The Design of CMOS Cellular Neural Network (CNN) Using the Neuron-Bipolar Junction Transistor (vBJT)

Chiou-Ling Yeh, Student Member, IEEE, and Chung-Yu Wu, Fellow, IEEE
E-mail: cywu@alab.ee.nctu.edu.tw

Abstract

In this paper, two CMOS implementations of the cellular neural network (CNN) are presented based on the neuron-bipolar junction transistor (vBJT) which consists of the parasitic PNP bipolar junction transistor and the spreading base resistor array in the CMOS process. In the first design, the vBJT is used to implement the neuron and weights of the cell. In the second design, it is used to implement the current summation and weights of the cell, and a diode structure is proposed to realize the neuron. For programmable capability, the resistor in the vBJT can be replaced by a tunable MOS resistor. The two kinds of circuits have been designed and fabricated in 0.6μm single-poly-triple-metal n-well CMOS process.

Introduction

The cellular neural network (CNN) is a massively parallel, locally connected, and nonlinear dynamic system. Its applications have been focused mainly on image processing, and also been explored by many researches.

So far, many hardware implementations of the cellular neural network have been proposed, including continuous-time [1] and discrete-time [2] design approaches. In the model of CNN, the only nonlinear element in a cell is a piecewise linear function (sigmoid function). Among previously reported continuous-time CNN integrated circuit design approaches, most of them implement the sigmoid function by operational transonductance amplifiers (OTAs) [3], current limiters [4], or other similar circuits with large area. The cloning templates were achieved mainly by current mirrors [1], analog multipliers [5], or digitally programmable current switch technology [4].

Recently, a new device structure called the neuron-bipolar junction transistor (vBJT) for the compact implementation of VLSI neural networks was proposed by Wu and Yen [6]. In the new device structure, the parasitic PNP bipolar junction transistor in the CMOS process is used to implement the neuron whereas the spreading base resistor array is used to realize the synapse weights for various neuron inputs. Based upon the neuron-bipolar junction transistor (vBJT), a simple structure of the cellular neural network with easily tunable weights is proposed. The nonlinear element and the tunable weights are both implemented in a simpler way than previous. Thus the hardware complexity of a cellular neural network can be reduced, and the efficiency of the chip area of the integrated circuit can be increased.

Structure Description

The neuron is the only nonlinear element in the one cell circuit, and forms the sigmoid function. In previous published implementations of the neuron structure, many of them realized the sigmoid function by operational transconductance amplifiers (OTAs), current limiters, or other similar approaches. In the following, we will propose another two approaches to realize the neuron with small circuits.

The first approach is to utilize the characteristics of the bipolar junction transistor (BJT) with the voltage-to-current converter to form the neuron structure. In this structure, the outputs of the neuron are weighted current signals which will flow into current summations directly. The circuit is shown in Fig.1(a), where the PNP bipolar junction transistor Q1 and the resistor R1 form the neuron-bipolar junction transistor (vBJT). The diode D1 is used to clamp the source voltage of the PMOS MP at Vdd–Vγ, where Vdd is the voltage of the power supply and is equal to 5V in this circuit, and Vγ is the cut-in voltage of the diode D1. The PMOS MP and NMOS MN behave as a voltage inverter. The emitter of Q1 is fixed at 2.5V by a virtual bias operational amplifier shown in Fig.2, which consists of two PMOSFETs and four NMOSFETs. When the gain of the OP is designed to be about 100 and the bandwidth is about 10MHz, the voltage at the node of the emitter will be fixed at about V_{com}.

Since the voltage between the emitter and the collector, V_{EC}, of Q1 is fixed at 2.5V, the PNP transistor will work either in the forward-active region or in the cutoff region. The emitter current is then proportional to the base current with a factor of (β+1), where β is the forward current gain

0-7803-5529-6/99/$10.00

of the bipolar junction transistor. When the state voltage V_x is higher than $Vdd-V\gamma-|V_{TP}|$, MP is turned off and the base current is a constant current provided by MN and the fixed bias Vb. When V_x decreases, the output voltage V_o will raise. The base current and the output current (emitter current) are then decided by the output voltage V_o and the resistance of R1. When V_o is larger than the voltage of (V_E-V_T), where V_T is the cut-in voltage of the Emitter-Base junction, the PNP transistor will be cut off and the output current is zero. The transfer curve of the emitter current I_E versus the state voltage V_x is drawn in Fig.1(b).

In CMOS processes, the PNP transistor in the neuron-bipolar junction transistor (νBJT) can be implemented by using the vertical parasitic PNP bipolar junction transistor whose collector is tied to the P-type substrate. The N-type well region forms the base of the transistor, and the P+ diffusion inside the N-well forms the emitter. The resistor in the neuron-bipolar junction transistor (νBJT) can be realized by using the N-well resistor of the base region. The weights are then controlled by the resistance of the N-well resistors. The cross-sectional view of a neuron-bipolar junction transistor (νBJT) in CMOS processes is illustrated in Fig.3. Although it will make the layout more compact, the weights realized by N-well resistors are fixed and unprogrammable. The application of the cellular neural network will be limited to specified one. For tunable weights, one MOSFET operated in the linear region can be used to replace the N-well resistor. Then, the weight can be tuned easily by controlling the gate voltage of this MOSFET. The diode D1 can be realized by a MOSFET with the gate and the drain connected together. The complete CMOS implementation of Fig.1(a) is shown in Fig.1(c).

The above implementation has two disadvantages. The first is the nonsymmetry in the stable current. Because the low stable current remains in zero no matter how many the weight is, it will have nonsymmetric effect. The second disadvantage is the lack of realizing negative weights. In many applications, the template requires negative weights to implement specific function. Since the base current of the PNP transistor can not flow in reverse direction, the negative weight can not be realized.

The second approach of implementing the neuron is proposed by using the diode structure. A diode with the N-type region connected to the ground can realize the right side of the sigmoid function, and a diode with the P-type region connected to the power supply can realize the left side. The circuits and the transfer curves are shown in Figs.4(a) and (b), respectively. The sigmoid function can be realized by combining the circuits in Figs.4(a) and (b), and let one diode be off when the other is on.

In order to make the two diodes not turn on simultaneously, it can be seen that the cut-in voltage of the diode must be higher than $\frac{Vdd}{2}$. Six diodes shown in Fig.5 are required to implement the sigmoid function if the power supply is 5V. The complete MOS implementation is shown in Fig.6.

The slop of the sigmoid function when the output is not saturated can be modified by adding a resistor in parallel with the diodes. A realization of the resistor is shown in Fig.7, which consists of two serially connected PMOSFETs with the gate connected to the drain and operated in the saturation region. The complete MOS implementation of the neuron is shown in Fig.8(a), and the V-I transfer curve is shown in Fig.8(b).

The weight of the template is determined by the resistor connected between the cell and it's neighboring cells. The N-well resistor can be the appropriate realization of the weight in the circuit of the cellular neural network. However, the value of N-well resistance is fixed and unprogrammable. There are two approaches in implementing the tunable resistor by the MOSFET. One is to use a MOSFET operated in the linear region, as described above. However, when V_{DS} increases to be comparable with V_{GS}-V_T, the MOSFET will behave as a nonlinear voltage-controlled resistor.

The other approach is shown in Fig.9, which consists of two parallel-connected MOSFETs operated in the linear region. One is the PMOSFET and the other is the NMOSFET. Assume the conductivity parameters of the two MOSFETs are the same. Then the relationship of current I and voltages V_1 and V_2 can be derived as

$$V_2 - V_1 = \frac{I}{2K\left(V_{rn} - V_{rp} - |V_{TP}| - |V_{TN}|\right)}$$

Thus, the resistance of the two MOSFETs is equal to $\frac{1}{2K\left(V_{rn} - V_{rp} - |V_{TP}| - |V_{TN}|\right)}$.

From the above analysis, the two parallel-connected MOSFETs behave as a linear resistor, and the resistance can easily be tuned by controlling the voltages of the gates of the two MOSFETs.

According to the above designs of the neuron and the tunable MOS resistor, two hardware implementations of the one cell circuit are proposed, as shown in Figs.10 and 11, respectively. In the first implementation of the one cell circuit shown in Fig.10, the resistor connected between V_o and the base region of the BJT determines the weight of the current either flowing into the neighboring cells or flowing as the feedback signal. The emitter current of each

BJT is either the output current or the feedback current. The emitter of each BJT in one cell is virtual biased by the simple operational amplifier shown in Fig2. When the resistance of the weighted resistors are 1R, 2R, 3R, and 4R, for example, the currents flowing through the bases are about $\frac{V_{com}-V_{EB}-V_o}{R}$, $\frac{V_{com}-V_{EB}-V_o}{2R}$, $\frac{V_{com}-V_{EB}-V_o}{3R}$, and $\frac{V_{com}-V_{EB}-V_o}{4R}$, respectively. In this structure, the output current which is the emitter current of the BJT is amplified by the BJT. Thus, the response time between the initial state and the final state will be smaller because of the larger input current. Nevertheless, the dimension of M1 in the current summation must be designed to have large value to prevent it from saturation due to too large input current. The currents from the voltage-to-current converters of both neighboring cells and itself are summed, flowing through the current mirror to change the current direction, and shifted by a dc bias current. Finally, the current will flow into the state node and the next change in output begins.

There are three disadvantages in the first implementation. The first is that the BJT consume too large area. As can be seen in Fig.10, each input current has one BJT. Furthermore, the feedback current also contains one BJT. The virtual bias operational amplifier consumes six MOSFETs although three of them can be shared as shown in Fig.2. Thus, in the application of large array structure, the complete array will consume too large area. The third disadvantage is that this circuit does not have negative weight.

In order to improve the above disadvantages, the second implementation of one cell circuit is proposed as shown in Fig.11. The output of the neuron is a voltage signal. The voltage-to-current converter is needed to transfer the voltage signal into the current signal. Our approach is to use the common source amplifier. As can be seem in Fig.11, MP1 is the common source amplifier. Because the output current of each cell may be positive or negative, MP2 can be used as a dc current shift. When the drain current of MP1 is larger than that of MP2, the output current is negative. On the other hand, if the drain current of MP1 is smaller than that of MP2, the output current will be positive. The weight is realized by adjusting the resistance of the tunable resistor and the weighted current will flow into the neighboring cells or flow as the feedback current. The currents flowing into the cell are summed at the base of the BJT and shifted a dc bias current in order to have the negative current flowing into the cell. This current is amplified by the BJT with a factor of (β+1), flows through a current mirror to change the flowing direction, and then is shifted by a dc bias current. This implementation of the voltage-to-current converter and the current summation has smaller area in one cell, and thus can be used in application with a large array.

Simulation Result

Since the input pictures usually come from the real world through a camera or some other optical equipment, there will always be some noise superimposed on the images of the objects. A small array of 5×5 cells with the function of noise removal is shown below. The black pixel "■" is defined as 1 (high level), and the white pixel "□" is defined as 0 (low level). The weights for noise removal are shown as follows.

0	1	0
1	2	1
0	1	0

The second type circuit of the cellular neural network shown in Fig.11 is adopted for this application. The simple Chinese characters of "一", "二", "五", "七", "工", "土", "上", "止", "山", and "日" are used in noise removal function. At time zero, the correct character with the noise is applied to the state nodes of the CNN as the initial state. After the transient time, the CNN reaches a steady state. The noise will be eliminated and the correct character is recognized. The simulation result is shown in Figs.12(a)-(j). Other simulations are shown in Figs.13(a) and (b).

A large array of 16×16 cells for noise removal function is also simulated. The weights are the same as the 5×5 cells described above. The Chinese character of "王" with a large amount of noise is used for demonstration. The initial state is shown in Fig.14(a). After the transient time, the correct character is recognized as shown in Fig.14(b). The HSPICE simulation results of node 50 and node 92 are shown in Figs.15(a) and (b), respectively.

The shadow detector proposed by Matsumoto, Chua, and Suzuki [7] creates the "shadow" of an object in an image. The feedback operator is obtained as

0	1	1

and the control operator is zero [8]. The second type circuit of the cellular neural network is adopted to implement the shadow detection. The simulation results are shown in Figs.16(a)-(e). The HSPICE simulation result of Fig.16(b) is shown in Fig.17. Finally, the shadow detection of the character "王" is shown in Fig.18.

Conclusions

A new cell structure based on the neuron-bipolar junction transistor (vBJT) is proposed and fabricated in 0.6μm single-poly-triple-metal n-well CMOS process. This

new CNN cell not only reduces the hardware complexity, but also increases the efficiency of the chip area of the integrated circuit. Functions of noise removal and shadow detection have been implemented by the new CNN structure through HSPICE simulation results.

References

[1] L. Yang, L. O. Chua, and K. R. Krieg, "VLSI Implementation of Cellular Neural Networks," *IEEE International Symposium on Circuits and Systems*, vol. 3, pp. 2425-2427, 1990.

[2] H. Harrer at al., "An Analog Implementation of Discrete-Time Cellular Neural Networks'" *IEEE Trans. Neural Networks*, vol. 3, pp. 466-476, 1992.

[3] J. M. Cruz and L. O. Chua, "A CNN Chip for Connected Component Detection," *IEEE Trans. on Circuits and Systems*, vol. 38, pp. 812-817, 1991.

[4] E. Y. Chou, B. J. Sheu, T. H. Wu, and R. C. Chang, "VLSI Design of Densely-Connected Array Processors," *IEEE International Conference on Computer Design : VLSI in Computers and Processors*, pp. 492-497, 1995.

[5] G. F. D. Betta, S. Graffi, Z. M. Kovacs, and G. Masetti, "CMOS Implementation of an Analogically Programmable Cellular Neural Network," *IEEE Trans. on Circuits and Systems*, vol. 40, pp. 206-215, 1993.

[6] C. Y. Wu and W. C. Yen, "The Neuron-Bipolar Junction Transistor(vBJT) – A New Device Structure for VLSI Neural Network Implementation," *IEEE International Conference on Electronics, Circuits and Systems*, vol. 3, pp. 277-280, 1998.

[7] T. Matsumoto, L. O. Chua, and H. Suzuki, "CNN Cloning Template : Shadow Detector," *IEEE Trans. on Circuits and Systems*, vol. 37. pp. 638-640, 1990.

[8] Y. Y. Lin, "The Expandable Analog VLSI Implementation of Cellular Neural Networks," *Master Thesis, Institute of Electronics College of Electrical Engineering, National Chiao Tung University*, 1993.

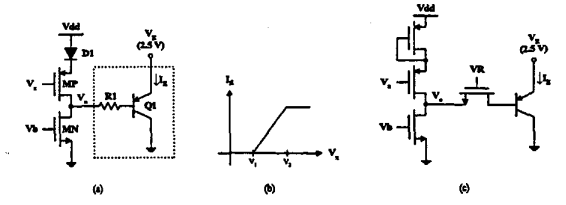

Fig.1(a) The circuit of the current-mode neuron.
(b) The transfer curve of the circuit shown in (a).
(c) The complete MOS implementation of the neuron.

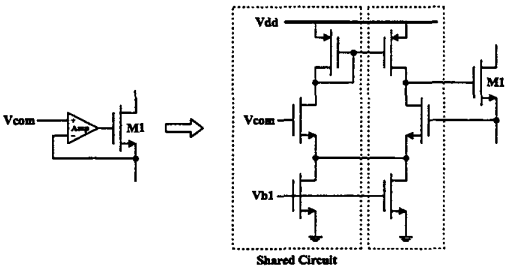

Fig.2 The structure of the operational amplifier.

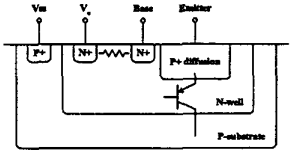

Fig.3 The cross-sectional view of a neuron-bipolar junction transistor.

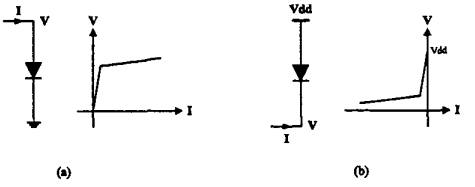

Fig.4 (a) The circuit and transfer curve of the diode which can realize the right side of the sigmoid function.
(b) The circuit and transfer curve of the diode which can realize the left side of the sigmoid function.

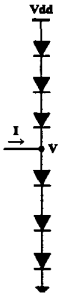

Fig.5 The realization of the neuron using the diode structure.

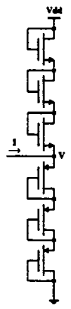

Fig.6 MOS implementation of the circuit shown in Fig.5.

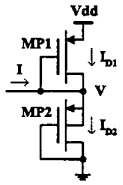

Fig.7 The MOS implementation of a resistor.

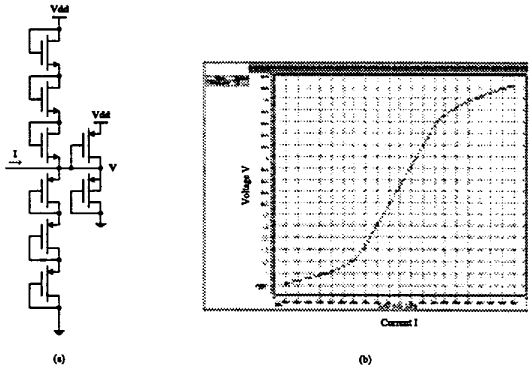

Fig.8 (a) The complete MOS implementation of the neuron.
(b) The transfer curve of the circuit shown in (a).

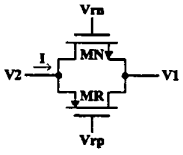

Fig.9 The circuit of a tunable MOS resistor.

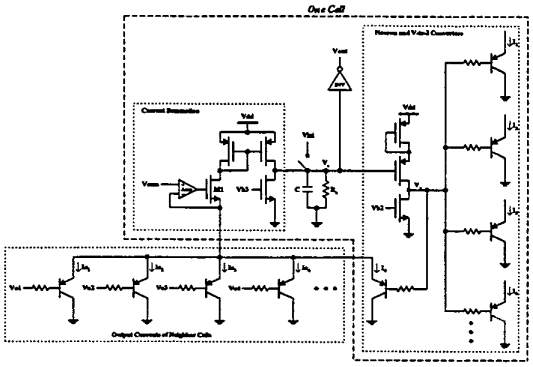

Fig.10 The first implementation of the one cell circuit.

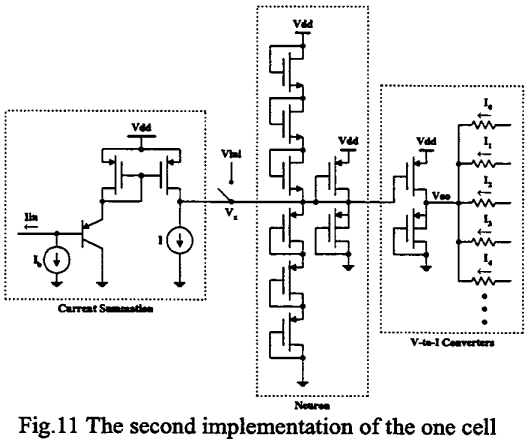

Fig.11 The second implementation of the one cell circuit.

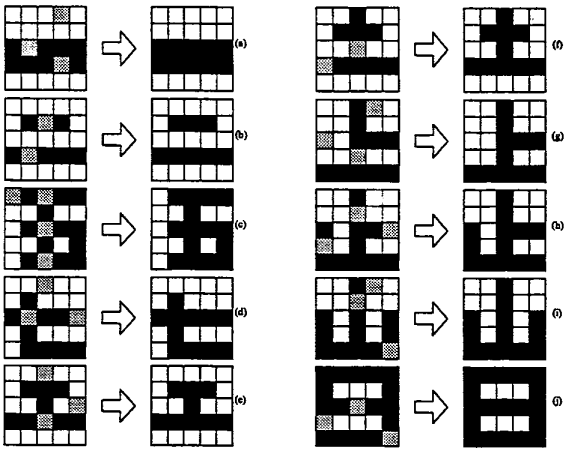

Fig.12 The input noisy images of Chinese characters "一", "二", "五", "七", "工", "土", "上", "止", "山", and "日" to the CNN for noise removal and their final outputs.

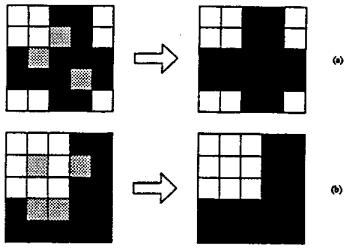

Fig.13 The input noisy images to the CNN for noise removal and their final outputs.

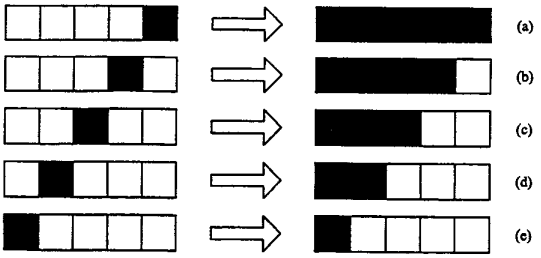

Fig.16 The input states of the CNN for shadow detection and their final outputs.

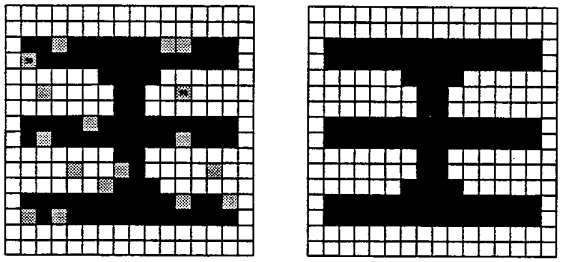

Fig.14 (a) The input noisy image of Chinese character "王" to the CNN for noise removal.
(b) The final output of the CNN.

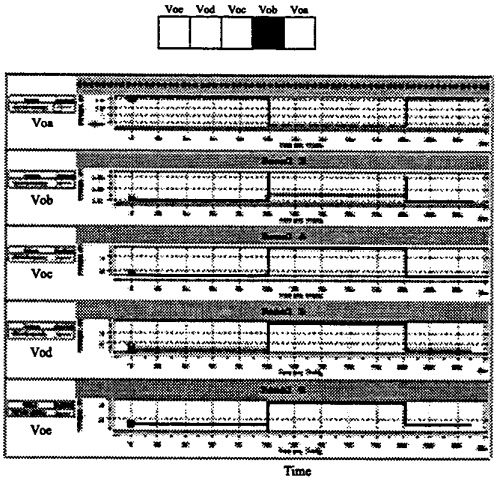

Fig.17 The transient waveforms of the shadow detection shown in Fig.18(b).

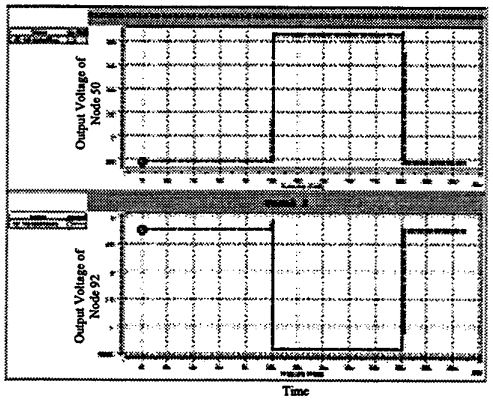

Fig.15 The transient waveforms of node 50 and node 92 in Fig.16(a).

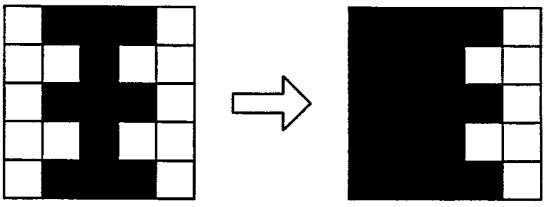

Fig.18 The character "王" and its final output for shadow detection.