A New Compact Neuron-Bipolar Junction Transistor (νBJT) Cellular Neural Network (CNN) Structure with Programmable Large Neighborhood Symmetric Templates for Image Processing

Chung-Yu Wu, *Fellow, IEEE,* and Wen-Cheng Yen

Abstract—Based on the basic device physics of the neuron–bipolar junction transistor (νBJT), a new compact cellular neural network (CNN) structure called the νBJT CNN is proposed and analyzed. In the νBJT CNN, both νBJT and lambda bipolar transistor realized by parasitic p-n-p BJTs in the CMOS process are used to implement the neuron whereas the coupling MOS resistors are used to realize the symmetric synapse weights among various neurons. Thus it has the advantages of small chip area and high integration capability. Moreover, the proposed symmetric νBJT CNN can be easily designed to achieve large neighborhood without extra interconnection. By adding a metal-layer optical window to the νBJT, the νBJT can be served as the phototransistor, and the νBJT CNN can receive optical images as initial state inputs or external inputs. The correct functions of the νBJT CNNs in noise removal, hole filling, and erosion have been successfully verified in HSPICE simulation. An experimental chip containing a 32×32 νBJT CNN and a 16×16 νBJT CNN with phototransistor design, has been designed and fabricated in 0.6-μm single-poly triple-metal n-well CMOS technology. The fabricated chips have the cell state transition time of 0.8 μs and the static power consumption of 60 μW/cell. The area density can be as high as 1270 cells/mm^2. The measurement results have also confirmed the correct functions of the proposed νBJT CNNs.

Index Terms—Cellular neural network, νBJT, large neighborhood.

I. INTRODUCTION

THE cellular neural network (CNN) proposed by Chua and Yang [1], is a special type of analog nonlinear processor array. Due to its continuous-time dynamics and parallel-processing feature, the CNN is very effective in real-time image processing applications such as noise removal, edge and corner detection, hole filling, connected component detection, shadowing, etc. Moreover, regularity, parallelism, and local connectivity in the CNN circuit architecture make it suitable for very large scale integration (VLSI) implementation. So far, several application-dedicated analog CMOS CNN chips with programmable template [2]–[9] or fixed template [10]–[12] have been reported.

It is known that VLSI implementation of neural networks has been a very interesting and challenging research area, which can enhance the performance of neural networks for various applications. To efficiently simplify the VLSI neural network structure for large-size network implementation on a single chip, some effort has been contributed to implement neural network functions using the basic physical characteristics of CMOS or bipolar devices [13]–[19]. Two basic device structures based on this approach have been proposed. One is the neuron-MOS (νMOS) device [13]. The other is the neuron–bipolar junction transistor (νBJT) [17]–[19]. In the neuron–bipolar device, basic neural functions are realized by the BJTs with multiple base terminals separated by base resistances. It has been applied to the implementation of Hamming neural network [17] and CNNs [18], [19].

In many CNN applications of image halftoning [20] and subcortical visual pathway [21], [22], the templates with more than one neighborhood, i.e., $r > 1$, are required. To realize large-neighborhood templates in CNN structures, template decomposition methods [23], [24] have been proposed to decompose them into several smaller single-neighborhood templates which can be implemented on CNN universal machine (CNNUM) [23], [25] or discrete-time CNN (DTCNN) through multiple CNN operations [24]. Generally, it is difficult to directly implement the large-neighborhood templates through single CNN operation.

In this paper, a new circuit structure is proposed to compactly implement CNNs with certain types of single- or large-neighborhood symmetric templates [19]. In the new structure called the neuron–bipolar CNN or νBJT CNN [19], the νBJTs are used as the neurons with the emitter current as the neuron output whereas the base resistances connected among the base terminals of νBJTs and realized by MOS devices, are used to realize the symmetric synapses in the A-template [1]. Due to the compact structure, the νBJT CNN has small chip area and high integration capability. In the νBJT CNN, the synapse values in the template can be adjusted through the gate voltages of MOS devices. The self-feedback function is compactly realized by incorporating a pMOS transistor with the νBJT. The resultant structure
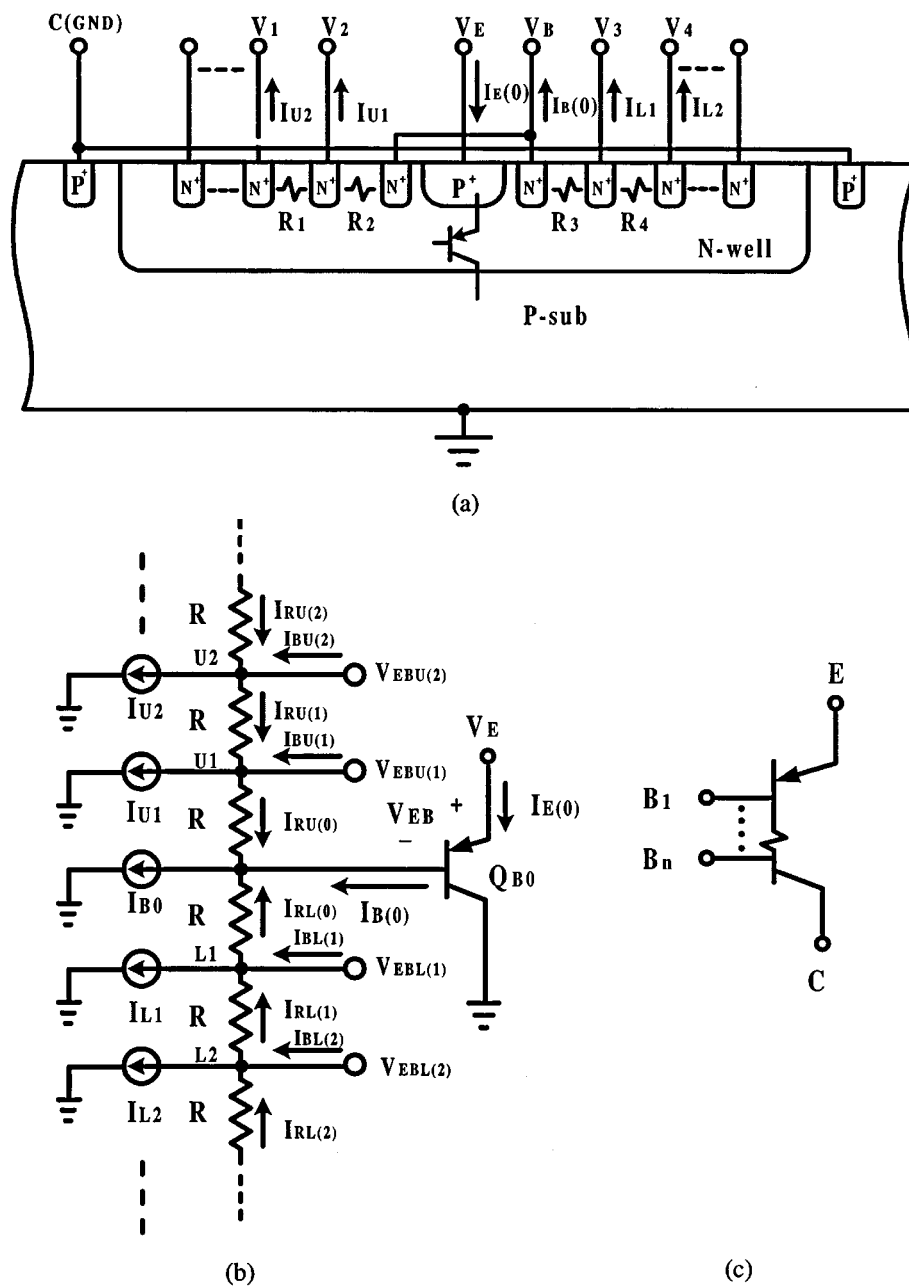
Fig. 1.   (a) The cross-sectional view. (b) The equivalent circuit. (c) The device symbol of the proposed neuron–bipolar junction transistor (νBJT).

is similar to that of the lambda bipolar transistor [26] and has a small chip area. The neuron input can be applied to the base of νBJT through the nMOS transistors. Since the neurons are realized by the νBJTs which can also be served as the phototransistors, the optical images can be input directly to the νBJT CNN without adding any extra sensor device. As the demonstrative examples on the applications of νBJT CNNs, the functions of noise removing, hole filling, and erosion have been successfully realized and verified.

In Section II, the structure of νBJT is described. In Section III, the VLSI implementation of symmetric νBJT CNN structures with single or large neighborhood are analyzed. Some application examples are also demonstrated for verification. In Section IV, the experimental results are presented. Finally, the conclusion is given.

II. NEURON–BIPOLAR JUNCTION TRANSISTOR (νBJT) STRUCTURE

The cross-sectional view and the equivalent circuit structure of the basic BJT realized in the n-well CMOS technology is illustrated in Fig. 1(a) and (b). As shown in Fig. 1(a), the vertical parasitic p^+-n-well-p-substrate p-n-p bipolar junction transistor with the collector biased at ground is used as the neuron. The neuron output signal is the emitter current $I_E(0)$ whereas the neuron state signal is the base voltage $V_B(0)$ or the base current $I_B(0)$. The input currents I_{U1}, I_{U2}, I_{L1}, and I_{L2} representing neuron input signals from external sources or other neurons, are applied to the four base terminals in the n-well base spreading resistance array R_1 to R_4. Thus the multi-input neuron structure can be compactly realized by simply extending

the base diffusion region. When all the input currents are zero, the standby base current I_{BO} keeps the νBJT in the active region. The input currents which may be positive or negative, are summed together with their synaptic weights at the base node to drive the νBJT to conducting or off region. The symbol of the νBJT is shown in Fig. 1(c). Since the basic operational principle of the νBJT is based on the majority carrier transportation of the BJT [15], the realized neuron structure becomes compact without complicated interconnection.

In the equivalent circuit of Fig. 1(b), the base of the νBJT is driven by I_{B0}, I_{U1}, I_{U2}, I_{L1}, and I_{L2} through the spreading resistance array. To develop a simple analytical model for the synaptic weights of νBJT, one-dimensional (1-D) uniform resistor array with the same resistance R is considered as shown in Fig. 1(b). Based on the theoretical model in [27] and [28] and some fundamental assumptions, the current $I_R(n)$ flowing through R at the nth node is derived in Appendix B. In Fig. 1(b), if the only excitation is I_{U1} with all other current-source excitations equal to zero, the contribution of I_{U1} to $I_B(0)$ can be expressed by using (B7) and (B12) in Appendix B as

$$I_B(0) \cong \frac{I_{U1}}{2}\left[1 - \frac{\tan\left(Z_{U1}\dfrac{N-1}{N}\right)}{\tan Z_{U1}}\right] \qquad (1)$$

where

$$Z_{U1} \equiv C_{U1}\frac{RN}{2V_T}$$

$$C_{U1} \equiv \frac{I_{U1}/2}{\tan\left(C_{U1}\dfrac{RN}{2V_T}\right)}$$

N is the total number of resistors R, and V_T is the thermal voltage. Similarly, the contribution of I_{U2} to $I_B(0)$ can be written as

$$I_B(0) \cong \frac{I_{U2}}{2}\left[\frac{\tan\left(Z_{U2}\dfrac{N-1}{N}\right) - \tan\left(Z_{U2}\dfrac{N-2}{N}\right)}{\tan Z_{U2}}\right]. \qquad (2)$$

By using the linear superposition principle and generalizing the expression, $I_B(0)$ can be approximated by

$$I_B(0) \cong I_{B0}$$
$$+ \frac{I_{Ui}}{2}\left[\frac{\tan\left(Z_{Ui}\dfrac{N-i+1}{N}\right) - \tan\left(Z_{Ui}\dfrac{N-i}{N}\right)}{\tan Z_{Ui}}\right]$$
$$+ \frac{I_{Li}}{2}\left[\frac{\tan\left(Z_{Li}\dfrac{N-i+1}{N}\right) - \tan\left(Z_{Li}\dfrac{N-i}{N}\right)}{\tan Z_{Li}}\right],$$
$$i = 1, 2, \ldots, N. \qquad (3)$$

Generally, the factor

$$\frac{\tan\left(Z_i\dfrac{N-i+1}{N}\right) - \tan\left(Z_i\dfrac{N-i}{N}\right)}{\tan Z_i}$$

is smaller for larger i and the contribution of I_{U2} to $I_B(0)$ becomes smaller than that of I_{U1}. This means that farther current excitation from Q_{B0} has smaller contribution to $I_B(0)$. The above degradation effects become more significant for larger R.

Besides receiving the inputs from other neurons, the νBJT in Fig. 1(b) can send its output current via the base node to other neurons as well. Applying the same theoretical model in Appendix B, the currents sent to the neurons at the nodes $U2$ and $U1$ are

$$I_{BU}(i) = [I_{B0} - I_B(0)]$$
$$\cdot \left[\frac{\tan\left(Z_0\dfrac{N-i+1}{N}\right) - \tan\left(Z_0\dfrac{N-i}{N}\right)}{\tan Z_0}\right]$$
$$i = 1, 2, \ldots, N-1 \qquad (4)$$

where

$$Z_0 \equiv C_0\frac{RN}{2V_T}$$

and

$$C_0 \equiv \frac{[I_{B0} - I_B(0)]/2}{\tan\left(C_0\dfrac{RN}{2V_T}\right)}.$$

As discussed before, $I_{BU}(i)$ is smaller for larger i.

From (3) and (4), it can be realized that the factor

$$\left[\frac{\tan\left(Z_i\dfrac{N-i+1}{N}\right) - \tan\left(Z_i\dfrac{N-i}{N}\right)}{\tan Z_i}\right]$$

is equivalent to the synaptic weight in the neuron. Since Z_i is dependent on C_i which is a nonlinear function of I_i, the value of weighting factor is also dependent on I_i. In the νBJT application on the CNN with $r = 1$, about 2.5 μA is chosen for I_i for $i = 1$ to realize the template coefficients. During the CNN operation period from the beginning to the point that all the transition neurons move across their critical states toward the final stable states, the change of I_i is within 28% which causes the variations of the synapse weighting factor being within 5% for $N = 16$ and $R = 230\,\text{k}\Omega$. Once the transition neurons pass the critical states, the template coefficients have no effects on the neuron states. In the νBJT application on the CNN with $r = 2$, the variation of I_i for $i = 2$ from 0.1 to 0.3 μA. This causes the variations of the synapse weighting factor being within 10% for $N = 16$ and $R = 230\,\text{k}\Omega$. The below 10% variations of template coefficients are tolerable in the νBJT CNN applications.

As may be seen from (3), the summation of the weighted inputs from other neurons is performed at the base node in the current mode. Moreover, the input excitation currents I_{Ui} and I_{Li} from farther neurons still can reach the excited neuron across
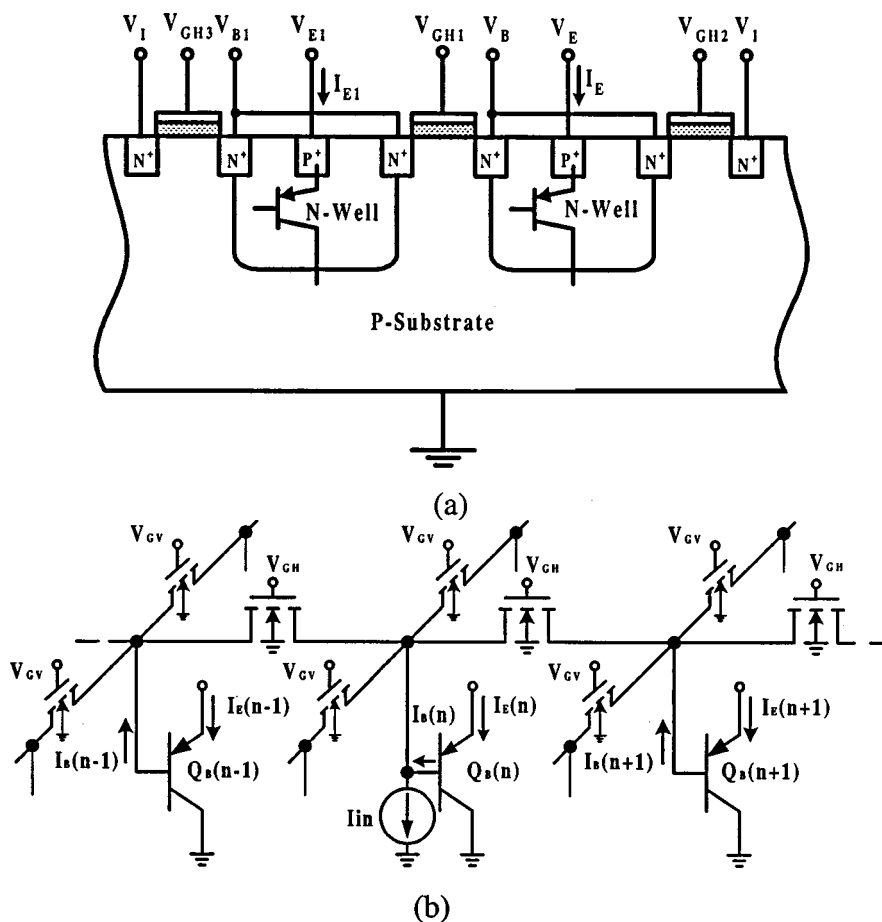
Fig. 2. (a) The cross-sectional view and (b) the equivalent circuit of the improved νBJT structure which uses the enhancement nMOSFETs to realize the base resistance array.

the nearest neuron without extra direct interconnection. Similarly, the neuron can send its weighted output currents via the base node to other neurons as may be seen from (4). For farther neurons, the master neuron still can source its weighted outputs without direct interconnection. This special feature is the major advantage of using a BJT instead of a MOSFET as the basic neuron. It makes the νBJT very suitable for large neural network implementation in VLSI.

To efficiently realize the resistor array of Fig. 1(b) in VLSI, the base spreading resistance is replaced by an enhancement-mode n-channel MOSFET which is inserted between the bases of two parasitic p-n-p BJTs in n-well CMOS process as shown in Fig. 2(a) [27]. Through the control of the gate voltages V_{GHi} and V_{GVi}, the inserted nMOSFET can be operated in either strong inversion region or subthreshold region to provide a wide range of resistance values to achieve the wide-range adjustment of synapse weights. Generally, the proposed νBJT structure in Fig. 2(a) has a smaller chip area than that in Fig. 1(a). The equivalent circuit of Fig. 2(a) is shown in Fig. 2(b) where the input current I_{in} which is applied to the base of the νBJT $Q_B(n)$, represents either initial state input or external input currents to the neuron.

To verify the characteristics of the νBJT of Fig. 2(a), an experimental chip of 98 $\times$ 1 νBJT array was designed and fabricated by 0.5 μm double-poly double-metal (DPDM) n-well
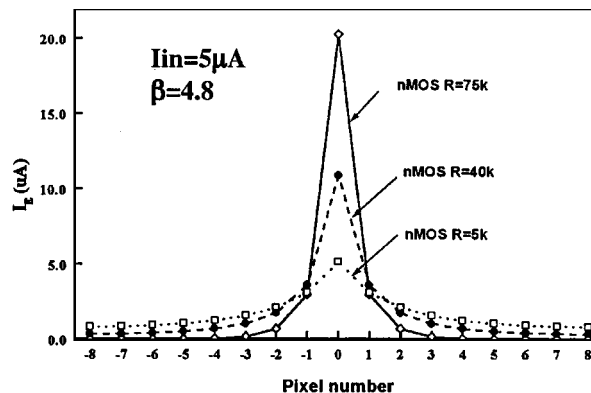

Fig. 3. The measured results of the fabricated 98 $\times$ 1 νBJT array with $I_{in} = 5\,\mu$A and $\beta = 4.8$ for different coupling MOS resistance values.

CMOS technology. A current source of 5 μA is applied to the base of one νBJT in the array. Fig. 3 shows the measured results of the emitter current I_E of each νBJT versus pixel position for different coupling MOS resistance values under the single-point stimulus of 5 μA. It can be seen that larger coupling resistance leads to faster decreasing rate of I_E and less effect of the stimulus on the father νBJTs. This means that the stimulus has no effect on farther BJTs if the coupling resistance is large enough. Thus the coupling resistor can be used to control the connected layers of neighborhood neurons in the CNN.
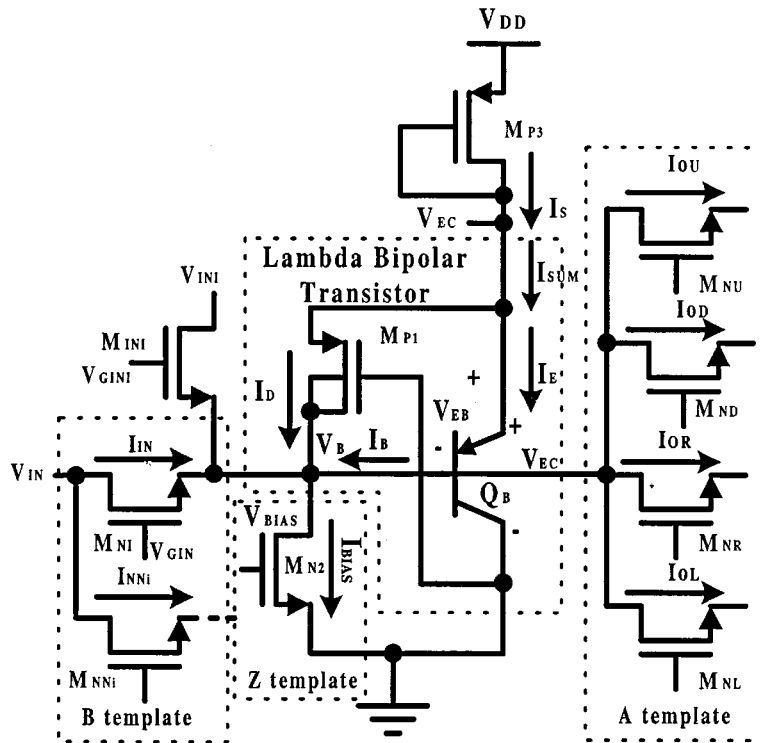
Fig. 4.   The complete cell circuit of one $\nu\lambda$BJT neuron in the νBJT CNN.

III. SYMMETRIC νBJT CNN STRUCTURES

A. νBJT CNN with Single Neighborhood

The basic cell circuit of the νBJT CNN is shown in Fig. 4 where the neuron is realized by the νBJT Q_B with the nMOS transistor M_{N2} biased by the gate voltage V_{BIAS} to generate the standby base current $I_{BO} = I_{\mathrm{BIAS}}$. Such a neuron is called the νBJT neuron. The neuron output current I_E flows through the load pMOS device M_{P3} to generate the neuron output voltage V_{EC}. The neuron state voltage is the base voltage V_B. The HSPICE simulated neuron output voltage V_{EC} versus neuron state voltage V_B is shown in Fig. 5. This transfer characteristic curve is similar to that in [1] except that a small nonlinearity exists. For different I_{BIAS}, V_B is different. Thus the current I_{BIAS} can also be used to realize the Z template [1] as will be described later.

In the νBJT neuron of Fig. 4, M_{P1} provides a positive feedback to Q_B so that the negative resistance is generated and the neuron has two stable states. Thus the νBJT CNN formed by $\nu\lambda$BJT neurons is of the monotonic binary-valued CNNs [29].

The self-feedback synapse in the CNN is realized by using the positive-feedback pMOS transistor M_{P1} with gate connected to ground and source (drain) connected to emitter (base) of Q_B. The structure of Q_B and M_{P1} is called the lambda bipolar transistor as proposed in [26]. In realizing the lambda bipolar transistor, M_{P1} can be compactly implemented in the n-well base region with its source shared with the emitter of Q_B and its n-well substrate with the base. Thus the substrate of M_{P1} is connected to its drain and the positive substrate bias exists [26]. Since the neuron structure combines νBJT with lambda bipolar transistor, it can be called the neuron–lambda–BJT neuron or $\nu\lambda$BJT neuron. As shown in Fig. 4, the input capacitance of the
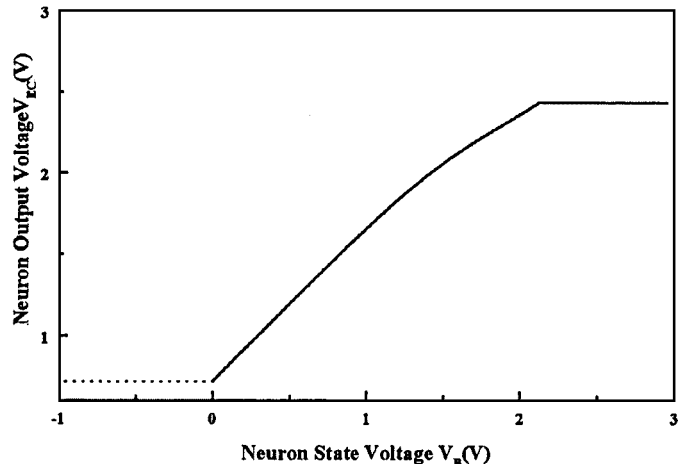

Fig. 5.   The transfer characteristic of neuron output voltage V_{EC} versus neuron state voltage V_B.

$\nu\lambda$BJT neuron is the capacitance seen at the base node, which is dominated by the base–emitter junction capacitance. The input resistance is the resistance seen at the base node, which is the input resistance of Q_B in parallel with the output resistance of M_{P1}.

The HSPICE simulated I_E–V_{EC} characteristic of the lambda bipolar transistor is shown in Fig. 6 where the curves of I_{BIAS} and I_D versus V_{EC} are also plotted. In the I_E–V_{EC} characteristic, I_E is equal to zero when V_{EC} is smaller than 0.6 V. In this case, Q_B and M_{P1} are off and I_{BIAS} is forced to zero. When V_{EC} is larger than 0.6 V, I_{BIAS} is greater than I_D and Q_B is turned on with I_E increased with V_{EC}. When V_{EC} is larger than the peak voltage V'_{ECP}, the increase of I_D is greater than that
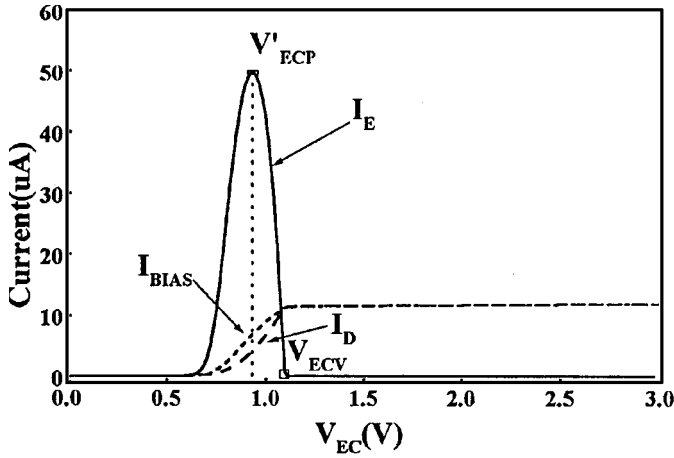
Fig. 6.   The HSPICE simulated currents I_E, I_{BIAS}, and I_D versus the voltage V_{EC} in the p-n-p lambda bipolar transistor.
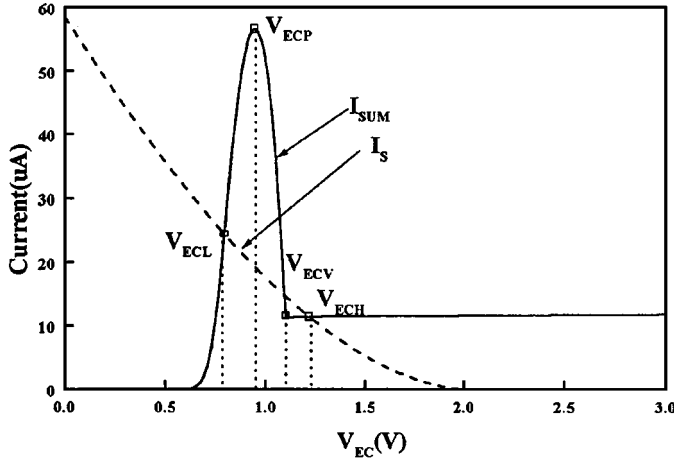

Fig. 7.   The HSPICE simulated transfer curves of the currents I_{SUM} and I_S versus the emitter voltage V_{EC} in the νBJT neuron with p-n-p lambda BJT.

of I_{BIAS} and thus both I_B and I_E are decreased with V_{EC}, creating a negative-resistance region. When V_{EC} is larger than the valley voltage V_{ECV}, I_D is equal to I_{BIAS} and Q_B is turned off with $I_E = 0$. It can be seen from Fig. 6 that the $\nu\lambda$BJT neuron has one stable state in the region 0.7 V $< V_{EC} < V'_{ECP}$ with Q_B ON and the other in the region $V_{EC} > V_{ECV}$ with Q_B OFF. Moreover, the self-feedback current I_D is proportional to the neuron output voltage V_{EC} between 0.6 V and V_{ECV}. But I_D is not linearly proportional to V_{EC} as in [1]. Since the νBJT CNN is a monotonic binary-valued CNN, the nonlinearities in both I_D–V_{EC} and neuron transfer characteristic of Fig. 5 are tolerable. Due to the local stability, the νBJT CNN can guarantee functionality [29].

The HSPICE simulated characteristics of the currents I_{SUM} and I_S in the $\nu\lambda$BJT neuron of Fig. 4 versus the emitter voltage V_{EC} is shown in Fig. 7 where the peak and valley voltages are V_{ECP} and V_{ECV}, respectively. It can be seen from Fig. 7 that the two stable points are located at V_{ECH} and V_{ECL} which are the intersection points of I_S and I_{SUM} in the positive-resistance region of I_{SUM}. In the stable state V_{ECL} (V_{ECH}), the source–gate voltage is low (high) and the self-feedback current I_D to the base is low (high). The corresponding neuron state voltages in both states are V_{BL} and V_{BH}. For $I_{\mathrm{BIAS}} = 12\,\mu\mathrm{A}$

and $V_{DD} = 3$ V, we have $V_{ECH} = 1.21$ V and $V_{ECL} = 0.80$ V from the HSPICE simulation.

The peak and valley voltages in the I_{SUM}-V_{EC} characteristic curve are important parameters. They can be expressed in terms of device parameters. At the peak voltage, Q_B is operated in the active region, M_{P1} is operated in the saturation region, and M_{N2} is operated in the linear region. I_D, I_{BIAS}, and I_{SUM} can be written as

$$I_D = K_p[(V_{EC} - |V'_{TP1}|)^2] \tag{5}$$

$$I_{\mathrm{BIAS}} = K_n[2(V_{\mathrm{BIAS}} - V_{TN2})(V_{EC} - V_{EB})$$
$$- (V_{EC} - V_{EB})^2] \tag{6}$$

$$I_{\mathrm{SUM}} = I_E + I_D = (1+\beta)I_B + I_D$$
$$= (1+\beta)(I_{\mathrm{BIAS}} - I_D) + I_D$$
$$= (1+\beta)I_{\mathrm{BIAS}} - \beta I_D \tag{7}$$

where K_n and K_p are given by

$$K_n = 1/2\,\mu_n C_{OX}(W/L)_n, K_p = 1/2\,\mu_p C_{OX}(W/L)_P.$$

In the above equations, $\mu_n(\mu_p)$ is electron (hole) mobility, C_{OX} is the capacitance per unit area, L is the channel length, W is the channel width, V'_{TP1} is the threshold voltage of M_{P1} under positive substrate bias V_{EB}, V_{TN2} is the threshold voltage of M_{N2}, and V_{EB} is the emitter–base voltage of Q_B. The peak voltage V_{ECP} is determined by the maximum point of I_{SUM}, which can be calculated from conditions

$$\frac{\partial I_{\mathrm{SUM}}}{\partial V_{EC}} = 0$$

and

$$\frac{\partial^2 I_{\mathrm{SUM}}}{\partial^2 V_{EC}} < 0.$$

By using (5)–(7) and assuming a constant β, V_{ECP} can be calculated as

$$V_{ECP}$$
$$= \frac{(1+\beta)K_{n2}(V_{\mathrm{BIAS}} - V_{TN2} + V_{EB}) + \beta K_{p1}|V'_{TP1}|}{(1+\beta)K_{n2} + \beta K_{p1}}. \tag{8}$$

From (8), it can be seen that V_{ECP} can be controlled by the ratio

$$(W/L)_n/(W/L)_P$$

and V_{BIAS}.

Similarly, the valley voltage V_{ECV} can be derived from the condition $I_D = I_{\mathrm{BIAS}}$ with M_{N2} operated in the saturation region. If I_{BIAS} is known, V_{ECV} can be written as

$$V_{ECV} = \sqrt{\frac{I_{\mathrm{BIAS}}}{K_{P1}}} + |V'_{TP1}|. \tag{9}$$

Substituting the parameter values into (8) and (9), we have $V_{ECP} = 0.93$ V and $V_{ECV} = 1.11$ V, which are consistent with the HSPICE simulate results.

The voltages V_{ECL}, V_{BL}, V_{ECH}, and V_{BH} can be characterized analytically by using the suitable device equations. The detailed derivations are given in Appendix A. With $V_{DD} = 3$

V and $I_{\mathrm{BIAS}} = 12\,\mu$A, the calculated $V_{ECH} = 1.22$ V and $V_{ECL} = 0.74$ V which are close to the HSPICE simulated values.

In Fig. 4, the input voltage V_{IN} of the neuron is sent to the base of Q_B through the nMOS transistor M_{NI}. It can also be sent to other neighboring neurons through the nMOS transistors M_{NNi} as the synapse weight control. In this way, the B template of the CNN can be realized. Using a similar structure, the initial state V_{INI} of the neuron can be sent to the base of Q_B through the nMOS transistor M_{INI} with the gate voltage $V_{GINI} = V_{DD}$. V_{INI} can be taken off from the base of Q_B by turning off M_{INI} with $V_{GINI} = 0$. The standby base voltage V_B is either V_{BL} or V_{BH} depending on the initial input voltage V_{INI}.

Besides the self-feedback, the neuron output current can be sent to the neighboring neurons from the base of νBJT Q_B through the nMOS transistors M_{NU}, M_{ND}, M_{NR}, and M_{NL} as the synapse weight control. Similarly, the outputs of neighboring neurons are sent to the base of Q_B through the same MOS devices and summed there to control the neuron state. The operational principle and basic theoretical model for this structure are described in the previous section. According to the derived model, the symmetric A-template of the CNN can be realized by the nMOS transistors with their gate voltages used to control the synaptic weights of A-template.

The symmetric A-template as realized by the nMOS transistors M_{NU}, M_{ND}, M_{NR}, and M_{NL} in Fig. 4, can be characterized in terms of the currents I_{OU}, I_{OD}, I_{OR}, I_{OL}, and I_D. In the stable state with V_{ECH}, I_B is nearly zero and part of I_D is shared by the currents I_{OU}, I_{OD}, I_{OR}, and I_{OL}. Thus the effective self-feedback current I_P is equal to $I_D - I_{OU} - I_{OD} - I_{OR} - I_{OL} = I_{\mathrm{BIAS}}$ rather than I_D. In this stable state, the required amount of the current I_{TH} to make a transition to the other stable state is $I_{TH} = I_S|_{V_{ECV}} - I_{\mathrm{BIAS}}|_{V_{ECH}}$. Thus the condition for the transition is

$$I_{OU} + I_{OD} + I_{OR} + I_{OL} \cong I_D - I_{\mathrm{BIAS}}|_{V_{ECH}}$$
$$\cong I_S - I_{\mathrm{BIAS}}|_{V_{ECH}} \geq I_{TH}.$$

In the stable state with V_{ECL}, the currents I_{OU}, I_{OD}, I_{OR}, and I_{OL} are either negative or equal to zero. In this case, the effective self-feedback current is equal to I_D which is very small as shown in Fig. 5. In this stable state, the required transition current I_{TL} is

$$I_{TL} = I_{\mathrm{BIAS}}|_{V_{ECP}} - \frac{I_{SP} - I_{DP}}{\beta + 1} - I_D|_{V_{ECP}}$$

where I_{SP} and I_{DP} are the values of I_S and I_D at the peak point $V_{EC} = V_{ECP}$. The condition for the transition is

$$|I_{OU} + I_{OD} + I_{OR} + I_{OL}| \cong I_{\mathrm{BIAS}} - I_B - I_D \geq I_{TL}.$$

To achieve the symmetric transition, the condition $I_{TH} = I_{TL}$ must be satisfied by adjusting I_{BIAS} via V_{BIAS}. In this design, $I_{\mathrm{BIAS}} = 12\,\mu$A is chosen to achieve symmetric transition with $Z = 0$. Decreasing (Increasing) I_{BIAS} leads to a negative (positive) value of Z. The semiempirical relation between I_{BIAS} and Z is $I_{\mathrm{BIAS}} = 12 + 0.6Z$.

0	I_{OU}	0
I_{OL}	I_P	I_{OR}
0	I_{OD}	0

Fig. 8. The synaptic coefficients of the A-template as represented by the currents I_P, I_{OR}, I_{OL}, I_{OU}, and I_{OD}.

TABLE I
SOME CNN TEMPLATES

Application	A	B	Z
Noise removal CNN	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	0
Hole filling CNN	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	-1
Erosion CNN	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	-4.5

From the above analysis, the synaptic coefficients of the A-template can be represented by the self-feedback current I_P and the four neighboring output currents I_{OU}, I_{OD}, I_{OR}, and I_{OL} as shown in Fig. 8. Since the self-feedback current is very small and the currents sent out to the neighboring neurons are much smaller than the input currents from them in the stable state V_{ECL}, the ratios I_P/I_{OU}, I_P/I_{OD}, I_P/I_{OR}, and I_P/I_{OL} are determined in the stable state V_{ECH}. The current ratios can be controlled by adjusting the gate voltages of the corresponding nMOS transistors M_{NU}, M_{ND}, M_{NR}, and M_{NL}, to change their resistances. The relation of currents to resistances can be approximately determined from (3) and (4) in Section II. In the simple structure of Fig. 4, only one nMOS transistor is used to realize the coupling path between two neurons. Thus only symmetric templates with positive coefficient sign can be realized.

The synaptic coefficients of B-template can be represented by the current I_{IN} to the master neuron and the currents I_{NNi} to the neighboring neurons as shown in Fig. 4, which can be adjusted by the corresponding gate voltages. In this way, the synaptic coefficients of B-template must have positive sign.

By using the cell circuit of the $\nu\lambda$BJT neuron of the Fig. 4, a two-dimensional (2-D) νBJT CNN array can be formed. To verify its function, three CNN applications with symmetric templates are tested in the νBJT CNN by using the HSPICE simulation.

In the noise removal CNN, the cloning template is given in the Table I where the central weight is two times larger than its four neighboring weights [1]. This template can be realized by making the self-feedback current I_P two times larger than the four output currents I_{OU}, I_{OD}, I_{OR}, and I_{OL} to the four neighboring cells. This can be achieved by controlling the resistance of nMOS transistors in Fig. 4 through their gate voltages. To implement the noise removal operation, first, the suitable gate
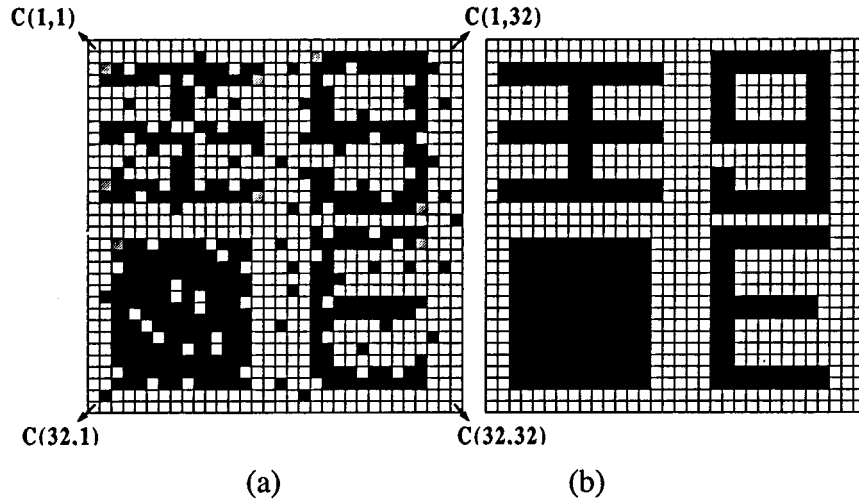
Fig. 9. (a) The initial image and (b) the final output image in the νBJT CNN under the noise removal operation.
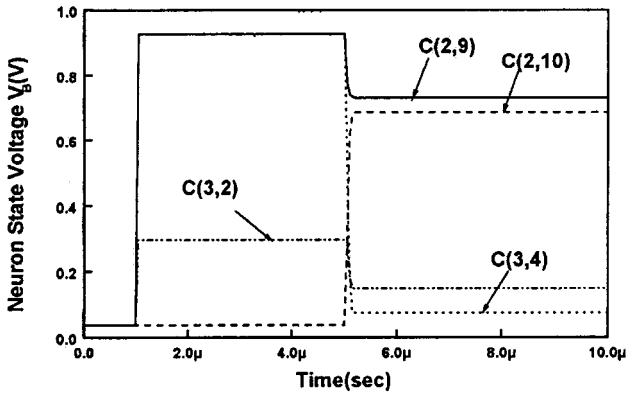

Fig. 10. The transient waveforms of the neuron state voltages V_B in different cells of the νBJT CNN in performing noise removal function.

Fig. 11. (a) The input image and (b) the final output image in the νBJT CNN under the hole filling operation.

voltages are applied to the gate of the MOS transistors realizing the template coefficients. Then the initial image pattern is applied to the input base node of the neuron as the initial condition. Secondly, the initial input is taken away by turning off M_{INI} in Fig. 4 and the νBJT CNN starts its operation. After the transient time, the νBJT CNN can reach a steady state. The transient time is dependent on the resistance and the capacitance in the $\nu\lambda$BJT neuron. The final steady state can be read out by sending out the state voltage V_B through a source follower as the output buffer so that V_B is not disturbed during readout.

Fig. 9(a) shows the initial noisy image used to test the noise removal capability of the proposed νBJT CNN. The image size is 32 × 32 pixels and the νBJT CNN has 32 × 32 cells. The HSPICE simulated output image from the νBJT CNN is shown in Fig. 9(b). It can be seen from Fig. 9(b) that the noise has been eliminated. Fig. 10 shows the HSPICE transient waveforms of neuron state voltages V_B in $C(2,9), C(2,10), C(3,2)$, and $C(3,4)$ cells where the states are kept constant by the initial inputs during 1 to 5 μs.

To test the hole-filling function of the νBJT CNN, both A and B templates [8], [22] in Table I are used. To realize the B-template, the input image is sent to the cell through the nMOS M_{NI}. Its gate voltage V_{GIN} is adjusted to make I_{IN} two times larger than the self-feedback current I_P in
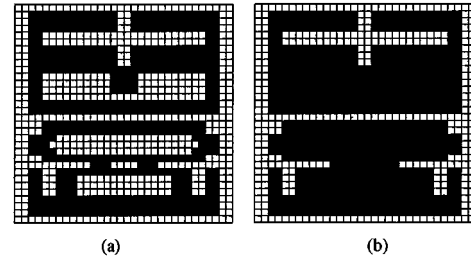
the A-template. $I_{BIAS} = 11.4\,\mu$A is used to realized the Z-template with $Z = -1$. The neuron states are all initialized to the black stable state with $V_B = 0$ V. For the white pixel, $V_B = 0.9$ V. Fig. 11(a) shows the input image containing four holes, which is sent to the νBJT CNN. The output image with the holes filled is shown in Fig. 11(b).

As a third example, the erosion operation is tested in the νBJT CNN. The erosion templates are given in Table I [22]. To implement the B-template, the nMOS transistors M_{NI} and M_{NNi} for $i = 4$ as shown in Fig. 4 should be used. $I_{BIAS} = 9.5\,\mu$A is used to realized the Z-template with $Z = -4.5$. Fig. 12(a) shows the input image used to test the image erosion operation. The initial states is $V_B = 0.4$ V. The HSPICE simulated output image from the νBJT CNN is shown in Fig. 12(b) which verifies the correct function of the νBJT CNN in the erosion operation.

B. νBJT CNN with Phototransistor Design

In the $\nu\lambda$BJT neuron of Fig. 4, the BJT Q_B can be served as the phototransistor by simply using a metal layer to define the optical window and cover the rest area [14]–[16], [27]. With the phototransistor design, the νBJT CNN can use the optical images as its initial state input of the neurons. Since no extra sensor devices are required and the devices associated with initial state input can be saved, the νBJT CNN with phototransistor design has small chip area and high integration capability. Similarly, the same νBJT CNN with phototransistor design can use the optical images directly as its external input if only the self-feedback coefficient exists in the B-template. The optical external
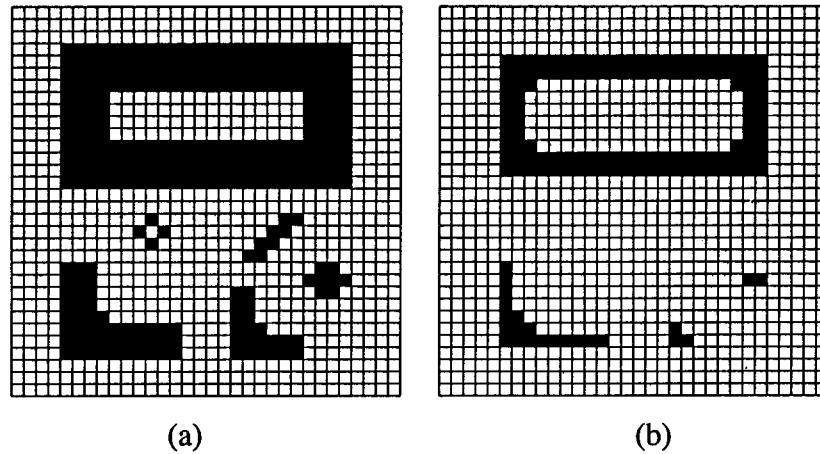
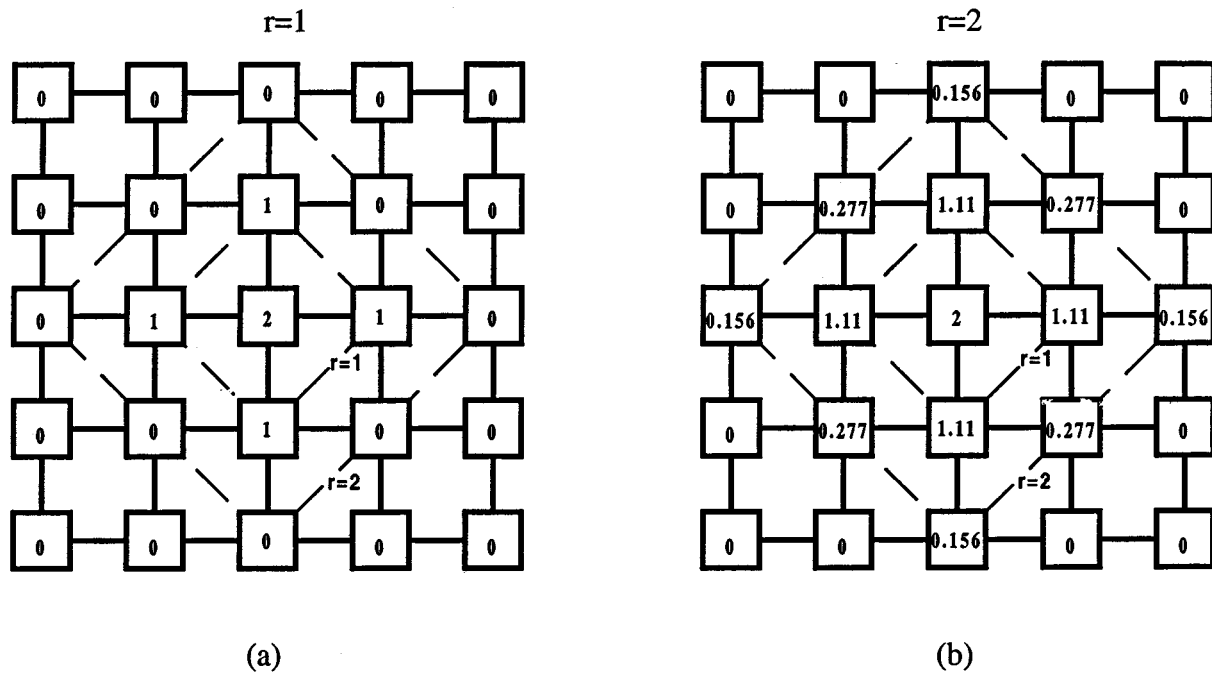Fig. 12.   (a) The input image and (b) the output image in the νBJT CNN under the image erosion operation.

Fig. 13.   The 7×7 template with number r of connected neighborhood equal to: (a) 1; (b) 2; and (c) 3.

input image is applied to the CNN right after turning off the optical initial-state input image. For larger self-feedback B-template coefficient, higher light intensity is used. If more than one coefficient exist in the B-template, another phototransistor is required.

C. νBJT CNN with Large Neighborhood

As shown in Fig. 3 and derived in (2) and (3), smaller coupling resistors lead to slower decreasing rate of the currents sending from one neuron to other neurons. Thus the farther neurons can receive the current from the master neuron through its neighboring neuron without extra interconnection. Based upon the above principle, the coupling resistor can be used to control the connected layers of neighboring neurons in the CNN. Fig. 13(a) and (b) shows the A-templates for the noise removal image processing with the number of neighborhood layers $r = 1$ and $r = 2$, respectively. In $r = 2$ template, the synaptic coef-

ficients decrease with the distance from the central coefficient. In the template with $r = 2$, the synaptic coefficients are determined from the output current of a neuron in the high stable state (white) to the first-neighborhood neuron in the transition point from low to high stable state and to the second-neighborhood neuron in the low stable state (black). For the template with $r = 2$ given in Fig. 13(b), the self-feedback current of the central neuron, its output current to the first-neighborhood neuron, and that to the second-neighborhood neuron are 4.08, 2.21, and 0.31 μA, respectively. The nMOS devices used to realize the template coefficients have the device dimension $W/L = 1\,\mu$m$/12\,\mu$m. The device voltages are $V_{DS1} = 0.51$ V and $V_{BS1} = -0.12$ V in the first neighborhood layer and $V_{DS2} = 0.07$ V and $V_{BS2} = -0.05$ V in the second layer. Thus the effective coupling resistances are 232 and 237 KΩ, respectively.

Using the A-template with $r = 1$ as shown in Fig. 13(a) and the input noisy image of Fig. 14(a) in the νBJT CNN, the output

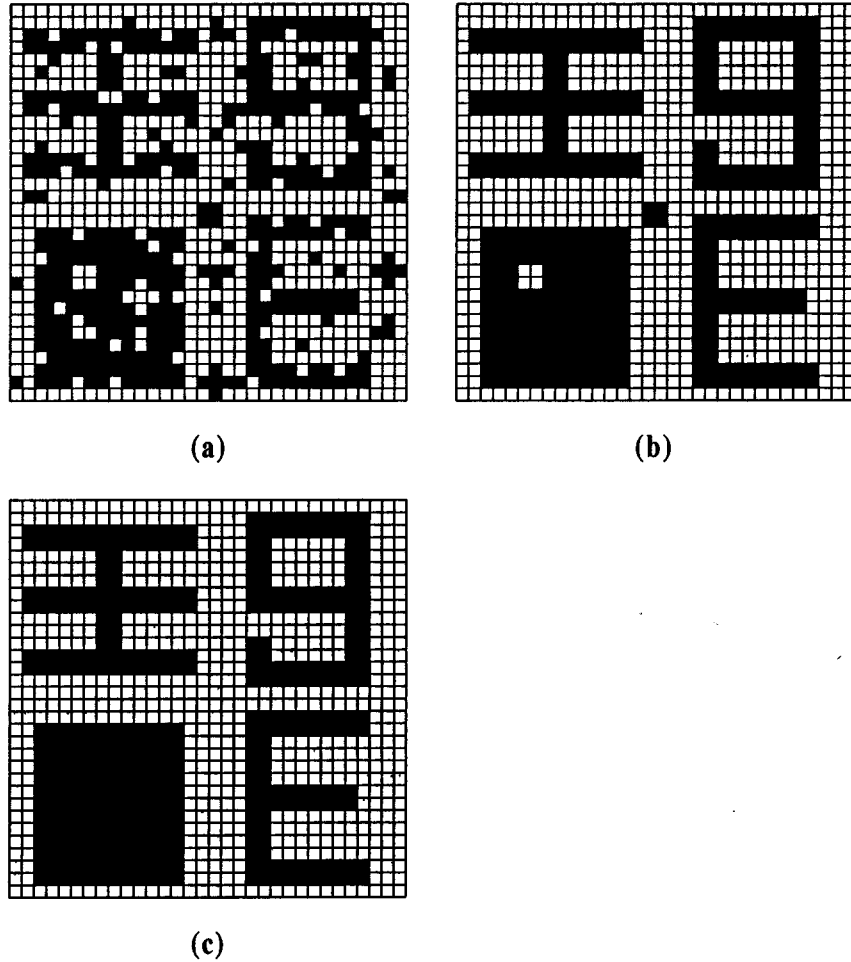(a)   (b)

(c)

Fig. 14.   With (a) the initial state image in the νBJT CNN for noise removal, the resultant output images are shown in (b) for $r = 1$; (c) for $r = 2$; and (d) for $r = 3$.

images is shown in Fig. 14(b) where the 4-pixel square black or white noise images are not removed even if the self-feedback coefficient is reduced from 2 to 1. But these noise images can be removed by using the A-template with $r = 2$ as shown in Fig. 13(b). Since in the A-template with $r = 2$, there is a larger spatial mask of 5×5, thus they have stronger local averaging effects which makes all the white (black) noisy pixels in the local region change to the black (white) ones when the total number of black (white) pixels is larger than that of white (black) pixels. From the above simulation results, it can be seen that the noise removal capability is enhanced for $r > 1$.

In the proposed νBJT CNN, simple MOS resistors are used to realize the A-templates with large neighborhood. Thus the realizable template coefficients in the large neighborhood layers must be smaller and those in the intermediate layers cannot be zero.

IV. EXPERIMENTAL RESULTS

Based on the cell circuits in Fig. 4, an experimental chip of the proposed symmetric νBJT CNNs with the array sizes of 32×32 and 16×16 as well as the 16×16 νBJT CNN with phototransistor design, has been designed and fabricated by using 0.6-μm single-poly triple-metal (SPTM) n-well CMOS technology. Due to its compact structure, a high cell density of 1270 cells per

square millimeter is achieved in the 32×32 νBJT CNN with five A-template coefficients, one B-template coefficient, and Z. Fig. 15 shows a photograph of the fabricated chips of 32×32 νBJT symmetric CNN, 16×16 symmetric νBJT CNN with $r = 2$, and 16×16 symmetric νBJT CNN with phototransistor design. In the 32×32 symmetric νBJT CNN experimental chip, both image noise removal and hole-filling operations are tested.

The image-noise removal function of the fabricated 32×32 νBJT CNN chip has been successfully verified with the fixed initial noisy image of Fig. 9(a) for noise removal and the fixed input image of Fig. 12(a) for hole filling. The fixed initial image is input to the chip simultaneously through M_{INI} as shown in Fig. 4 whereas the fixed input image through V_{IN} and M_{IN}. To read out the neuron state voltage V_B, a source follower is used as the output buffer for each cell. To save the wiring, only 16 cells are read out at a period of 5 μs. The measured characteristics of the 32×32 νBJT CNN experimental chip are summarized in Table II. Fig. 16 shows the measured currents I_{SUM} and I_S versus the voltage V_{EC} in the fabricated p-n-p $\nu\lambda$BJT neuron. Due to fabricated process variations, about 10% deviation between SPICE simulation and measured results is observed. Fig. 17 shows the measured output waveforms of the neuron state voltage V_B in the cells $C(2,9), C(2,10), C(3,2)$, and $C(3,4)$ cells with the initial noisy image of the Fig. 9(a).
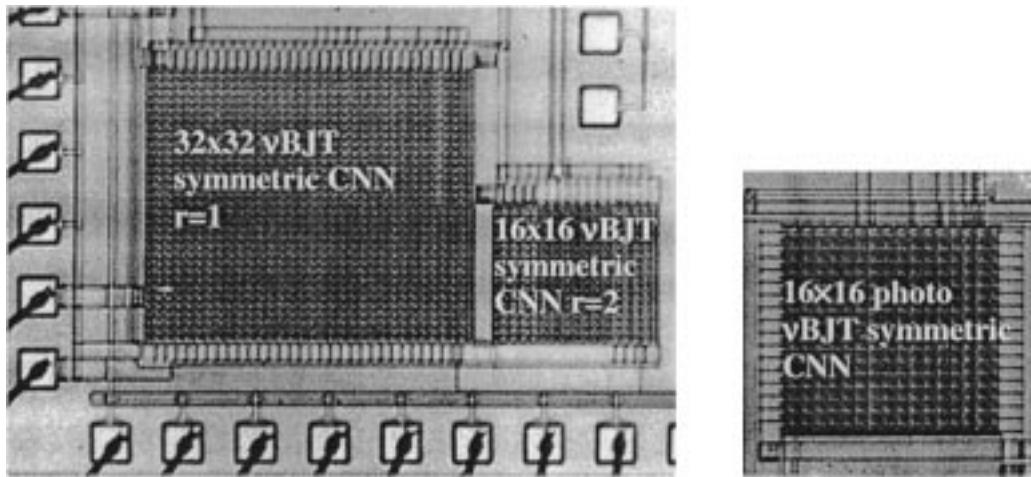
Fig. 15.   The chip photograph of 32×32 νBJT CNN and 16×16 νBJT CNN with phototransistor design.

TABLE II
THE SUMMARY ON THE CHARACTERISTICS OF THE FABRICATED
νBJT CNN CHIP

| Technology	0.6μm Single Poly Triple Metal N-well CMOS
Resolution	32×32 cells
Single pixel area	22μm $\times 25\mu$m
vBJT Q_B and pMOS M_{P1} area	15μm $\times 15\mu$m
CNN array size(not including pad)	850μm $\times 980\mu$ m
Power supply	3V
Total quiescent power dissipation	60mW
Dynamic power dissipation of the array	55mW ~ 75mW (Depending on image input current)
Current gain of BJTs	17.5
State transition time	0.8μs
Minimum readout time of a pixel	1 μs
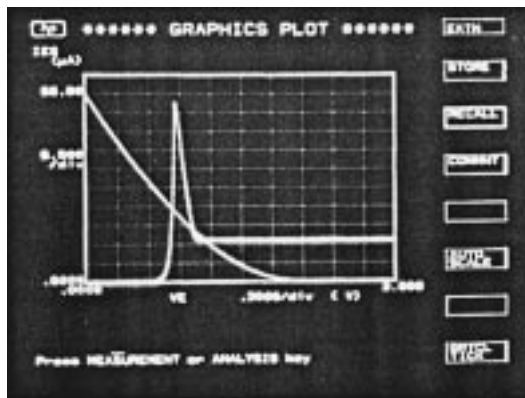
Fig. 16.   The measured currents I_{SUM} and I_S versus the voltage V_{EC} in the fabricated p-n-p $\nu\lambda$BJT neuron.

Fig. 17.   The measured waveforms of the neuron state voltage V_B in the νBJT CNN under noise removal operation.

Fig. 18.   The measured emitter current I_E of the fabricated bipolar phototransistor with the light illumination turned off during the sweep of V_{EC}.

It can be seen from Fig. 9(a) that the state transition time of the cell is 0.8 μs. Thus the minimum readout time is 1 μs.

In the fabricated 16×16 νBJT CNN array with phototransistor design and the cell circuits in Fig. 4, the third metal layer is used to define the optical window for the transistor Q_B and cover the rest part of cell circuit. The same metal layer is used to define the input image pattern by putting the optical window only in the white pixels. The size of the optical window is 16 μm $\times$ 16 μm whereas the base area is 15 μm $\times$ 15 μm. Fig. 18 shows the measured output emitter current of the fabricated p-n-p phototransistor with the light illumination turned off to complete
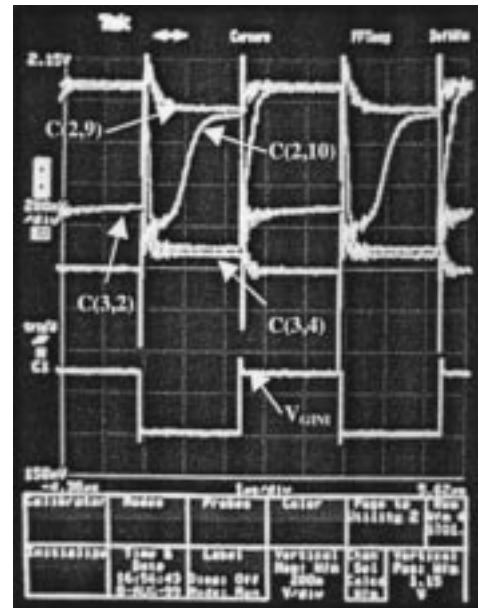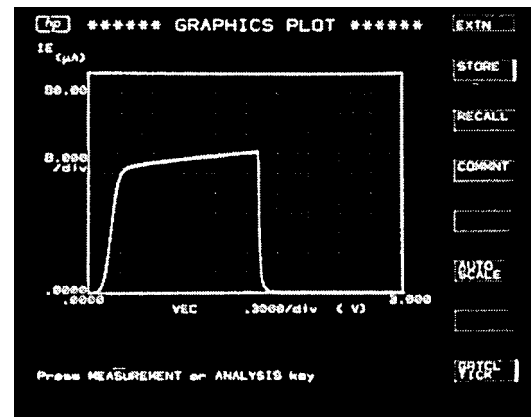
darkness during the sweep of V_{EC}. The measured dark current is about 60 pA whereas the illuminated current is 65 μA. In this
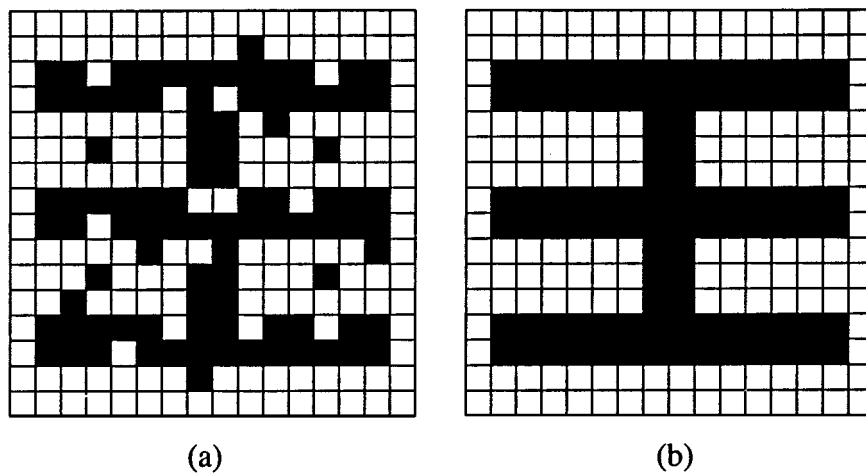
Fig. 19.   (a) The initial state optical image incident to the fabricated νBJT CNN chip with phototransistor design for noise removal and (b) its final output image.
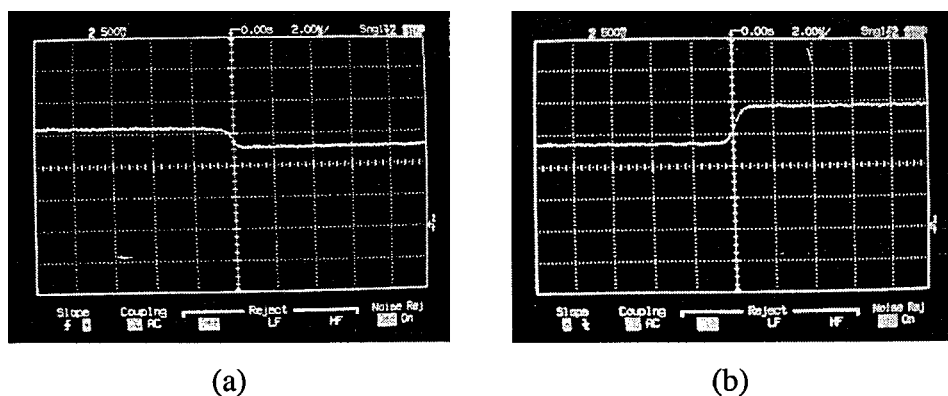

Fig. 20.   The measured waveforms of the neuron state voltage V_B of (a) the cell $C(2, 10)$ and (b) the cell $C(3, 4)$ in the νBJT CNN with phototransistor design under the noise removal operation on the initial states image of Fig. 19(a).

case, the dynamic range is close to 120 dB. The measured large bright-to-dark current ratio provides an enough wide range for input optical images with different optical intensity. The current gain is about 17.5 for the parasitic vertical p-n-p phototransistor.

Fig. 19(a) shows the initial state input optical image incident to the fabricated 16×16 νBJT CNN chip with phototransistor design. Since the image pattern has been defined on-chip by creating the optical window of the third metal layer on the white pixels, a light source incident on the chip can provide the input image to the chip. It can be seen from the output image shown in Fig. 19(b) that the noise has been eliminated. Fig. 20(a) and (b) shows the measured waveforms of the state voltage V_B of the cells $C(3, 4)$ and $C(2, 10)$ in the νBJT CNN with phototransistor design under the noise removal operation on the initial-state image of Fig. 19(a). The characteristics of the fabricated 16×16 νBJT CNN chip with phototransistor design are summarized in Table III.

The image noise removal function of the fabricated 16×16 symmetric νBJT CNN chip with $r = 2$ has been experimentally verified with the initial noisy image of Fig. 21(a) where the 4-pixel square black noise image is created. By using the A-template with $r = 2$ as shown in Fig. 13(b), the noise can be removed as shown in the measured output image of Fig. 21(b). The measured waveforms of the neuron state voltage V_B in the cells $C(12, 3)$ and $C(13, 3)$ of the 4-pixel square black noise

TABLE III
THE SUMMARY ON THE CHARACTERISTICS OF THE FABRICATED 16×16 νBJT CNN CHIP WITH PHOTOTRANSISTOR DESIGN

Technology	0.6μm Single Poly Triple Metal N-well CMOS
Resolution	16×16 cells
Pixel area	30μm $\times$ 35μm
vBJT Q_B and pMOS M_{P1} area	20μm $\times$ 20μm
Optical window size	16μm $\times$ 16 μm
Fill factor	0.35
CNN array size(not including pad)	630μm $\times$ 650μmm
Power supply	3V
Total quiescent power dissipation	15mW
Dynamic power dissipation of the array	12mW ~ 20mW (Depending on image light intensity)
Current gain of BJTs	17.5
State transition time	0.8μs
Minimum readout time of a pixel	1 μs

pixels as well as the cells $C(13, 5)$ and $C(14, 6)$ of the normal black pixels are shown in Fig. 21(c). It can be seen that the noisy black cells become white with higher V_B whereas the normal black cells keep their lower V_B value and remain black.

In the fabricated νBJT CNN chip, the current gain β of BJTs is not completely matched due to process variations. One of the dominant factors for β mismatch is the base width. Since the parasitic p-n-p BJTs in n-well CMOS process has a wide base width, the resultant β value is 17.5 and the β mismatch is low. The measured global variations are 3%–6% on the same wafer
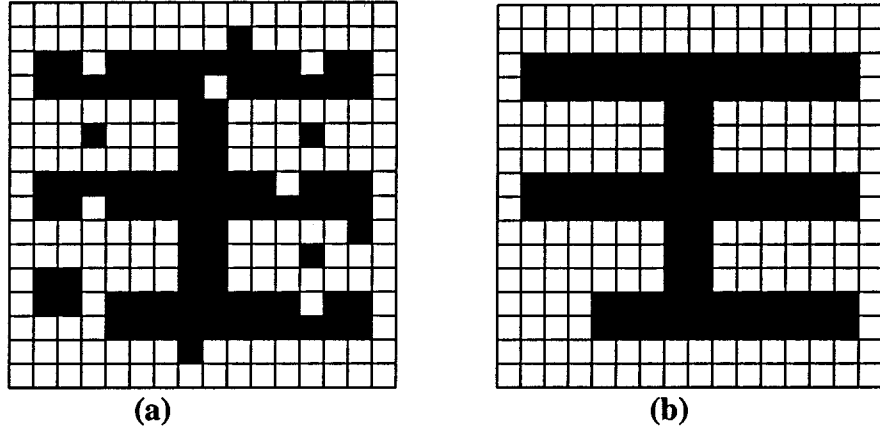
(a)   **(b)**

Fig. 21.   (a) The initial noisy image with the 4-pixel square black noise, (b) the measured final output image, and (c) the measured V_B waveforms of the selected cells in the fabricated 16×16 symmetric νBJT CNN with $r = 2$ under noise removal operation.

and 2%–4% in the same chip. Thus the β variation has negligible effects on the characteristics of the νBJT CNN structure.

The chip area of the νBJT Q_B and pMOS M_{P1} of Fig. 4 can be reduced to a minimum value of 7 μm $\times$ 8 μm in 0.6 μm SPTM n-well CMOS technology, where the emitter area is 1.5 μm $\times$ 1.5 μm with only a minimum metal contact. Thus the overall chip area of the same CNN cell as that in the 32 $\times$ 32 symmetric νBJT CNN can be further reduced to 16 μm $\times$ 18 μm, which is equivalent to a high cell density of 2430 cells/mm^2. As compared to 3000 cells/mm^2 in the CNN proposed in [30] with mixed-signal single-neighborhood template coefficients and hard-limited neuron transfer characteristics realized in 0.25 μm double-poly hexagonal-metal CMOS technology, the cell density of the symmetric νBJT CNN is in the same range.

V. CONCLUSION

A new CNN structure called the neuron–bipolar CNN (νBJT CNN) is proposed and analyzed. In the νBJT CNN, the lambda bipolar transistor is incorporated with the νBJT to form the $\nu\lambda$BJT neuron. Based on the basic device physics, simple MOS resistor array is used in the νBJT to realize the symmetric synapse weights of the A-template. Thus the νBJT CNN has a compact structure which leads to small chip area and high packing density. Through the adjustment of MOS resistance by controlling the gate voltage, the νBJT CNN can easily extend its neighborhood layer size without extra interconnection Moreover, the phototransistor design can be easily applied to the νBJT CNN to enable optical inputs as the neuron initial inputs or external inputs. Thus the chip area can be further reduced. The noise removal, hole filling, and erosion functions have been successfully verified through both simulation and measurement in the symmetric νBJT CNN with the sizes of 32 $\times$ 32 or 16 $\times$ 16.

Future research will focus on the improvement of νBJT CNNs in realizing asymmetric templates, with positive and negative coefficients. Since the proposed νBJT CNN has a soft-limited transfer characteristics and the self-feedback device can be turned off, further research on the applications of

gray scale image processing as well as other image processing will also be explored.

APPENDIX A

A. V_{ECL} and V_{BL}

At the point of V_{ECL}, M_{N2} is operated in the linear region, M_{P1} and M_{P3} are operated in the saturation region, and Q_B is operated in the active region. We have

$$I_S = K_{p3}[(V_{DD} - V_{ECL} - |V_{TP3}|)^2] \tag{A1}$$

$$\begin{aligned} I_{\text{SUM}} &= I_E + I_D = (1 + \beta)I_{\text{BIAS}} - \beta I_D \\ &= (1 + \beta)K_{n2}[2(V_{\text{BIAS}} - V_{TN2})(V_{ECL} - V_{EB}) \\ &\quad - (V_{ECL} - V_{EB})^2] - \beta K_{p1}[(V_{ECL} - |V'_{TP1}|)^2]. \end{aligned} \tag{A2}$$

In (A1) and (A2), $(V_{ECL} - V_{EB})^2$ and $(V_{ECL} - |V'_{TP1}|)^2$ can be neglected. Thus V_{ECL} and V_{BL} can be derived by using the relation $I_{\text{SUM}} = I_S$. The results are

$$\begin{aligned} V_{ECL} &\cong \frac{\omega - \sqrt{\omega^2 + 4\omega(V_{DD} - |V_{TP3}| - V_{EB})}}{2} \\ &\quad + (V_{DD} - |V_{TP3}|) \end{aligned} \tag{A3}$$

$$V_{BL} \cong V_{ECL} - V_{EB,\text{act}} \tag{A4}$$

where

$$\omega = \frac{2(1 + \beta)K_{N2}(V_{\text{BIAS}} - V_{TN2})}{K_{P3}}.$$

B. V_{ECH} and V_{BH}

At the point of V_{ECH}, M_{P1} is operated in linear region, M_{N2} and M_{P3} are operated in saturation region, Q_B is operated cutoff region. In the case

$$I_{\text{BIAS}} = I_D = I_S$$

By using the MOS device equations, we have

$$
K_{n2}[(V_{\text{BIAS}} - V_{TN2})^2]
= K_{p3}[(V_{DD} - V_{ECH} - |V_{TP3}|)^2] \tag{A5}
$$

$$
\begin{aligned}
K_{n2}[(V_{\text{BIAS}} - V_{TN2})^2]
&= K_{p1}[2(V_{ECH} - |V'_{TP1}|)(V_{ECH} - V_{BH}) \\
&\quad - (V_{ECH} - |V'_{TP1}|)^2].
\end{aligned} \tag{A6}
$$

From (A5) and (A6), V_{ECH} and V_{BH} can be written as

$$
V_{ECH} = \left[-\sqrt{\frac{K_{n2}}{K_{p3}}}(V_{\text{BIAS}} - V_{TN2}) + (V_{DD} - |V_{TP3}|) \right] \tag{A7}
$$

$$
V_{BH} = -\frac{\dfrac{K_{n2}}{K_{p1}}(V_{\text{BIAS}} - V_{TN2})^2 + (V_{ECH} - |V'_{TP1}|)^2}{2(V_{ECH} - |V'_{TP1}|)} + V_{ECH}. \tag{A8}
$$

APPENDIX B

Consider the 1-D BJT and resistor array shown in Fig. 1(b) where each node is connected to the BJTs Q_{Bi} with the emitter–base voltage $V_{EB}(i)$ and the current $I_B(i)$ for $i = 1, 1, 2, \ldots$. To model its operation, the following basic assumptions are used.

1) The array resistors have the same resistance R which is independent of the flowing current.
2) The upper and lower subarrays are symmetrical and the total number N of resistors is very large.
3) The lumped array can be approximated by a continuous one.
4) The common-base current gain α of all BJTs in the array is constant.
5) The leakage current is neglected.

Assume that the only excitation is the current I_{BO}. The emitter–base voltage $V_{EB}(n)$ of the BJT at the nth node of a subarray can be expressed as

$$
\begin{aligned}
V_{EB}(n) &= V_{EB}(0) - \sum_{i=1}^{n} I_R(i)R \\
&\cong V_{EB}(0) - \int_1^n I_R(i)R \, di
\end{aligned} \tag{B1}
$$

where
$V_{EB}(0)$ is the emitter–base junction voltage of the reference,
$I_R(i)$ is the current flowing through R at the ith node,
n is the node number from 1 to $N - 1$, and
N is the total node number in the each side of linear array.

Based on the third assumption given above, the summation expression can be substituted by the integration as in (B1). Since all BJTs are biased in the active region, the emitter current $I_E(n)$ at the nth pixel can be expressed as

$$
\begin{aligned}
I_E(n) &= I_S \left[\exp\left(\frac{V_{EB}(n)}{V_T} \right) - 1 \right] \\
&\cong I_S \exp\left[\frac{V_{EB}(n)}{V_T} \right]
\end{aligned} \tag{B2}
$$

where I_S is the reverse-saturation current and V_T is the thermal voltage. The current $I_R(n)$ flowing through R at the node n is given by

$$
\begin{aligned}
I_R(n) &= \sum_{i=n}^{N-1} I_B(i) = \sum_{i=n}^{N-1} (1 - \alpha)I_E(i) \\
&\cong \int_n^{N-1} (1 - \alpha)I_E(i) \, di
\end{aligned} \tag{B3}
$$

where $I_B(i)$ is the base current of BJT at the nth node and α is the common-base current gain of BJTs.

Differentiating $I_R(n)$ with respect to n in (B3) and assuming that the integration of $I_E(i)$ with respect to i at the $(N - 1)$th node is nearly independent of n, we have

$$
\begin{aligned}
\frac{dI_R(n)}{dn} &= -(1 - \alpha)I_E(n) \\
&= (1 - \alpha)I_S \exp\left[\frac{V_{EB}(0) - \int_1^n I_R(i)R \, di}{V_T} \right]
\end{aligned} \tag{B4}
$$

where the expression of $I_E(n)$ in (B2) has been used with $V_{EB}(n)$ given in (B1). Differentiating (B4) with respect to n and using the fact that $V_{EB}(0)$ and $I_R(1)$ are nearly independent of n, we have

$$
\frac{d^2 I_R(n)}{dn^2} = \left[-\frac{R}{V_T} I_R(n) \right] \frac{dI_R(n)}{dn}. \tag{B5}
$$

This is a second-order nonlinear differential equation. The solution is

$$
I_R(n) = C_1 \tan\left[C_1 \frac{R}{2V_T}(C_2 - n) \right] \tag{B6}
$$

where C_1 and C_2 are arbitrary constants determined by boundary conditions.

As shown in Fig. 1(b), Since the upper and lower subarrays are symmetrical, the boundary conditions are

$$
\begin{aligned}
I_R(0) &= [I_{BO} - I_B(0)]/2 \cong I_{BO}/2 \\
&= C_1 \tan\left[C_1 \frac{R}{2V_T} C_2 \right]
\end{aligned} \tag{B7}
$$

$$
\begin{aligned}
I_R(N - 1) &= -(1 - \alpha)I_E(N - 1) \\
&= -\left. \frac{dI_R(n)}{dn} \right|_{n = N-1}.
\end{aligned} \tag{B8}
$$

Substituting (B6) into (B8), we have

$$\sin\left[C_1\frac{R}{V_T}(C_2 - N + 1)\right] = \frac{C_1 R}{2V_T}. \qquad \text{(B9)}$$

Given $I_R(0)$, the constants C_1 and C_2 can be solved from (B7) and (B9) by using the numerical method. It is found that the value of C_2 is approximately equal to N if N is sufficiently large.

In order to obtain the analytical solution of $I_E(n)$, C_2 in (B7) is approximated by N and the constant C_1 can be expressed as

$$z\tan z = \frac{RN}{2V_T}I_R(0) \qquad \text{(B10)}$$

where z is defined as

$$z = C_1\frac{RN}{2V_T}. \qquad \text{(B11)}$$

Using $C_2 = N$ and substituting (B10) and (B11) into (B6), $I_R(n)$ can be rewritten as

$$I_R(n) = \frac{2V_T}{RN}z\tan\left[z\left(1 - \frac{n}{N}\right)\right]$$
$$= I_R(0)\left[\frac{\tan\left(z\frac{N-n}{N}\right)}{\tan z}\right]. \qquad \text{(B12)}$$

ACKNOWLEDGMENT

REFERENCES

[1] L. O. Chua and L. Yang, "Cellular neural networks: Theory and applications," *IEEE Trans. Circuit Syst.*, vol. 35, pp. 1147–1180, Oct. 1988.

[2] R. Domínguez-Castro, S. Espejo, A. Rodríguez-Vázquez, R. A. Carmona, P. Földesy, Á. Zarándy, P. Szolgay, T. Szirányi, and T. Roska, "A 0.8-μm CMOS two-dimensional programmable mixed-signal focal-plane array processor with on-chip binary imaging and instructions storage," *IEEE J. Solid-State Circuits*, vol. 32, pp. 1013–1026, July 1997.

[3] J. M. Cruz and L. O. Chua, "A 16×16 cellular neural network universal chip: The first complete single-chip dynamic computer array with distributed memory and with gray-scale input–output," *Analog Integr. Circuits Signal Processing*, vol. 15, no. 3, pp. 3–14, 1998.

[4] E. Y. Chou, B. J. Sheu, and R. C. Chang, "VLSI design of optimization and image processing cellular neural networks," *IEEE Trans. Circuits Syst.—I*, vol. 44, pp. 12–20, Mar. 1997.

[5] P. Kinget and J. Steyaert, "A programmable analog cellular neural network CMOS chip for high speed," *IEEE J. Solid-State Circuits*, vol. 30, pp. 235–243, Mar. 1995.

[6] M. Anguita, F. J. Pelayo, F. J. Fernandez, and A. Prieto, "A low-power CMOS implementation of programmable CNNs with embedded photosensors," *IEEE Trans. Circuits Syst.—I*, vol. 44, pp. 149–152, Feb. 1997.

[7] M. Salerno, F. Sargeni, and V. Bonaiuto, "A 6×6 cells interconnection-oriented programmable chip for CNN," *Analog Integr. Circuits Signal Processing*, vol. 15, no. 3, pp. 15–26, 1998.

[8] G. F. Dalla Betta, S. Graffi, Zs. M. Kovács, and G. Masetti, "CMOS implementation of an analogically programmable cellular neural network," *IEEE Trans. Circuits Syst.—II*, vol. 40, pp. 206–215, Mar. 1993.

[9] A. Paasio, A. Dawidziuk, K. Halonen, and V. Porra, "Fast and compact 16 by 16 cellular neural network implementation," *Analog Integr. Circuits Signal Processing*, vol. 12, no. 3, pp. 59–70, 1998.

[10] S. Espejo, A. Rodríguez-Vázquez, R. Domínguez-Castro, J. L. Huertas, and E. Sánchez-Sinencio , "Smart-pixel cellular neural networks in analog current-mode CMOS technology," *IEEE J. Solid-State Circuits*, vol. 28, pp. 895–905, Aug. 1994.

[11] M. Anguita, F. J. Pelayo, F. J. Fernandez, and A. Prieto, "Area efficient implementations of fixed-template CNNs," *IEEE Trans. Circuits Syst.—I*, vol. 45, pp. 968–973, Sept. 1998.

[12] J. E. Varrientos, E. Sánchez-Sinencio, and J. Ramírez-Angulo, "A current-mode cellular neural networks implementation," *IEEE Trans. Circuits Syst.—II*, vol. 40, pp. 147–156, Mar. 1993.

[13] T. Shibata and T. Ohmi, "An intelligent MOS transistor featuring gate-level weighted sum and threshold operations," in *IEDM Tech. Dig.*, Dec. 1991, pp. 919–922.

[14] C. Y. Wu and C. F. Chiu, "A new structure for the silicon retina," in *IEDM Tech. Dig.*, Dec. 1992, pp. 439–442.

[15] ——, "A new structure of the 2-dimensional silicon retina," *IEEE J. Solid-State Circuits*, vol. 30, pp. 890–897, Aug. 1995.

[16] H. C. Jiang and C. Y. Wu, "A 2-D velocity- and direction-selective sensor with BJT-based silicon retina and temporal zero-crossing detector," *IEEE J. Solid-State Circuits*, vol. 34, pp. 241–247, Feb. 1999.

[17] C. Y. Wu and W. C. Yen, "The neuron-bipolar junction transistor (νBJT)-a new device structure for VLSI neural network implementation," in *Proc. Int. Conf. Electronics, Circuits and Systems*, vol. 3, Sept. 1998, pp. 277–270.

[18] W. C. Yen and C. Y. Wu, "A new compact neuron-bipolar cellular neural network structure with adjustable neighborhood layers and high integration level," in *Proc. IEEE Int. Symp. Circuits Systems*, vol. 4, June 1999, pp. 505–508.

[19] ——, "A new compact programmable νBJT cellular neural network structure with adjustable neighborhood layers for image processing," in *Proc. Int. Conf. Electronics, Circuits and Systems*, vol. 2, Sept. 1999, pp. 713–716.

[20] K. R. Crounse, T. Roska, and L. O. Chua, "Image halftoning with cellular neural networks," *IEEE Trans. Circuits Syst.—II*, vol. 40, pp. 147–156, Apr. 1993.

[21] T. Roska, J. Hámori, E. Lábos, K. Lotz, L. Orzó, J. Takács, P. L. Venetianer, Z. Vidnyánsky, and Á. Zarándy, "The use of CNN model in the subcortical visual pathway," *IEEE Trans. Circuits Syst.—I*, vol. 40, pp. 1822–1895, Mar. 1993.

[22] L. O. Chua, *CNN: A Paradigm for Complexity (World Scientific Series on Nonlinear Science.* Singapore: World Scientific, 1998, vol. 31.

[23] L. Kék and Á. Zarándy, "Implementation of large-neighborhood nonlinear templates on the CNN universal machine," *Int. J. Circuit Theory Appl.*, vol. 26, pp. 551–566, Nov./Dec. 1998.

[24] M. H. ter Brugge, J. H. Stevens, J. A. G. Nijhuis, and L. Spaanenburg, "Efficient DTCNN implementations for large-neighborhood functions," in *Proc. 5th IEEE Int. Workshop Cellular Networks and Their Applications*, Apr. 1998, pp. 88–93.

[25] T. Roska and L. O. Chua, "The CNN universal machine: An analogic array computer," *IEEE Trans. Circuits Syst.—II*, vol. 40, pp. 163–173, Mar. 1993.

[26] C. Y. Wu and C. Y. Wu, "An analysis and the fabrication technology of the LAMBDA bipolar transistor," *IEEE Trans. Electron Devices*, vol. ED-27, pp. 414–419, Feb. 1980.

[27] C. Y. Wu and H. C. Jiang, "An improved BJT-based silicon retina with tunable image smoothing capability," *IEEE Trans. VLSI Syst.*, vol. 72, pp. 241–248, June 1999.

[28] ——, "The modeling and design of the BJT-based silicon retina for image smoothing and edge detection," in *Proc. 3rd Australian and New Zealand Conf. Intelligent Information Systems*, vol. 1, Nov. 1995, pp. 232–235.

[29] I. Fajfar and F. Bratkovic, "Design of monotonic binary-valued cellular neural networks," in *Proc. 4th IEEE Int. Workshop Cellular Networks and Their Applications*, June 1996, pp. 321–326.

[30] A. Paasio, A. Kananen, K. Halonen, and V. Porra, "A QCIF resolution binary I/O CNN-UM chip," *J. VLSI Signal Processing*, vol. 23, pp. 281–290, Nov./Dec. 1999.

Chung-Yu Wu (S'76–M'88–SM'96–F'98) was born in 1950. He received the M.S. and Ph.D degrees from the Department of Electronics Engineering, National Chiao-Tung University, Taiwan, R.O.C. in 1976 and 1980, respectively.

From 1980 to 1984, he was an Associate Professor in the National Chiao-Tung University. During 1984–1986, he was a Visiting Associate Professor in the Department of Electrical Engineering, Portland State University, Portland, OR. Since 1987, he has been a Professor in the National Chiao-Tung University.

Dr. Wu was a recipient of the IEEE Third Millennium Medal, the Outstanding Academic Award by the Ministry of Education in 1999, the Distinguished Researcher in 1999, and the Outstanding Research Award in 1989–1990, 1995–1996, and 1997–1998, by the National Science Council, the Outstanding Engineering Professor by the Chinese Engineer Association in 1996, and the Tung-Yuan Science and Technology Award in 1997.

From 1991 to 1995, he served as Director of the Division of Engineering and Applied Science in the National Science Council. He is now the Centennial Honorary Chair Professor at the National Chiao-Tung University. He has published more than 200 technical papers in international journals and conferences. He also has 18 patents including nine U.S. patents. Since 1980, he has served as a consultant to high-tech industry and research organization. He has built strong research collaborations with high-tech industries. His research interests focus on low-voltage low-power mixed-mode circuits and systems for multimedia applications, hardware implementation of visual and auditory neural systems, cellular neural networks, and RF communication circuits and systems. He served as a Guest Editor of the Multimedia Special Issue for IEEE TRANSACTION ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY in August–October, 1997. He also served as Associate Editor for the IEEE TRANSACTIONS ON VLSI SYSTEMS and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—PART II.

He served on the Technical Program Committees of IEEE, ISCAS, ICECS, and APCCAS. He served as the VLSI Track Co-Chair of the Technical Program Committee of ISCAS'99. He served as General Chair of IEEE APCCAS'92. He also served as the Chair of Neural Systems and Applications Technical Committee and Chair of Multimedia Systems and Applications Technical Committee of the IEEE CAS Society. He was one of the Society representatives in the Steering Committee of IEEE TRANSACTIONS ON MULTIMEDIA. Currently, he serves as Associate Editor for the IEEE TRANSACTIONS ON VLSI SYSTEMS and IEEE TRANSACTIONS ON MULTIMEDIA. He is the Distiguished Lecturer of the CAS Society and one of the society representatives in the Neural Network Council. He is a member of Eta Kappa Nu and Phi Tau Phi Honorary Scholastic societies.

Wen-Cheng Yen was born in Taichung, Taiwan, R.O.C., in 1968. He received the B.S. degree from the Department of Electrical Engineering, Tamkang University, Taipei, Taiwan, in 1993 and the M.S. degree from the Institute of Electronics, National Chiao-Tung University, Hsinchu, Taiwan, in 1995. He is currently working toward the Ph.D. degree at the same institute.

His research interests include cellular neural networks, signal processing, VLSI design, and RF communication circuits.