VLSI Implementation of Synaptic Weighting and Summing in Pulse Coded Neural-Type Cells

Gyu Moon, *Student Member, IEEE,* Mona E. Zaghloul, *Senior Member, IEEE,* and Robert W. Newcomb, *Fellow, IEEE*

Abstract— **This paper presents the hardware realization for synaptic signal weighting and summing using pulse coded neural-type cells (NTC's). The basic information processing element (NTC) encodes the information into the form of pulse duty cycles using voltage-controlled resistors for which a pulse duty cycle modulation technique is newly proposed. Summation is executed by a simple capacitor circuit as a current integrator. Layouts for and measurements on a fabricated integrated design are included.**

Index Terms—**Neural networks, VLSI implementation.**

I. Introduction

ANALOG pulse neural circuit techniques [1]–[9] offer a good trade-off between digital and analog design, where either pulse amplitude or frequency is used for encoding information. Recently several authors proposed electronic neural circuits which employ well-developed pulsed communication techniques [2]–[6], [9]. In this paper, we discuss the use of a "neural-type cell" (NTC) [10]–[14] in the design of synaptically weighted signal summation.

Neural-type micro systems, which were introduced in [10] and [11], draw one's attention owing to their similarity in signal processing to that of biological nervous systems as well as their simple structure and small size. Among basic components which compose the neural-type micro systems, we adopt, in this work, an NTC as a basic processing element. The fully integrated NTC [12]–[14] is composed of nine transistors and one capacitor. In hardware implementation, it can use small silicon area, which is an important criterion in the field of neural network hardware realization because of the need for an enormous numbers of neurons to realize practical neural systems.

Inspired by biological models, we develop here a simple integrated circuit structure for a neuron with synaptic weighting and summing. The NTC and its associated circuits function like a biological neuron with synaptic junctions. The NTC is an electronic analogy of a biological soma; it initiates reactions, with a given external voltage (stimulus), by generating a stream of electrical (biochemical) pulse waves. Three inverters are added for digitizing the pulse waves, where a threshold level is determined by an inverter logic threshold voltage [27]. The weighting, analogous to that of synaptic junctions, is realized by variable voltage-controlled resistors. It is well known [15], [31] that, in artificial neural networks, a computation of weighting and summing is to be performed. Weighting signals in voltage-level neural network implementations are computed through a multiplication process. Instead, we adopt a voltage-controlled resistor for performing weighting on the pulse duty cycle. For the summation, a simple capacitor circuit is adopted to add charge.

Manuscript received July 1, 1991; revised October 31, 1991. This work was supported by the NSF under Grants MIP-90-01658 and MIP-89-21122 and by the ONR under Grant N0001490J114.

G. Moon and M. E. Zaghloul are with the Department of Electrical Engineering and Computer Science, George Washington University, Washington, DC, 20052.

R. W. Newcomb is with the Electrical Engineering Department, University of Maryland, College Park, MD 20742.

IEEE Log Number 9105808.

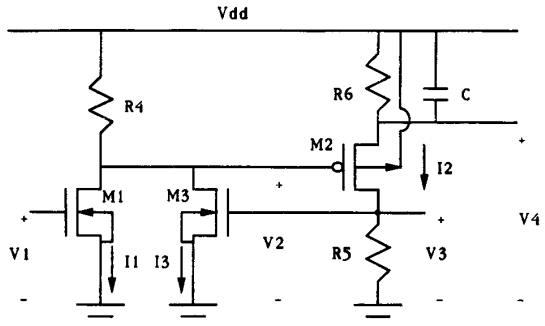


Fig. 1. Circuit diagram of neural-type cell.

We introduce a new pulse duty cycle modulation (PDCM) technique. PDCM is a modulation technique on a pulse stream whose duty cycle contains information. The duty cycle is controlled by a voltage-controlled resistor and is later converted into a dc voltage form. In this technique there is no need for a clock to synchronize pulses or to adjust pulse width. Thus, with this technique along with the NTC, we are able to build a very simple electronic neuron of small size. This PDCM technique is used on the output pulse stream of the NTC and is controlled by changing a voltage variable resistance in the NTC. The voltage-controlled resistor is composed of two enhancement-type CMOS transistors [16]. Weighting signals are summed by means of switching transistors and a capacitor.

In Section II, we review the basic operation of the NTC. In Section III, we briefly describe the CMOS voltage-controlled linear resistor. In Section IV, we discuss the circuits for synaptic weighting and their use in signal summing. A simple capacitor circuit with large RC time constant is employed as a charge integrator. Considerations for VLSI design along with chip design, layout, and measurements on a fabricated chip are discussed in Section V.

II. Basic Structure of an NTC

The CMOS circuit for the NTC under consideration is shown in Fig. 1. It consists of three transistors ($M1$–$M3$),

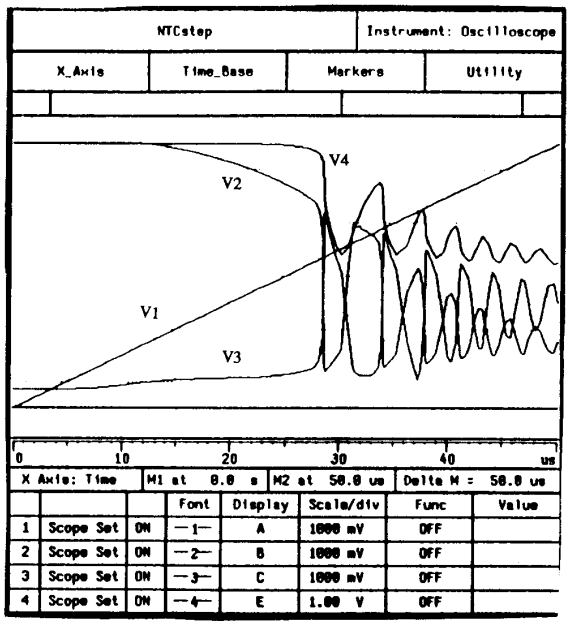


Fig. 2. SPICE output of NTC.

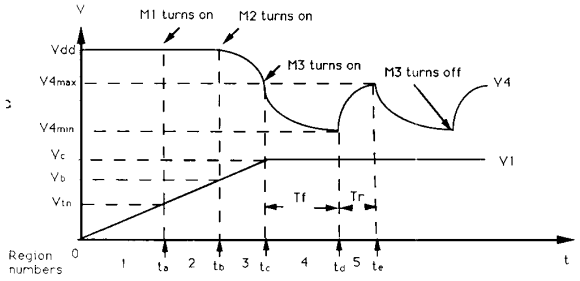
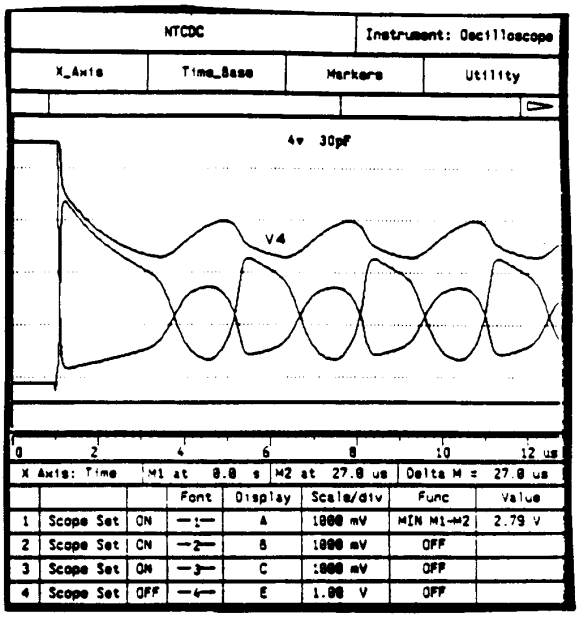


Fig. 3. Input ($V1$)–output ($V4$) relation of NTC.

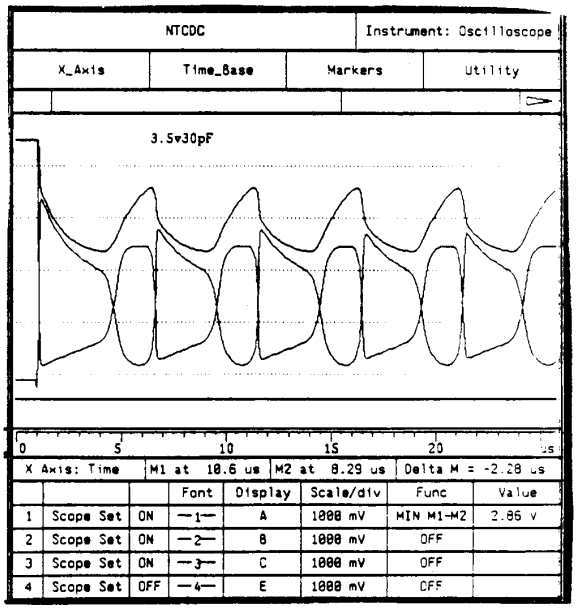
TABLE I
DIFFERENT REGIONS DEPENDING ON THE OPERATING CONDITION

	Region Numbers				
	1	2	3	4	5
		LIN	LIN		LIN
M1	OFF	SAT	SAT	LIN	SAT
M2	OFF	OFF	SAT	LIN	SAT
M3	OFF	OFF	OFF	LIN	OFF

"LIN" denotes linear operation, "SAT" saturation operation, and "OFF" cutoff.



(a)



(b)

Fig. 4. SPICE output of NTC with two constant inputs: (a) 4 V and (b) 3.5 V.

three resistors ($R4$–$R6$), and a load capacitor (C). The NTC can simply be viewed as a kind of analog voltage-controlled oscillator which is operating based on nonlinear hysteresis characteristics [13], [14]. In response to input voltage, it generates a stream of analog waves (not digitized pulses) whose shapes are controlled by the input voltage level. Vdd denotes the source (of 5 V) power. $V1$ is the voltage of the input node and $V3$ is the voltage of the output node. As the input increases above a certain level, the output signals will start to oscillate with a frequency which is approximately proportional with the input level (stimuli). Simulation results for this cell are shown in Fig. 2. Fig. 3 shows the behavior of the voltages $V1$ and $V4$ separately with respect to time for a saturating-ramp input. Note that $V4$ is chosen, rather than $V3$, because of its advantages for analysis of the circuit. But knowing the behavior of $V4$, we know the behavior of the

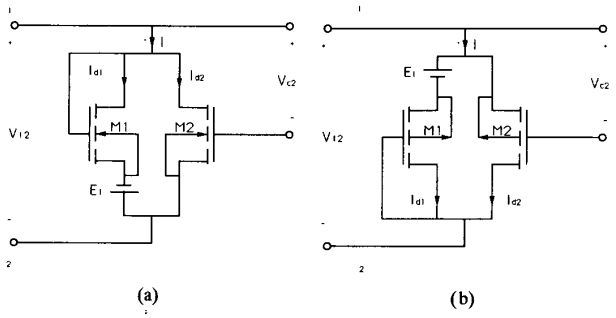


Fig. 5. Enhancement-mode (a) NMOS and (b) PMOS linear resistors.

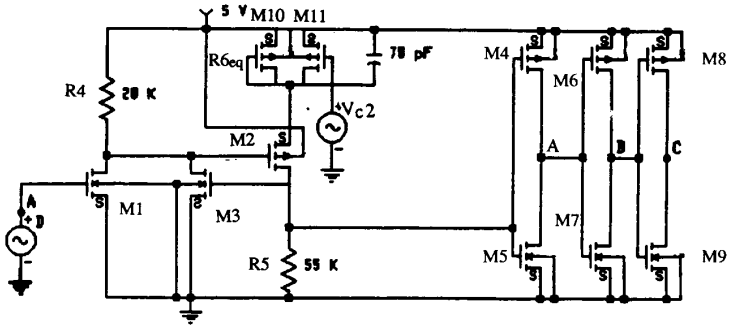


Fig. 6. NTC with an MOS resistor and buffered inverters.

output $V3$. Here, we partition the signals into five important regions along the time axis. Table I lists the different operational conditions for transistors $M1$, $M2$, and $M3$ of Fig. 1. From these operational conditions, we can derive the approximate analytical equations for the oscillation frequency of the NTC [13] as explained below. These equations are derived by using standard current equations [17] for transistors in linear, saturation, and cutoff modes of operation as cataloged in Table I. Nonideal effects, such as device mismatches, temperature variations, and parameter shifts, are not taken into consideration here since these have little effect on the circuit operation while these will make analysis much more complicated.

Assume that input $V1$ is increasing in time. Until it reaches the threshold voltage of the transistor $M1$ (V_{tn}), $M1$ stays cut off, causing the voltage $V2$ to be in the high (Vdd) state. Transistors $M2$ and $M3$ will also be cut off. From this, we get region 1 in Table I, where the three transistors are all cut off. At the time $t = t_a$, the input voltage $V1$ reaches the threshold voltage of $M1$. As a result, $M1$ turns on, and current starts to flow through $R4$ and $M1$. This induces a certain amount of voltage drop across $R4$, forcing $V2$ to decrease from the high state. However, at this instant, the gate-to-source voltage of the p-type transistor, $M2$, is still smaller in magnitude than its threshold level; thus, $M2$ stays in the cutoff region. So we get the second region shown in Table I. As the input increases further, more voltage difference will appear at the gate-to-source terminals of $M2$, and this will eventually cause $M2$ to turn on. At this time, the output $V4$ will start to fall (see Fig. 3). At the time that $M2$ turns on, the turn-on condition, $V_{gs,2} = V_{tp}$, gives the corresponding input voltage level as

$$V1 = V_b = V_{tn} + \sqrt{\frac{(-2V_{tp})}{K_n(W1/L1)R4}} \tag{1}$$

where K_n is a process gain factor of n-type transistors [18], and V_{tn} and V_{tp} are threshold voltages for n-type and p-type transistors, respectively, all of which are assumed to be enhancement mode. W/L represents the geometric size of each transistor.

In region 3, as the input voltage increases, the gate-to-source voltage of the transistor $M2$ also increases, and this will yield more drain current through $M2$. Hence, the voltage drop across the resistor $R5$ increases until it reaches the threshold level of the feedback transistor $M3$. At the time that $M3$ turns on ($t = t_c$), $V4$ starts to oscillate, which is triggered by the hysteresis characteristic of the cell [19], [20]. At this moment, $t = t_c$, $V4$ is at its highest point in the pulsing, or oscillation, mode. This high value, $V_{4\max}$, which depends on the bias condition of $M1$, can be found by equating the voltage drop across $R5$ to V_{tn} of $M3$. There are two cases, as follows;

Case 1: $M1$ is in saturation region:

$$V4_{\max} = Vdd - \left[\frac{KnW1}{2L1}(V1 - V_{tn})^2\right] * R4 - V_{tp} + \left(\frac{2V_{tn}}{K_pR5}\frac{L2}{W2}\right)^2. \tag{2}$$

Case 2: $M1$ is in linear region: $V4_{\max}$ is given by (3) [shown at the bottom of the page], where K_p is a process gain factor of p-type transistors [18].

In region 4, as the feedback transistor turns on, a surge of current will flow through $M3$, resulting in a large voltage drop across $R4$. Thus, transistor $M1$ will be operating in its linear region with small drain-to-source voltage. This will increase the gate-to-source voltage of $M2$, forcing it to operate in its linear region ($|V_{ds}| < |V_{gs}| - |V_t|$). We can, therefore, regard transistor $M2$ as operating as a simple ohmic device (resistor) whose equivalent resistance is approximately represented by [18]

$$R_{\mathrm{eq},M2} = \left[K_P \frac{W2}{L2}(|V_{gs2,\mathrm{ave}}| - |V_{tp}|)\right]^{-1} \tag{4}$$

where $V_{gs2,\mathrm{ave}}$ is an average value for the gate-to-source voltage in region 4.

During this period in region 4, the load capacitor, C, will slowly accumulate charge and the drain current through $M2$ will decrease. Therefore, $V3$ will fall until it reaches V_{tn} again. By replacing $M2$ with an equivalent resistor, we can get an RC time constant for the output stage where R is the Thevenin equivalent approximate resistance seen by C. Using this time constant, and by the condition $V_3 = V_{tn}$, we get an approximate falling transient time, T_f, as

$$T_f = -RC \ln\left\{\frac{V_{tn} - \frac{R5*Vdd}{R_{\mathrm{eq},M2}+R5+R6}}{\frac{R5*V4_{\max}}{R_{\mathrm{eq},M2}+R5} - \frac{R5*Vdd}{R_{\mathrm{eq},M2}+R5+R6}}\right\} \tag{5}$$

where

$$R = (R_{\mathrm{eq},M2} + R5) * R6/(R_{\mathrm{eq},M2} + R5 + R6). \tag{6}$$

Next we find $V4_{\min}$ by using the exponentially decaying $V4(t)$ and putting $t = T_f$:

$$V4_{\min} = (R_{\mathrm{eq},M2} + R5) * \left\{\left(\frac{V4_{\max}}{R_{\mathrm{eq},M2}+R5} - \frac{Vdd}{R_{\mathrm{eq},M2}+R5+R6}\right) \cdot \exp\left(-\frac{T_f}{RC}\right) + \frac{Vdd}{R_{\mathrm{eq},M2}+R5+R6}\right\}. \tag{7}$$

In region 5, when $M3$ turns off, the circuit configuration again becomes the same as in the previous region 3, and $V4$ starts to rise until it hits $V4_{\max}$. Since $M2$ is now in its saturation region, a current source can replace it for circuit simplification. So, using the above two values $V4_{\max}$ and $V4_{\min}$, as end points, and applying KCL at the node $V4$, we can derive first order differential equation for $V4$. Solving for the approximate rising transient time, Tr:

$$T_r = -C * R6 \ln\left(\frac{V4_{\max} - Vdd + I2_{\mathrm{ave}} * R6}{V4_{\min} - Vdd + I2_{\mathrm{ave}} * R6}\right) \tag{8}$$

where the current source value

$$I2_{\mathrm{ave}} = \frac{K_p W2}{2L2}(|V_{gs2,\mathrm{ave}}| - |V_{tp}|)^2. \tag{9}$$

From the above, the frequency of oscillation is now given by adding (5) and (8):

$$f = 1/T = 1/(T_r + T_f). \tag{10}$$

Thus we can say that the frequency varies with input stimuli $V1$, via (2) or (3). Note also there should be some ranges

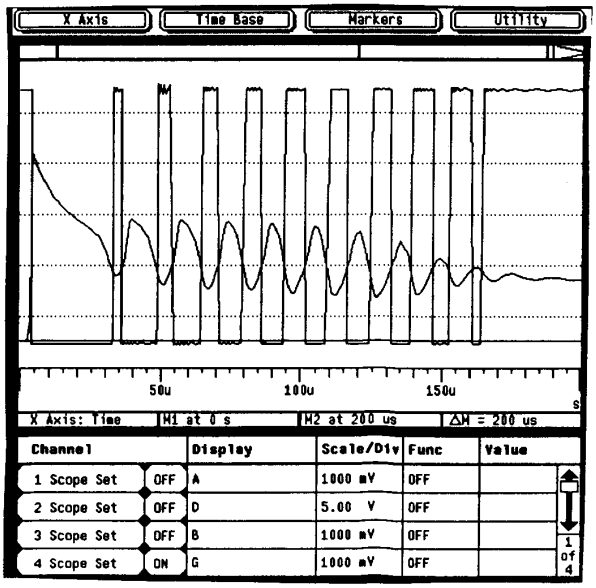


Fig. 7. Example of PDC variation for V_{c2} varying linear in time from -8 V to -20 V.

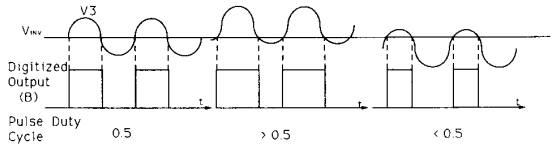
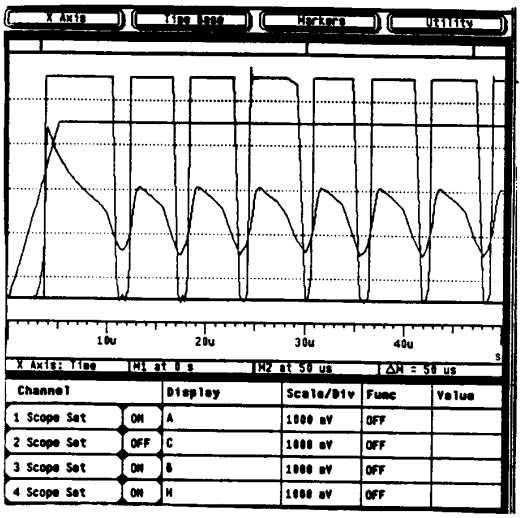
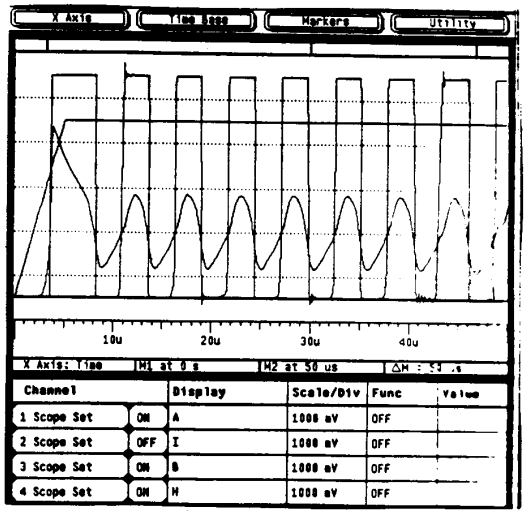


Fig. 8. Pulse duty cycle modulation.

$$V4_{\max} = \frac{\frac{1}{R4} + K_n\frac{W1}{L1}(V1 - V_{tn}) - \left\{\left[\frac{1}{R4} + K_n\frac{W1}{L1}(V1 - V_{tn})\right]^2 - 2K_n\frac{W1Vdd}{L1R4}\right\}^{1/2}}{K_n\frac{W1}{L1}} - V_{tp} + \left(\frac{2V_{tn}}{K_pR5}\,\frac{L2}{W2}\right)^2 \tag{3}$$



(a)



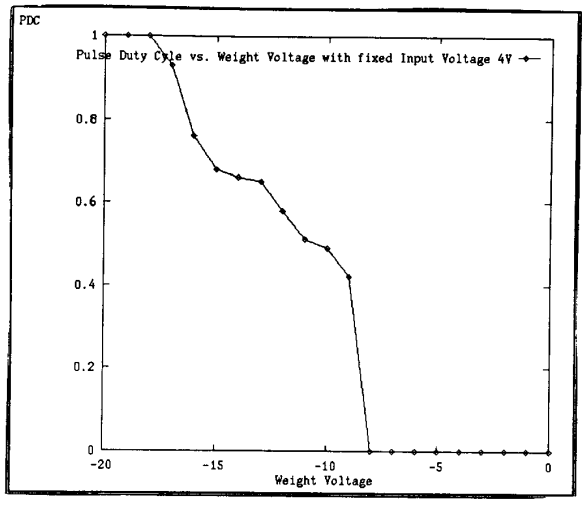
(b)

Fig. 9. Pulse duty cycle changes of node B in Fig. 6 with (a) $V_{c2} = -11.25$ V and (b) $V_{c2} = -10$ V when $V1 = 4$ V.

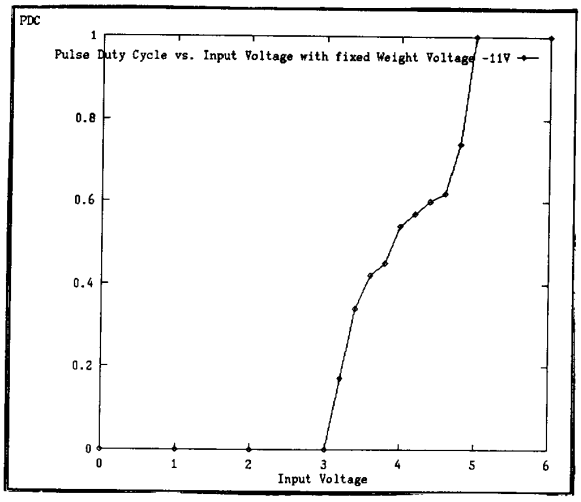
for $V1$ to satisfy an inequality condition GND $< V4_{\min} < V4_{\max} < Vdd$ [13].

Fig. 4 shows simulation results of the NTC with two fixed input voltages: (a) 3.5 V and (b) 4 V. The simulation results of (b) is used in order to compare and check the analytical equations above. For this we used the values with MOSIS parameters [21]:

$$Vdd = 5 \text{ V}, \quad Kn = 100\text{e} - 6, \quad Kp = 50\text{e} - 6,$$
$$W1/L1 = 8/8 \ (\mu\text{m}), \quad W2/L2 = 48/8,$$
$$W3/L3 = 20/8, \quad R4 = 20 \text{ K},$$
$$R5 = 55 \text{ K}, \quad R6 = 90 \text{ K},$$



(a)



(b)

Fig. 10. PDC versus (a) weight voltage and (b) input voltage from the second inverter.

$$C = 30\text{e} - 12, \text{ and input } V1 = 4 \text{ V}.$$

SPICE gives $Vb = 1.4$, $V4_{\max} = 3.49$, $V4_{\min} = 2.79$, $T_f = 1.4$ μs, $T_r = 1.5$ μs, and period $= 2.9$ μs (see Fig. 4). By comparison, the analytical equations (1), (2), (3), (5), and (8) give $Vb = 1.6$, $V4_{\max} = 3.03$, $V4_{\min} = 1.95$, $T_f = 2.0$ μs, $T_r = 0.6$ μs, and period $= 2.6$ μs. Thus, the analytical equations do give reasonable numbers for initial design purposes, as is seen by comparing with the results of the SPICE simulations.

III. Voltage-Controlled CMOS Resistors [16]

From the analytic equations in the previous section, we know that the output pulse shape (width and frequency) is described as a function of input voltages (stimulus), the shape factors of the three transistors, and three resistance values. If

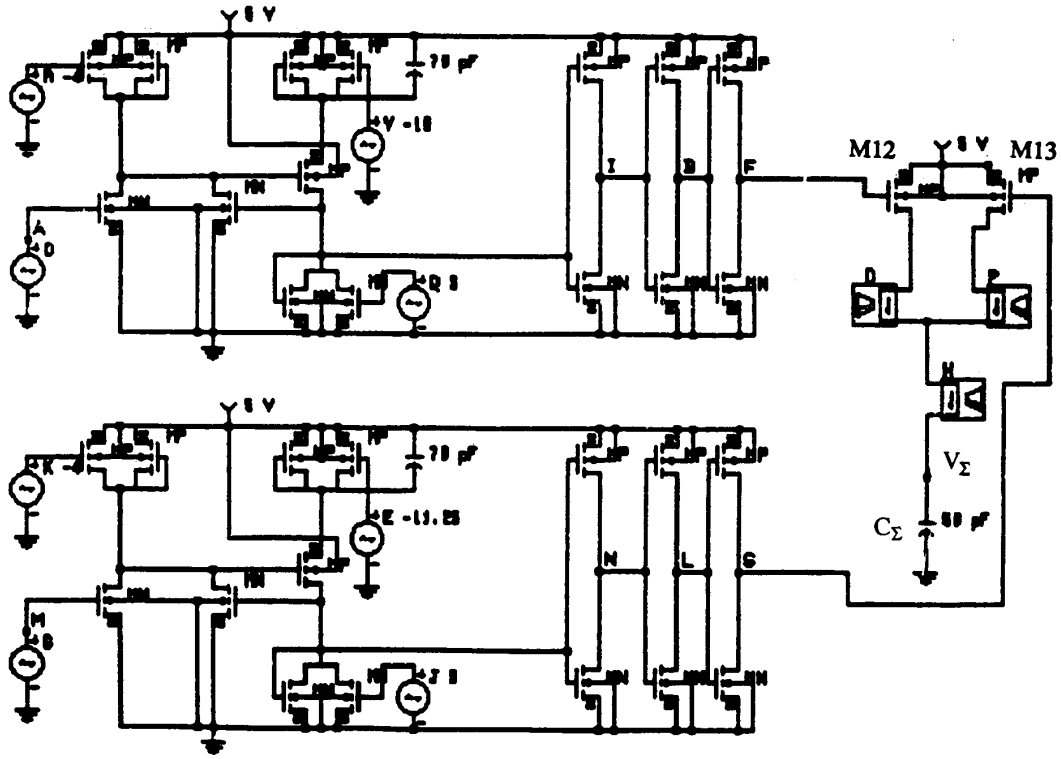


Fig. 11. CMOS circuit diagram for the summation of two excitatory weighting signals.

we replace the passive resistors with MOS voltage-controllable ones, then we will be able to control output oscillating waves simply by changing the resistance value. In this way, not only the input voltage but also the variable conductance value of the MOS resistor will determine the output waveforms. We consider this MOS resistor as a synaptic junction where weighting is controlled by the control voltage for the resistor as explained in the next section.

Several ways of realizing voltage-controlled resistance have recently been proposed by various authors [22]–[26]. The main idea in their work is to cancel the effect of nonlinearities by manipulating the sum or difference of the current using two or more MOS devices. However, the techniques have limitations, in the sense that some are applicable only for NMOS technology [22], [23] with depletion-type transistors, some have limited dynamic ranges [23], [25], and others have complex auxiliary circuitry for bias [24], [26]. In order to comply with today's popular CMOS technology with relatively large dynamic range, a new technique was proposed in [16].

Consider two enhancement-mode MOS transistors connected as shown in Fig. 5, with independent voltage source E_1. An NMOS structure is shown in (a) and a PMOS is in (b). The PMOS structure works in the same way as the NMOS one, except with opposite polarities. In this case $M1$ in the NMOS resistor is operating in its saturation region owing to its gate-to-drain connection. Transistor $M2$ will be operating in its linear region if we apply a suitably large voltage V_{c2} (larger than V_{12}) on its gate-to-source terminal. If the second-order terms of the two drain current are adjusted to be the same but with different sign, by simply using the same shape factor, $W1/L1 = W2/L2$, for both transistors, then the total current, $I = I_{d1} + I_{d2}$, will be

$$I = \begin{cases} K\left\{(V_{c2} - V_{tn})V_{12} - \frac{V_{12}^2}{2}\right\}, \\ \qquad 0 \le V_{12} \le V_{tn} - E_1 \\ K\left\{(V_{c2} - 2V_{tn} + E_1)V_{12} + \frac{(V_{tn} - E_1)^2}{2}\right\}, \\ \qquad V_{tn} - E_1 \le V_{12} \le V_{c2} - V_{tn} \end{cases} \tag{11}$$

where V_{tn} is the threshold voltage of n-type transistors and equal K factors. $K = K_n(W1/L1) = K_n(W2/L2)$ can be taken for granted assuming two matching MOS devices with no process parameter variations between two closely placed transistors. K_n is the process gain factor, as denoted in Section II. Note that the dynamic range of this structure can be controlled by choosing suitable values of V_{c2} and E_1 depending on the range of the applied voltage V_{12} across the resistance in the circuit. The most convenient value of $E_1 = 0$ V is chosen in the linear region; the equivalent resistance will be

$$R_{\text{eq}} = \frac{1}{\{K(V_{c2} - 2V_{tn})\}} \tag{12}$$

which is controlled by the gate voltage V_{c2}.

In Section IV, we will describe how to use the above voltage-controlled resistor as a synaptic junction by replac-

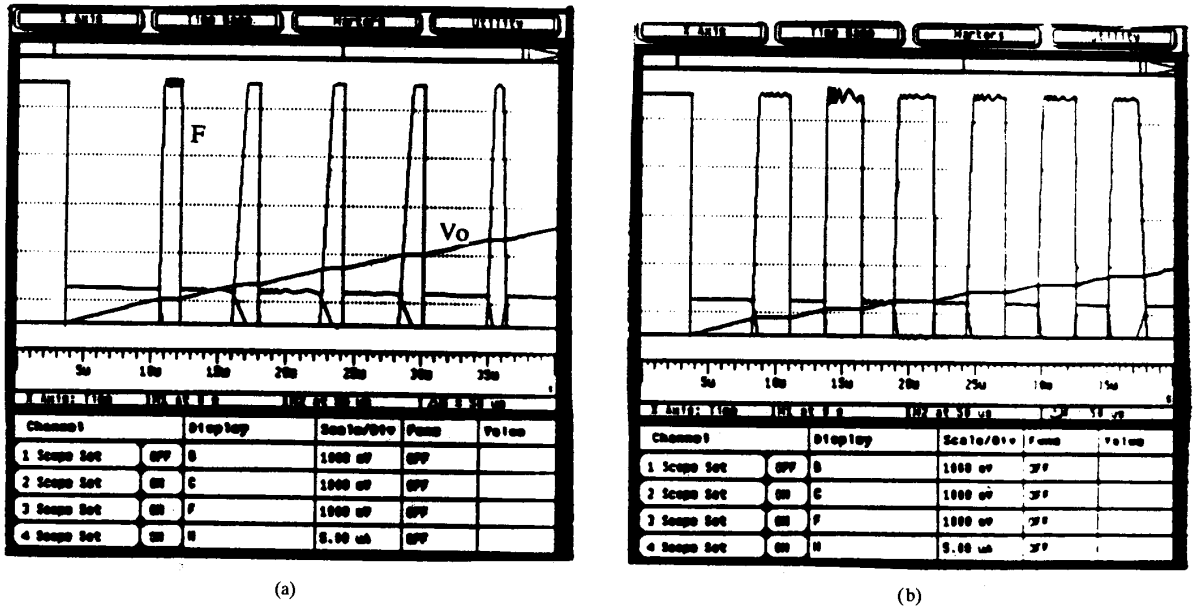


Fig. 12. Charge accumulation for Fig. 11 with the upper NTC acting alone and two values of weighting control voltage: (a) $V_{c2} = -11.25$ V; (b) $V_{c2} = -10$ V.

ing $R6$ in Fig. 1. This will result in the pulse duty cycle modulation technique described below.

IV. Synaptic Neuron—Weighting and Summing

Fig. 6 shows the NTC circuit of Fig. 1 with $R6$ replaced by the PMOS voltage-controlled linear resistor of Fig. 5. Notice that $R6$ is replaced with the PMOS resistor ($M10$, $M11$) without voltage source E_1 ($E_1 = 0$ for circuit simplicity). Three inverter stages ($M4$–$M9$) are also added to the NTC output. The first inverter ($M4$, $M5$) is for standardizing the output pulse wave signal into a square wave with high value of Vdd (5 V) and a low of GND (0 V). When the output pulse wave signal is larger than the first inverter's threshold level, $V_{\rm INV}$ [27], designed as 1.5 V by the W/L ratios of $M4$ and $M5$, the inverter digitizes the pulse wave signal as GND and otherwise as Vdd. The second and the third inverters are for large driving capability for local or long-distance transmission in the network.

As the gate control voltage V_{c2} of $R6$ changes, the resistance value of $R6$ will also change. This is clear from (12), assuming that transistor gain factor K and threshold voltage V_{tn} are constant. The equivalent resistance value of $R6_{\rm eq}$ is inversely proportional to the gate control voltage, V_{c2}. Thus, according to (5)–(10) in Section II, the change of V_{c2} results in the modifications of the shape of the voltage $V4$ through the change of $R6_{\rm eq}$. Note that V_{c2} also affects the output voltage $V3$ because the output stage of the NTC can be viewed as a simple voltage divider where three devices ($R6_{\rm eq}$, $M2$, and $R5$) are connected in series. Consequently, the dc offset voltage of $V3$, around which the signal will swing, can then be approximately expressed by

$$V_{\rm offset} = R5/(R6_{\rm eq} + R_{{\rm eq},M2} + R5) \tag{13}$$

for operation in the linear region of the transistor $M2$. In (13) $R6_{\rm eq}$ is the equivalent resistance value of (12), and $R_{{\rm eq},M2}$ is the small-signal equivalent resistance of the NTC transistor $M2$ of Fig. 1, as described in (4). It follows from the above equation that a decrease of the value of the equivalent resistance $R6_{\rm eq}$, caused by an increase of V_{c2}, will yield an increase of the offset voltage, $V_{\rm offset}$. Now, if we transform $V3$ through the first inverter, we will obtain at node A a digitized pulse stream whose shape is controllable by V_{c2} of $R6_{\rm eq}$.

We define in this work the pulse duty cycle (PDC) of a pulse stream over a (possibly variable) time interval T as

$$\mathrm{PDC}(T) = \frac{\sum_{i=1}^{n} \mathrm{PW}(i)}{T} \tag{14}$$

where $\mathrm{PW}(i)$ is the ith pulse width in the stream, assumed to be of n pulses, in time interval T. This is for a pulse stream analogous to the duty cycle of a single pulse [30]. As a result of (14), $0 \leq \mathrm{PDC}(T) \leq 1$ and PDC represents the average value of pulses in the time period T. Thus we can say that the closer PDC is to 1, the denser a pulse stream is. In order to have a meaningful value of PDC for a given pulse stream, the time T should be chosen larger than any of the $\mathrm{PW}(i)$.

As seen in Fig. 7 the resultant pulse streams of an NTC, Fig. 6, will have different PDC values with different V_{c2}; thus, we call this technique pulse duty cycle modulation (PDCM). Fig. 8 illustrates how this technique works. Notice that as the offset voltage of $V3$ increases with a given inverter logic threshold voltage, the PDC monotonically increases. This technique has its advantages in terms of signal processing, because there is no need for premanipulation or preprocessing for signal modulations, in contrast to the conventional pulse modulation techniques [5], [29], i.e., pulse frequency

modulation (PFM), pulse amplitude modulation (PAM), pulse duration or density modulation (PDM), and pulse position modulation (PPM).

From the above considerations, the weighting process can be controlled by the gate voltage of the equivalent resistor $R6_{eq}$; consequently, this gate voltage determines the PDC of the output pulse through the change of the offset voltage, V_{offset}. This results in a monotonic change of PDC with respect to V_{c2}. From (12) and (13), if we arbitrarily choose maximum and minimum values of V_{c2} as bounds for allowed weight change, then this structure works as an electronic neuron with controllable synaptic weight. Fig. 9 shows the simulation results of these operations for different gate control voltages, i.e., different $R6_{eq}$ values. The waveforms of $V3$ and the corresponding pulse stream from node B of Fig. 6 are traced for two different V_{c2}'s when the same input (4 V) is applied to both cases. Notice that the V_{offset} for $V_{c2} = -11.25$ of (a) is higher than that for $V_{c2} = -10$ of (b) because of the smaller equivalent resistance value of $R6_{eq}$ (see (12)). As a result, higher PDC is acquired with the more negative V_{c2} of (a). This shows that we can use the gate control voltage of the equivalent resistor $R6_{eq}$ as a tool for modifying the PDC. Fig. 10 shows the variations of PDC on node B in Fig. 6 with respect to (a) weight voltage (V_{c2}) and (b) input voltage ($V1$) as found from SPICE simulations. Notice that PDC is changing monotonically with respect to these two variables. Although PDC is not exactly linear relative to either $V1$ or V_{c2}, the configuration in Fig. 6 will enable us to have a weight control in a very simple way without using a multiplier for weight multiplication. Besides, in this technique, we are free of the need for synchronization, chopping clocks, or pulse width manipulation, as required for certain other pulse coded neural networks [5]–[7]. It takes full advantage of the inherent analog property of the NTC and converts the output signal into standardized form for ease of handling. It is expected, therefore, that this structure will consume relatively small area and have a large noise margin owing to its pulse-based operation.

For the summation of several weighting cell signals, we simply add signal charges into a capacitor, C_Σ of Fig. 11, where several cells are connected in parallel through switching transistors ($M12$, $M13$). The gate nodes for these switching transistors are tied to the individual outputs of the NTC's so that these NTC output voltages are converted into currents via switching transistor's transconductances to be summed in the capacitor. The total amount of charge versus time is then a representation of PDC versus time for the system. The switching transistors are operating either in their saturation region, when the gate voltage is 0 V, or in their cutoff region, when the gate voltage is 5 V. In order to design these transistors and the capacitor, we desire a large time constant, leading to a linear charging of the capacitor. Specifically, the switching transistors' sizes and the capacitor value are designed to give a time constant which is much greater (by a factor of at least 10) than the maximum pulse width. Fig. 12 shows simulation results for two different excitatory weightings. Notice that charge is accumulated into the capacitor via the switching transistors and that the slope of the capacitor voltage, V_Σ, versus time is inversely proportional to the PDC value because p-type transistors are used for the switches. An inhibitory function can also be realized by an NMOS switching transistor tied to ground to subtract (discharge) charges from the capacitor. Finally, Fig. 13 shows V_Σ as the summation of the two weighting signals of the NTC's in Fig. 11 with the upper $V_{c2} = -11.25$ and the lower $V_{c2} = -10$, simultaneously, as per Fig. 12. Note that the slope of the summation signal is the sum of each slope in Fig. 12.

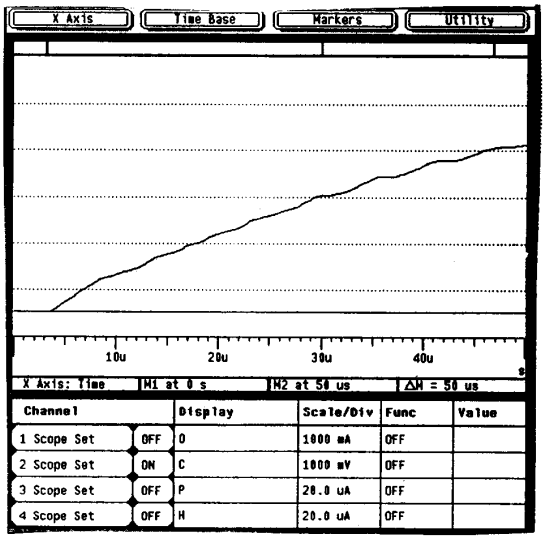


Fig. 13. Current summation for Fig. 11 with input of 4 V to both NTC's and $V_{c2,upper} = -11.25$ V and $V_{c2,lower} = -10$ V (as per Fig. 12).

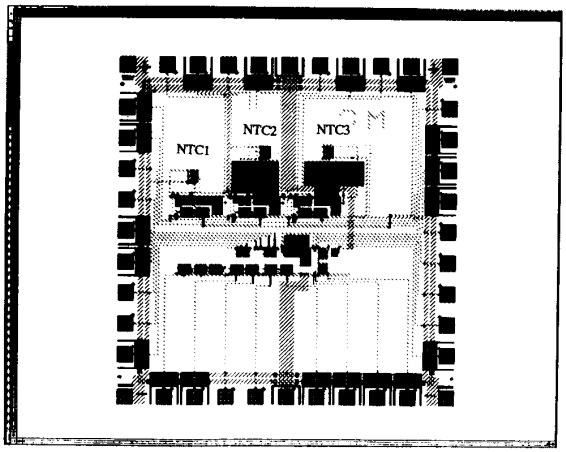
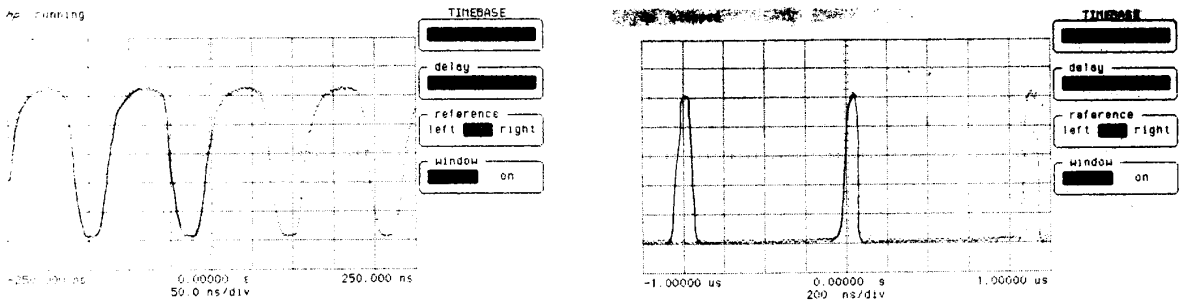


Fig. 14. Chip layout.

V. VLSI Implementation—Chip Design and Measurements

The circuit for the NTC's of Fig. 11 was designed and integrated via MOSIS [21] using the silicon gate p-well CMOS technology with 2 μm minimum feature size. Not only the resistor $R6_{eq}$ but also $R4$ and $R5$ are implemented through



(a) Vc2=-13.5V (b) Vc2=-14.0V

Fig. 15. Chip measurements; output pulses from chip with different weighting signals. (a) $V_{c2} = -13.5$ V. (b) $V_{c2} = -14.0$ V.

the voltage-controlled resistor, and as a result we have more flexibility of control over the circuit operation. The process gain factors, K, of the pull-up and pull-down of the first inverter ($M4$, $M5$) are adjusted so that its inverter logic threshold voltage is 1.5 V, which is as close as possible to the offset voltage of the signal $V3$. The first inverter is also designed close to minimum size in order to detect the $V3$ signal with minimum loading effect. The second inverter ($M6$, $M7$) is an intermediate buffer, and the last inverter ($M8$, $M9$) is designed to have relatively large size so that it can drive other neurons through the adaptive feedback neural network connections present in most neural networks. The chip layout is shown in Fig. 14. Along with a set of basic test components, the chip contains three test NTC's and their associated inverters (NTC1–NTC3) with different circuit sizes; each occupies about 100 μm × 150 μm. As a basic building block, each cell is designed to be square with the same height for future integrations. The chip is a Tiny Chip [21] with 40 pins and the die size is 2250 μm × 2220 μm. Fig. 15 shows the measurements of oscillation from NTC2 of the chip with two different weighting control signals: (a) $V_{c2} = -13.5$ V and (b) $V_{c2} = -14.0$ V. The PDC was measured from the output of the third inverter and is (a) 0.65 for $V_{c2} = -13.5$ and (b) 0.11 for $V_{c2} = -14$ when $V1 = 4.5$. Since we used the output of the third inverter we should use $1 -$ PDC of Fig. 10, in which we see that case (b) should have less PDC than case (a), as it does. However the simulation results show only the qualitative trend.

VI. Conclusions and Discussions

In this paper we have developed the hardware to realize weighting and summing signals in our pulse-coded neural networks. In particular, we have presented a novel design and implementation of synaptic weighting and summing using a neural-type cell (NTC). The NTC has its merits in hardware implementation owing to its simple structure, small size, and high speed of operation (μs order of oscillation). Weighting is controlled by control voltages of equivalent resistors composed of two enhancement-mode MOS transistors. Although the weighting process is not exactly linear, as seen from Fig. 10, it provides a very simple structure for the weighting. The pulse duty cycle modulation technique is introduced for extracting information from a stream of pulses. In this way, one processing element, an NTC and its inverters, consists of 15 transistors and a capacitor. Summation of weighting signals is accomplished by the use of a single capacitor as charge integrator fed by several NTC's. With its small size, this technique can be applied to adaptive pulse coded neural networks with learning capabilities on the chip. The chip realized is of relatively large size since it is a test chip. Currently the error correction circuits with gradient descendant method are under investigation for adaptive feedback. Other choices for this error correction might use a simple charge transfer circuit, such as that in [28]. The NTC along with inverters is a type of VCO for which the output duty cycle as a mathematical function of the control voltage is yet to be determined. In fact, we do not treat here the mathematical theory of pulse-coded neural networks. The use of the NTC for realizing neural networks where the information is contained in the pulse duty cycle is in its infancy. Thus, there are many open problems, one key of which is the development of a suitable mathematical theory.

Acknowledgment

The authors wish to thank the reviewers for their careful reading and many valuable comments.

References

[1] D. Corso and L. Reyneri, "Mixing analog and digital techniques for silicon neural networks," in *Proc. IEEE ISCAS* (New Orleans, LA), May 1990, pp. 2446–2449.
[2] J. Meador, A. Wu, C. Cole, N. Nintunze, and P. Chintrakulchai, "Programmable impulse neural circuits," *IEEE Trans. Neural Networks,* vol. 2, pp. 101–109, Jan. 1991.
[3] A. F. Murray and A. W. Smith, "Asynchronous VLSI neural networks using pulse stream arithmetic," *IEEE J. Solid State Circuits,* vol. 23, pp. 688–697, 1988.
[4] A. Murray, "Pulse arithmetic in VLSI neural network," *IEEE MICRO,* vol. 9, pp. 64–74, Dec. 1989.
[5] J. E. Tomberg and K. Kaski, "Pulse-density modulation technique in VLSI implementations of neural network algorithms," *IEEE J. Solid State Circuits,* vol. 25, pp. 1277–1286, Oct. 1990.
[6] A. Murray and A. Smith, "Asynchronous arithmetic for VLSI neural systems," *Electron. Lett.,* vol. 23, no. 12, pp. 642–643, June 1987.

[7] A. Murray, D. Corso, and L. Tarassenko, "Pulse-stream VLSI neural networks mixing analog and digital techniques," *IEEE Trans. Neural Networks,* vol. 2, pp. 193–204, Mar. 1991.
[8] E. Sanchez-Sinencio, "Neural network circuit implementations" (Guest Editorial) *IEEE Trans. Neural Networks,* vol. 2, pp. 192, Mar. 1991.
[9] C. Cole, A. Wu, and J. Meador, "A CMOS impulse neural network," in *Proc. Colorado Microelectronics Conf.* (Colorado Springs, CO), Mar. 1989, pp. 16–24.
[10] R. W. Newcomb, "Neural-type microsystems circuit status," in *Proc. IEEE ISCAS* (Newport Beach, CA), 1983, pp. 97–100.
[11] R. W. Newcomb, "MOS neuristor lines," in *Constructive Approaches to Mathematical Models,* Cliffman and G. Fix, Eds. New York: Academic Press, 1979, pp. 87–111.
[12] G. Moon, M. Zaghloul, and R. Newcomb, "IC layout for an MOS neural-type cell," in *Proc. 32nd Midwest Symp. Cir. Syst.* (Urbana-Champaign, IL), Aug. 1989, pp. 482–484.
[13] G. Moon, "VLSI implementation of a neural-type cell," Master's thesis, Dept. of Electrical Engineering and Computer Science, George Washington University, Sept. 1990.
[14] G. Moon, M. Zaghloul, M. Savigny, and R. Newcomb, "Analysis and operation of a neural-type cell," in *Proc. IEEE ISCAS* (Singapore), June 1991.
[15] R. Lippmann, "An introduction to computing with neural nets," *IEEE ASSP Magazine,* pp. 4–22, Apr. 1987.
[16] G. Moon, M. Zaghloul, and R. Newcomb, "An enhancement-mode MOS voltage-controlled linear resistor with large dynamic range," *IEEE Trans. Circuits Syst.,* vol. 37, pp. 1284–1288, Oct. 1990.
[17] R. L. Geiger, P. E. Allen, and N. R. Strader, *VLSI Design Techniques for Analog and Digital Circuits.* New York: McGraw-Hill, 1990, pp. 143–158.
[18] J. Mavor, M. Jack, and P. Denyer, *Introduction to MOS LSI Design.* Reading, MA: Addison-Wesley, 1982, pp. 29–39.
[19] N. El-leithy and R. W. Newcomb, "Hysteresis in neural-type circuits," in *Proc. IEEE ISCAS* (Espoo, Finland), June 1988, pp. 993–996.
[20] M. Savigny, G. Moon, M. Zaghloul, N. El-Leithy, and R. Newcomb, "Hysteresis turn-on-off voltages for a neural-type cell," in *Proc. 33rd IEEE Midwest Symp. Circ. Syst.* (Calgery. Canada), Aug. 1990, pp. 37–40.
[21] *MOSIS User Manual,* The Information Science Institute of the University of Southern California USC/ISI in Marina del Rey, CA, 1988.
[22] J. Babanezhad and G. Temes, "A linear NMOS depletion resistor and its application in an integrated amplifier," *IEEE J. Solid State Circuits,* vol. SC-19, no. 6, pp. 932–938, 1984.
[23] I. Han and S. B. Park, "Voltage-controlled linear resistor by two MOS transistors and its application to active *RC* filter MOS integration," *IEEE Trans. Circuits Syst.,* vol. 72, pp. 1655–1657, Nov. 1984.
[24] K. Nay and A. Budak, "A voltage-controlled resistance with wide dynamic range and low distortion," *IEEE Trans. Circuits Syst.,* vol. CAS-30, pp. 770–772, Oct. 1983.
[25] Y. Tsividis, "Continuous-time MOSFET-C filters in VLSI," *IEEE J. Solid State Circuits,* vol. SC-21, pp. 15–30, Feb. 1986.
[26] M. Ismail, S. Smith, and R. Beal, "A new MOSFET-universal filter structure for VLSI," *IEEE J. Solid State Circuits,* vol. 23, pp. 183–194, Feb. 1988.
[27] D. Puckness and K. Eshraghian, *Basic VLSI Design.* Englewood Cliffs, NJ: Prentice-Hall, 1988, pp. 32–33.
[28] F. Ibrahim and M. Zaghloul, "Design of modifiable-weight synapse CMOS analog cell," in *Proc. IEEE ISCAS* (New Orleans, LA), 1990, pp. 2975–2978.
[29] A. Carlson, *Communication Systems.* New York: McGraw-Hill, 1986, pp. 308–314.
[30] D. Bell, *Solid-State Pulse Circuits.* Reston, 1988, p. 8.
[31] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Nat. Acad. Sci. U.S.,* vol. 79, pp. 2554–2558, Apr. 1982.

Gyu Moon (S'90) received the B.S. degree in control and instrumentation engineering from the Seoul National University, Korea, in 1982 and the M.S. degree in electrical engineering from the George Washington University, Washington, DC, in 1990. He is currently pursuing the Ph.D. degree in the Department of Electrical Engineering and Computer Science at GWU.

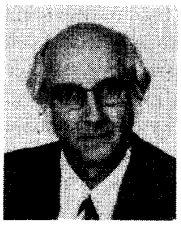
From 1982 to 1988 he worked for the Electronics and Telecommunications Research Institute, Korea, as a member of technical staff in the fields of VLSI design/CAD. During that period, he spent a year and half at VLSI Technology Inc., San Jose, CA, as a visiting ASIC design engineer. Since 1988, he has been a graduate teaching and research assistant at GWU. His research interests include VLSI/ASIC design, CAD/CAE, and neural networks applications.

Mr. Moon is the recipient of a Korean American Scholarship Award and a Hyundai Scholarship Award.

Mona E. Zaghloul (M'81–SM'85) received the B.Sc. degree in electrical engineering from Cairo University, Egypt, in 1965. She then received the M.A.Sc. degree in electrical engineering, the M.Math degree in applied analysis and computer science, and the Ph.D. degree in electrical engineering from the University of Waterloo, Waterloo, Ont., Canada, in 1970, 1971, and 1975, respectively.

From 1975 to 1980, she held positions at Aalborg University, Denmark, and the Computer Sciences Corporation, U.S. In 1980 she joined the George Washington University, Washington, DC, where she has been engaged in teaching and research on nonlinear circuit theory, integrated circuit analysis and design, computer-aided design and testing of VLSI circuits, and the VLSI implementation of neural networks. She is a Full Professor in the Department of Electrical Engineering and Computer Science. She has published more than 70 articles in journals, conference proceedings, and book chapters.

Dr. Zaghloul is the general chairman of the 35th Midwest Symposium on Circuits and System, which will be held in Washington, DC, in August 1992.



Robert W. Newcomb (S'52–M'56–F'72) was born in Glendale, CA, in June 1933. He obtained the B.S.E.E. from Purdue in 1955, the M.S. in electrical engineering from Stanford in 1957, and the Ph.D. from the University of California, Berkeley, in 1960.

After serving on the tenured faculty at Stanford, he moved to the University of Maryland to update the graduate program. He has held visiting appointments in Belgium, Malaysia, and Spain and presently directs the Microsystems Laboratory at the University of Maryland. He has been a member of a number of IEEE groups, including the Neural Networks Council and the Society for Social Implications of Technology. In the 1960's he began research on micromotor fabrication and pulse coded neural network circuit design. His recent research has concentrated upon circuit realizations of neurophysiologically realistic neural-type systems, semistate theory and its use in nonlinear and neural systems design, and the determination of autoacoustic emission parameters for the ear.