第 21 卷第 11 期　　半 导 体 学 报　　Vol. 21, No. 11
2000 年 11 月　　CHINESE JOURNAL OF SEMICONDUCTORS　　Nov., 2000

Low Power Design of High Speed CMOS Pulse Stream Neuron Circuit*

CHEN Ji-wei(陈继伟) and SHI Bing-xue(石秉学)

(*Institute of Microelectronics, Tsinghua University, Bejing* 100084, *China*)

Abstract: A new pulse stream neuron circuit is presented, which can be obtained in the digital CMOS process and combines both the merits of digital circuits and analog ones. The output is expressed by the frequency of the pulses with transfer characteristic, which is correspondent with the ideal sigmoid curve perfectly. Moreover, the pulse-active strategy is introduced into the design of this CMOS pulse stream neuron circuit for the first time in order to reduce the power dissipation, which is applicable to the low-power design of mixed-signal circuits, too. A simple technical process and compact architecture make this circuit work at a higher speed and with lower power dissipation and smaller area.

Key words: artificial neural networks; pulse stream; pulse-active strategy; CMOS; low power
EEACC: 2570D; 1295; 1280; 1260
CLC number: TN386.1　**Document code**: A　**Article ID**: 0253-4177(2000)11-1064-05

1 Introduction

The great information processing power of human being's neural systems has attracted a lot of attention of those who are dedicated to the implementation of Artificial Neural Networks(ANNs), which are expected to be of the same computation and learn ability, and able to process information with the data and examples in the real world instead of the prearranged algorithms[1]. More and more valuable results have been obtained during the in-depth research in this field in the past decades[2].

It is possible to simulate the behavior of neural networks with software based on certain models in virtue of the rapid development of computer technology, but the speed limit due to the serial computation of the software makes it difficult to deal with the process of mass data. In contrast, ANNs implemented with hardware have arrested more and more

* Project Supported by National Natural Science Foundation of China Under Grant No. 69636030.
CHEN Ji-wei(陈继伟) Ph. D. Candidate. His work focuses on the VLSI implementations of artificial neural networks and the design methodology of mixed-signal circuits. E-mail: chjw@dns.ime.tsinghua.edu.cn
SHI Bing-xue(石秉学) Professor. His present research focuses on the design of neural networks, fuzzy logic and mixed-signal circuits. E-mail: shbx@dns.ime.tsinghua.edu.cn
Received 28 March 2000, revised manuscript received 29 June 2000

researcher's attention because its computation speed falls little with the increase of the scale due to its inherent parallelity, and the VLSI implementation of ANNs has become a focus because of the maturation of VLSI technology and the scale of integrated circuits which is getting larger and larger[3,4].

A single neuron can receive the inputs from external environment or the outputs from the neurons in the preceeding layer through weighted synapses, and then generate its output to external environment or the neurons in the posterior layers according to the activation function. A three-layer ANNs can perform any functional computations theoretically[5]. For the forward neural networks, the sigmoid function is adopted as an activation function popularly due to its continuous and symmetric derivative, which is suitable for the error back-propagation learning algorithm. It is expected as:

$$S_{\text{out}} = \frac{1}{1 + e^{-\alpha(\text{activation} - \theta)}} \tag{1}$$

In a variety of VLSI architectures, digital circuits are accurate but bulky, while analog circuits are compact but susceptible. The mixed-signal circuits are welcome because they combine the merits of them. Recently, more attention has been focused on the pulse stream circuits due to its virtues, which can turn an analog input into a digital signal pulse stream, with the interval of pulses representing the activation of the input, as well as keep the signal in analog state when processing and digital state when transferring[6,7].

The implementation of the neuron with sigmoid activation function is very important to the performance of ANNs because most probably, the circuits can not work properly with unqualified weights, which are obtained through the off-chip learning, due to the mismatch of the actual activation function with the ideal sigmoid one. Furthermore, it is desirable that the neuron circuit can be implemented in a simple process and of a compact conformation with high speed and low power dissipation for the large-scale integration and application.

A compact neuron circuit with a sigmoid activation function is presented in this paper, which is compatible with the digital CMOS process. We have proposed a pulse-active strategy to reduce the power dissipation. Due to the CMOS process we adopted, the circuit has lower currents than BiCMOS counterparts, thus it has lower power consumption and can be implemented easily. It also has a higher speed compared with other pulse stream circuits, which makes this circuit a promising candidate for the VLSI implementation of great scale neuron networks.

2 Circuit Architecture

The neuron circuit is shown in Fig. 1, whose blocks in the dotted line frame 1, 2 and 3 are active current generator I_{sigmoid}, constant current source I_{const} and pulse generator, respectively. Compared with the circuit proposed by Haycock in Ref. [7], this circuit does not contain input buffer and input phase-inverter stage, thus having a lower power and

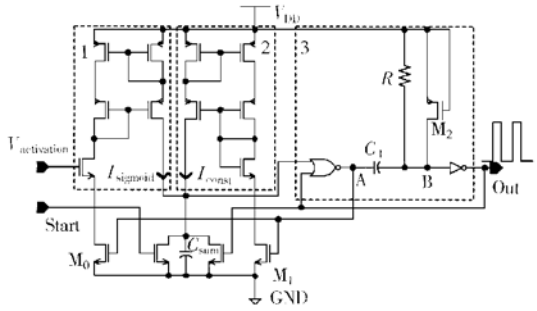

FIG. 1 Neuron Circuit

area. Being a pure CMOS circuit, it can be implemented more easily than the BiCMOS process adopted in Ref. [7]. Moreover, CMOS circuit has a lower current than BiCMOS circuit, so the power dissipation can be reduced further.

The sigmoid activation is achieved by the charge of capacitor C_{sum}. The time of driving the voltage on C_{sum} from zero to the threshold voltage V_{thres} in NOR circuit can be calculated by Equation (2):

$$\frac{1}{C_{sum}}\int_0^T (I_{const} + I_{sigmoid})\,dt = V_{thres} \tag{2}$$

Because active current $I_{sigmoid}$ is the function of active voltage $V_{activation}$, and $V_{activation}$ is a time-variant signal, $I_{sigmoid}$ is also time-variant. If $V_{activation}$ keeps constant or varies little in the time segment from zero to T, formula (2) can be turned into formula(3):

$$\frac{1}{T} = \frac{I_{const}}{C_{sum}V_{thres}} + \frac{I_{sigmoid}}{C_{sum}V_{thres}} \tag{3}$$

According to Nyquist sampling principle, the frequency of the pulses should not be smaller than twice of the input signal's highest frequency, so the smaller T is, the faster the speed would be, at which the circuit can work. Should $I_{sigmoid}$ be a sigmoid function of $V_{activation}$, $\frac{1}{T}$ will also be a sigmoid function of it. Should the operation frequency increase linearly, $I_{sigmoid}$ and I_{const} will decrease linearly while C_{sum} decrease squarely. Since power dissipation P is proportional to the current I, capacitor C is proportional to the area, the speed of the circuit will increase while the power and the area are reduced.

The pulse generator consists of digital cells of NOR and INV, capacitor C_1, resistor R and PMOS M_2. For the pulse duration, the voltage at node B is low, M_2 is cut off and the charge time is determined by C_1 and resistor R; while at the pulse interval, V_B is driven to a level that is higher than V_{DD}, by C_1, which turns M_2 on, so that V_B can decrease faster. When the voltage on C_{sum} reaches the threshold voltage V_{thres} of the NOR, the V_A is lowered, as makes V_B low and OUT high, thus the pulse is generated, C_{sum} will be discharged. Source V_{DD} will not charge C_1 through R until V_B reaches the threshold voltage of the inverter, and the pulse will be finished when the OUT is lowered. Since there is no information of the activation voltage in the pulse width, narrower pulses can create faster speed and smaller area of the capacitor C_1 and resistor R. In fact, the pulse width under different activation voltage needs not to be identical. Improved Wilson current mirrors are adopted in the design of the active current generator $I_{sigmoid}$ and the constant current source circuit, which can guarantee the consistency of the currents.

Furthermore, compared with Ref. [7], a pulse-active strategy has been introduced in

this circuit to decrease the power dissipation. C_{sum} does not play a role in the pulse duration as the computing unit, since the charge stored on it must be dismissed thoroughly to ensure the precision of the pulse interval. And a pass path should be guaranteed from C_{sum} to the ground due to the leakage current. $I_{sigmoid}$ and I_{const} in this pulse duration will cause additional power dissipation. So the activation of the active current $I_{sigmoid}$ generator and the constant current source circuit is ensured by the turn-on of M_0 and M_1 in the charge period of C_{sum} as well as the deactivation guaranteed in the pulse duration period to reduce the power dissipation. Moreover, the control signal of M_0 and M_1 is V_A, so the control circuit is very simple and effective without any additional external control signal adopted. When switching the two currents, there exists some unsteadiness. According to our experiment, the time of these two currents becoming steady is no more than 5ns, which is much less than the minimum pulse interval of 143ns. For this pulse stream neuron circuit, its speed is expected as its maximum pulse interval, so the effect of this strategy on the circuit's speed is rather small and can be ignored. The signal Start is used to start working the circuit. In fact, only a small time segment, in which the Start is kept in the high level, will satisfy this requirement. In addition, there is no additional bias voltage and bias current in this circuit.

3 Simulation Results

This circuit is simulated using 0.8μm standard digital CMOS process with the HSPICE. It can be concluded that this circuit has a near ideal transfer characteristic with the maximum normalized standard deviation of 3.3% from Fig. 2 and Fig. 3.

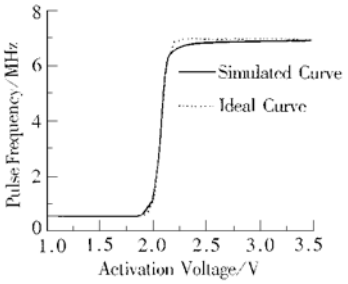

FIG. 2 Neuron Transfer Characteristic

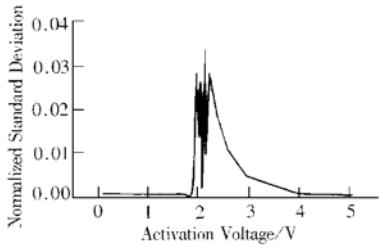

FIG. 3 Normalized Standard Deviation

An experiment has been done to verify the effectiveness of the pulse-active strategy, and the results are plotted in Fig. 4. It can be seen that the power dissipation using the pulse-active strategy is smaller than that not using it during most of the pulse duration time. The maximum DC power dissipation and the average one in one pulse period with the pulse-active strategy are 72 and 52.5μW respectively, while they are 90.2μW and 55.6μW, respectively, without the pulse-active strategy. It is proved that the pulse-active

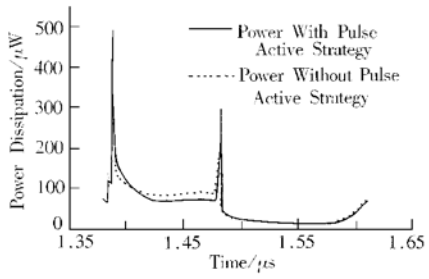

FIG. 4 Power Dissipation in a Pulse Period

strategy is highly effective for the low power design of both this circuit and the general mixed-signal circuits.

Some parameters of this circuit are compared with those of the circuit in Ref. [7], with the results listed in Table 1. Obviously, our circuit has smaller power dissipation and smaller area, which can work at a much higher speed. All those results show that the circuit presented in this paper is a good neuron one with excellent performance in many aspects.

Table 1 Parameters of the Circuit in This Paper and the Circuit in Ref. [7]

	Neuron in This Paper	Neuron in Ref. [7]
Process	0.8μm CMOS	1.2μm BiCMOS
Area of 10 Neurons	200μm×100μm (Estimated)	400μm×190μm
Sigmoid Output	Pulse Frequency	Pulse Period
Pulse Width	Unfixed	Fixed
Maximum Pulse Width	0.21μs	0.5μs
Maximum Pulse Interval	1.78μs	10.5μs
Maximum Power Dissipation	73μW	100μW

4 Conclusion

Due to the good performance of the neuron circuit, the pulse stream circuit is proved to be attractive method for the implementation of ANNs. The results show that the pulse stream circuit can realize the complex functions of the ANNs with larger and larger scale. Very large scale ANNs chip with on-chip learning ability is the research direction of the pulse stream neural networks since the learning ability is a main merit of ANNs, and the merits of neural networks have been shown on this kind of chip, such as high parallelity, flexible adaptability in different environments, etc. More efforts should be made on above researches sequentially.

References

[1] E. Sanchez-Sinencio and C. Lau, Artificial Neural Networks, New York: IEEE Press, 1992.
[2] Francoise Fogelman Soulie and Patrick Gallinari, Industrial Applications of Neural Networks, Singapore: World Scientific, 1998.
[3] S. M. Fakhraie and K. C. Smith, VLSI-Compatible Implementations for Artificial Neural Networks, Boston: Kluwer Academic Publishers, 1997.
[4] LIN Gu and SHI Bing-xue, "A Novel Programmable, Scalable Hamming Neural Network", Chinese Journal of Semiconductors, 2000, **21**(1): 69—75(in Chinese).
[5] R. P. Lippman, "Pattern Classification Using Neural Networks", IEEE Communication Magazine, November, 1989.
[6] Alan Murry and Lionel Tarassenko, Analogue Neural VLSI: A Pulse Stream Approach, International Thomson Computer Press, London, 1994.
[7] R. J. Haycock and T. A. York, "Hardware Implementation of a Pulse-Stream Neural Network", IEE Proc. Circuits Devices Syst., 1998, **145**(3): 141—147.