Analogue adaptive neural network circuit

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

Indexing terms: Circuit theory and design, Transistors

Abstract: Current integrated circuits realising neural networks take up too much area for implementing synapses. The paper presents a one-transistor (1T) synapse circuit that uses a single MOS transistor, which is more efficient for VLSI implementation of adaptive neural networks, compared to other synapse circuits. This 1T synapse circuit can be used to implement multiply/divide/sum circuits to realise an adaptive neural network. The feasibility of using this circuit in adaptive neural networks is demonstrated by a 4-bit analogue-to-digital converter circuit, based on the Hopfield modified neural network model, with an analogue LMS adaptive feedback. DC and transient studies show that 1T synapse circuits with an analogue adaptive feedback circuit can be used more efficiently for VLSI implementation of adaptive neural networks.

1 Introduction

Recently, VLSI implementation of neural networks [1, 2] has received widespread attention, due to better speed performance for the model realised by hardware compared to that realised 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 storage capacity of a neural network, no matter what learning rules have been used [6]. In reality, the number of neurons that can be placed in a neural net (NN) chip is limited by the size of the synaptic connection area. A method for reducing 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 transistor 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 analogue adaptive feedback circuit. The feasibility of using this 1T synapse circuit will be demonstrated by a 4-bit analogue-to-digital converter circuit, using the Hopfield modified neural network model [7] with an LMS adaptive feedback.

Paper 8201G (E10), first received 16th January 1990 and in revised form 22nd April 1991

The authors are with the Department of Electrical Engineering, National Taiwan University, Room 526, 1 Roosevelt Road, Sec. 4, Taipei, Taiwan 107

2 The multiply/divide/sum circuit

In the Hopfield modified neural network model, a neuron performs a sigmoid function of its input, which is the summation of other neurons' outputs multipled by the weights stored in the synapses. Many synapse circuits have been proposed and implemented. Until now, the synapse circuit realised by a resistor has proved to be the most efficient. However, realisation of a large number of resistors requires special processing technologies [8]. Fundamentally, the resistor synapse is used to transform the neuron's output voltage into an input current for another neuron with a weight. Here, a single MOS transistor with a zero threshold voltage, which can be easily realised 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. Fig. 1 shows the multiply/divide circuit built with the 1T synapse circuits. With an NMOS synapse circuit at the

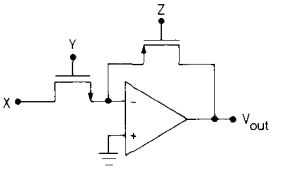


Fig. 1 *Sum/multiply circuit*

input and a PMOS one in the feedback path, the circuit provides a good linearity property for multiply/divide operations. Fig. 2 shows the SPICE simulated characteristics of the multiply and divide operations. For one input x, with a swing between -1 V and $+1$ V, another input y, with a positive swing, and the other input z, with a negative swing, the output $v_0 = kxy/z$ has a swing from -1 V to 1 V. Within the output dynamic range, both the multiply and divide operations have an acceptable linearity for adaptive neural network applications.

3 The analogue adaptive circuit

Adaptation has been broadly used in signal processing, control and telecommunication systems [9]. Usually, the adaptive neural networks, proposed in trainable pattern recognition [10], are realised by either computer software or digital circuits, with analogue-to-digital and digital-to-analogue conversions required. During the training

period, many iterations are required for convergence of the weight functions, which determine the conductances of the synapses [6]. These data conversion requirements

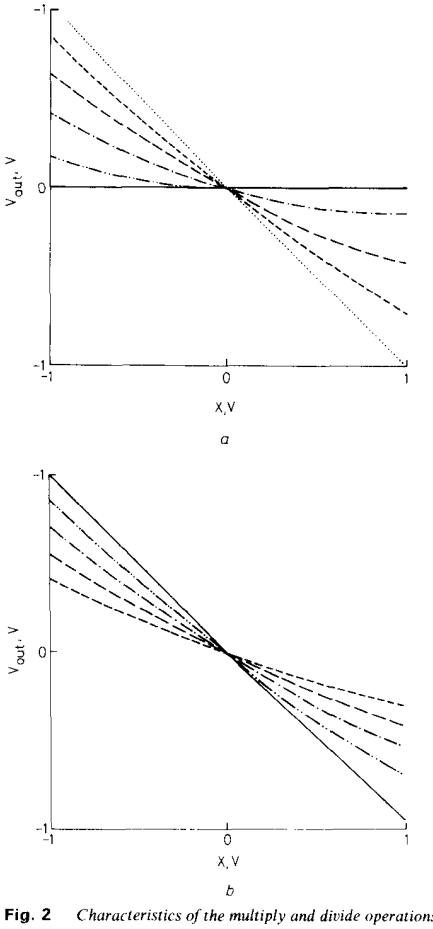


Fig. 2 *Characteristics of the multiply and divide operations*
a $V_{OUT} = kXY/Z$, $Z = -5$ V *b* $V_{OUT} = kXY/Z$, $Y = 5$ V
$Y = 5$ V; $Z = -20/1$ V
$Y = 4$ V; $Z = -20/2$ V
$Y = 3$ V; $Z = -20/3$ V
$Y = 2$ V; $Z = -20/4$ V
$Y = 1$ V; $Z = -20/5$ V
$Y = 0$ V

severely limit not only the complexity of the patterns that can be trained but also the storage capacity of the neural networks. Fig. 3 shows an analogue adaptive feedback circuit. This analogue adaptive circuit is derived from the digital LMS algorithm [9] $w_{k+1} = w_k + (\alpha/x^2)\varepsilon_k x$, 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 the convergence speed; and ε_k is the error, which is defined as the difference between the desired output and the analogue output [9].

The transistor MN_1 is a synapse circuit, with the source end connected to the input $-x$ and the gate controlled by a weight function w. With the $-d$ connected to the positive input, the operational amplifier OA_1 produces the error, which is the difference between the desired output, and the product of the input and the weight. Then, the error is integrated through the OA_2 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 analogue adaptive feedback circuit, the weight function is

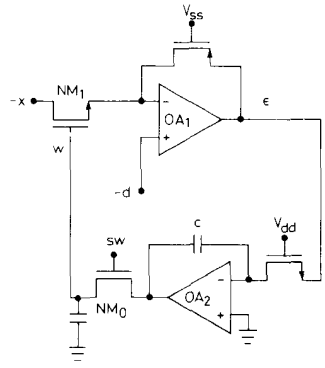


Fig. 3 *Analogue adaptive circuit*

written as

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

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

where w_0 is the initial value of the weight vector, and $w(t)$ and $x(t)$ are the current values of the weight and the input, respectively; β_0 and β are the beta for the transistors MN_0, MN_1, respectively; C is the integrator capacitor; and R is the conductance of the transistor in the negative feedback path of operational amplifier -- OA_1. 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 analogue adaptive circuit designed for two different values of μ_0, which are determined by the corresponding circuit parameters. A

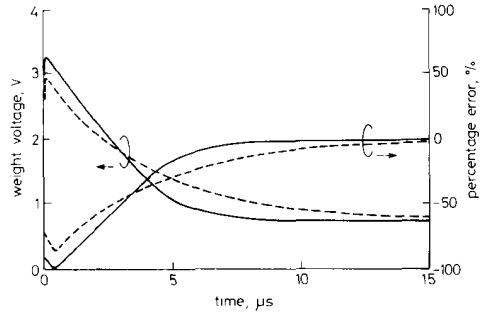


Fig. 4 *Weight and error of the analogue adaptive circuit during the training transient for two different values of* μ
$x = 5.0$ V; $d = 0.5$ V
$\mu = 2.4G$
$\mu = 1.2G$

larger μ_0 results in a quicker response. However, a larger μ_0 may also lead 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 w, for different d. A smaller d provides larger swings at the input and the weight. For

neural net applications, the available swings at the input and the weight of the adaptive circuit are sufficient.

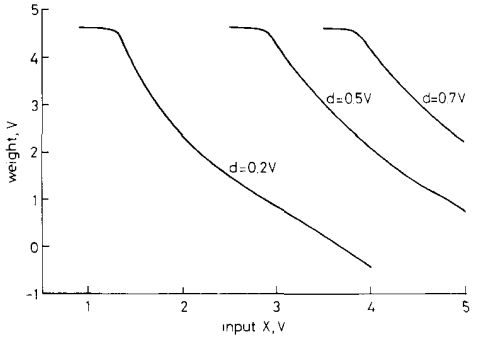


Fig. 5 *Voltage swings of the input x and the weight for different values of d*

4 The ADC circuit

To show the applicability of the 1T synapse circuit, with the analogue 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 zero threshold voltages have been used in the synapse array. In fact, the synapse array can still work with MOS devices having nonzero threshold voltages for the ADC with the analogue adaptive feedback.

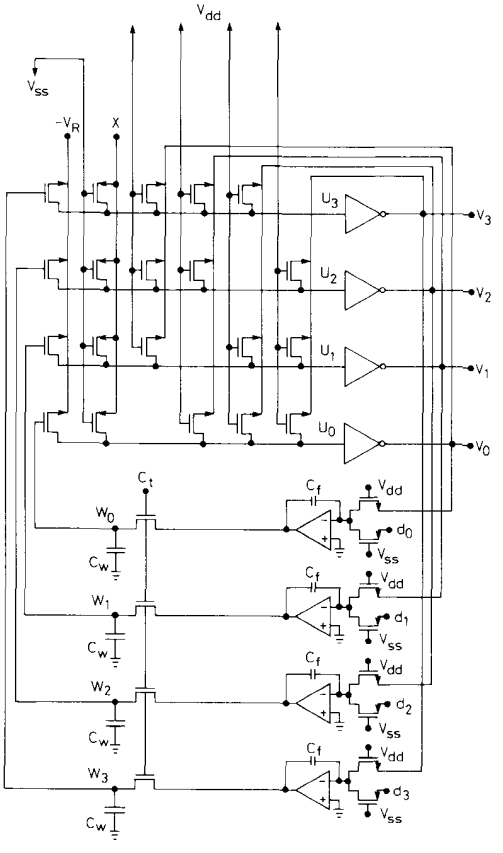


Fig. 6 *4-bit ADC circuit using Hopfield's modified neural net model with the analogue adaptive mechanism*

The conductance T_{ij} of each synapse circuit is determined by the formulas described below. The derivation of T_{ij} is based on the circuit point of view instead of the minimum entropy method [7]. Consider the Hopfield modified neural model in an analogue-to-digital conversion circuit, as shown in Fig. 6. Synapses T are connected between the neuron's input u_i and neuron's output $V_{j,j\neq i}$, the input signal x, and the reference voltage $-V_R$. For an input u_i greater than zero, the neuron produces an output $V_i = -V_{BB}$, which is 1. For u_i less than zero, the neuron produces an output $V_i = 0$, which is 0. Since the input impedance of the neuron is very large, the input u_i to the neuron can be expressed as follows:

$$u_i = \frac{\sum_{j\neq i} T_{ij} V_j + T_{xi} x - T_{Ri} V_R}{\sum T_{ij} + T_{Ri} + T_{xi}} \tag{2}$$

Consequently, the characteristics of the neuron can be modelled as

$$\sum_{j\neq i} T_{ij} V_j + T_{xi} x - T_{Ri} V_R > 0 \quad V_i = -V_{BB}$$

$$\sum_{j\neq i} T_{ij} V_j + T_{ix} x - T_{iR} V_{iR} V_R < 0 \quad V_i = 0$$

For an N-bit neural net ADC, applying a linearly increasing waveform from 0 V to V_H at the input x, there are two cases associated with the active and passive changes in the neuron outputs, which are defined below.

4.1 Active-change case

As the voltage at the neuron input u_i, which is proportional to the signal input x, is larger that the threshold voltage of the neuron, the neuron output changes from '0' to '1' — this is defined as the active change. For example, as the output pattern changes from '0111' to '1000', the third-bit output has an active change. In fact, an active change in the ith-bit output will trigger passive changes in the output of all less-than-ith bits. Hence, for $V_{j,j\leq i} = 0$, the ith bit output may have an active change.

4.2 Passive-change case

An active change in one bit may trigger another bit's output to change from '1' to '0', which is defined as the passive change for that bit. For example, as the output pattern changes '0111' to '1000', the zero-, first- and second-bit outputs have passive changes triggered by the third-bit active change. When evaluating the characteristics of the ADC circuit, only the decisive active change should be considered. In the following, the neuron's I/O characteristics and the associated active change will be used to derive the formula for T_{ij}, T_{xi}, T_{Ri}.

First, consider the first active change for each bit. For the zero bit, as x reaches Δ, which is defined as $V_H/2^N$, the output pattern will change from '0000' to '0001'. In this case, $T_{Ro} V_R = T_{x0} \Delta$. For the first bit, as $x = 2\Delta$, the output changes from '0010' to '0011'. Under this situation, $T_{10} V_{BB} + T_{R1} V_R = T_{x1} 2\Delta$. For the second bit, as x is 4Δ, the output switches from '0011' to '0100', and $\sum_{j<2} T_{2j} V_{BB} + T_{R2} V_R = T_{x3} 4\Delta$. Similarly, we may obtain the formula for the ith bit, as follows:

$$\sum_{j<i} T_{ij} V_{BB} + T_{Ri} V_R = T_{xi} 2^i \Delta \tag{3}$$

Now, consider the active changes after the first change for each bit. For the zero bit at the second active change, as x reaches 3Δ, the output pattern changes from '0010'

to '0011'. Now the characteristic equation is as follows: $T_{01} V_{BB} + T_{R0} V_R = T_{x0} 3\Delta$, which can be rearranged:

$$T_{01} V_{BB} = T_{x0} 2\Delta \tag{4}$$

At the third active change, as $x = 5\Delta$, the output changes from '0100' to '0101'. Consequently, $T_{02} V_{BB} + T_{R0} V_R = T_{x0} 5\Delta$, which can be further rearranged to give

$$T_{02} V_{BB} = T_{x0} 4\Delta \tag{5}$$

At the fourth change, since $x = 7\Delta$, the output switches from '0110' to '0111', and $(T_{01} + T_{02})V_{BB} + T_{R0} V_R = T_{x0} 7\Delta$, which is $(T_{01} + T_{02})V_{BB} = T_{x0} 6\Delta$. Note that this equation can be derived from the last two equations. Consequently, we may obtain the following formula:

$$T_{0j} V_{BB} = T_{x0} 2^j\Delta \quad j \neq 0 \tag{6}$$

Based on the formula for the zero bit, the formula for other bits at each active change can be obtained.

$$T_{ij} V_{BB} = T_{xi} 2^j\Delta \quad j > i \tag{7}$$

Therefore

$$T_{Ri} = \frac{1}{V_R}\left[T_{xi} 2^i\Delta - \sum_{j<i} T_{ij} V_{BB} \right] \tag{8}$$

$$T_{ij} = \frac{1}{V_{BB}} T_{xi} 2^j\Delta \quad j > i \tag{9}$$

Assuming that for $j < i T_{ij} = (1/V_{BB})T_{xi} 2^j\Delta$,

$$\sum_{j<i} T_{ij} = \frac{1}{V_{BB}} T_{xi}\Delta \sum_{j=0}^{i-1} 2^j = \frac{1}{V_{BB}} T_{xi}\Delta(2^i - 1) \tag{10}$$

$$T_{Ri} = \frac{1}{V_R}\left[T_{xi} 2^i\Delta - T_{xi}\Delta(2^i - 1)\right] \tag{11}$$

Further assuming that $T_{ij} = T_{ji}$, $(1/V_{BB})T_{xi} 2^j\Delta = (1/V_{BB})T_{xj} 2^i\Delta$. Consequently, $T_{xi} 2^j = T_{xj} 2^i$ and $T_{xi} = K2^i$. So, we have derived the following formulas:

$$T_{xi} = C \frac{2^N}{V_H} 2^i \tag{12}$$

$$T_{ij} = C \frac{1}{V_{BB}} 2^{i+j}$$

$$T_{Ri} = C \frac{1}{V_R} 2^i$$

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 5 V supply connected to all gates, the synapse circuits can be programmed by properly choosing the W/L ratios for their specified T. Fig. 7 shows the layout of the synapses array based on a 2 μm, 1-layer metal, 1-layer polysilicon, N-well CMOS process, for the ADC

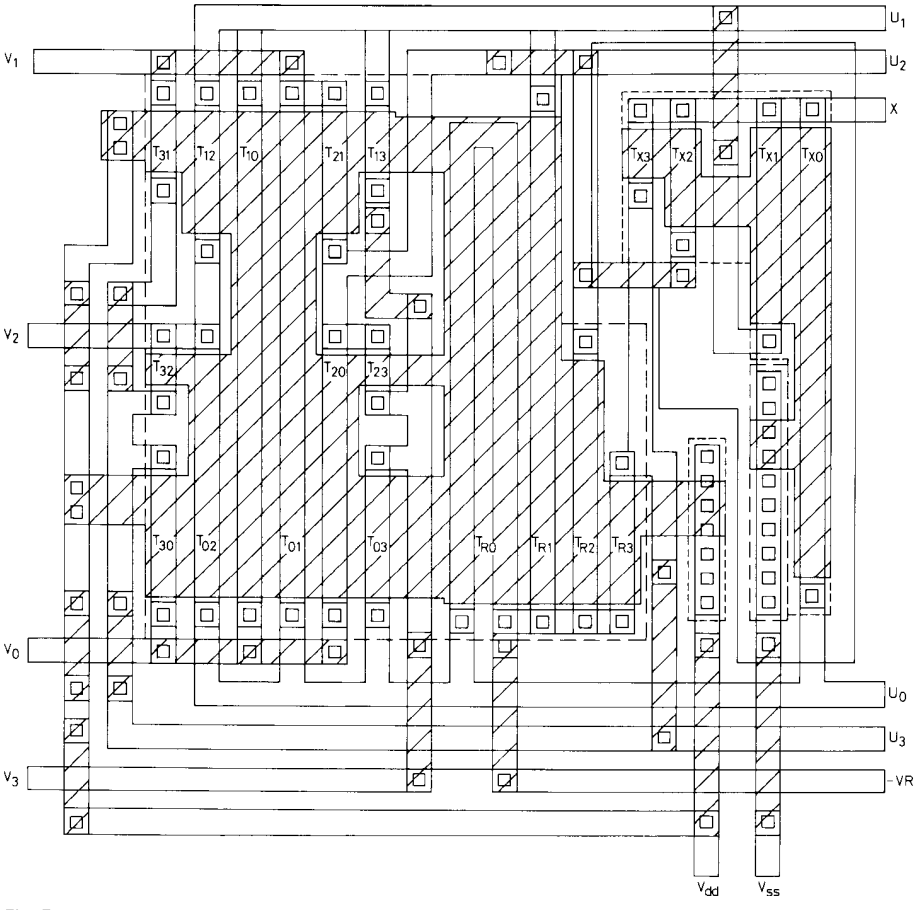
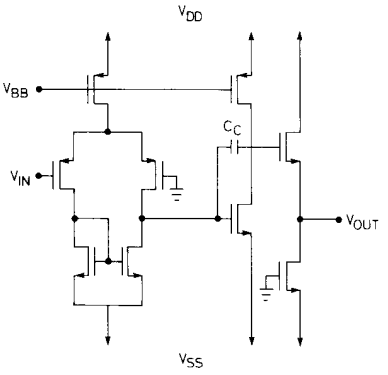


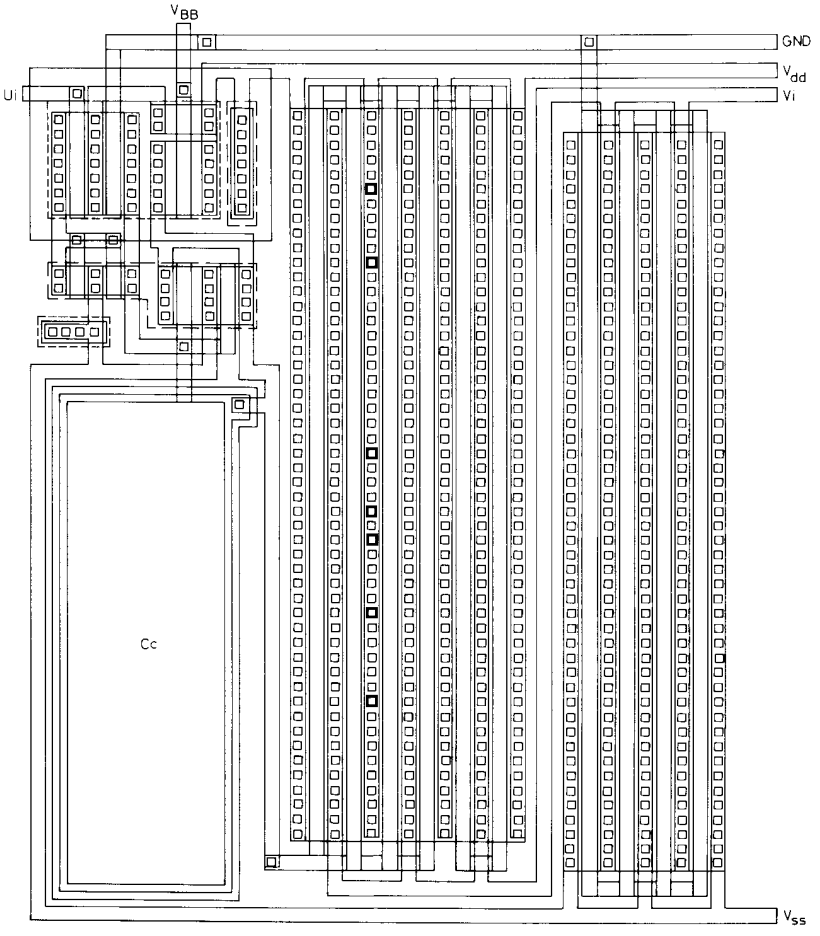
Fig. 7 *Layout of the synapse array for the ADC using adaptive neural net*

using adaptive neural net. The T_{ij} synapses are permuted vertically from left to right, followed by the T_{Ri} synapses and the T_{xi} synapses. Power supplies for V_{SS} and V_{DD} are connected from the bottom right. Inputs to the synapses are from neuron outputs V_0, V_1, V_2 and V_3 at left. The outputs of the synapse array are u_0, u_1, u_2 and 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 synapase array occupies a silicon area of 140 μm × 140 μm. The 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 realised by the CMOS operational amplifier, with an output stage serving 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 3000, which is high enough 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 2.9 pF is used to provide a stable frequency response. The unity gain bandwidth is about 1 MHz. Fig. 8*b* shows the layout of the neuron circuit. It occupies a silicon area of 200 μm × 240 μm. The power dissipation of the neuron circuit is about 2 mW. Owing to the varia-



a



b

Fig. 8 *Neuron circuit*
a diagram
b layout

tions during the fabrication process, the neuron circuit may have an unpredictable, nonzero 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].

Fig. 9 shows 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 3 V. When the adaptation is on, switches SW_0–SW_3 are closed, and a repetitive training pattern of a linearly increasing waveform is imposed at the input x. The adaptive strategy used here is to adjust the T_{Ri} 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 optimised. After four repetitions of the input pattern, the four weights associated with synapses T_{x0}, T_{x1}, T_{x2} and T_{x3} converge to their final values. Then, switches SW_0–SW_3 are opened, and the A/D converter begins to work with the adapted conductances stored in the capacitors C_0–C_3.

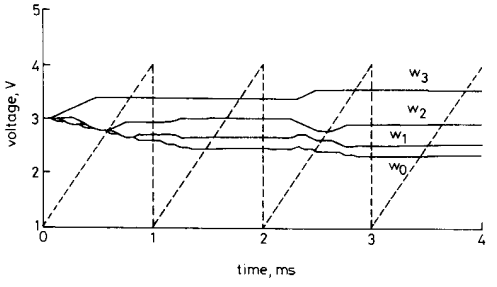


Fig. 9 *Transients of the ADC with the adaptive synapse circuit during the training period*

——— weight waveform
- - - - input waveform

Together with the 1T synapse circuits, the analogue adaptive feedback circuits have been used to compensate for the nonzero thresholds associated with the synapses and the neurons.

Fig. 10 shows the DC performance of the ADC with the adaptive 1T synapse circuit. Fig. 10*a* shows the neuron output voltages (V_0, V_1, V_2) against input (x) curves at DC for output swings from 0 to -5 V and from 0 to -1 V, respectively. The resolution of the A/D converter, with the analogue adaptive circuit within the input dynamic range (0–4 V), is much more uniform than the ADC without it. Fig. 10*b* shows the binary values of neuron outputs against input (x) curves, with and without the analogue adaptive feedback circuit. Without the adaptive feedback circuit, the 4-bit ADC with the 1T synapse circuits shows an unacceptable performance. With the adaptive feedback, the I/O transfer curve is linear.

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

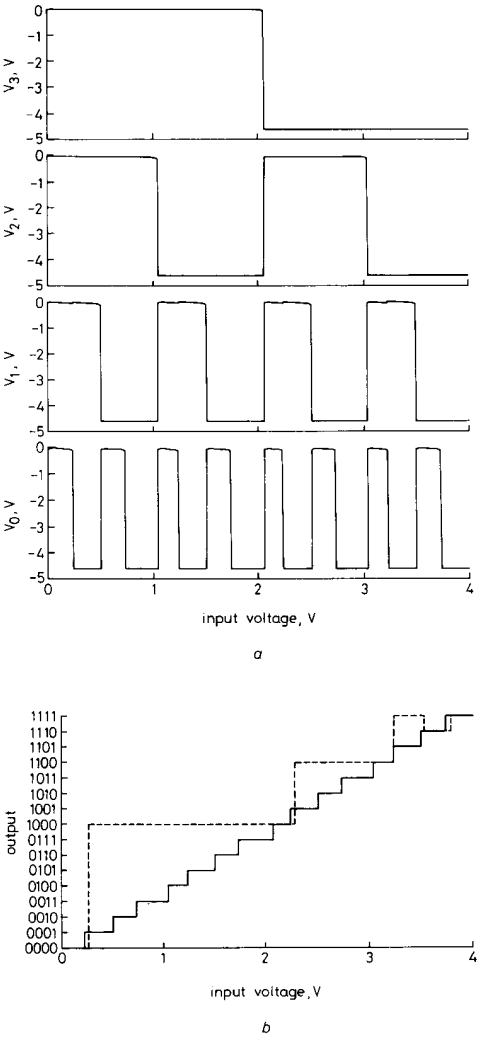


Fig. 10 *DC performance of the ADC*

a neuron output voltage (V_0, V_1, V_2) against input voltage x
b binary values of neuron outputs against input (x) curves of the ADC, with and without the analogue adaptive circuit
——— with adaptive
- - - - without adaptive

5 Conclusion

The 1T synapse circuits with an adaptive feedback, which are more efficient for VLSI implementation of adaptive neural networks, have been described. The versatilities of using 1T synapse circuits with the adaptive feedback have been demonstrated by the 4-bit analogue-to-digital converter circuit.

6 Acknowledgments

This work is supported under ROC National Science Council Contract NSC79-0404-E002-21.

7 References

1 GRAF, H.P., JACKEL, L.D., and HUBBARD, W.E.: 'VLSI implementation of a neural network model', *IEEE Computer*, 1988, pp. 41–49

2 VERLEYSEN, M., SIRLETTI, B., VANDEMEULEROECKE, A.M., and JESPERS, P.G.A.: 'A high-storage capacity content-addressable memory and its learning algorithm', *IEEE Trans.*, 1989, **CAS-36,** pp. 762–766

3 VERLEYSEN, M., SIRLETTI, B., VANDEMEULEROECKE, A.M., and JESPERS, P.G.A.: 'Neural networks for high-storage content addressable memory: VLSI circuit and learning algorithm', *IEEE J. of Solid-State Circuits*, 1989, pp. 562–569

4 BOAHEN, K.A., POULIQUEN, P.O., ANDREOU, A.G., and JENKINS, R.E.: 'A heteroassociative memory using current-mode MOS analog VLSI circuits', *IEEE Trans.*, 1989, **CAS-36,** pp. 747–755

5 REED, R.D., and GEIGER, R.L.: 'A multiple-input OTA circuit for neural networks', *IEEE Trans.*, 1989, **CAS-36,** pp. 767 770

6 LIPPMAN, R.P.: 'An introduction to computing with neural nets', *IEEE ASSP Magazine*, April 1987

7 TANK, D.W., and HOPFIELD, J.J.: 'Simple neural optimization networks: An A/D converter, signal decision circuit, and a linear programming circuit', *IEEE Trans.*, 1986, **CAS-33,** pp. 533–541

8 HOWARD, R.E., *et al.*: 'An associate memory based on electronic neural network architecture', *IEEE Trans. on Electron Devices*, 1987, pp. 1553–1556

9 WIDROW, B., and STEARNS, D.D.: 'Adaptive signal processing' (Prentice-Hall, 1985)

10 WIDROW, B., *et al.*: 'Neural networks for adaptive filtering and adaptive pattern recognition', *IEEE Computer*, March 1988, pp. 25–39

11 WIDROW, B., *et al.*: 'Layered neural nets for pattern recognition', *IEEE Trans.*, 1988, **ASSP-,** pp. 1109–1118

12 LEE, B.W., and SHEU, B.J.: 'Design of a neural-based A/D converter using using modified Hopfield network', *IEEE Journal of Solid-State Circuits*, 1989, pp. 1129–1135

13 HOPFIELD, J.J.: 'Neurons with graded response have collective computational properties like those of two state neurons', *Proc. of the National Academy of Science*, 1981, pp. 3088–3092

14 CHIANG, M.L., and KUO, J.B.: 'A one-transistor synapse circuit for neural network VLSI'. Digest of International Symposium on Electron Device Material Symposium, Taiwan, Nov. 1990

15 CHIANG, M.L.: 'The study and design of neural network integrated circuits'. Master Thesis, National Taiwan University, 1990