A One-Transistor Synapse Circuit with an Analog LMS Adaptive Feedback for Neural Network VLSI

T. C. Lu, M. L. Chiang and J. B. Kuo

Rm. 526, Dept. of Electrical Eng., National Taiwan University
Roosevelt Rd. Sec. 4, #1, Taipei, Taiwan
FAX:886-2-363-8247, Telephone:886-2-363-5251, E-mail:t7503007@twntucc1.bitnet

Abstract

Current integrated circuits realizing neural networks consume too much area for implementing synapses. This paper presents a one-transistor(1T) synapse circuit, which uses a single MOS transistor, more efficient for VLSI implementation of adaptive neural networks, as compared to other synapse circuits. This 1T synapse circuit can be used to implement multiply/divide/sum circuits for realizing an adaptive neural network. The feasibility of using this circuit in adaptive neural networks is demonstrated by a 4-bit analog-to- digital converter circuit based on the Hopfield's modified neural network model with an analog LMS adaptive feedback. DC and transient study shows that the 1T synapse circuits with an analog adaptive feedback circuit can be more efficiently used for VLSI implementation of adaptive neural networks.

I.Introduction

Recently, VLSI implementation of neural networks [1],[2] receives widespread attention owing to better speed performance for the model realized by hardware compared to by software. In most neural network ICs, most of the chip area is occupied by a large number of complicated synaptic connections. For example, Graf[1] uses four switches and two resistors to construct a synapse and nearly 90% of the chip area is occupied by the synapses. Although in other applications[2]-[5], simpler synapse circuits have been proposed, they are not concise enough for efficient VLSI implementation. It has been pointed out that the number of neurons in a chip determines the the storage capacity of a neural network no matter what learning rules have been used[6]. In reality, the number of neurons can be placed in a NN chip is limited by the size of the synaptic connections area. How to reduce the size of the synapse circuits has been the major task in designing neural network ICs. In this paper, a concise synapse circuit using a single MOS transitor will be described. It will also be shown that the multiply/divide/sum circuit implemented with this 1T synapse circuit has a good linearity within the output dynamic range for the neural networks with an analog adaptive feedback circuit. The feasibility of using this 1T synapse circuit will be demonstrated by a 4-bit analog-to-digital converter circuit using the Hopfield's modified neural network model[7] with an LMS adaptive feedback.

II. The Multiply/Divide/Sum Circuit

In Hopfield's modified neural network model, a neuron performs a sigmoid function of its input, which is the summation of other neurons' outputs multiplied by the weights stored in the synapses. Many synapse circuits have been proposed and implemented. Until now, the synapse circuit realized by a resistor is the most efficient. However, realization of a large amount of resistors requires special processing technologies[8]. Fundamentally, the resistor synapse serves as a role to transform the output voltage of one neuron's output to an input current for another neuron with a weight. Here, a single MOS transistor with a zero threshold voltage, which can be easily realized by adjusting the doses of the channel implants in a CMOS fabrication process, operating in the triode region has been used to replace the resistor for the synapse circuit. This 1T MOS synapse circuit performs as a programmable resistor with its conductance determined by the W/L ratio and the gate voltage. Figure 1 shows the multiply/divide circuit built with the 1T synapse circuits. With an NMOS synapse circuit at the input and a PMOS one in the feedback path, the circuit provides a good linearity property for multiply/divide operations. Figure 2 shows the SPICE simulated characteristics of the multiply and the divide operations. For one input, x, with a swing between -1V and +1V and another input, y, with a positive swing and the other input, z, with a negative swing, the output, $v_o = kxy/z$, has a swing from -1V to 1V. Within the output dynamic range, both the multiply and divide operations have an acceptable linearity for adaptive neural network applications.

III.The Analog Adaptive Circuit

Adaptation has been broadly used in signal processing, control and telecommunication systems to remove non-ideal effects [9]. The adaptive circuit in neural networks is used to compensate for the 1T synapse circuits and the neuron circuit with non-zero thresholds. Usually, adaptive neural networks, proposed in trainable pattern recognition [10], are realized by either computer software or digital circuits with analog-to-digital and digital-to-analog conversions required. During the training period, many iterations are required for convergence of the weight functions, which determine the conductances of the synapses [6]. This data conversion requirements severely limit not only the complexity of the patterns can be trained but also the storage capacity of the neural networks. Fig. 3 shows an analog adaptive feedback circuit. This analog adaptive circuit is derived from the digital LMS algorithm [9]: $w_{k+1} = w_k + \frac{\alpha}{x^2}\epsilon_k x$, where w_{k+1} is the next value of the weight vector, w_k is the current value of the weight vector, x_k is the current input vector, α is a parameter to adjust, and ϵ_k is the error, which is defined as the difference between the desired output and the analog output [9].

CH 3006-4/91/0000 – 1303 $1.00 © IEEE

The transistor MN1 is a synapse circuit with the source end connected to the input $-x$ and the gate controlled by the weight function w. With the $-d$ conncected to the positive input, the op amp $OA1$ produces the error which is the difference between the the desired output, d and the product of the input and the weight. Then the error is integrated through the $OA2$ circuit as the new weight for the synapse circuit. During the training period, the switch SW is on and the weight w converges to a value such that $xw = d$. Based on the analog adaptive feedback circuit, the weight function is written as:

$$w(t) = w_0 + \mu_0 \int_0^t (\frac{d}{\beta} - R[(w+x)x - \frac{1}{2}\frac{x^2}{\beta}])dt$$

$$\mu_0 = 5\beta_0\beta/C$$

where w_0 is the initial value of the weight vector, $w(t)$, $x(t)$ are the current values of the weight, the input respectively. β_0, β are the beta for the transistors $MN0, MN1$ respectively, and C is the integrator capacitor. Fig. 4 shows the SPICE simulation results of the weight and the percentage error in terms of the maximum error during the training period, of the analog adaptive circuit designed for two different μ_0's, which are determined by the corresponding circuit parameters. A larger μ_0 results in a quicker response. However, a larger μ may also leads to instability. Generally speaking, in spite of the nonlinearity of the adaptive circuit, the weight converges to its final value within a few microseconds [11]. Fig. 5 shows the available swings of the input x and the weight for different desired outputs, d. A smaller d provides a larger swings at the input and the weight. For the neural net applications, the available swings at the input and the weight are sufficient for circuit design.

IV.The ADC Circuit

In order to show the applicability of the 1T synapse circuit, with the analog adaptive feedback, a 4-bit ADC circuit using Hopfield's modified neural network model [7], [12] has been implemented with the 1T synapse circuits as shown in Fig. 6. P-channel and N-channel MOS devices with a zero threshold voltage have been used in the synapse array. In fact, the synapse array can still work with MOS devices having non-zero threshold voltages for the ADC with the analog adaptive feedback.
The conductance T_{ij} of each synapse circuit is determined by the formulas described below. $T_{xi} = C\frac{2^N}{V_H}2^i$, $T_{ij} = C\frac{1}{V_{BB}}2^{i+j}$, $T_{Ri} = C\frac{1}{V_R}2^i$, where V_{BB} is the swing of the neuron output, V_H is the swing of the input x, and N is the number of bits.

With a $5V$ supply connected to all gates, the synapse circuits can be programmed by properly choosing the W/L ratios for their specified T's. Fig. 7 shows the layout of the synapses array based on a $2\mu m$, 1-layer metal, 1-layer poly, N-well CMOS process, for the ADC using adaptive nerual net. The T_{ij}'s synapses are permuted vertically from left to right, followed by the T_{Ri}'s and the T_{Xi}'s. Power supplies of V_{SS} and V_{DD} are connected from the bottom right. Inputs of the synapses are from the left, V_0, V_1, V_2, V_3, which are the neuron outputs. The ouputs of the synapse array are u_0, u_1, u_2, u_3, at the right. Input signal, x, and reference voltage, $-V_R$, are also from the right. The hatch area shows the polysilicon layer, which is used for gates of the synapses and cross-overs. The synapse array occupies a silicon area of $140\mu m \times 140\mu m$. Overall layout of the synapse array is simple and efficient, which shows the strength of the 1T synapse circuit for neural net VLSI.

The neuron performing a sigmoid function with a large gain is realized by the CMOS op amp with an output stage served as a level shifter and an output drive as shown in Fig. 8(a). The small signal dc gain of the neuron circuit is about 5000, which is high enough such that the energy maxima and minima locate at the corners of the hypercube where all the neuron outputs are "0" and "1" exactly [13]. The compensation capacitor C_C of $2pF$ is used to provide a stable frequency response. The unity gain bandwidth product is about $2MHz$. Fig. 8 (b) shows the layout of the neuron circuit. It occupies a silicon area of $200\mu m \times 240\mu m$. The power dissipation of the neuron ciruit is about $2mW$. Due to the variations during the fabrication process, the neuron circuit may have an un-predictable non-zero threshold voltage, whose effects can be removed by adjusting the weights of the synapses connected between the input of the neuron and the reference voltage $-V_R$ controlled by the adaptive circuit during the adaptation period [14], [15].

Figs. 9 and 10 show the transients of the ADC with the adaptive 1T synapse circuit during the training period. As shown in Fig. 9, initially, the synapse weights are set at $3V$. When is the adaptation is on, switches SW_0-SW_3 are closed, a repetitive training pattern of a linearly increasing waveform is imposed at the input x. The adapt strategy used here is to adjust the T_{Ri}'s of four bits such that the nonideal effects associated with the neuron circuits can be reduced and the ADC characteristics for the overall input swing can be optimized. After four reptitions of the input pattern, the four weights associated with synapses T_{X0}, T_{X1}, T_{X2}, T_{X3} converge to their final values. Then, switches $SW0$ - $SW3$ are open, and the A/D converter begins to work with the adapted conductances stored in the capacitors C_0 - C_3.

Fig. 10 shows the dc performance of the ADC with the adaptive 1T synapse circuit. Fig. 10 (a) show the neuron output voltages (V_0, V_1, V_2) vs. input (V_{in}) curves at dc for output swings from 0 to -5V and from 0 to -1V respectively. The resolution of the A/D converter with the analog adaptive circuit within the input dynamic range (0-4V) is much more uniform as compared to the ADC without it. Fig. 10 (b) shows the binary values of neuron outputs vs. input (x) curves with and without the analog adaptive mechanism. Without the adaptive feedback circuit, the 4-bits ADC with the 1T synapse circuits show an unacceptable performance. With the adaptive feedback, the I/O transfer characteristics are linear.

So far, the adaptation of synapses for removing the non-zero neuron offsets has been considered. In fact, the nonlinearity in the conductance of the MOS transistor and the non-ideal neuron outputs, which affects the T_{ij}'s, can be critical in the performance of the neural nets, too. Consequently, the T_{ij}'s need also be trained individually. We may expand the adaptation period to cover the individual training of each synapses other than offset correction related synapses.

VI. References

[1] H. P. Graf, L. D. Jackel, and W. E. Hubbard, "VLSI implementation of a neural network model," *IEEE Computer* pp.41-49, March 1988

[2] M. Verleysen, et. al.,"A high-storage capacity content-addressable memory and its learning algorithm," *IEEE Trans. on Circuits and Systems*, pp.762-766, May 1989

[3] M. Verleysen, et. al., "Neural networks for high-storage content addressable memory: VLSI circuit and learning algorithm," *IEEE Journal of Solid – State Circuits*, pp.562-569, June 1989

[4] K. A. Boahen, et. al., "A heteroassociative memory using current-mode MOS analog VLSI circuits," *IEEE Trans. on Circuits and Systems*, pp.747-755, May 1989

[5] R. D. Reed and R. L. Geiger, "A multiple-input OTA circuit for neural networks," *IEEE Trans. on Circuits and Systems*, pp.767-770, May 1989

[6] R. P. Lippman, "An introduction to computing with neural nets," *IEEE ASSP Magazine*, April 1987

[7] D. W. Tank and J. J. Hopfield, "Simple neural optimization networks: An A/D converter, signal decision circuit, and a linear programming circuit," *IEEE Trans. on Circuits and Systems*, pp.533-541, May 1986

[8] R. E. Howard et. al., "An associate memory based on electronic neural network architecture," *IEEE Trans. on Electron Devices*, pp.1553-1556, Nov. 1987
[9] B. Widrow and D. D. Stearns, *Adaptive Signal Processing*, " *Prentice – Hall*, 1985
[10] B. Widrow et. al., "Neural networks for adaptive filtering and adaptive pattern recognition," *IEEE Computer*, pp. 25-39, March 1988
[11] B. Widrow et. al., "Layered neural nets for pattern recognition," *IEEE Trans. on Acoustics, Speech and Signal Processing* pp.1109-1118, July 1988
[12]B. W. Lee and B. J. Sheu, "Design of a neural-based A/D converter using modified Hopfield network," *IEEE Journal of Solid – Circuits*, pp.1129-1135, Aug. 1989
[13]J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two state neurons," *Proceeding of the National Academy of Science* pp.3088-3092, 1981
[14]M. L. Chiang and J. B. Kuo, "A one-transistor synapse circuit for neural network VLSI", to be published in *IEEE Trans. on Circuits and Systems*
[15]M. L. Chiang and J. B. Kuo, "A CMOS 1T synapse circuit for neural network VLSI," to be published in *IEEE Digest of* 1990 *Internationa Symposium on Circuits, and Systems*

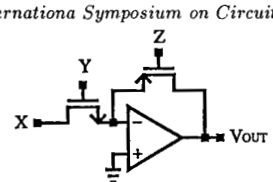


Fig. 1. The sum/multiply circuit.

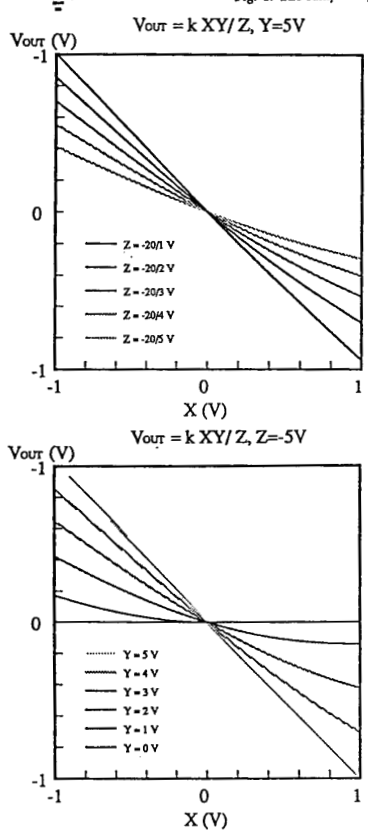


Fig. 2. Characteristics of the multiply and the divide operations.

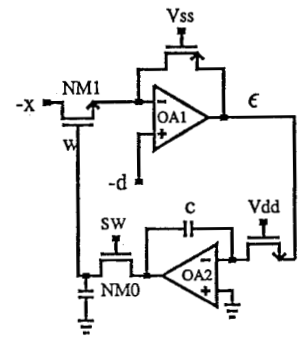


Fig. 3. The analog adaptive circuit.

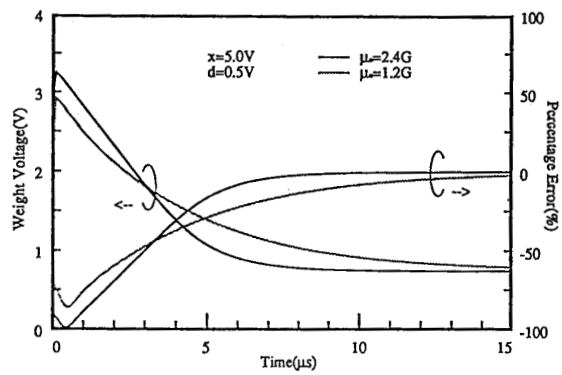


Fig. 4. The weight and the error of the analog adaptive circuit during the training transient for two different μ's.

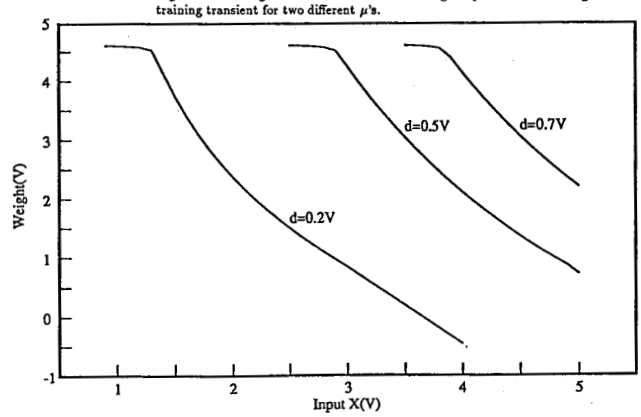


Fig. 5. The voltage swings of the input x and the weight for different d's.

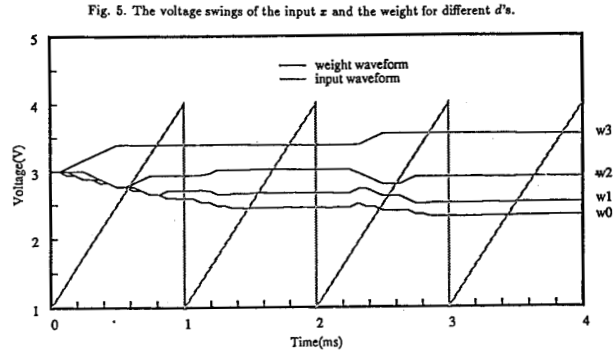


Fig. 9. Transients of the ADC with the adaptive synapse circuit during the training period.

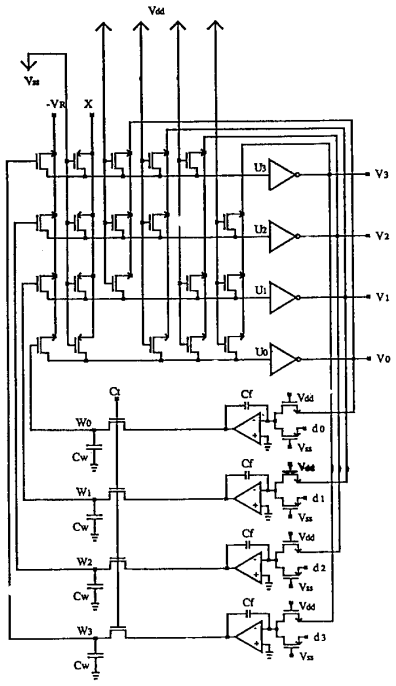


Fig. 6. The 4-bit ADC circuit using Hopfield's modified neural net model with the analog adaptive mechanism.

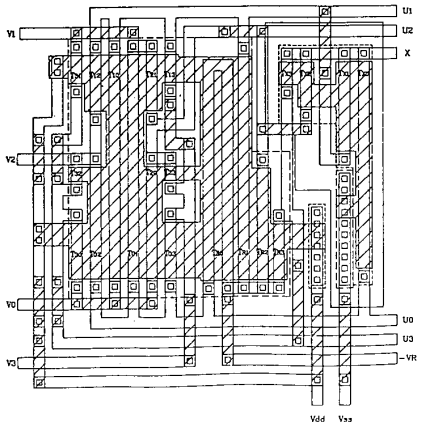


Fig. 7. Layout of the synapse array for the ADC using adaptive neural net.

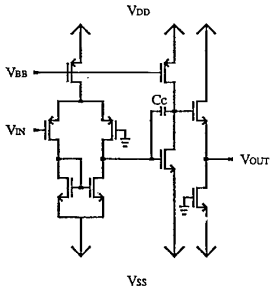
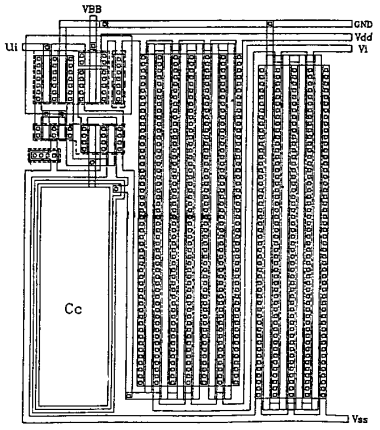


Fig. 8. (a)The neuron circuit.



(b)Layout of the neuron circuit.

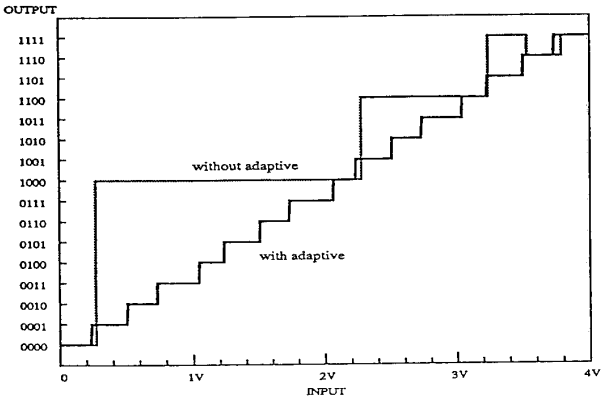
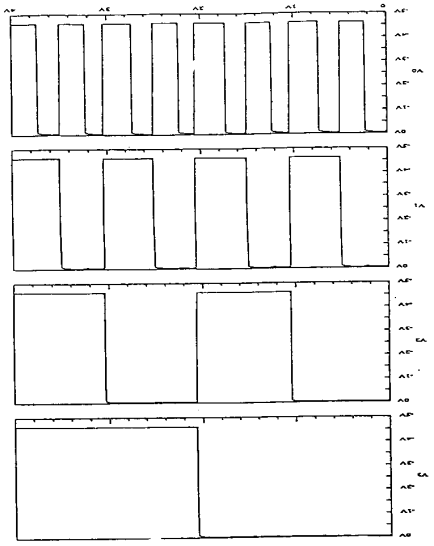


Fig. 10. DC performance of the ADC. (a)The neuron output voltage (V_0, V_1, V_2 vs. input voltage (x).



(b) The binary values of neuron outputs vs. input (x) curves of the ADC with and without the analog adaptive circuit.