An Improved Neural Processing Element Using Pulse Coded Weights

G. Moon+, M. E. Zaghloul+, and R. W. Newcomb*

+Department of Electrical Engineering
and Computer Science
The George Washington University
Washington, DC 20052

*Microsystems Laboratory
Electrical Engineering Department
University of Maryland
College Park, MD 20742

Abstract — A technique is presented to allow both inhibition and excitation weights to take place on the same input lines of pulse coded neural processing element (NPE). This is achieved by using different ranges of pulse duty cycles for excitation and inhibition. The pulse coded NPE contains blocks of neural type cells (NTC), a summation, a nonlinear threshold logic, and a learning block for adaptive weight feedback. Modifications on the Neural-Type Cell are presented to achieve this as well as wider oscillation ranges. Simulation results verify the functions for each block. The results allow for adapting weights through sign changes leading to more flexibility in design and less power consumption. Along with its functional block-based structure, the NPE, as an artificial neuron, can be applied in many different configurations of neural networks. A Layout for a configuration for a Winner-Take-All network was carried out and sent to MOSIS for IC fabrication.

1. Introduction

In this paper, an improved NPE using NTCs is presented. The summation block in the NPE is modified for adapting weight for both excitation and inhibition. Thus, this modified design eliminates the need for a dedicated hardwired weighting scheme and this will lead us to have more flexibility in the design. This improvement also includes a modified NTC to achieve an input polarity change so as to be proportional with the output, along with a wider oscillation range.

Rigorous studies have been done on hardware implementation of the artificial neural networks (ANN) [1-4]. Using neural-type cells [NTC] [5], a new way of pulse coded weighting and summing was introduced in [6]. Based on this preliminary work, a new structure of a single neural processing element (NPE) was presented in [7]. In this scheme, each NPE has multiple inputs and single output. Both input and output are analog continuous variable in voltage while weights are expressed in terms of pulse duty cycle through pulse duty cycle modulation (PDCM) technique [7]. Thus this structure is expected to comply the basic functional characteristics of the well-known classical model of an artificial neuron [8,9], yet hold similarities with biological neurons with its spiking signals.

On the other hand, we found some drawbacks in the previous work. In the previous work [7], the summing output voltage $Y_s(T)$ across a capacitor C_s [see Fig. 4] in a time interval of T was expressed in terms of pulse duty cycle (PDC) of each pulse as,

$$Y_s(T) = \frac{g_m \alpha}{C_s}\left[\sum_{p=1}^{N}\left\{1 - Y_{PDC}^{ex,p}(T)\right\} - \sum_{q=1}^{M} Y_{PDC}^{in,q}(T)\right] + Y_{si} - I \qquad (1)$$

where $Y_{PDC}^{ex,}$ and $Y_{PDC}^{in,}$ are PDCs of excitatory and inhibitory weights, respectively. Y_{si} is initial voltage value across C_s. N and M are the numbers of weights for excitatory and inhibitory inputs, respectively, and I represents the contribution through external inputs.

As can be seen in eq. (1) and in [7], due to its hardwired connections for weights, we did not have flexibility of changing signs of the weights between excitations and inhibitions. In another word, an NTC worked as either an excitatory weight or an inhibitory one, but not both. Besides this, the structure was found to consume a large current (power) when both P- and N-type transistors are on and thus forming a direct DC path from the power line to the ground.

In addition, in the design of the synaptic junctions, one would prefer to have the output of the neuron to be proportional to both inputs and weights as described in [8,9]. In the original structure of the NTC [5], however, this could not be realized because of the input signal polarity difference with respect to the output. In order to satisfy this basic property of the artificial neuron, in this paper, we change the first stage of the NTC.

2. Structure of the Neural Processing Element(NPE)

Figure 1 shows the modified block diagram of an NPE. As can be seen, the NPE consists of four functional blocks: NTCs for synaptic junctions, a summation, a nonlinear threshold, and an adaptive weight feedback block. The NTC generates a pulse stream in response to a given analog input (stimulus), whose duty cycle is controlled by both the input and weight signals. In the summation block, the duty cycles of the pulses from several different neurons are converted into a voltage across a capacitor C_s, through the

0-7803-1254-6/93$03.00 © 1993 IEEE

PDCM technique introduced in [6]. This analog output voltage of the summation block is considered representing a weighted summation. In the threshold block, comparison between the weighted output and a given threshold value(voltage) is executed using a well-known differential stage. Finally, we also have a learning block where an adaptive weight feedback is executed. The feedback scheme depends on applications and several different schemes for different applications were experimented. Examples are winner-take-all, Hopfield, and auto-tracking schemes. This feedback scheme is expected to be easily integrated on the chip with the rest of the blocks.

2.1. Modified Neural-Type Cell for Synaptic Weighting

Figure 2 shows circuit diagrams of (a) a previous NTC [7] and (b) the newly modified one. Two major modifications are done: First, interchanging R1 and M1, second, introducing three more transistors (M8-10) at the input port. Interchanging R1 and M1 results in a new analytical equation of the voltage V2 as

$$V2 = \frac{K_1R4\left(V1-V_{tn}+\frac{1}{K_1R4}\right)+\sqrt{\left[\left(K_1R4\left(V1-V_{tn}+\frac{1}{K_1R4}\right)\right)^2-(K_1R4(V1-V_{tn}))^2\right]}}{K_1R4} \quad (2)$$

where K_1 is the transistor gain factor and V_{tn} is the threshold voltage of the N-type transistor of M1.

Now, as the input X to M1 increases, V2 also increases, which is in the opposite direction of the previous NTC. This will take care of the polarity problem of the previous structure.

The second modification is to widen the input oscillation range of the NTC [10,11]. To do this, three new transistors ,M8-10, are added to serve to linearly scale the original small oscillation range into a wider one. Figure 3 shows simulation results of this new structure in (b) compared with the old one in (a). As can be seen in the Fig. 3(b), the PDC, is now inversely proportional to input X, which will make the output Ys in Fig. 4 (b) be proportional with the input. Also from Fig. 3 a much wider oscillation range is seen to be acquired.

2.2. Improved Summation Block

Figure 4 shows the circuit diagrams of a previous design in (a) and an improved one in (b) for the summation block. The new summation block is a simple inverter with a charge feed through minimization transistors (M3, M4) with a large RC time constant. If a signal with pulse duty cycle (PDC) larger than 0.5 is applied the output will be close to GND(0V) level, acting as an inhibitory weight, otherwise the output will be close to Vdd and acting like an excitatory one. The circuit sizing is done in such a way that when PDC is equal to 0.5 the summation output is half of Vdd. Thus, with this new scheme, we are now able to control the sign of the weights in a very simple way and have considerable flexibility in the design. Also introducing two long channel feed-through transistors in the summation block significantly decreases the direct current flow from the power to the ground line. From the simulations, this current measured in the previous structure was 8.99uA, and decreases down to 1.64uA in the improved structure.

3. Simulation Results and Layout

Simulation results for a two input neuron case are shown in Fig. 5. In Fig. 5(a), two inputs X1, X2 and two weights W1, W2 are 3V, 4V, 3.0V and 3.5V, respectively. Only Ys and YD (high or low) output [Fig. 1] are shown. The threshold voltage is set as 1.5V. In Fig. 5(b), the result with inputs X1, X2 of 1V and 2V, respectively, with the same weights is shown. As expected, with the same weights, the case (b) has lower steady-state output (1.3V) than that of the case (a) (1.8V). Figure 6 shows the MOSIS tiny chip (1.7mm x 1.7mm) layout of a circuit with five fully-connected NPEs. Capacitors are realized by using gate capacitors. As can be seen, the layout was done in modular pattern: each functional block has a rectangular structure with the same height, and inputs and outputs run vertically for connections to other modules. Therefore, almost all types of artificial neural networks can be implemented using NPEs by different combinations of basic functional blocks. A layout was done by using CMOS P-well technology with 2um minimum feature size.

4. Conclusions

A new CMOS structure of an artificial neuron is presented. Modifications were done on the NTC and the summation block which lead us to have less power and more flexibility in weight adaptation. An NPE with single input contains 22 transistors, two resistors, and two capacitors and it has a size of ~180 mil^2. For each added input we use one more NTC (16 transistors and one capacitor). Different techniques for realizing capacitors, poly-to-poly structure, for example, are under consideration to minimize the silicon area. Although this technique does not supply an exact linear control over the weighted summation, it allows us to implement the mathematical model of an artificial neuron in a very simple way using pulse coding and without using operational amplifiers for weight multiplications and summations. Thus, by keeping the essence of the previous PDCM technique, we retain its robustness while improving the NPE's behavior as an artificial neuron. With its modular structure, this NPE can be used in hardware realization for many different configurations of the neural networks. For example, the layout for Fig. 6 can be configured as a winner-take-all, a Hopfield, or an auto-tracking networks with corresponding feedback schemes.

REFERENCES

[1] J. Meador, A. Wu, C. Cole, N. Nintunze, and P. Chintrakulchai, "Programmable impulse neural circuits," *IEEE trans. Neural Networks*, vol. 2, no. 1, pp. 101-109. January 1991.

[2] J. E. Tomberg and K. Kaski, "Pulse-density modulation technique in VLSI implementations of neural network

algorithms," *IEEE JSSC*, vol. 25, no. 5, pp. 1277-1286, October 1990.

[3] A. Murray, D. Corso, and L. Tarassenko, "Pulse-stream VLSI neural networks mixing analog and digital techniques," *IEEE trans. Neural Networks*, vol. 2, no. 2, pp. 193-204 March 1991.

[4] D. Corso and L. Reyneri, "Mixing analog and digital techniques for silicon neural networks," *Proc. IEEE ISCAS*, New Orleans, Louisiana, May 1990, pp. 2446-2449.

[5] R.W. Newcomb, "Neural-type microsystems circuit status", *Proc. IEEE ISCAS*, Newport Beach, CA 1983, pp. 97-100.

[6] G. Moon, M. Zaghloul, and R. Newcomb, "VLSI Implementation of Synaptic Weighting and Summing in Pulse Coded Neural-Type Cells," *IEEE Trans. on Neural Networks*, vol. 3, no. 3, pp. 394-403, May 1992 .

[7] G. Moon, M. Zaghloul, "CMOS Design of Pulse Coded Adaptive Neural Processing Element Using Neural-Type Cells,"*Proc. of IEEE International Symposium on Circuits and Systems*, San Diego, California, pp. 2224-2227, May, 1992.

[8] R. Lippmann, "An introduction to computing with neural nets," *IEEE ASSP magazine*, vol. 4, no. 2, pp. 4-22, April 1987.

[9] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Nat. Acad. Sci. U.S.*, vol. 79, pp. 2554-2558, April 1982.

[10] M. Savigny, G. Moon, M. Zaghloul, N. El-Leithy, and R. Newcomb, "Hysteresis turn-on-off voltages for a Neural-Type Cell", *Proc. 33rd IEEE Midwest Symp. Circ. Syst.*, Calgery, Canada, August 1990. pp. 37-40.

[11] G. Moon, "VLSI implementation of a Neural-Type Cell". *Master's Thesis*, Dept. of Electrical Engineering and Computer Science, The George Washington University, September 1990.

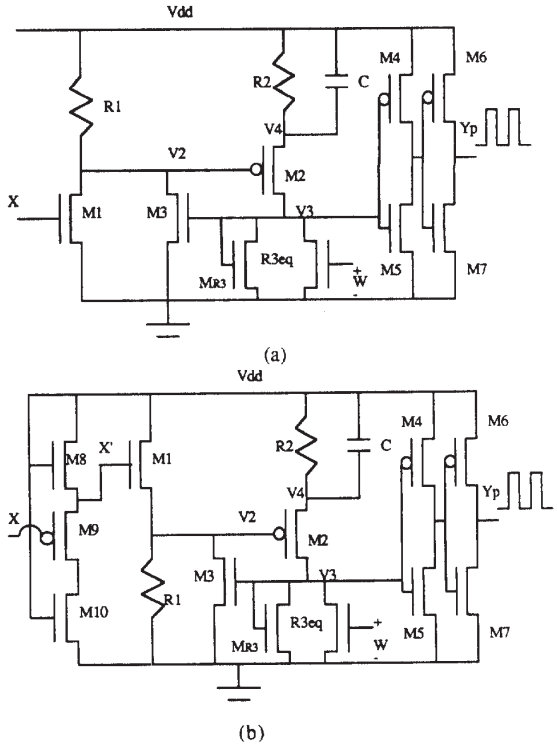

Fig. 2. Circuit diagrams of (a) a previous NTC, and (b) the new NTC.

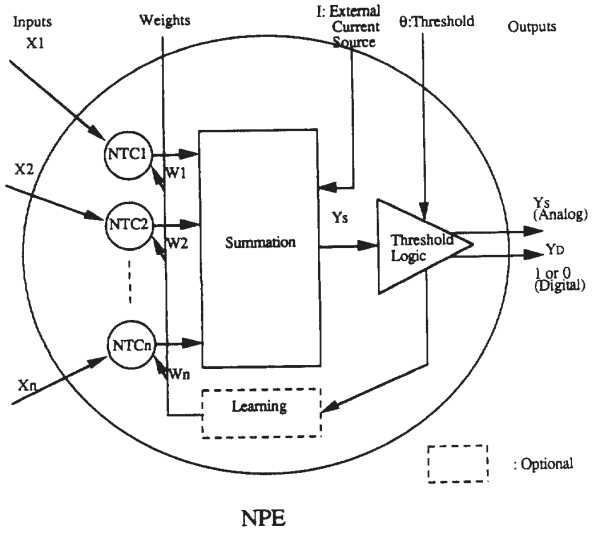

Fig. 1. Functional block diagram of an Neural Processing Element (NPE).

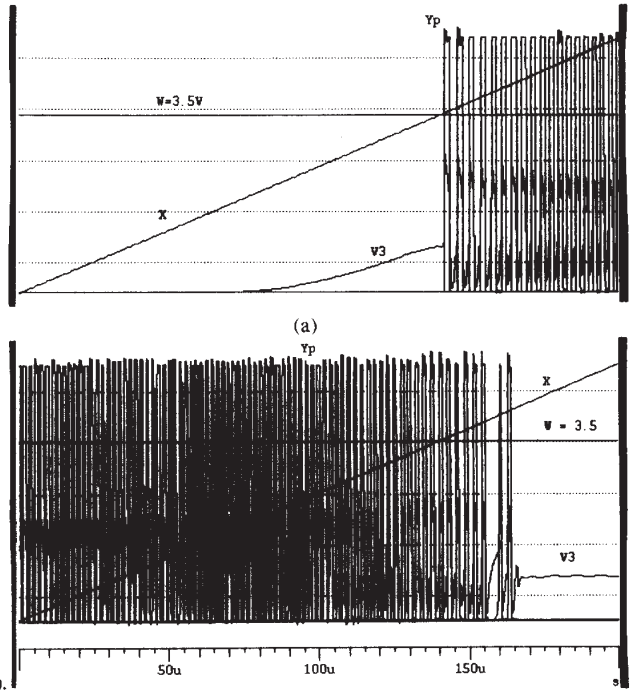

Fig. 3. SPICE results of (a) a previous NTC, and (b) the new NTC. (1V/X-div)

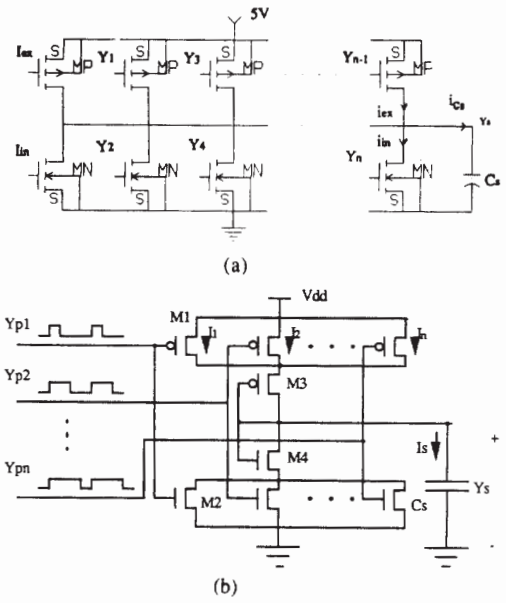

Fig. 4. Circuit diagrams of (a) a previous, and (b) an improved summation blocks.

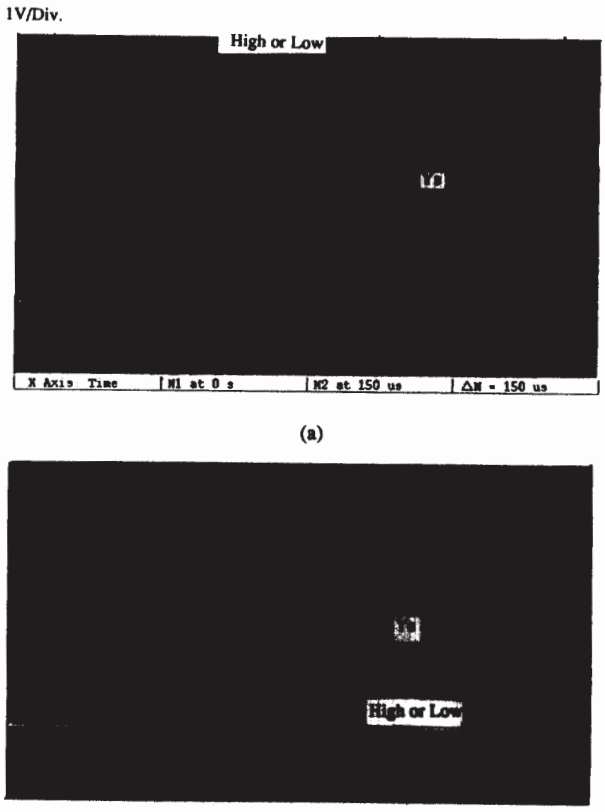

Fig. 5. Simulation results for a two input neuron case with same threshold of 1.5V.
(a) X1 = 3V, X2 = 4V, W1 = 3V, and W2 = 3.5V.
(b) X1 = 1V, X2 = 2V, W1 = 3V, and W2 = 3.5V.

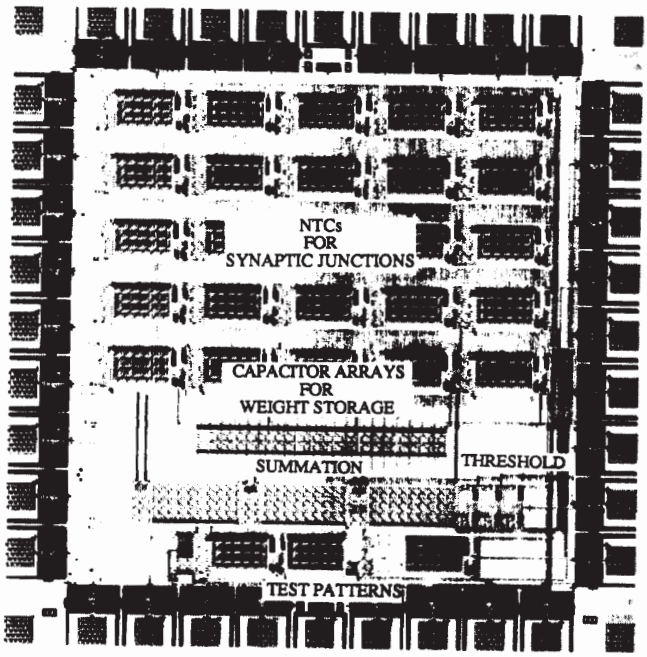

Fig. 6. Layout of five fully-connected NPEs.