A VLSI Implementation of Mixed-Signal mode Bipolar Neuron Circuitry

Dong Pan, and Bogdan. M. Wilamowski
University of Idaho, College of Engineering
800 Park Blvd. Suite 200
Boise, ID 83712

Abstract —**Neuron circuits have parallel operation features. VLSI implemented Neuron networks are suitable for high speed and low power consumption applications. Digital implementations have good noise immunity while analog neuron circuits have smaller size. This paper presents a mixed-signal neuron design. It uses digital input, output, and weight signals while keeps analog internal operation. Thus, this circuit has both good noise immunity and small size features. Clock signal is used to synchronize the neuron circuit operation. Simulation shows it has sigmoid activation function. This circuit is suitable for being used in feed-forward type neural networks.**

I. Introduction

Although most neural networks can be implemented by software through microprocessors, VLSI based neural network shows promising future in the high speed and low power consumption fields due to its parallel operation nature. Meanwhile, many interests were paid to integrate the entire neural network circuit to CMOS VLSI chip, which will effectively reduce both cost and size of the implementation.

As the basic component of the neural network, a single neuron can be designed as digital circuits, as analog circuits, or as mixed-signal mode circuits. The digital neuron circuit[1] uses many adders and multipliers. It performs mathematics operation like a simple microprocessor with the digital input and weight signals. The digital neuron circuit can achieve accurate activation function while the complex structure may limit its application. The analog neuron circuit [2][3][4] uses analog input and analog weight signals. It generates analog output signals. Comparing to its digital counterpart, the analog design is more compact. However, the analog signals are easily affected by noise and it is difficult to store the analog weight signals. The mixed-signal neuron circuit [5] keeps the analog input and output signal while uses digital weight signal. This solution is much more practical in the circuit implementation. However, the noise interference and the size issue are still not solved. Another type of mixed-signal neuron circuit uses digital pulses as the input/output signals [6][7]. It can be designed with compact size and good noise immunity. But it is hard to achieve good activation functions.

This paper presents a neuron circuit with digital input, digital output and digital weight signals. Unlike the pure digital neuron circuit, which uses digital adder and multiplier blocks, the internal operation of this circuit is analog. Therefore it has good noise immunity and compact size. The preferred activation function can be obtained.

Signal definition and circuit diagram are presented in the section II. After illustrating the basic operation mechanism, detail circuits of each block are discussed in Sections III and IV. Then, the simulation results of the neuron circuit and the entire neural network are presented.

II. Circuit Diagram and Signal Difinition

The circuit diagram of the neuron cell is shown in Fig.1. It includes synapse circuits, a weight control circuit, an initialization circuit, a capacitor, a comparator, and a feedback circuit. The number of the synapse circuits is same

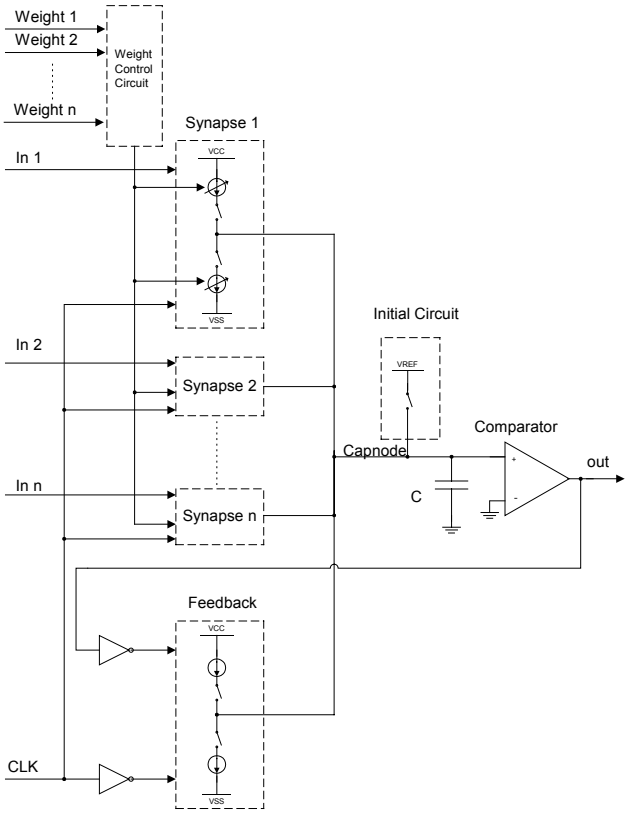


Fig. 1. Circuit diagram of a neuron cell.

as the number the input signals of the neuron. The operation of this circuit is based on charging and discharging the capacitor C, which is similar to the operation of dual slope A/D converter.

The voltage or current levels usually carry input and output values of the analog or mixed signal neuron circuits. The voltage or current levels can be easily processed in the neuron with pre-defined activation function. However, the input and the output values can also be described by other variables that can be applied in the neuron circuit. In this paper, the relative timing of the selected digital signal with the clock signal is used to define the magnitude of the values while the initial value (high or low) of the select signal is used as the sign.

As the signals shown in Fig.2, the magnitude of the input and output value are proportional to the timing relative to CLOCK, which is t_{in} and t_{out}. The sign of the bipolar input and output signals are defined by its initial value (V_{in_init} and V_{out_init}) when the CLOCK changes. Thus, we can define the input and output as follows

$$D_{in} = \begin{cases} \dfrac{t_{in}}{T/2} & \text{if } V_{in\_init} = VCC \\ -\dfrac{t_{in}}{T/2} & \text{if } V_{in\_init} = VSS \end{cases} \qquad (1)$$

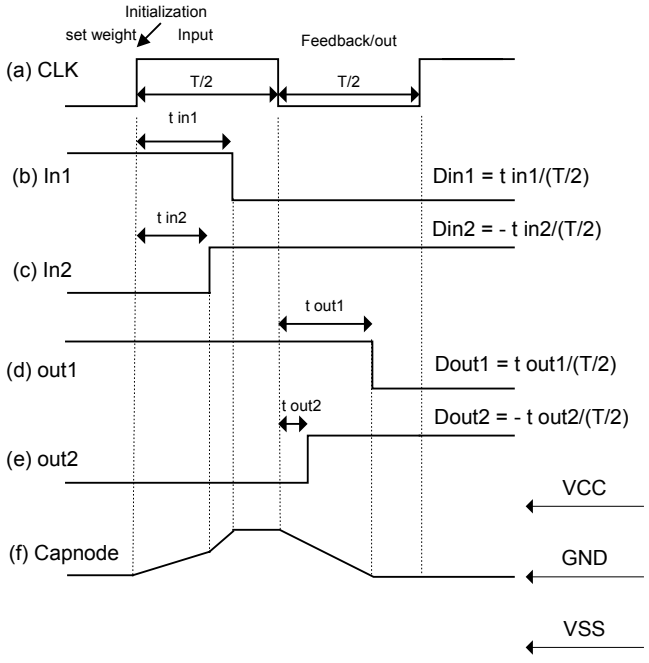


Fig.2.Signal definition of the neuron circuit: (b)positive input, (c) negative input, (d) positive output, (e) negative output, and the operation of a neuron circuit (a), (b),(c), (d), and (f).

$$D_{Out} = \begin{cases} \dfrac{t_{out}}{T/2} & \text{if } V_{out\_init} = VCC \\ -\dfrac{t_{out}}{T/2} & \text{if } V_{out\_init} = VSS \end{cases} \qquad (2)$$

in which T is the period of the clock signal (about 1-2ms). If the signal does not change during the certain clock phase (input signal when clock is high and output signal when clock is low), t_{in} and t_{out} will be recognized as T/2. Thus, the maximum and the minimum value of D_{in} and D_{out} will be 1 and –1.

III. Neuron Operation

Similar to dual slope A/D converter, the operation of the neuron circuit is based on charging and discharging the capacitor, which is synchronized by the CLOCK signal. The operation of a neuron circuit with two synapses, a positive input (IN1), a negative input (IN2) and positive weight signals is shown in Fig.2. It can be divided as the following steps:

1). Set weight: When the CLOCK is low, the weight signals of all the synapses are generated by the weight control circuit.

2). Initilization: When the CLOCK rises, the voltage of the top plate of the capacitor ($V_{CAPNODE}$) will be set to the predefined value VREF. The example shown in Fig.2 uses GND as VREF level.

3). Input charging/discharging: When the CLOCK is high, all the synapses start to charge or discharge the capacitor through their own variable current sources, which are controlled by the weight signals. The sign of the input and the weight determine whether to charge or discharge the capcitor. The charging and discharging will be terminated when the input changes (either rising or falling) or the CLOCK becomes low. The voltage of the output terminal(OUT1) is generated by the comparator, which compares the voltage of CAPNODE and GND.

4). Negative feedback charging/discharging: This stage is trigered by the falling edge of the CLOCK signal. A fixed current source will charge or discharge the capcitor. V_{out_init} determine whether to charge or discharge the capacitor. The $V_{CAPNODE}$ will finally achieve GND becasue of the negative feedback nature. Once the $V_{CAPNODE}$ reachs GND, the OUT1 state will change and it will further stop the process of charging/discharge the capaciator.

Since the weight signals are set in the phase before input charging and discharging, it can be hiden the feedback phase. Thus, the entire operation can be completed in one cycle of clock. The next set of input data will be evaluated from the rising edge of next clock cycle.

Considering the input charging/discharging stage, if we assume the input current source is stable, we can get:

$$V_{C\_fall} = V_{ref} + \frac{I_{in1}t_{in1} + I_{in2}t_{in2} + \cdots I_{inn} \times t_{inn}}{C} \quad (3)$$

in which V_{C_fall} is the $V_{CAPNODE}$ value when the CLOCK is falling. I_{in1}, I_{in2}, …, and I_{inn} are the current sources of the synapse 1, synapse 2, …, and synapse n. t_{in1}, t_{in2}, …, and t_{inn} are the input time of the synapse1, synapse2, …, and synapse n. C is the capacitor value.

If the feedback current source ($I_{feedback}$) is fixed, then the magnitude of the output value will be:

$$\begin{aligned} |D_{Out}| &= \frac{t_{out}}{T/2} = \frac{C \times V_{C\_fall}}{I_{feedback} \times T/2} \\ &= \frac{CV_{ref} + I_{in1}t_{in1} + I_{in2}t_{in2} + \cdots I_{inn} \times t_{inn}}{I_{feedback} \times T/2} \\ &= \frac{CV_{ref} + |I_{in1}|D_{in1} \times T/2 + \cdots |I_{inn}|D_{inn} \times T/2}{I_{feedback} \times T/2} \\ &= \frac{CV_{ref}/(T/2) + |I_{in1}|D_{in1} + \cdots |I_{inn}|D_{inn}}{I_{feedback}} \end{aligned} \quad (4)$$

Considering THD $=-CV_{ref}/(T/2)$ and SUM$= |I_{in1}|D_{in1} + \ldots + |I_{inn}|D_{inn}$, the above equation can be written as

$$|D_{Out}| = (1/I_{feedback}) \times (SUM - THD) \quad (5)$$

The output value is proportional to the SUM in case the THD is zero (when VREF is equal to GND). However, due to the VCC and VSS voltage limitation in the real circuit, the maximum $|D_{out}|$ will be the less than $C*VCC/(I_{feedback}*(T/2))$ and 1. At the same time, the value of input controlled current source will change due to the channel length modulation effect and the non-saturation operation as shown in the next section. Thus, the activation function has sigmoid feature, which is required by the neuron operation network circuits.

IV. Detail Circuit Configuration

A. Synapse

Synapse circuit is shown in Fig.3. It accepts the analog weight control signal V_{WTIN} and stores in the capacitor C1. V_{WTIN} controls MP1 and MN1, which operate like current sources. Control signals PUP_ and PDN are used to turn on or turn off these current sources.

As mentioned before, these current sources are used to charge or discharge the following capacitor when CLOCK is high. Besides CLOCK signal, the states of the PUP_ and PDN signals are also controlled by the input signal, the initial input signal (voltage level while CLOCK is rising), and the sign of the weight signal W_{sign}(H is negative and L is positive). Fig. 4 (a) is the truth table of the PUP_ and PDN. Fig. 4 (b) shows the PUP_ and PDN signals generation circuit.

The current generated by MP1 and MN1 will not keep constant when the voltage of CAPNODE approaches VCC and VSS. MP1 and MN1 will operate in non-saturation region when the |Vds|<|Vgs|-|Vt|. Furthermore, even in saturation region, the existence of the channel length

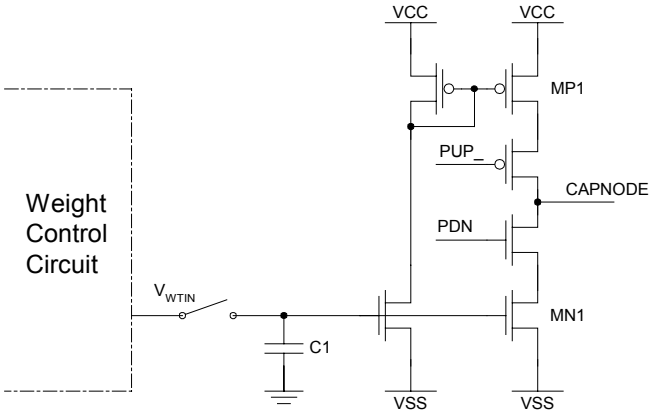
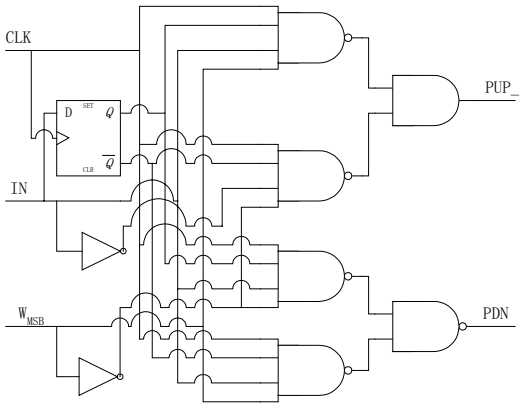


Fig.3. Synapse circuit.

CLK	Initial input when CLK rising	Input signal	MSB of Weight	PUP_	PDN
L	X	X	X	H	L
H	L	L	L	L	L
H	L	L	H	H	H
H	L	H	X	H	L
H	H	L	X	H	L
H	H	H	L	H	H
H	H	H	H	L	L

(a)



(b)

Fig.4. Synapse control: (a) signal true table and (b) circuit

modulation parameter λ will make the current variation when the $V_{CAPNODE}$ changes as shown in the follow:

$$I_D = \frac{\beta}{2}(V_{gs} - V_t)^2[1 + \lambda(V_{ds} - V_{ds,sat}) \qquad (6)$$

Different transfer curves can be obtained by choosing different lengths of the CMOS transistors, which changes the value of λ.

As shown in (5), D_{out} is proportional to $V_{CAPNODE}$. At the same time, $V_{CAPNODE}$ affects the value of the I_{MP1} and I_{MN1}. Thus, it is possible to get a sigmoid activation function.

B. Weight Control Circuit

Digital signals are preferred in weight control circuits than analog signals due to the robustness. Pure digital weight signals can easily be adopted by replacing the MP1 and MN1 in Fig.3 by transistor stacks as shown in Fig.5. However, for a normal 10 bit weight control signals, the largest transistor used will be 2^{10} * min size. A neural network usually has at least several tens of neurons and each neuron has several synapses. The total size of the neural network circuit will be too large to integrate into a VLSI chip.

A weight control circuit is used in this neuron circuit to reduce the transistor size. As shown in Fig.6, a weight control circuit is a simple D/A converter. It converts the digital signals to a voltage control signal V_{WTIN}. V_{WTIN} will be stored in the capacitor in the synapse. Switch S1 can be used to determine whether this synapse circuit is being selected or not. One or several the weight control circuits can be used to set the weights for all neurons if they can charge the capacitor C1 much faster than the clock period. Thus, the total size of the neural network will be significantly reduced.

Leakage is the biggest problems for this type of the weight control circuit. V_{CAP} in Fig.7 (a) is equal to V_{WTIN} when the transmission gate is turned on. When the transmission gate is off, possible leakage current may flow through gate of MN1 (I1), through the capacitor (I2), and through transmission gate (I3). The leakage current I1 and I2 are usually much less than I3, which is caused by the sub-threshold current. Fig.7(b) shows a improved circuit to reduce the sub-threshold leakage current. V1 is force to GND when the transmission gate is off. Sub-threshold leakage current will thus be minimized since both V_{gs} and V_{sg} of the MN2 and MP2 are less than zero.

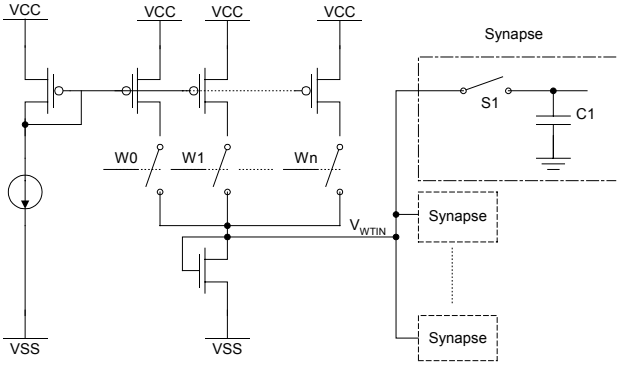
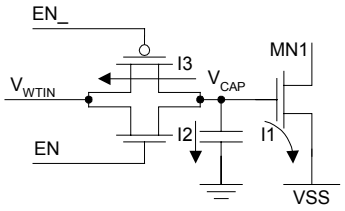
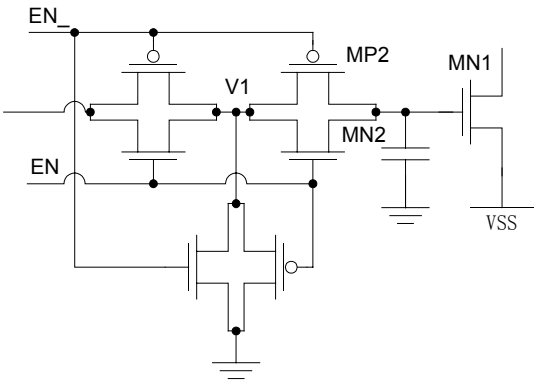


Fig. 6. Weight Control Circuit.



(a)



(b)

Fig.7. Switch circuit: simplified diagram (a), and actual diagram (b).

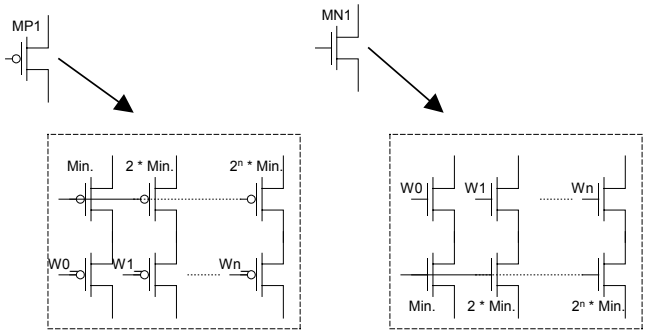
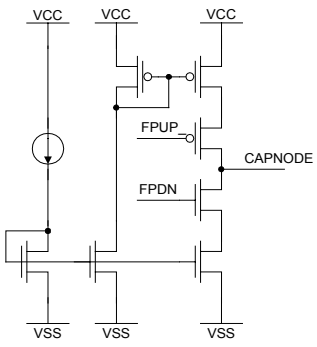
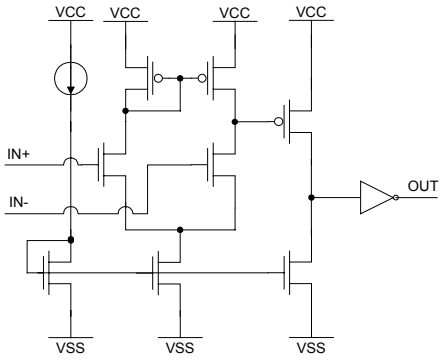


Fig.5. Method of using digital weight signals.



(a)



(b)

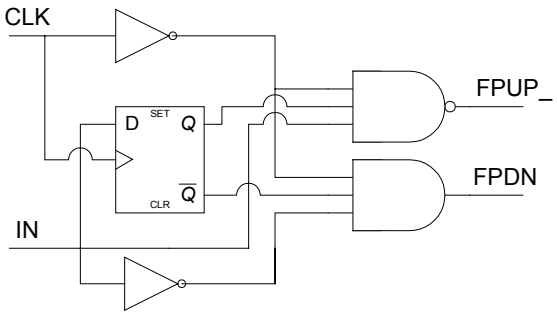
Fig. 8. Feedback circuit (a), and comparator circuit (b).

C. Feedback Circuit

The feedback circuit is shown in the Fig.8(a). Unlike the synapse circuit, which uses weight controlled current sources, the feedback circuit uses constant current source to charge or discharge the capacitor. The negative feedback will finally charge the capacitor to GND and further change the comparator output signal state. Control signals FPUP_ and FPDN are generated by the CLOCK, output signal of the comparator, and the initial output signal (output state while CLOCK is falling). As the truth table and the circuit shown in Fig.9, FPUP_ and FPDN will disables the charge or discharge process once the comparator output signal toggles.

CLK	Initial OUT when CLK falling	OUT signal	FPUP_	FPDN
H	X	X	H	L
L	L	L	H	H
L	L	H	H	L
L	H	L	H	L
L	H	H	L	L

(a)



(b)

Fig. 9. Feedback control: truth table (a), and circuit (b).

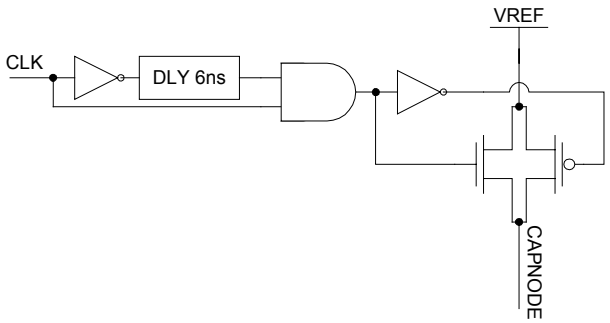


Fig.10. Initialization circuit.

D. Comparator and Initialization Circuit

A simple 2-stage comparator is used for comparing the $V_{CAPNODE}$ and GND level as shows in Fig.8(b). The initialization circuit (Fig.10) resets the $V_{CAPNODE}$ to VREF before starts the normal operation. It is a simple one-pulse generation circuit triggered by the CLOCK rising edge. Since the pulse width (5-10ns) is much shorter than the CLOCK cycle (1-2ms), the time used for initialization can be ignored.

V. Simulation Result and Network Operation

Fig.11 shows the simulation result of the activation function for the proposed bipolar neuron circuit. The input D_{in} and output D_{out} are defined in (1) and (2). The limit of both D_{in} and D_{out} are +/-1. It has sigmoid transfer function which is required for the proper operation of the neuron. The simulation uses 2ms clock period as the operation frequency. The input charging and discharging will happen in the first half cycle and the output will be generated in the next half cycle. Because all the inputs of the neuron must be available at the same time, this type of neuron is not suitable for feedback related neural network circuit.

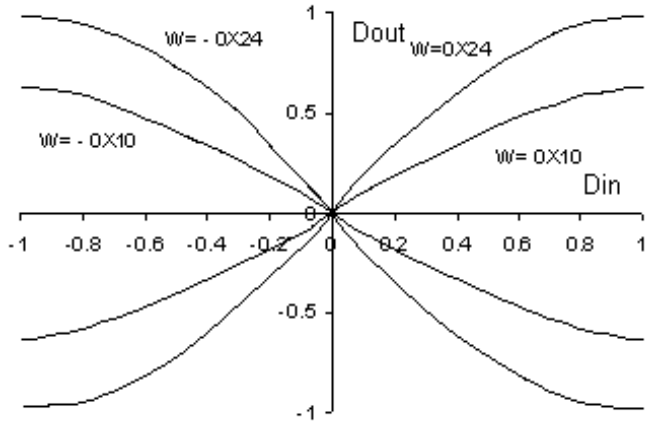


Fig.11. Simulation result of the activation function.

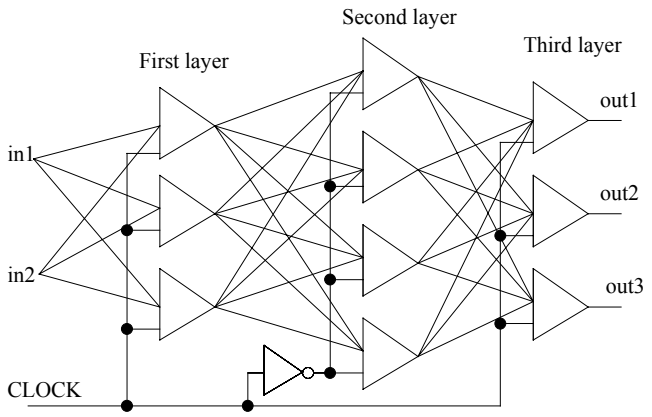
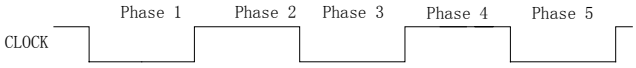


Fig. 12. Diagram of a feed-forward neural network circuit.

Fig.12 is the diagram of a feed-forward type neural network, which is suitable for using proposed neuron circuits. As shown in Fig.13, all the operation is synchronized by the CLOCK signal. The operation of the neural network has pipeline features. All the neurons in the same layer are operating at the same time. The neurons in the next layer will operate half cycle later. They receive the outputs of neurons in the previous layer as their inputs. Besides the input and the feedback charging/discharging operation, weight signals are generated by the weight control circuit in the phase previous to the input charging and discharging (may overlapped with the feedback charging/discharging phase). Thus, the time for processing the first set of input data is 2.5 clock cycles. After that, only half cycle is needed for processing one set of data.



	Phase1	Phase2	Phase3	Phase4	Phase5
Neurons in first layer	Set weights	input	feedback, output, set new weights	new input	feedback, output , set new weights
Neurons in second layer		Set weights	input	feedback, output , set new weights	input
Neurons in third layer			Set weights	input	feedback, output , set new weights

Fig. 13. Operation timing of feed-forward neural network.

The size of analog or mixed type neuron circuit is usually determined by the size of the passive components if it doesn't use huge size MOSFETs. This circuit uses only one charging and discharging capacitor (10pf-30pf) and several weight storage capacitors (1 pf each). It is suitable for being integrated into a VLSI chip because of the compact size.

VI. Conclusion

A mixed signal mode bipolar neuron circuit using digital weight, input, and output signals is presented. It features both good noise immunity and small size. The simulation result indicates that the circuit has sigmoid activation function. It is suitable for being used in the feed-forward neural networks circuit.

Acknowledgement

This work was supported by NSF-Idaho EPSCoR and National Science Foundation grant EPS-0132626.

References

[1] Tim Schoenauer, Sahin Atasoy, Nasser Mehrtash, and Heinrich Klar, "NeuroPipe-Chip: A Digital Neuro-Processor for Spiking Neural Networks," *IEEE Transactions on Neural Networks*, Vol. 13, No. 1, Jan. 2002, pp. 205-213.

[2] B. M. Wilamowski, J. Binfet, and M. O. Kaynak, "VLSI Implementation of Neural Networks," *International Journal of Neural Systems (IJNS)*, Vol. 10, No. 3, pp. 191-197, June, 2000.

[3] Jin Liu, Martin A. Brooke, and Kenichi Hirotsu, "A CMOS Feedforward Neural-Network Chip With On-Chip Parallel Learning for Oscillation cancellation," *IEEE Transactions on Neural Networks*, Vol. 13, No. 5, Sep. 2002, pp. 1178-1186.

[4] C.Lu, and B.Shi, "Circuit Realization of an Programmable Neuron Transfer Function and Its Derivative," *Proceedings of the 2000 international Joint Conference on Neural networks*, pp. 47-50.

[5] Vincent F.Koosh and Rodney M. Goodman, "Analog VLSI Neural Network With Digital Perturbative Learning," *IEEE Transactions on Circuits ans Systems*, Vol. 49, No.5, May. 2002, pp. 359-368.

[6] Yasuhiro Ota, and Bogdan M. Wilamowski, "Analog Implementation of Pluse-Coupled Neural," *IEEE Transactions on Neural Networks*, Vol. 10, No.3, May. 1999, pp. 539-544.

[7] Bo Liu, and James F. Frenzel, "A CMOS Neuron for VLSI Circuit Implementation of Pulsed Neural Networks," *Proceedings of the 28th annual conference of IEEE industrial Electronics society*, 2002.