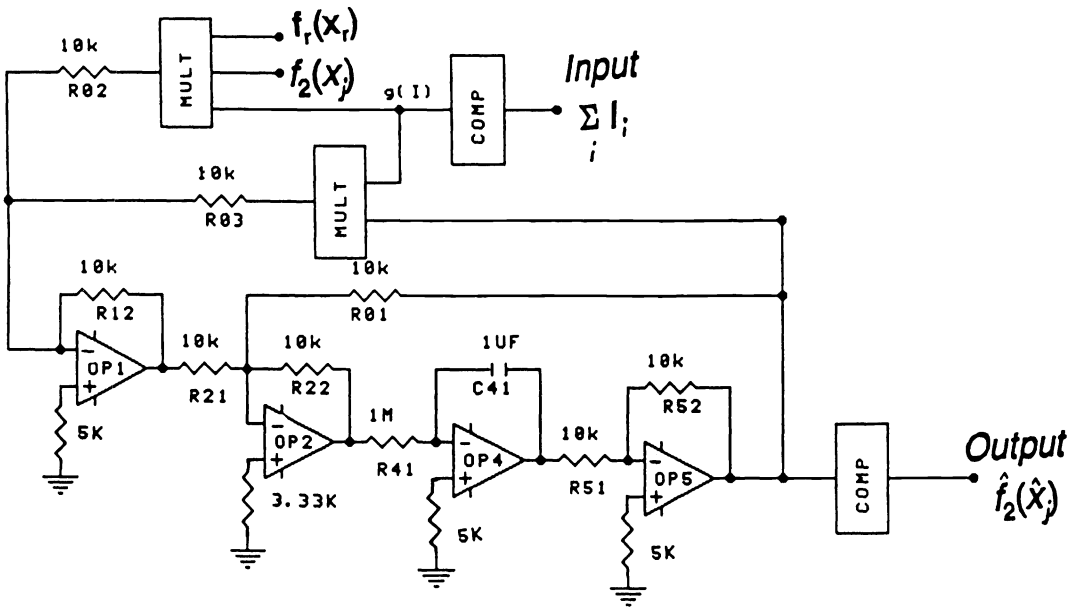


Figure 5: Circuit diagram for an arbitrary node $\hat{v}_j$ in the second layer of F_2 field.

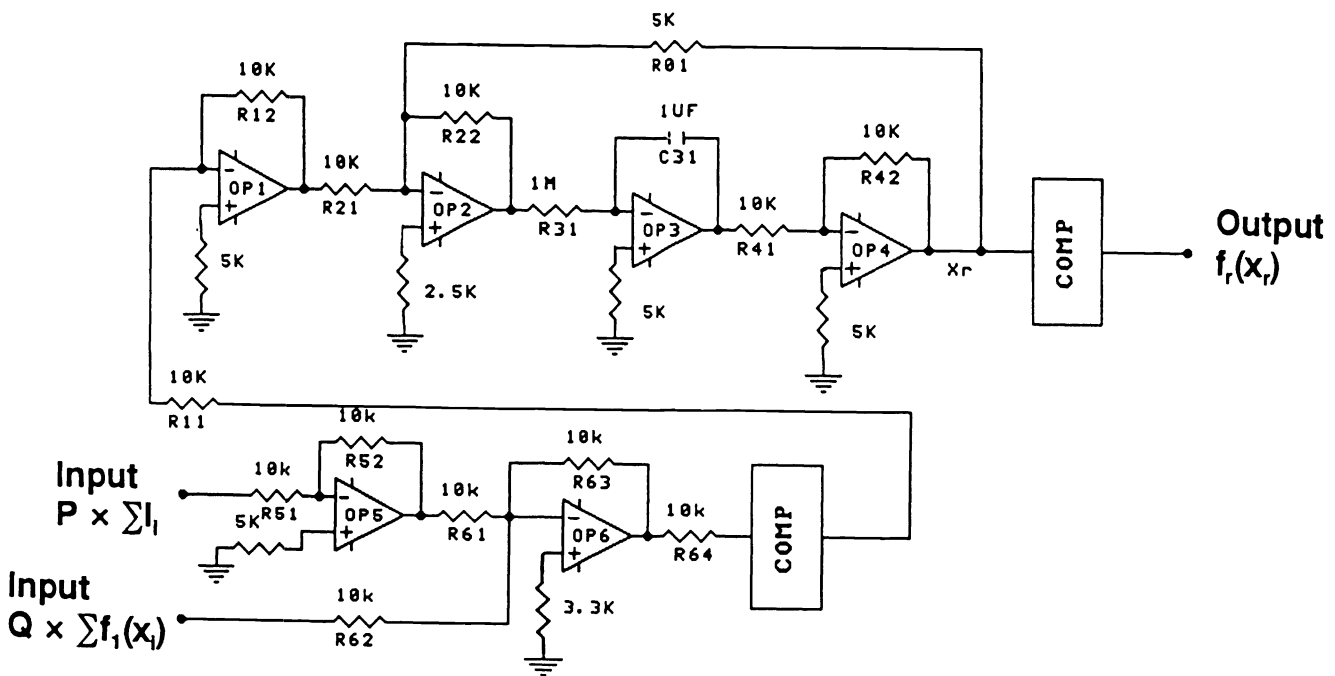


Figure 6: Circuit diagram for the reset node x_r.

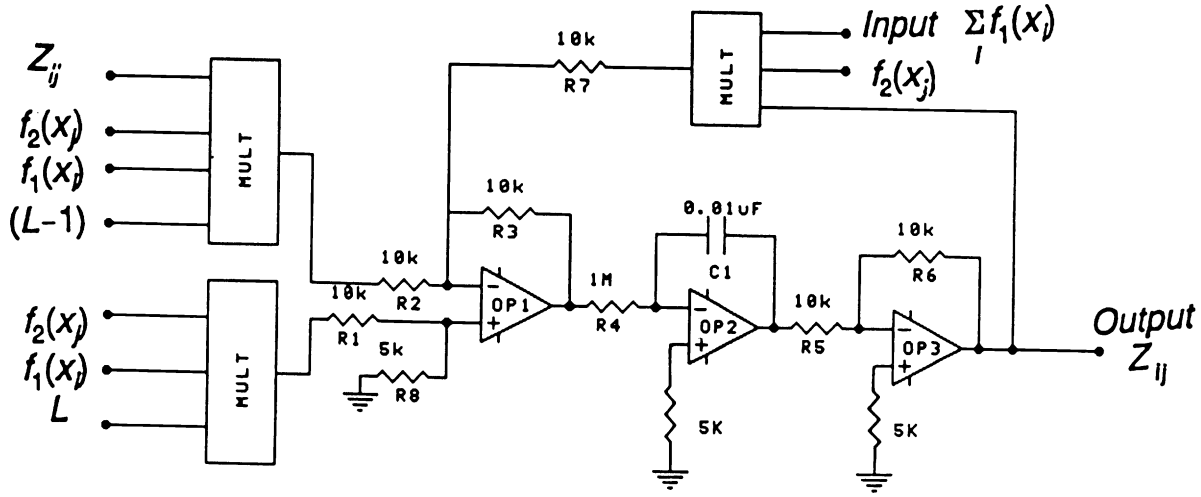


Figure 7(a): Circuit of bottom-up long-term memory (LTM) trace (z_{ij})

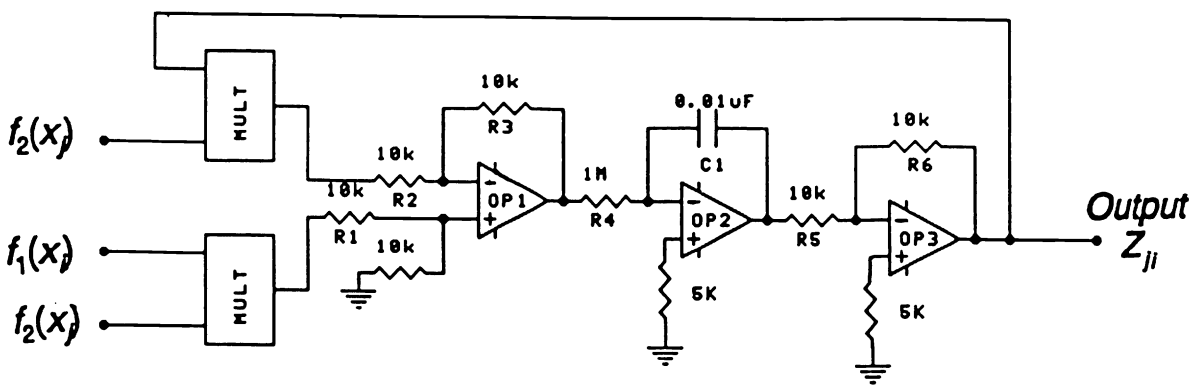


Figure 7(b): Circuit of top- down long-term memory (LTM) trace (z_{ji})

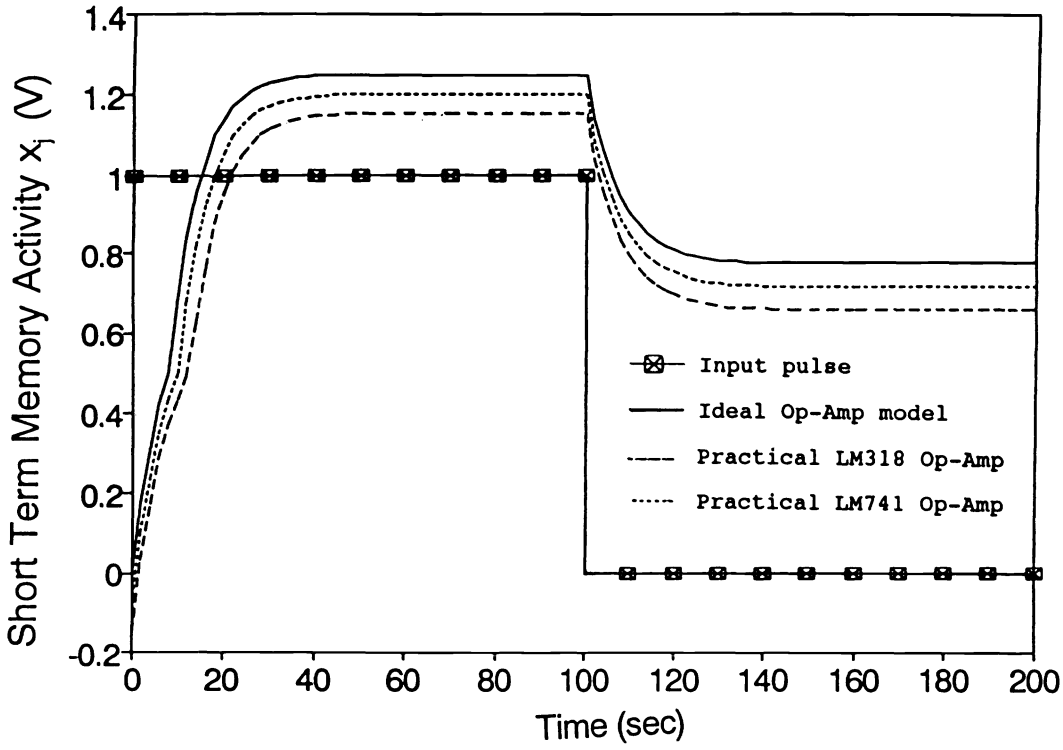


Figure 8: PSpice simulation results for the activity of an arbitrary node v_j in the first layer of F_2 field

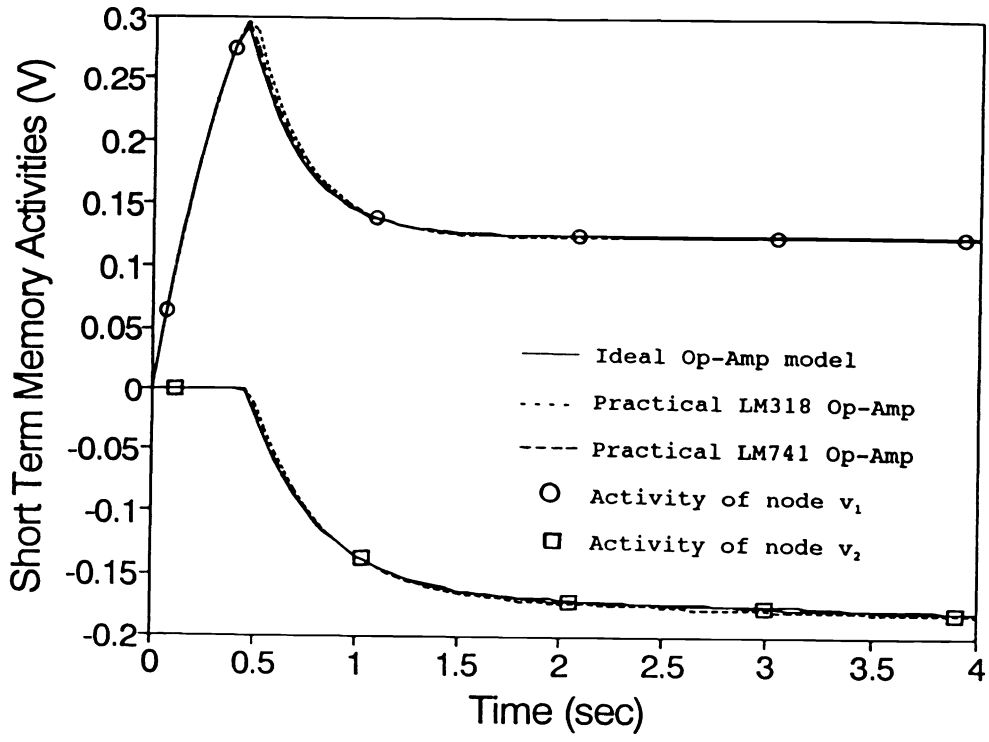


Figure 9(a): Activities of nodes v_1 and v_2 during the presentation of the first pattern I^1 (10)

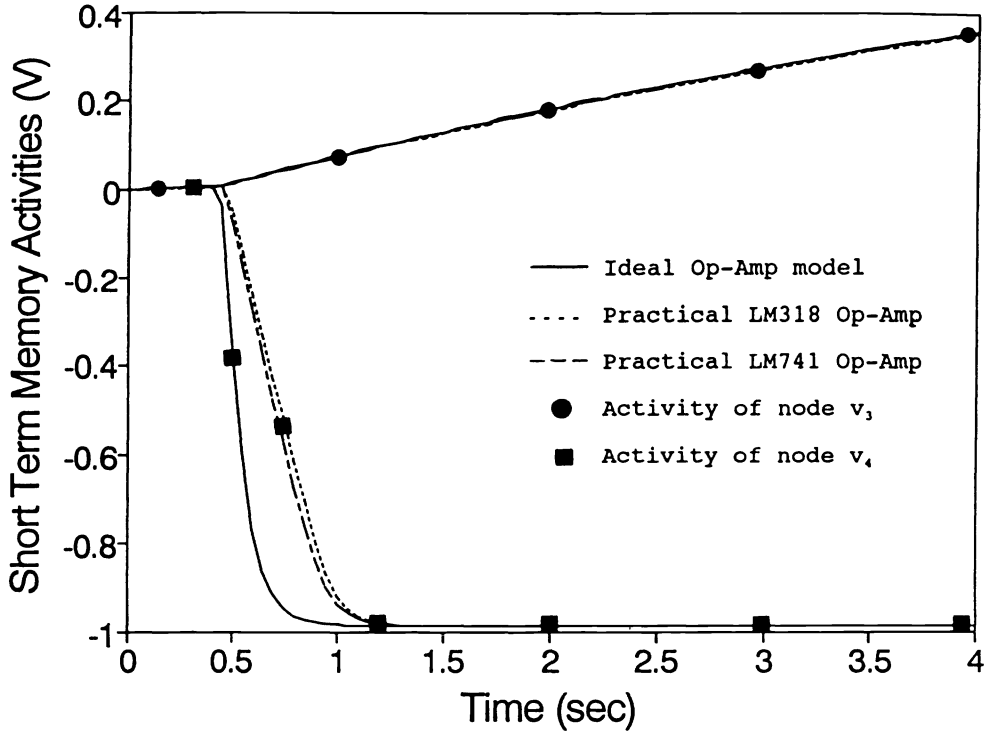


Figure 9(b): Activities of nodes v_3 and v_4 during the presentation of the first pattern I^1 (10)

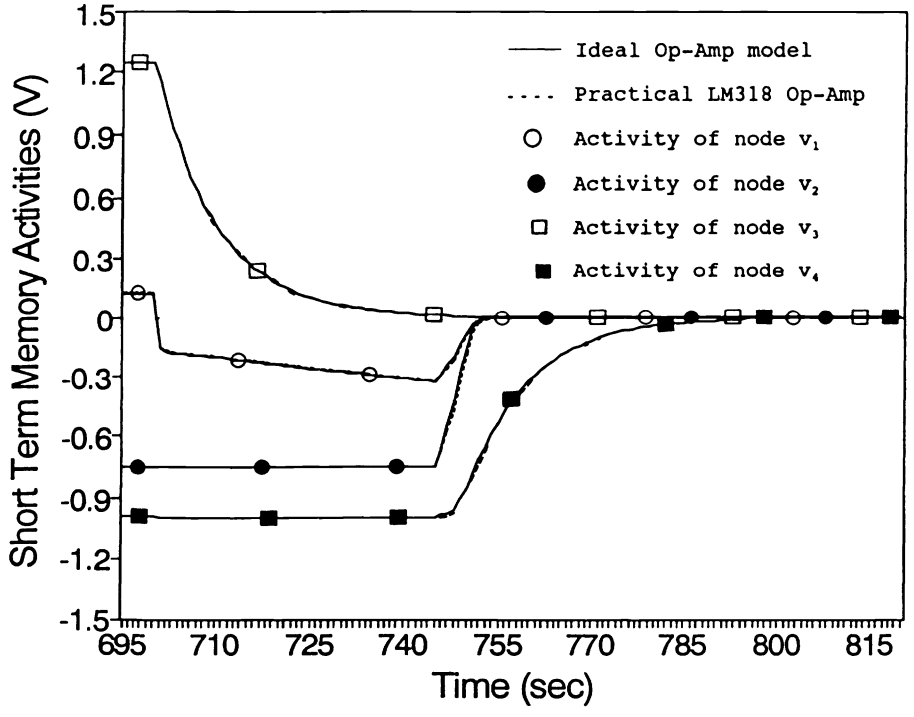
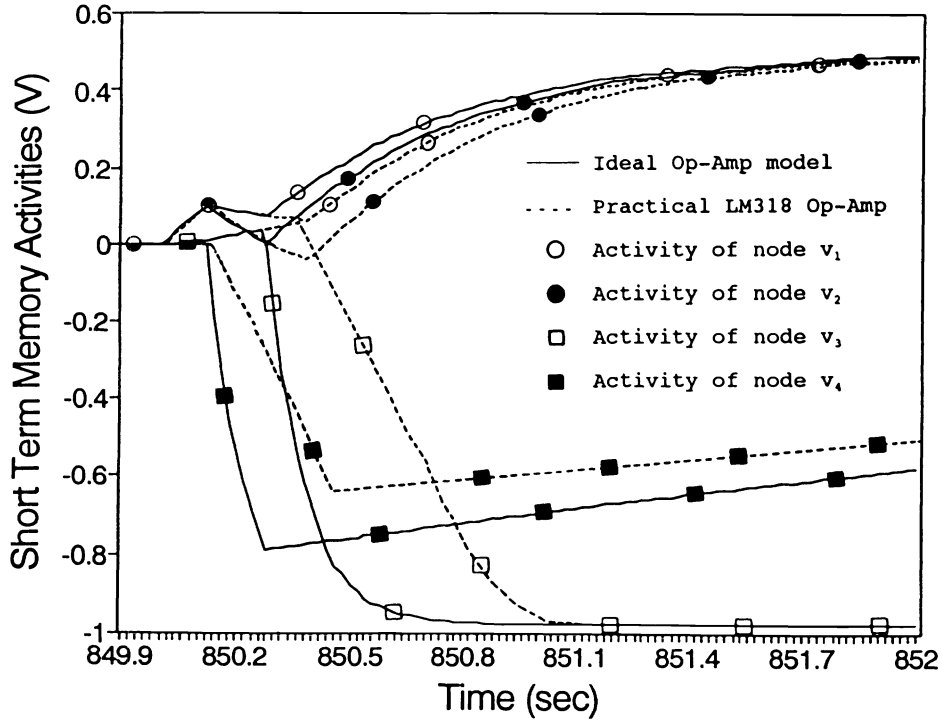
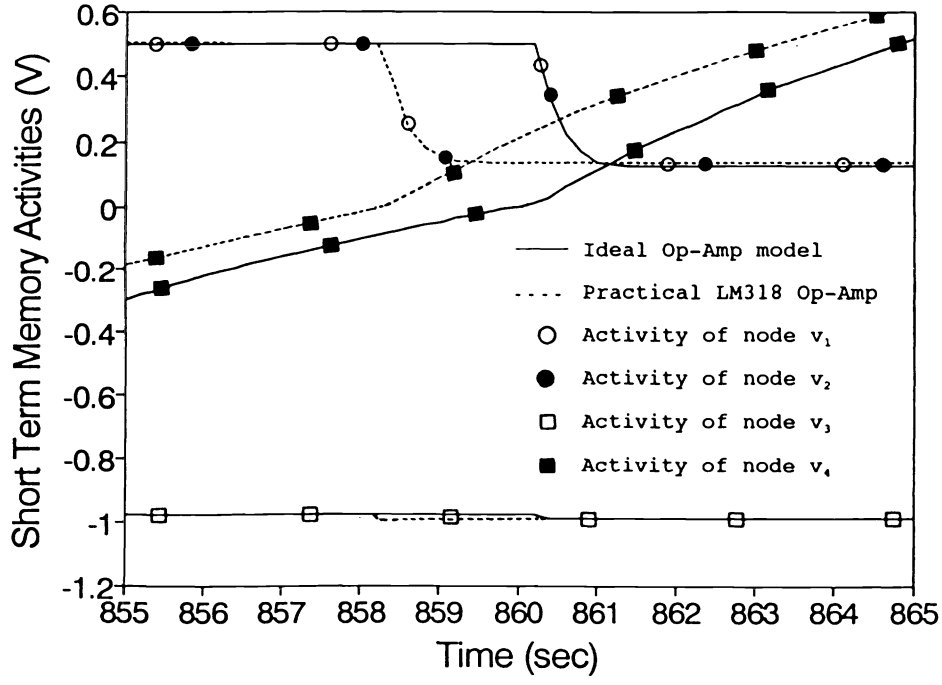


Figure 10: PSpice simulation results during the presentation of the second pattern I^2 (00).



(a)



(b)

Figure 11: PSpice simulation results during the presentation of the third pattern I^3 (11) illustrating (a) node v_4 is eventually selected within the time interval of [850, 852], (b) node v_4 becomes active within the time interval of [855, 865].