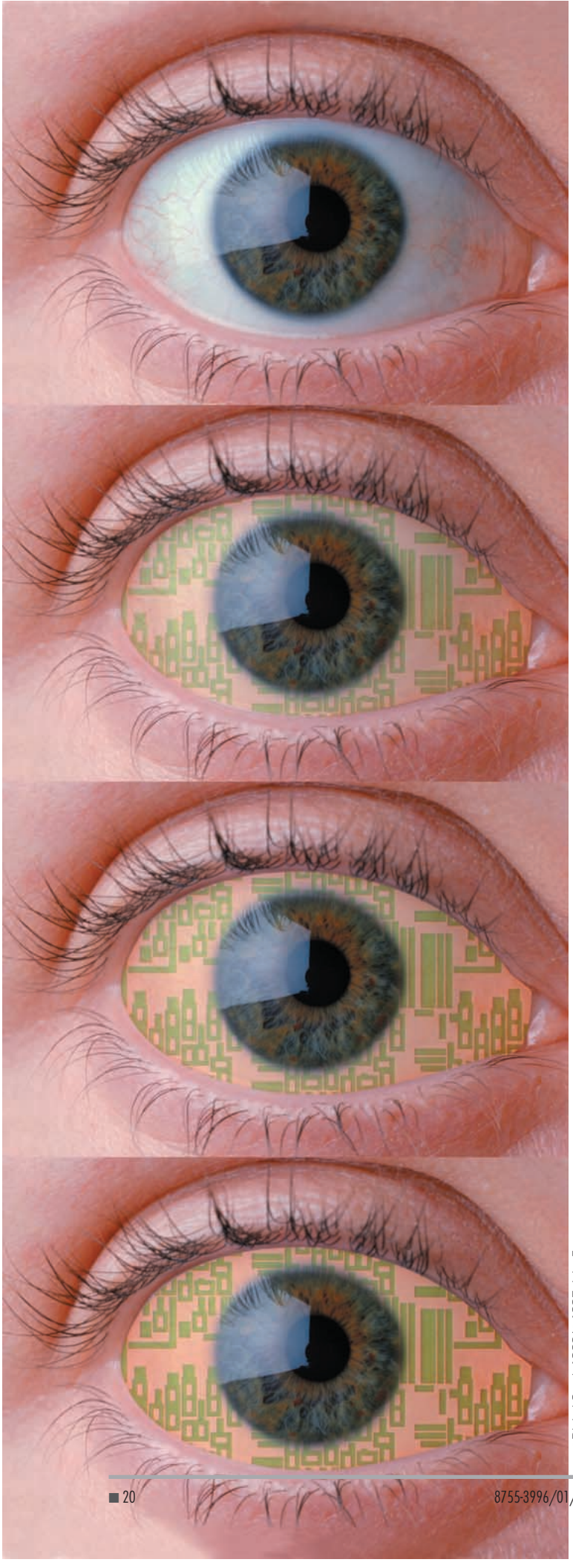


In the Blink of a Silicon Eye

Neuron-Bipolar Junction Transistor (νBJT) Circuits Enable the Formation of Cellular Networks and Human Neural Sensors Such as the Retina

By Chiu-Hung Cheng, Chung-Yu Wu, Bing Sheu, Li-Ju Lin, Kuan-Hsun Huang, Hsin-Chin Jiang, Wen-Cheng Yen, and Chiao-Wei Hsiao

The analysis and modeling of human neural systems has long been a very important research field where many concepts can be applied to many applications in intelligent machines performing human-like speech, image, or other neural signal processing. The resultant technological models have led to the development of new computational methodologies and artificial neural networks (ANNs). It is always a great challenge to design neuro-morphic or neuro-inspired ANNs to mimic both structures and functions of the brain as well as neural sensors such as the eyes, ears, nose, tongue, skin, etc. This is mainly due to complicated structures and dedicated functions of the brain and neural sensors. It is known that a human brain may consist of 10^{11} neurons with over 10^3 interconnections per neuron. It performs dedicated functions in a highly parallel processing manner. In human eyes, the retina has millions of cells performing image sensing, smoothing, feature extraction, and dynamic processing.

8755-3996/01/$10.00 ©2001 IEEE

Increasing research efforts have focused on solving the challenges of realizing that feasible ANNs should have a simple structure suitable for microchip realization as well as mimic some specific neural functions. One of the successful examples is the cellular neural network (CNN) proposed by Chua and Yang [1], which is a specific type of "analogic" nonlinear processor array. Due to its continuous-time dynamics and parallel-processing feature, the cellular network is very effective in real-time image processing applications such as noise removal, edge and corner detection, hole filling, connected component detection, and shadowing. Moreover, regularity, parallelism, and local connectivity in the cellular network architecture make it suitable for VLSI implementation.

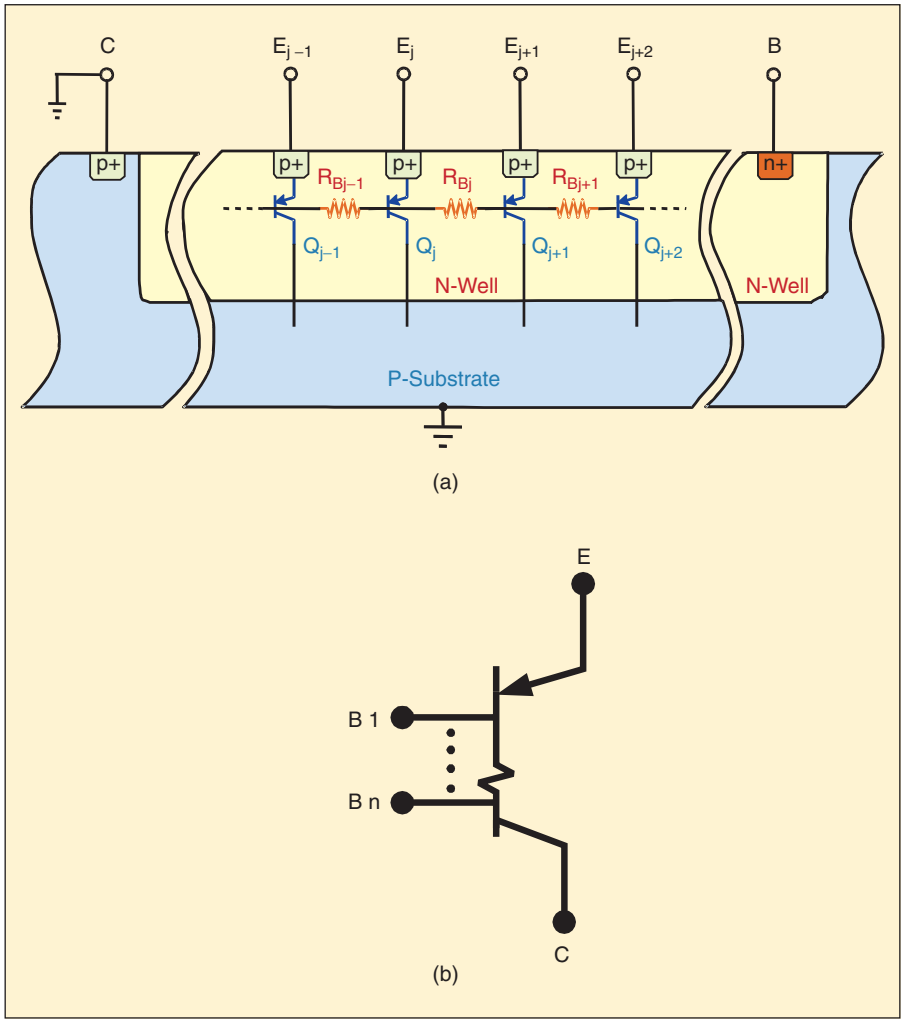
> The three major functions of the retina are photoinput sensing, edge extraction, and moving-object feature extraction.

Besides the structure simplification for practical realization, the basic components to realize cellular networks or neural sensors should consist of compact microelectronic devices or nanoelectronic devices. Thus, many neurons or cells can be implemented within a reasonable size to achieve the complex and dedicated neural functions. This means that fundamental device physics needs to be applied to realize neural behavior.

In this article, a novel device structure called the neuron-bipolar junction transistor (νBJT), for the compact implementation of large-neighborhood cellular networks and silicon retinal processing components is described. The basic device physics of νBJTs is based on the carrier diffusion within the common base region so that a νBJT in an array structure can be coupled to another without direct interconnection. Through this mechanism, νBJTs can be used to form neuron cells in the cellular networks and large-neighborhood structures can be realized in a small chip area. Similarly, the base carrier diffusion mechanism of a νBJT array can be applied to the realization of the smoothing function of horizontal cells in the retina. Thus, a compact silicon retina can be designed for different applications of real-time image processing.

Device Structure of Neuron-Bipolar Junction Transistors

The device structure of neuron-bipolar junction transistors (νBJT) is shown in Fig. 1(a) where the parasitic pnp BJT in N-well CMOS technology is used as an example. All the pnp BJTs have the same N-well base region and form an array. If the transistor Q_j is turned on, the base carriers (i.e., electrons) can diffuse through the base spreading resistance R_B to other νBJTs so that they are turned on. The amount of carriers diffusing to the νBJTs depends on the value of R_B and the distance to the fired νBJT Q_j. Similarly, other νBJTs can diffuse their base carriers to Q_j. Once again, the amount of carriers reaching Q_j depends on R_B and the distance. Equivalently, the νBJT Q_j is a neuron that interconnects with other neurons to send its state output to them or receive multiple state inputs from them. The device symbol of a νBJT is shown in Fig. 1(b).



1. (a) The cross-sectional view of the proposed neuron-bipolar junction transistor (νBJT). (b) The symbol of νBJT.

Since the spread emitter current of a vBJT transistor is determined by the carrier diffusion in the base, such current can be controlled by the base resistance value R_B. A large R_B leads to small spread current and narrow spread range. A controllable R_B can be realized as the drain-source resistance R_{DS} of an NMOS device operated in the linear region as shown in Fig. 2.

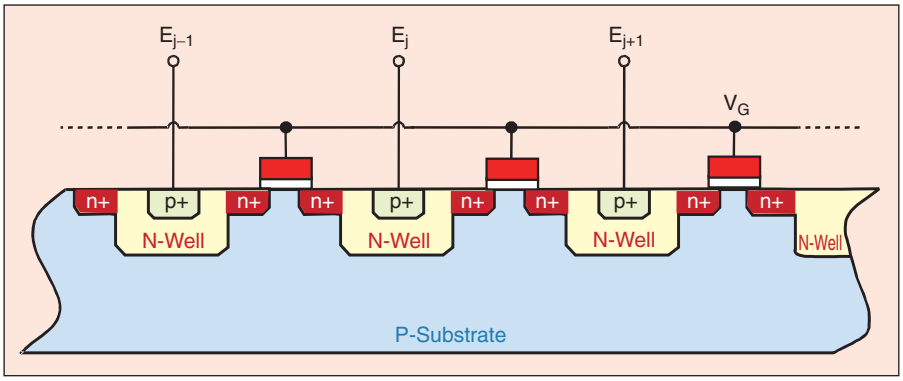
The HSPICE-simulated and model-calculated emitter currents of 32 vBJTs in a linear array of vBJTs are shown in Fig. 3 [2] where one vBJT is turned on with 0.4-μA base current drive. It can be seen from Fig. 3 that other vBJTs have smaller emitter currents caused by base carrier diffusion. The range of current distribution can be controlled by R_{DS} through the gate voltage V_G.

vBJT Smoothing Network for a Silicon Retina

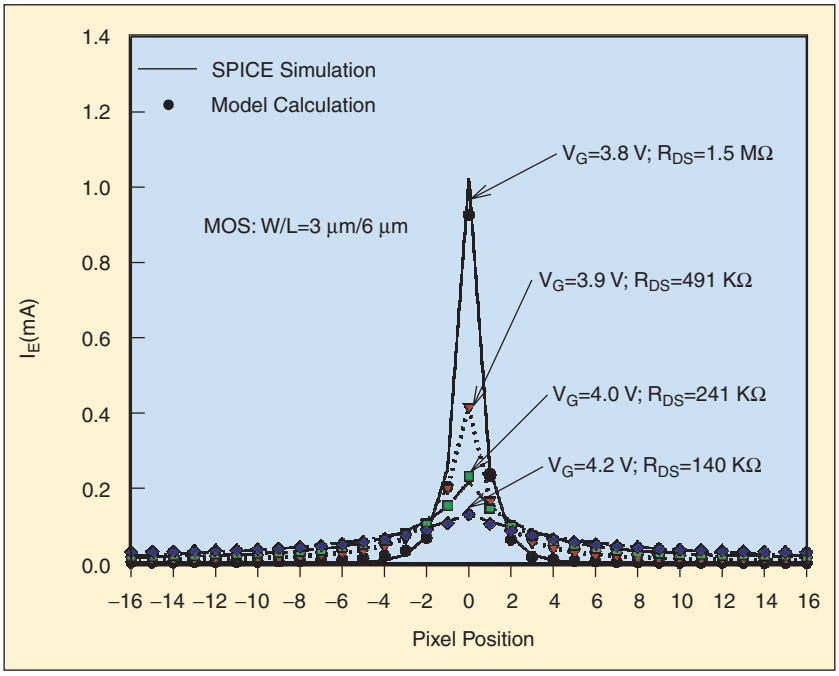
The retina is the early processing element in the visual nervous system of the vertebrate. It can perform three major functions in image processing, and it sends out suitable signals to the brain. The three major functions are photoinput sensing, edge extraction, and moving-object feature extraction.

In the outer-plexiform layer of the retina, there are three major cells, namely the photoreceptor, the horizontal cell, and the bipolar cell [3]. The major function of photoreceptors is to transduce light into electrical signals whereas that of horizontal cells is to provide a spatial and temporal smoothing for the signals from the photoreceptors. The bipolar cells process the signals from both photoreceptors and horizontal cells. It is believed that the outputs of bipolar cells are proportional to the difference between photoreceptor signals and horizontal-cell signals so that the image edge can be extracted. Since the smoothing function of horizontal cells is a local operation instead of a global one, edge details in both light and dark areas of high-contrast images can be sensed clearly.

In the retina, the edge of an object is detected by the contrast between the surface of the object and the background. The general principle of edge detection can be understood by using the conceptual curves in Fig. 4. In Fig. 4, the response of the photoreceptors has a high (low) current in the light (dark) area whereas the response of horizontal cells is the smoothing of the photoreceptor response. By performing the subtraction operation between the two responses, an abrupt difference from negative to positive at the edge can be obtained. By detecting the zero-crossing point, the edge of the image can be detected reliably [4].



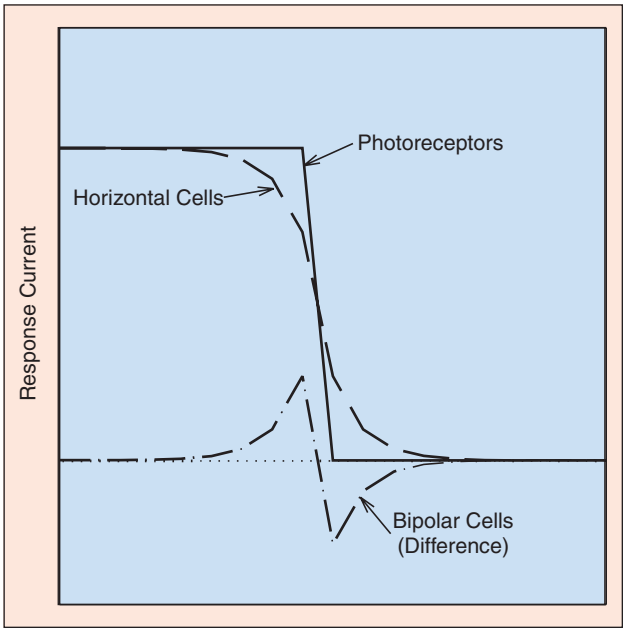
2. The pnp vBJT array with n-channel MOSFETs as controllable resistors.



3. The SPICE simulation and model calculation results on the emitter current characteristics of the proposed vBJT linear array with 0.4 μm single-point stimulus under different FET gate biases with single-point stimulus.

One compact and efficient way to incorporate the advantages of the retina in image processing is to realize the retina directly in a silicon microchip, called the silicon retina. Thus, the computation cost can be reduced and the processing speed can be enhanced. Moreover, high-performance real-time image processing could be achieved.

The vBJT can be applied to the compact realization of photoreceptors and horizontal cells in the retina. In the proposed vBJT silicon retina architecture [2]-[4], the parasitic pnp BJTs in N-well CMOS technology with open bases are used as phototransistors to sense the light and generate photocurrents like the photoreceptors. The vBJT array with all bases connected together through MOSFETs as adjustable resistors is used as the smoothing network, which mimics the smoothing function of horizontal cells. The structure of a two-dimensional (2-D) vBJT retina is shown in Fig. 5, where each basic cell consists of an isolated pnp phototransistor as

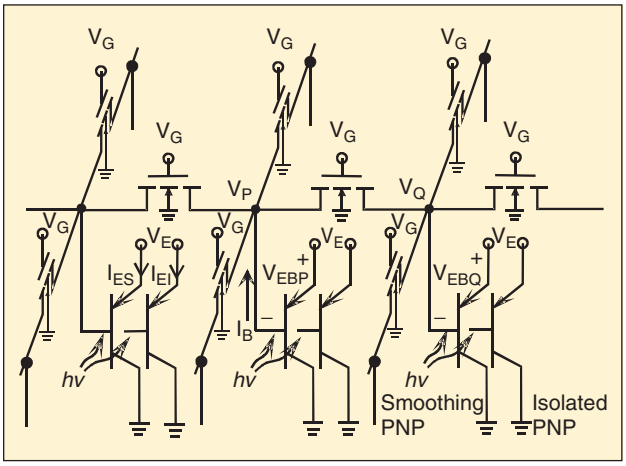


4. The conceptual response currents versus pixel position in the retina.

a photoreceptor and a smoothing pnp phototransistor as a horizontal cell. When the light is incident upon one cell, it is simultaneously incident upon the two floating bases of both the smoothing vBJT and isolated photo-BJT. Thus, the electron-hole pairs are generated at and nearby the depletion region of the two base-collector junctions. Due to the electric field in the deletion region, the generated electrons are swept into the base region whereas the holes are swept into the collector. The excess electrons in the base region are diffused out through the MOS resistor array. The results of the carrier diffusion and distribution make the emitter currents of the vBJT array perform the smoothing function of the retina. The proposed vBJT structure is compact and fully compatible with CMOS technology. Thus, it can be easily integrated with other integrated circuits to form a VLSI microsystem.

In many applications, input images often have wide-ranging variations of intensity or contrast. This requires an adjustable smoothing range. To achieve adjustable smoothing, the gate voltage of the enhancement-mode NMOS in the smoothing network of Fig. 5 can be controlled. When the gate voltage is high, the NMOS device is operated in the strong-inversion region and has a small drain-source resistance. On the other hand, when the gate voltage is low, the NMOS device is operated in the subthreshold region and the drain-source resistance is large. Therefore, the MOS device can provide a wide range of resistance values to achieve the wide-range adjustment of the smoothing range.

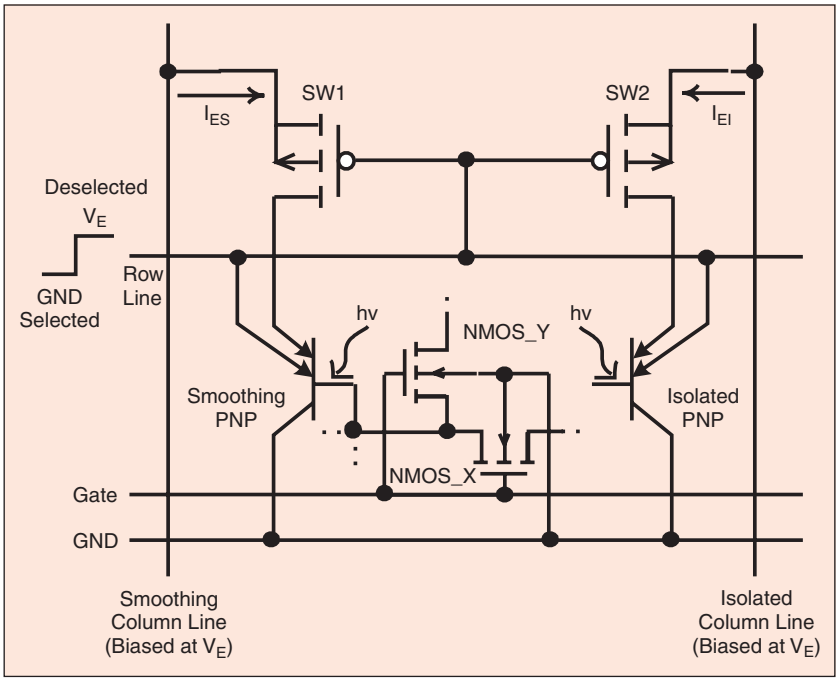
In the design of the focal-plane-array (FPA) of the proposed vBJT silicon retina,



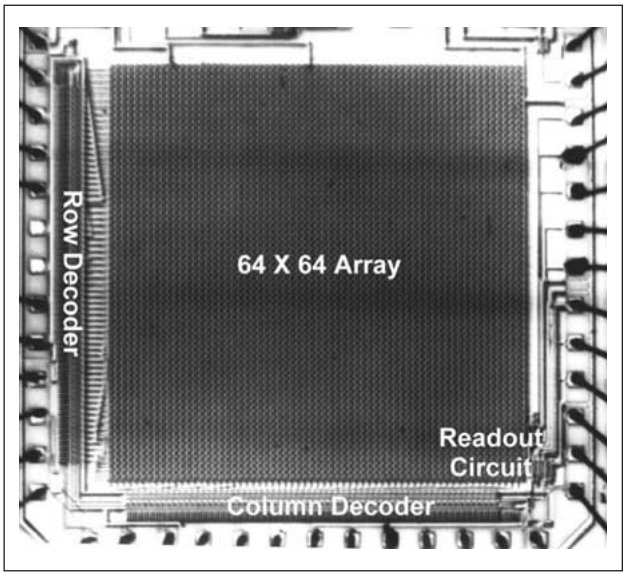
5. The equivalent circuit of the proposed 2-D vBJT silicon retina.

compact row/column selection circuits are required to achieve a small pixel area. Figure 6(a) shows the basic cell structure of the FPA of a vBJT silicon retina. Both isolated and smoothing BJTs at each pixel have extra emitters directly connected to the row line. Two PMOS devices are connected in series with the active emitters of BJTs to serve as the row switches SW1 and SW2 controlled by the row line. Through the switches, the emitter currents can be sent to smoothing and isolated column lines biased at V_E. In each cell, four NMOSFETs are connected from the smoothing BJT to the four neighbors to form the proposed 2-D BJT smoothing network. Equivalently, each cell has two NMOSFETs as shown in Fig. 6. All the gates of NMOSFETs are connected to the common gate line with the adjustable voltage V_G.

When the row line is selected through the row decoder, its voltage is set at ground. The extra emitters act as collectors of



6. The basic cell circuit of the FPA of the proposed 2-D vBJT silicon retina.

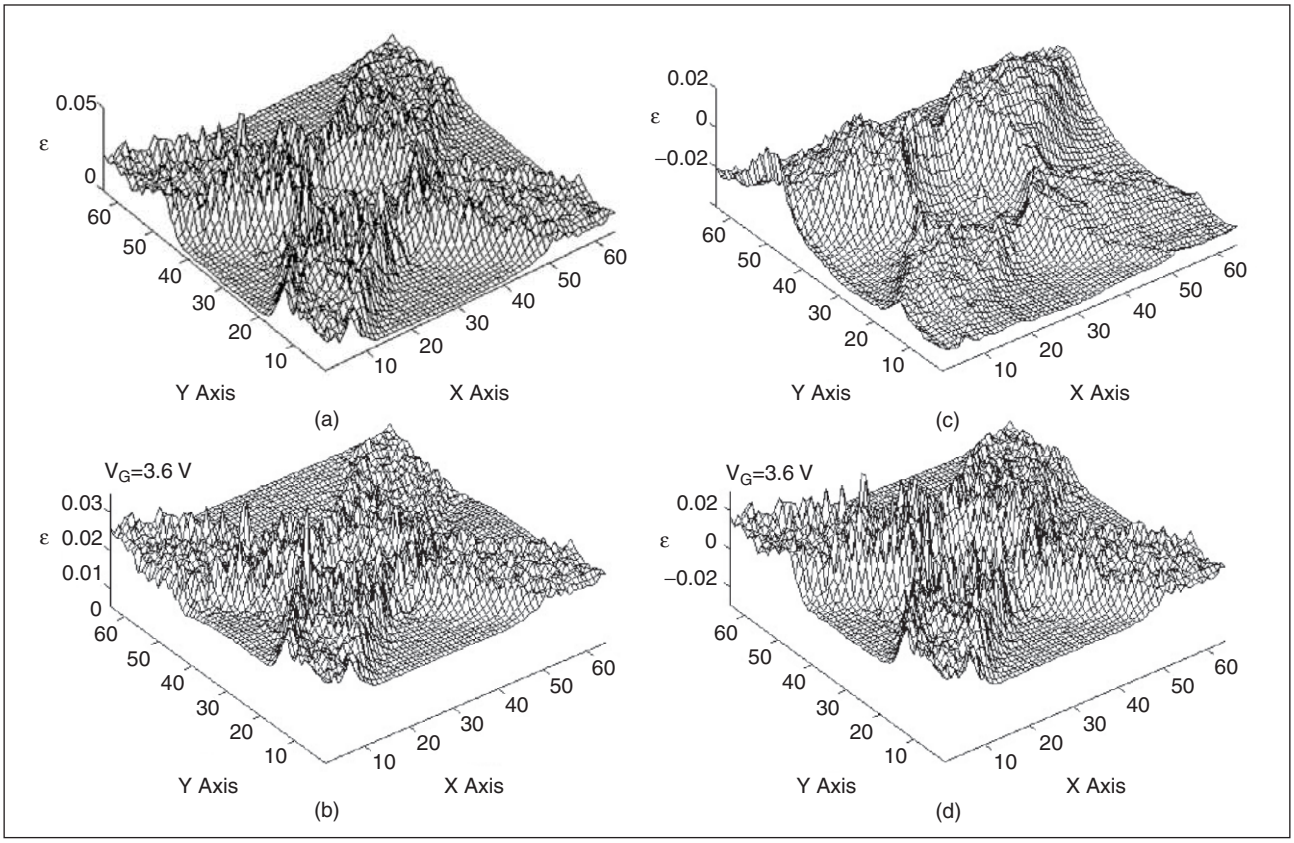


7. The chip photograph of the fabricated 64 × 64 FPA of the vBJT silicon retina.

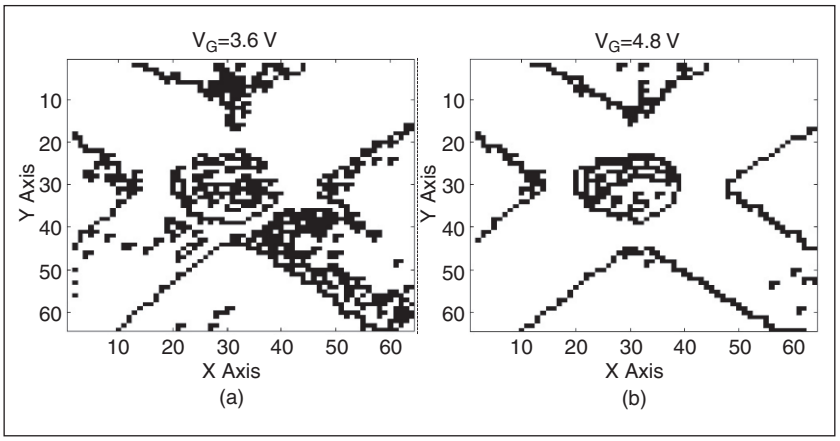
pnp BJTs and SW1 and SW2 are closed. Thus, the two emitter currents I_{EI} and I_{ES} of isolated and smoothing pnp BJTs at the selected pixel can be sent to the isolated and smoothing column lines, respectively. When the row line is de-selected, its voltage is raised to the emitter voltage V_E. In this case, SW1 and SW2 are open and the normal emitters of pnp BJTs are disconnected from the column lines and become floating. However, the extra emitters connected to the row selection line are connected to V_E. With all the de-selected emitters of BJTs connected to V_E, the BJTs in the smoothing network can be kept in the normal state so that the normal smoothing operation can be performed.

An experimental chip of the proposed BJT-based silicon retina has been designed and fabricated in a 0.5-µm N-well double-poly, double-metal CMOS technology. Figure 7 shows the chip and basic cell photographs of the fabricated 64 × 64 FPA of the BJT-based silicon retina. In the quiescent state without image inputs, the power dissipation of the sensor array in the silicon retina is very small, being that of a 64 × 64 open-base BJT. Thus the total quiescent power dissipation is determined by the peripheral circuit and is about 45 mW. In the illuminated state, the power dissipation of the sensor array is caused by the photocurrents of the BJTs and depends on the image light intensity. The typical total active power dissipation of the silicon retina FPA in the illuminated state is about 48 mW to 75 mW.

To verify the tunable smoothing function of the proposed BJT smoothing network, a nonregular cross pattern with nonuniform light intensity is projected on the experimental 2-D chip with the gate bias V_G of NMOSFETs tuned at different values. Figure 8(a) shows the measured output responses of the isolated BJTs as the photoreceptors in the fabricated silicon retina and Fig. 8(b) shows the measured output responses of the BJTs in the proposed vBJT smoothing network with V_G at



8. The measured output responses of (a) isolated BJT array; (b) vBJT smoothing network with V_c = 3.6 V; (c) silicon retina with V_c = 3.6 V; and (d) silicon retina with V_c = 4.8 V in the fabricated 2-D vBJT silicon retina FPA chip incident by a nonregular cross pattern with nonuniform light intensity.



9. The mathematically found zero-crossing positions of the measured output responses of the proposed BJT-based silicon retina as shown in: (a) silicon retina with V_G = 3.6 V and (b) silicon retina with V_G = 4.8 V.

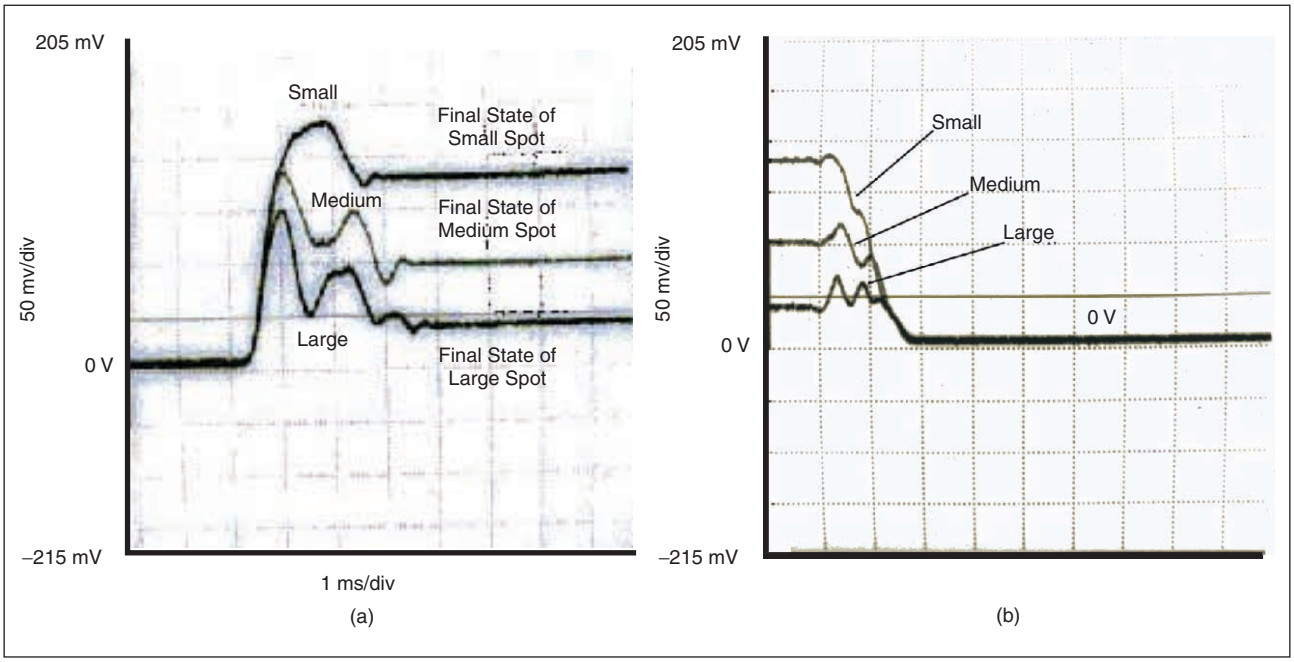
3.6 V, respectively. These responses represent those of the horizontal cells in the retina. When V_G is equal to 3.6 V, the NMOSFETs are operated in the subthreshold region and the channel resistance is the largest. So the pattern has the smallest smooth area as shown in Fig. 8(b).

Figure 8(c) and (d) shows the measured output responses of the proposed BJT-based silicon retina with V_G = 3.6 V and 4.8 V, respectively, which represent the responses of the bipolar cells in the retina. Figure 9(a) and (b) shows the mathematically found zero-crossing positions of the outputs of the BJT-based silicon retina as shown in Fig. 8(c) and (d), respectively. The zero-crossing points of the measured output voltages of the BJT-based silicon retina can be used to identify the edges of objects. As shown in Fig. 9, when the smooth area is smaller, the smaller variation of light intensity can be identified. In other words, the smaller contrast edges can be identified.

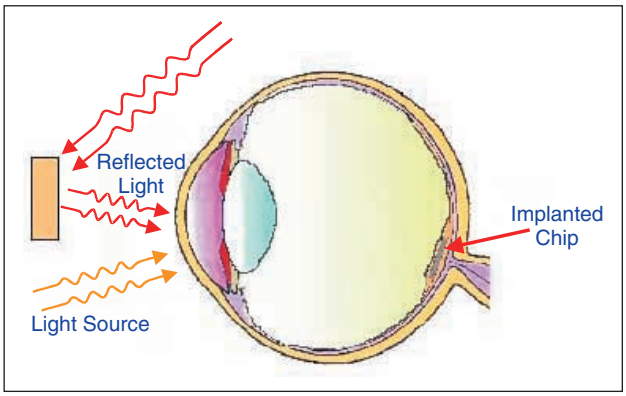
Figure 10(a) and (b) shows the measured output temporal responses of a single pixel in the fabricated FPA chip of the BJT-based silicon retina under three incident flash-light spot patterns with nearly the same intensity but different spot diameters. It can be seen in Fig. 10(a) that the positive pulse is generated when the flash-light source is turned on. The pulse has higher peak for smaller flash spot.

When the light is turned off, the measured output temporal responses of the fabricated silicon retina decays to the off state as shown in Fig. 10(b). The small negative pulse is not observed because the large off transition time of the flash-light source allows both isolated and smoothing BJT to be turned off simultaneously. The peaks in Fig. 10(b) are induced by the closing time of the mechanical shutter used in this measurement.

One of the most important applications of the silicon retina is the bionic retina implant system [6]. Patients who have diseases of the retina lose their vision. The purpose of the bionic retina implant system is to implant a silicon retina chip to replace, at least part of, the diseased retina so that the blind patient with retinal disease may restore vision. Basically, the developing architecture is conceptually shown in Fig. 11 where a silicon retina chip is implanted into the eyeball. The light reflected by an object or from an additional light source is used to generate the power supply for the chip by using the on-chip photovoltaic de-



10. The measured output (a) turn-on and (b) turn-off temporal responses of a single pixel in the improved 2-D νBJT silicon retina under different incident flash-light patterns with the same intensity and different sizes.



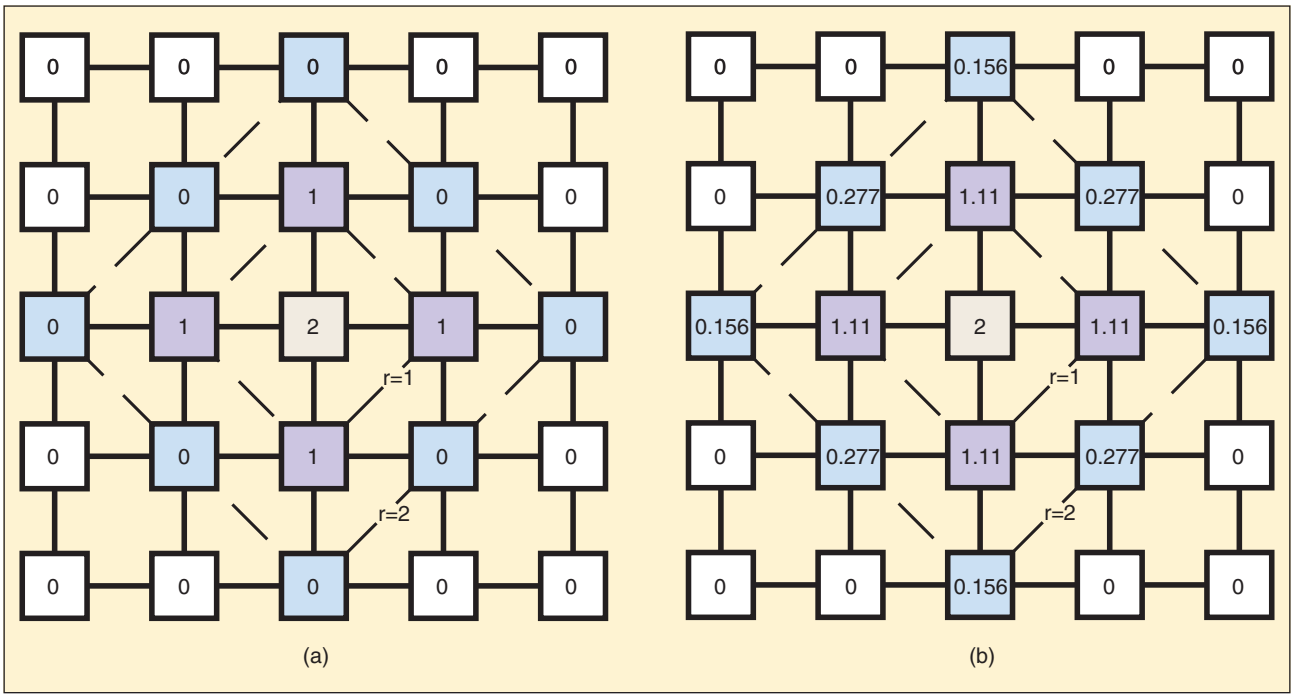
11. The conceptual diagram of silicon retina implant system.

vices. Thus, no external battery is required. Meanwhile, the retinal processing circuit is also integrated on the microchip to generate retinal pulses that are sent to ganglion cells. Through the ganglion cells, the pulses are sent to brain for vision. The silicon retina implant system will be a great challenge in the bioelectronics area.

νBJT Applications on Large-Neighborhood CNNs

The general architectures of cellular networks with the number *r* of connected neighborhood layers equal to 1 and 2 are shown in Fig. 12(a) and (b), respectively. In Fig. 12(a), the central neuron cell only receives the four state outputs of the nearest four neighboring cells with unity weighting factor and its own state outputs with the weighting factor of 1. The values shown in the cells are coefficients of the A templates, which correlate one cell's state input with other ones of neighboring cells. Only nearest neighboring cells interact with the cell. Thus, the number of connected neighborhood layers is equal to 1. If the state outputs of the second layer of neighboring cells is also input to the central cell with template coefficients as shown in Fig. 12(b), we have $r = 2$. Such a cellular network is called the large-neighborhood cellular network. Generally, it is difficult to directly implement the large-neighborhood cellular network because there are too many interconnections per cell. The resultant cell area would not be bearable.

To compactly implement cellular networks with single- and large-neighborhood templates [7], νBJTs are used as neurons with their base regions connected together through MOS resistors or some gain stages [8]. The base carrier diffusion mechanism is used for large-neighborhood connection. The base currents, which serve as state inputs or output currents, are decreased through the control of MOS resistors and/or amplified through gain stages. In such architectures, symmetrical or asymmetrical templates can be realized by adjusting the gate voltages of MOS devices. The cellular networks realized by νBJTs are called the νBJT cellular networks. In a νBJT cellular network, 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 is similar to that of the lambda bipolar transistor [9] and has a small chip area. The neuron input can be applied to the base of a νBJT through the NMOS transistors. Due to the compact structure, the νBJT cellular network requires a small chip area and has high integration capability. Since the neurons are realized by νBJTs, which can also be served as the phototransistors, the optical images can be entered directly to the νBJT cellular network without adding any separate sensor device.



12. The A templates with the number r of connected neighborhood layers equal to (a) 1 and (b) 2.

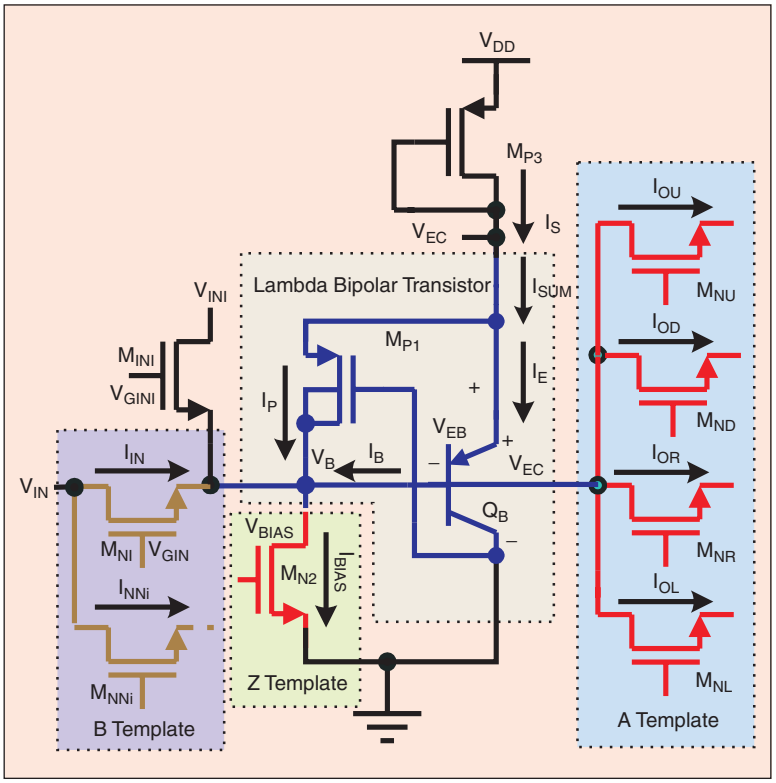
νBJT CNN with Symmetrical Templates

The basic cell circuit of the νBJT cellular network with a symmetrical template is shown in Fig. 13, 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 $I_{BO} = I_{BIAS}$. Such a neuron is called the νBJT neuron. The neuron's 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. For each I_{BIAS} value, the corresponding V_B value is unique. 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. 13, 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 cellular network is of the monotonic type of binary-valued cellular networks.

The self-feedback synapse of the A template in the cellular network is realized by the current I_P of the positive-feedback PMOS transistor M_{P1} with its gate connected to the ground and source (drain) connected to the emitter (base) of Q_B. The structure of Q_B and M_{P1} is called the lambda bipolar transistor as reported in [9]. 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 shared with the base of Q_B. Thus, the substrate of M_{P1} is connected to its drain and the positive substrate bias exists [9]. Since the neuron structure combines νBJT with the lambda bipolar transistor, it can be called the neuron-lambda-BJT neuron or νλBJT neuron. As shown in Fig. 13, the input capacitance of the νλBJT 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 synaptic coefficients of the A template can be realized by adjusting the self-feedback current I_P and 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. 13 through their gate voltages. The synaptic coefficients of the 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. 13, which can be adjusted by the corresponding gate voltages. In this way, the synaptic coefficients of the B template must have positive sign.



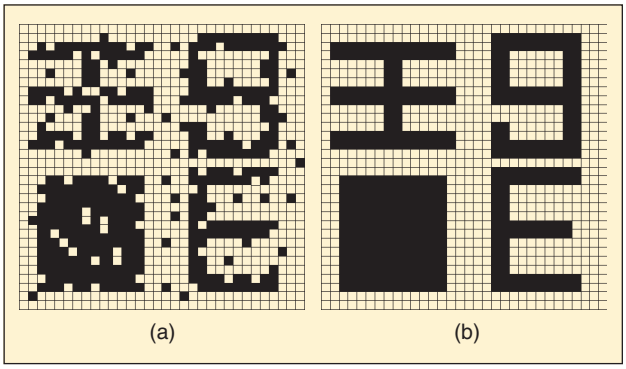
13. The complete cell circuit of one νλBJT neuron in the νBJT CNN.

By using the circuit of the νλBJT neuron as shown in Fig. 13, a 2-D νBJT cellular network can be formed. The fixed (Dirichlet) boundary condition is used to the νBJT cellular network. To verify its function, three applications with symmetrical templates are tested in the νBJT cellular network by using the HSPICE circuit simulation.

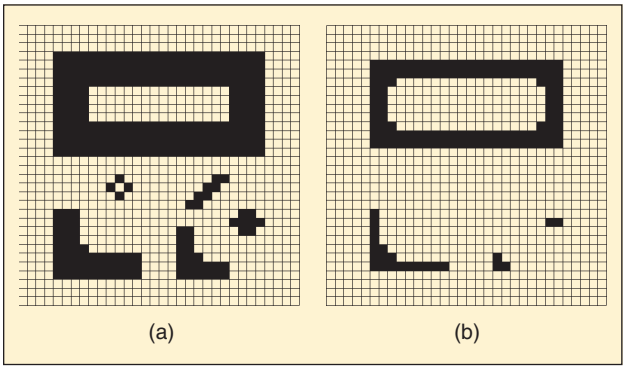
In the noise-removal operation, the cloning template is listed in Table 1 where the central weight is two times larger than its four neighboring weights [1]. To implement the noise removal operation, first the suitable gate voltages are applied to the gate of the MOS transistors realizing the template 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. 13, and the νBJT cellular network starts its operation. After the transient time, the νBJT cellular network can reach a steady state. The transient time depends on the resistance and the capacitance in the νλBJT neuron. The final steady state can be read out by sending out the state voltage V_B. Figure 14(a) shows the initial noisy image used to test the noise removal capability of the proposed νBJT cellular network. The image size is 32×32 pixels and the νBJT cellular network has 32×32 cells. The HSPICE-simulated output image

Table 1. Symmetrical 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
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



14. (a) The initial image and (b) the final output image in the νBJT CNN under the noise removal operation.



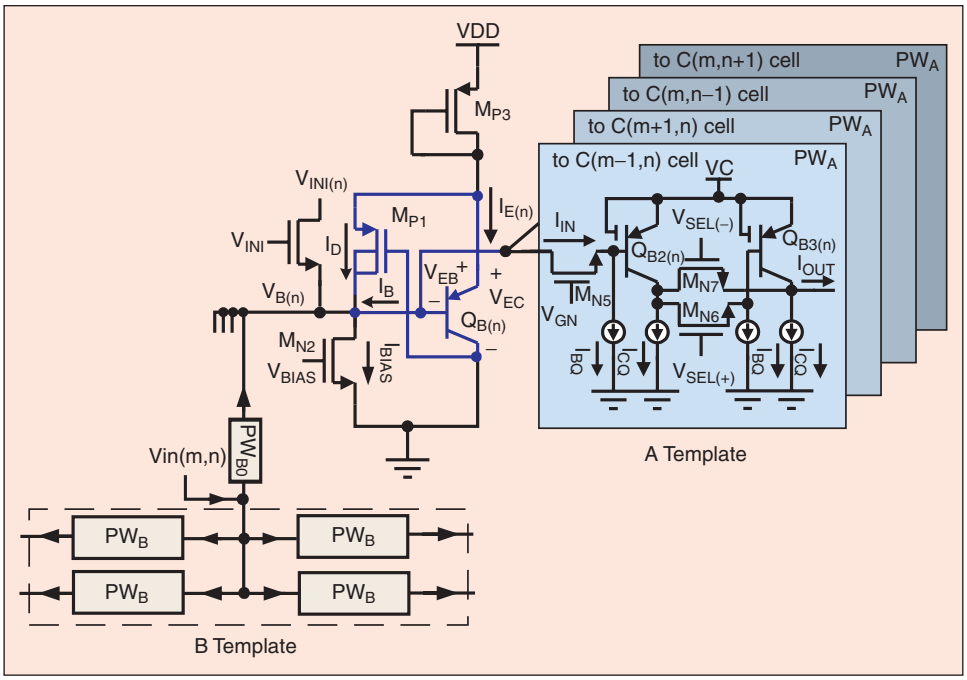
15. (a) The input image and (b) the output image in the νBJT CNN under the image erosion operation.

from the νBJT cellular network is shown in Fig. 14(b). It can be seen from Fig. 14(b) that the noise has been eliminated.

As a second example, the erosion operation is tested. The erosion templates are listed in Table 1 [7]. To implement the B template, the NMOS transistors M_{NI} and M_{NNi} for $i = 4$ as shown in Fig. 13 should be used. Figure 15(a) shows the input image used to test the image erosion operation. The HSPICE-simulated output image from the νBJT cellular network is shown in Fig. 15(b), which verifies the correct function of the νBJT cellular network in the erosion operation.

Smaller coupling resistors between neurons lead to a 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 cellular network. Figure 12 shows the A templates for the noise removal image processing with the number of neighborhood layers $r = 1$ and $r = 2$, respectively. In the $r = 2$ template, the synaptic coefficients 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).

By using the A template with $r = 1$ as shown in Fig. 12(a) and the input noisy image of Fig. 14(a) in the νBJT CNN, the output image is shown in Fig. 14(b) where the four-pixel square black or white noise images are not removed even if the self-feedback coefficient is reduced from 2 to 1. However, these noise images can be removed by using the A template with $r = 2$ as shown in Fig. 12(b). Since in the A template with $r = 2$ there is a larger spatial mask of 5×5, they have stronger local averaging effects that make 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 $r > 1$.



16. The circuit structure of the fully programmable νBJT CNN with A and B templates.

νBJT CNN with Asymmetrical Templates

To realize the asymmetrical template coefficients with positive or negative signs, a new synapse structure for the νBJT cellular network is proposed as shown in Fig. 16. In Fig. 16, the simple coupling MOS devices among the basic cell are replaced by the fully programmable weight stages PW_A, PW_B, and PW_{B0}. The PW_A stages realize the A template whereas the PW_B and PW_{B0} stages realize the B template.

The A template is realized by the stages PW_A, PW_{A1}, and PW_{A2}. The neuron output current is sent to the neighboring neurons through the stage

PW_{A2} that performs the current-to-current conversion with a suitable positive gain.

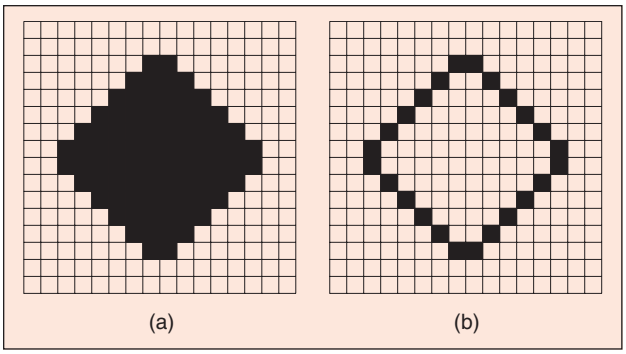
Each of the fully programmable weight stages PW_A, PW_B, and PW_{B0} in Fig. 16 has the same circuit structure. It consists of one MOS transistor M_{N5} with the gate voltage V_{GN} as the adjustable coupling resistor; two MOS transistors M_{N6} and M_{N7} as switches controlled by $V_{SEL(+)}$ and $V_{SEL(-)}$, respectively; and two common-emitter amplifiers with PNP lateral BJTs Q_{B2} and Q_{B3}; collector bias current source I_{CQ}; and base bias current source I_{BQ}. The absolute value of the weight can be adjusted through the gate voltage V_{GN} of the coupling MOS resistor M_{N5}. To realize the sign of the weights, the control voltages $V_{SEL(-)}$ and $V_{SEL(+)}$ are used to turn on M_{N7} for negative weights and M_{N6} for positive weights. Therefore, two control signals are needed for each weight. If the input current I_{IN} is zero, both Q_{B2} and Q_{B3} are in the active region with $I_{BQ} = I_{BI}$ and $I_{CQ} = \beta_L I_{BQ} = \beta_L I_{BI}$ so that the output current I_{OUT} is zero. If the input current I_{IN} applied to the base of Q_{B2} is positive (negative) and $V_{SEL(+)}$ is 3 V, I_{OUT} denoted as $I_O(+)$ is positive (negative) as a souring (sinking) current flowing out of (into) Q_{B3}. On the other hand, if I_{IN} is positive (negative) and $V_{SEL(-)}$ is 3 V, I_{OUT} denoted as $I_O(-)$ is negative (positive) as a sinking (souring) current flowing into (out of) Q_{B3}. In the circuit shown in Fig. 16, the parameters are set as $V_{DD} = 3$ V, $V_c = 1.15$ V, $I_{BQ2} = I_{BQ3} = 0.3$ µA, and $I_{CQ2} = I_{CQ3} = 20$ µA. The current gain β_L is 60.5.

By using the above-mentioned circuitry, a large-neighborhood cellular network can be realized. Given the template coefficients W_{ATr} of the rth neighborhood layer, W_A, W_{A1}, and W_{A2} of the stages PW_A, PW_{A1}, and PW_{A2} can be determined, respectively. Similar to the A template, the B template can be realized by W_B, W_{B0}, W_{B1}, and W_{B2} of the stages PW_B, PW_{B0}, PW_{B1}, and PW_{B2}, respectively.

Two image-processing functions are applied to the test of the νBJT cellular network with an asymmetrical single-neighborhood template. First, Fig. 17(a) shows the input image of a filled diamond pattern used in the edge-detection operation of the proposed fully programmable νBJT cellular network with a single neighborhood. The size of the νBJT cellular network is 16 × 16. The HSPICE-simulated output image is shown in Fig. 17(b). It can be seen from Fig. 17(b) that the edge of the diamond pattern has been extracted. Secondly, to test the shadow-projection operation, both A and B templates [9] in Table 2 are used. To realize the B template, the input image $V_{in(m,n)}$ is sent to the νBJT cells through the PW_{B0} stage and the neighboring cells through the PW_B stage. Both the PW_{B0} and PW_B stages are the same as the PW_A stage. The gate voltage V_{GIN} is adjusted to make I_{OUT} to be the same as the self-feedback current I_P in the A template. Each boundary cell is in the fixed condition. The neuron states are all initialized to the black stable state.

The proposed νBJT structure is compact and fully compatible with CMOS technology. Thus, it can be easily integrated with other integrated circuits to form a VLSI microsystem.

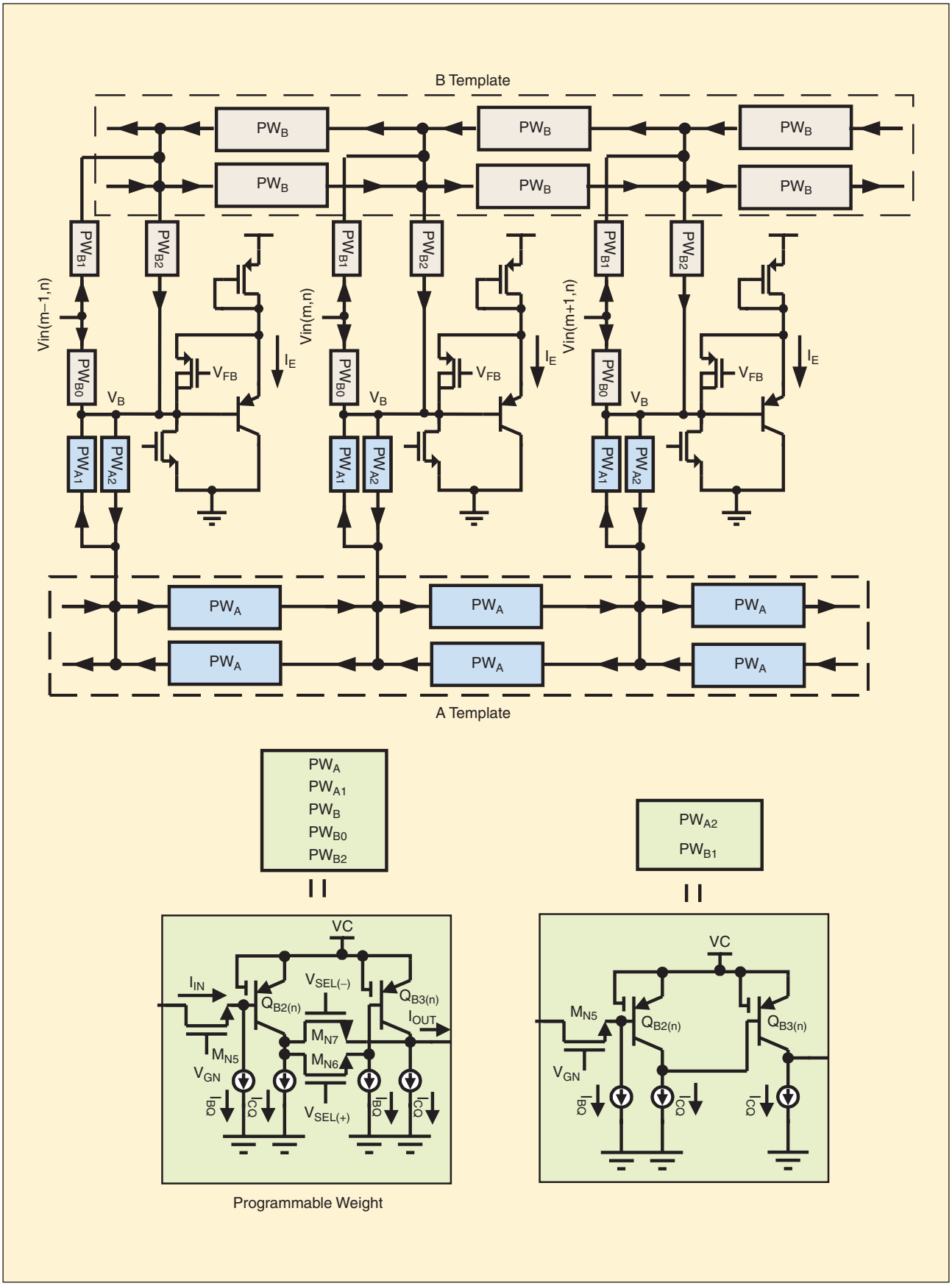
The proposed general structure of the νBJT cellular network with a large neighborhood is shown in Fig. 18, where the programmable weights of A and B templates are realized by the fully programmable stages that have the same structure as shown in



17. (a) The input image to the proposed νBJT CNN for edge detection; (b) the final output state of the νBJT CNN.

Table 2. Asymmetrical CNN Templates

Application	A	B	Z
Edge Detection CNN	$\begin{bmatrix} 0 & -0.25 & 0 \\ -0.25 & 2 & -0.25 \\ 0 & -0.25 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	–1
Corner Detection CNN	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	–8.5
Muller-Lyer Arrowhead Illusion CNN	$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} -0.1 & -0.1 & -0.1 & -0.1 & -0.1 \\ -0.1 & -0.1 & -0.1 & -0.1 & -0.1 \\ -0.1 & -0.1 & 1.3 & -0.1 & -0.1 \\ -0.1 & -0.1 & -0.1 & -0.1 & -0.1 \\ -0.1 & -0.1 & -0.1 & -0.1 & -0.1 \end{bmatrix}$	–2.8



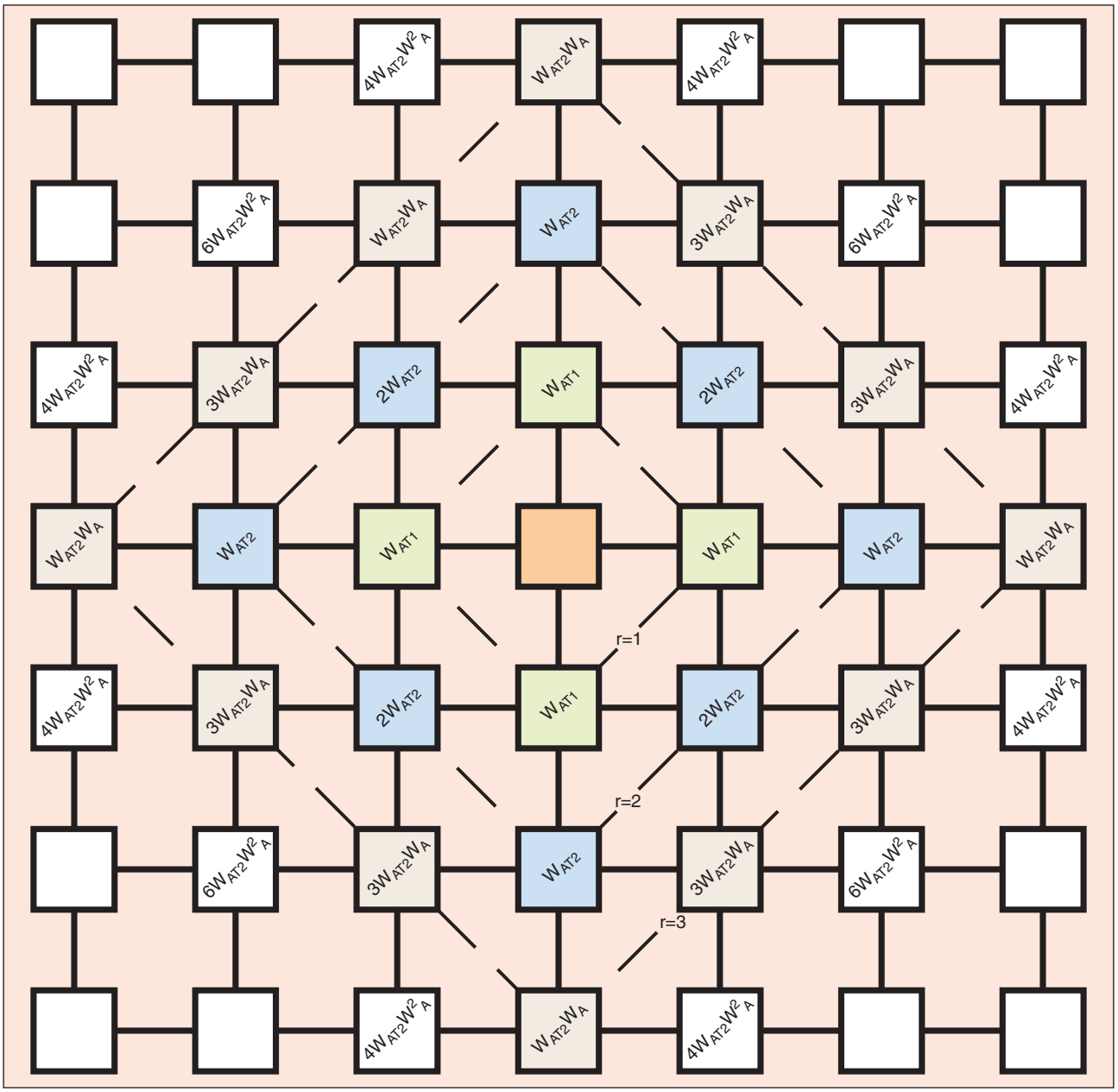
18. The general νBJT CNN structure with large neighborhood.

Fig. 16. The fully programmable stage in Fig. 16 is used to realize the stages PW_A, PW_{A1}, PW_{B0}, PW_B, and PW_{B1} as repeatedly shown in Fig. 18. The positive programmable stage without sign-selection MOS switches is used to realize the stages PW_{A2} and PW_{B1}, as shown in Fig. 18.

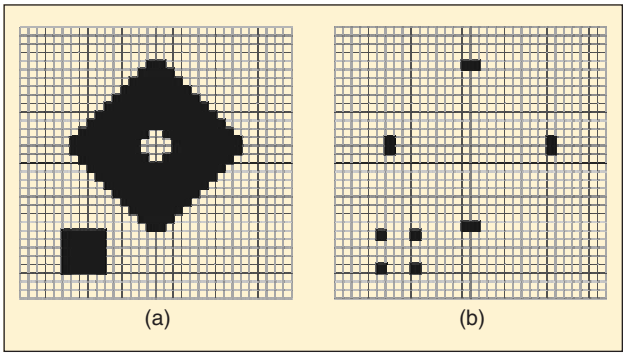
To verify the correct functions of the proposed BJT cellular network structure with a large neighborhood, two examples are realized and simulated in an HSPICE circuit simulator. Each boundary cell is in the fixed condition. The first example is for the corner-detection operation. Both the A and B templates are listed in Table 2 [7], where the A template has only one self-feedback weight. As compared with the template information shown in Fig. 19, the B template in Table 2 has two neighborhood layers with the negative weight $W_{BT1} = -1$ and $W_{BT2} = -1$ or 0. The positive central weight is eight times larger than those of the eight negative neighboring weights [7]. From Table 2, $W_B = 1$, W_{B1},$W_{B2} = -1$, and $W_{B0} = 9.5$.

Figure 20(a) shows the input image used to test the image corner-detection operation. The HSPICE-simulated output image from the vBJT cellular network is shown in Fig. 20(b), which verifies the correct function of the vBJT cellular network in the corner-detection operation. It should be noted that although the realized coefficients in the B template have some difference as compared with those in Table 2, the corner-detection function is still correct. Thus the difference is tolerable.

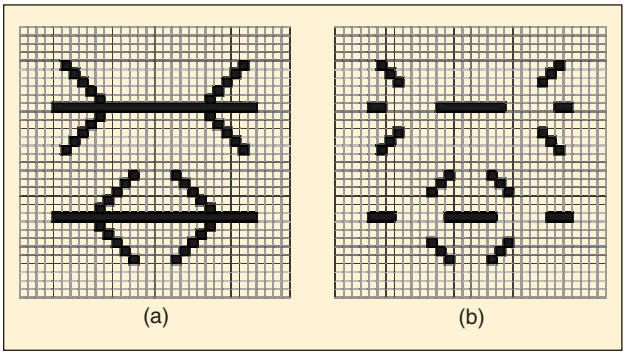
The second example is the vBJT cellular network for the Muller-Lyer arrowhead-illusion function. Figure 21(a) shows the 32 $\times$ 32 input image used to test the Muller-Lyer arrowhead illusion



19. The realizable large neighborhood A template for CNNs with four nearest neighborhood cells.



20. (a) The input image and (b) the final output image in the νBJT CNN under the corner detection operation.



21. (a) The input image and (b) the HSPICE simulated final output image in the νBJT CNN under the Muller-Lyer illusion operation.

function of the realized νBJT cellular network. In Fig. 21(a), there are two horizontal lines. One has diverging arrowheads at the ends whereas the other has converging arrowheads. Though the two lines are of the same length, the line with the converging arrowheads appears decidedly shorter. The HSPICE-simulated output image is shown in Fig. 21(b). It can be seen from Fig. 21(b) that the horizontal line with converging arrowheads becomes shorter as expected. It should be noted that although the realized coefficients in the B template have some difference as compared with those in Table 2, the Muller-Lyer arrowhead-illusion function is still correct. Thus the difference is tolerable.

Conclusions

In this article, a new device structure called the νBJT was presented. It has been successfully applied to the design of a silicon retina and large-neighborhood CNNs. The νBJT-based smoothing array for the silicon retina has a simple and compact structure, which is suitable for the VLSI implementation. It can be integrated with other CMOS retinal signal processing circuits to form smart sensor systems. On the other hand, based upon the νBJT structure, new compact νBJT CNN structures with programmable large neighborhood templates for image processing have also been successfully designed in CMOS technology. The proposed νBJT CNN is designed to implement large neighborhoods without adding too many extra interconnections and circuits. Using the proposed structure, the template coefficients with two neighborhood layers are fully realizable.

In the future, the research on efficient physical structures for the implementation of the bionic silicon retina functions or smart retinal processing will progress extensively toward nanoscale devices or integration. Further research on large-neighborhood νBJT CNNs will also advance toward the nanoelectronics regime with the feasible development of a powerful nanoscale CNN universal machine(CNNUM).

Chiu-Hung Cheng is a Ph.D. student at National Chiao Tung University in Hsinchu, Taiwan. *Dr. Chung-Yu Wu* is a Chaired Professor at National Chiao Tung University and also serves as Founding Chair of Technical Committee on Nanoelectronics and Giga-Scale Systems for the IEEE Circuits and Systems Society. *Dr. Bing Sheu* is with Nassda Corporation and is 2001 Past President of the IEEE Circuits and Systems Society. *Li-Ju Lin* and *Kuan-Hsun Huang* are Ph.D. students and *Chiao-Wei Hsiao* is an M.S. student with the Integrated Circuits and Systems Lab at National Chiao Tung University. *Hsin-Chin Jiang* is the section manager of the Analog IP Technology Section in the IP Technology Department of U2000/SoC Technology Center (STC) at the Industrial Technology Research Institute (ITRI) in Taiwan. *Wen-Cheng Yen* is an engineer at Terax Communication Technologies Inc. in Hsinchu, Taiwan.

References

1. L.O. Chua and L. Yang, "Cellular neural networks: Theory," *IEEE Trans. Circuits Sys.*, vol. 35, pp. 1257-1272, Oct. 1988.
2. C.Y. Wu and H.C. Jiang, "An improved BJT-based silicon retina with tunable image smoothing capability," *IEEE Trans. VLSI*, vol. 7, no. 2, pp. 241-248, Jun. 1999.
3. C.A. Mead, *Analog VLSI Implementations of Neural Systems*. Reading, MA: Addison Wesley, pp. 239-246, 1989.
4. A Hummel, "Representations based on zero-crossings in scale-space," in *Proc. IEEE Computer Vision and Pattern Recognition Conf.*, pp. 204-209, Jun. 1986.
5. C.Y. Wu and C.F. Chiu, "A new structure of 2-D silicon retina," *IEEE J. Solid-State Circuits,* vol. 30, no. 8, pp. 890-897, Aug. 1995.
6. M. Schwarz, R. Hauschild, B.J. Hosticka, J. Huppertz, T. Kneip, S. Kolnsberg, L. Ewe, and T. Hoc Khiem, "Single-chip CMOS image sensors for a retina implant system," *IEEE Trans Circuits Syst. II*, vol. 46, no. 7, pp. 870-877, Jul. 1999.
7. L.O. Chua, *CNN: A Paradigm for Complexity* (World Scientific Series on Nonlinear Science, vol. 31). Singapore: World Scientific, 1998.
8. C.Y. Wu and W.C. Yen, "A new compact neuron-bipolar junction transistor (νBJT) cellular neural network (CNN) structure with programmable large neighborhood symmetric templates for image processing," *IEEE Trans. Circuits Syst. I*, vol. 48, no. 1, pp. 12-27, Jan. 2001.
9. 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.

CD■