A Compact Low-Power CMOS Analog FSR Model-Based CNN

Edson P. SANTANA[1], Raimundo Carlos S. FREIRE[2], and Ana Isabela A. CUNHA[3]

[1]Instituto Federal da Bahia, Vitória da Conquista, BA, Brazil
[2]Departamento de Engenharia Elétrica, Universidade Federal de Campina Grande, Campina Grande, PB, Brazil
[3]Departamento de Engenharia Elétrica, Universidade Federal da Bahia, Salvador, BA, Brazil

e-mail: edsonp@ifba.edu.br

ABSTRACT

A compact low-power CMOS analog circuit implementation of a Cellular Neural Network based on Full Signal Range Model (FSR-CNN) is presented. The required operations in cell definition are synapses (multiplication and summation) and saturated integration. In each synapse, a new multiplier architecture is employed with voltage and current inputs and a current output, which allows sharing building blocks and using continuously programmable weight values. Feasibility and usefulness of the proposed FSR cell architecture is verified through the simulation of two applications: the connected component detector and the border extractor.

Index Terms: Cellular Neural Network, Four-Quadrant Multiplier, Analog Focal Plane Image Processing.

1. INTRODUCTION

On the development of real time image processing arrays, Cellular Neural Networks (CNN) stand out as a promising alternative. Similarly to other neuromorphic systems [1]-[2], CNNs have been used to artificially implement the first steps of visual processing [3], finding applications from retinal prosthesis to robotic sensors and featuring a good trade-off between operating speed and versatility [4].

An analog CNN has been defined by Chua et al. in [5] as a dense matrix of neuron cells whose model includes synapses (analog multiplication and summation), integration, limiting and spatial interaction with neighbor cells. The Full Signal Range (FSR) cell model [6]-[7] provides reduction in circuit complexity, as well as robustness and augmented processing speed.

Since the early nineties several implementation approaches have been applied to analog CNN using continuous-time and discrete-time models. Continuous-time implementations comprise current-mode [8], gm-C [9], and OTA-based [9] techniques. More recently, Gálan et al. in [3] and Dominguez-Matas et al. in [10] have shown that the use of voltage inputs and a binary-weighted current output in each synapse provides a dense summation at a single node, simplifies the integration scheme and avoids high density current replication. However, the inherently limited set of discrete weights allowed encumbers template optimization and learning process.

In this work we propose an analog FSR-CNN implementation using a new multiplier architecture with a voltage input, a current input and a current output with continuous range weight values, which reduces circuit complexity and power consumption.

2. FSR CELL MODEL

A standard Cellular Neural Network (CNN) consists of a rectangular array of cells whose dynamics and interrelation with neighbors are defined by [5]:

$$\dot{x}^c = -x^c + \sum_{d \in S_r(c)} \left(A_d^c y_d + B_d^c u_d \right) + D^c \qquad (1)$$

where x^c and u_d are the state and input variables, respectively, D^c is the threshold, A_d^c and B_d^c are the feedback and input synaptic operators, respectively, S_r is the neighborhood of radius r and $y_d = f(x^c)$ is the output variable given by a limiting function of x^c, e.g. $2f(x^c) = |x^c + 1| - |x^c - 1|$ (standard nonlinearity). Index c denotes a generic cell and d denotes a specific position (i,j) in the neighborhood relative to the cell c, where i and j are integer numbers (e.g. $d = i,j = 0,-1$ denotes a position corresponding to the neighbor cell at the same line and at the immediately left column with respect to cell c). In the case of space-invariant CNN, which covers most practical applications, the synaptic operators are represented by only two coeffi-

cient matrices, 3x3 in the case of r = 1: the feedback cloning template ($\bar{A}$) and the input cloning template ($\bar{B}$).

The FSR cell model introduced by Rodríguez-Vázquez et al. in [6] and analyzed concerning stability and convergence properties by Espejo et al. in [7] differs from standard cell model mainly due to the fact that state variable values are constrained to the cell input range, eliminating the need for a cell output nonlinear limiter. Dynamics of the cell block diagram depicted in Fig.1 is governed by the following expressions [6]-[7]:

$$\tau \dot{x}^c = g\left(x^c\right) + \sum_{d \in S_r(c)} \left(\hat{A}_d^c y_d + B_d^c u_d\right) + D^c \qquad (2.a)$$

$$g(x) = \lim_{m \to +\infty} \begin{cases} -m(x+1), & \text{if } x \le -1 \\ 0, & \text{if } -1 \le x \le +1 \\ -m(x-1), & \text{if } x \ge +1 \end{cases} \qquad (2.b)$$

where τ is the integration time constant, $x^c(0)$ is the initial value of the state variable at cell c, $g(x)$ is a clamping function and $\hat{A}_d^c = A_d^c$, except for $\hat{A}_{0,0}^c = A_{0,0}^c - 1$.

The main advantage of the FSR cell model over standard cell model is that the dynamic ranges of state and output variables are equal thus leading to more compact and robust circuit implementations [7].

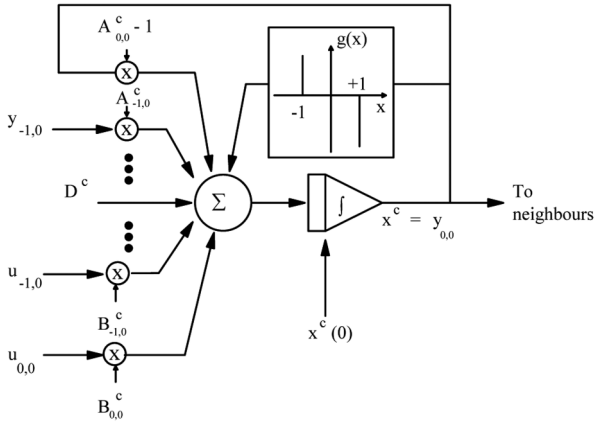


Figure1. Block diagram of FSR cell.

3. CIRCUIT DESCRIPTION AND DESIGN ISSUES

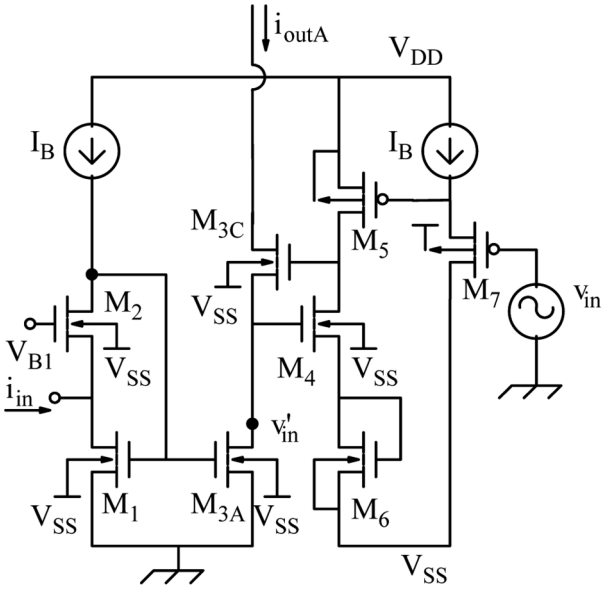
A. Proposed Multiplier: Architecture and Principle of Operation

Fig.2(a) depicts the core of the multiplier we have proposed in [11] to accomplish the synapses in our FSR cell. Transistors M_1 and M_{3A} operate in the beginning of nonsaturation, where a linear relationship between drain current and drain-source voltage is assumed. Moreover, M_1 and M_{3A} share gate, bulk and source terminals, thus performing source transconductance mirroring [12]. Since M_2 operates in saturation with constant gate voltage and constant drain current, and disregarding short channel effects as well as current source nonidealities, the drain-source voltage of M_1, v_{DS1}, may be assumed constant. Therefore, the source transconductances of M_1 (g_{ms1}) and M_{3A} (g_{ms3A}) are proportional to $i_{in} + I_B$, the sum of input and bias currents, respectively. Drain current of transistor M_3, $i_{outA} = g_{ms3A} v'_{in}$, is thus proportional to the product of drain-source voltage and current $i_{in} + I_B$.

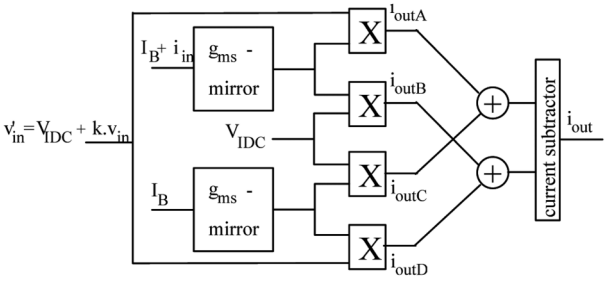
Transistors M_{3C}, M_4, M_5, M_6 and M_7, which operate in saturation, provide $v'_{in} = v_{IDC} + k v_{in}$, with constant V_{IDC} and k, and a high impedance output terminal. M_{3C} decouples the output terminal from the input voltage v_{in}, which is level shifted from the gate of M_7 up to the gate of M_5. Series connected transistors M_5 and M_6 are designed for a constant difference between gate potentials. Another level shift is accomplished between the gates of M_4 and M_6, which present equal aspect ratios. Therefore, V_{IDC} is the total level shift from the gate of M_7 to the gate of M_4 and k is due to the body effect. The design expressions regarding this set of transistors are presented in Appendix.

The output current, therefore, results

$$i_{outA} = \frac{\left(I_B + i_{in}\right)}{V_{DS1}} \left(V_{IDC} + k v_{in}\right) \qquad (3)$$



(a)



(b)

Figure 2. Proposed multiplier. (a) Core circuit. (b) Cancellation scheme.

In order to obtain a four-quadrant multiplier, the undesirable terms of the output current in (3) should be canceled. In the cancellation scheme of Fig.2(b), each input signal (v_{in} or i_{in}) is applied to only one node, not needing replication or inversion. The current subtractor here adopted is a cascoded version of the circuit used by Machado et al. in [12], as illustrated in Fig.3. The schematics of a general purpose four-quadrant multiplier using the proposed core circuit is depicted in Fig.4.

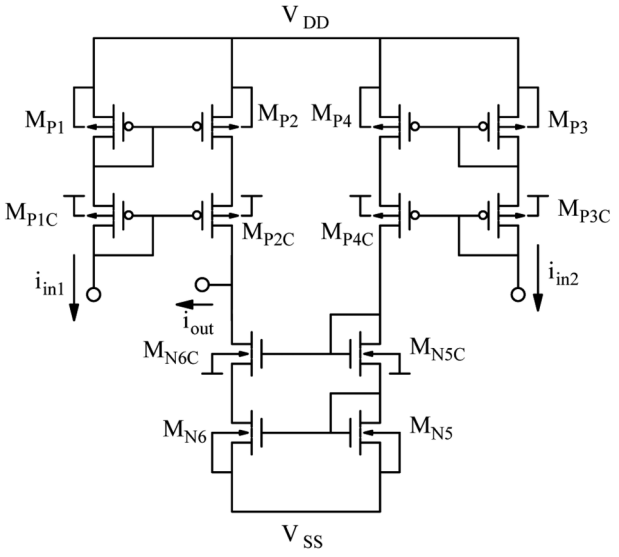


Figure 3. Current Subtractor.

B. Proposed Multiplier: Simulation Results

The proposed four-quadrant multiplier of Fig.4 has been designed in CMOS IBM 0.13 µm technology, adopting V_{DD} = -V_{SS} = 0.6 V, I_B = 500 nA, and the following extracted parameters of Advanced Compact MOSFET (ACM) model [18]: n_N = 1.379, n_P = 1.295, V_{T0N} = 0.362 V, V_{T0P} = 0.356 V, I_{SQN} = 353.38 nA and I_{SQP} = 49.51 nA. Long channel devices have been applied in order to reduce short channel effects and mismatching. Designed aspect ratios are listed in Table I.

Table II summarizes the main features of the designed four-quadrant multiplier, obtained through simulation on Mentor Graphics Tools [17] using IBM-8RF DM design kit.

Table I. Dimensions of transistors in Fig.3(b).

MOSFET	W/L (µm/µm)
$M_{1X(Y)}$, $M_{3A(B)(C)(D)}$	0.95 / 1.50
$M_{2X(Y)}$	5.39 / 1.50
$M_{3CA(B)(C)(D)}$	15.00 / 1.50
$M_{4(A)(B)(C)(D)}$, $M_{6(A)(B)(C)(D)}$	0.60 / 15.90
$M_{5(A),(B),(C),(D)}$	0.60 / 2.87
M_7	19.94 / 0.60

Table II. Simulation results.

Parameter	Value
Active area (µm²)	764
Power consumption (µW)	20
Bandwidth (MHz)	1.4
RMS output current noise (nA)	1
THD (dB)	< -34

Fig.5 depicts the simulated DC transfer characteristics of the proposed multiplier using (a) constant input current and (b) constant input voltage. It should be noted that the curves depart very slightly from the expected linear behavior for one constant input: the worst offset (maximum output current value for one zero input) is 0.73 nA in case (a) and -1.30 nA in case (b) and the maximum slope variation is 5.5 % in case (a) and 8 % in case (b). Symmetry is most affected in the case of varying input current, which is not critical since input currents represent constant, though programmable, coefficients in a CNN application.

In order to analyze the linearity performance concerning each input, the total harmonic distortion (THD) has been evaluated through simulation for: (i) a constant input current of 160 nA and a sinusoidal input voltage of 1 kHz and amplitude varying from 1 to 16 mV; (ii) a constant input voltage of 16 mV and a sinusoidal input current of 1 kHz and amplitude varying from 10 nA to 160 nA.

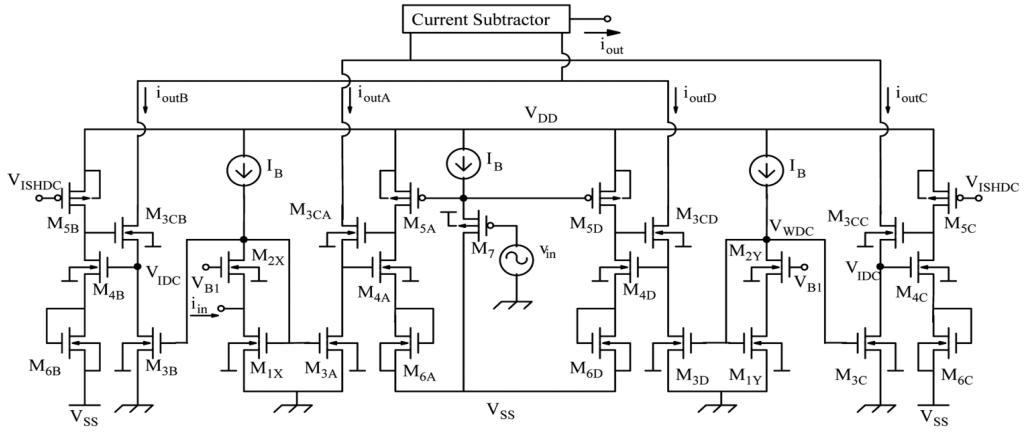


Figure 4. Four-quadrant multiplier.

The results are illustrated in Fig.6. THD is less than approximately –43 dB (0.7 %) for the entire range of input voltages and less than approximately –34 dB (2 %) for the entire range of input currents.

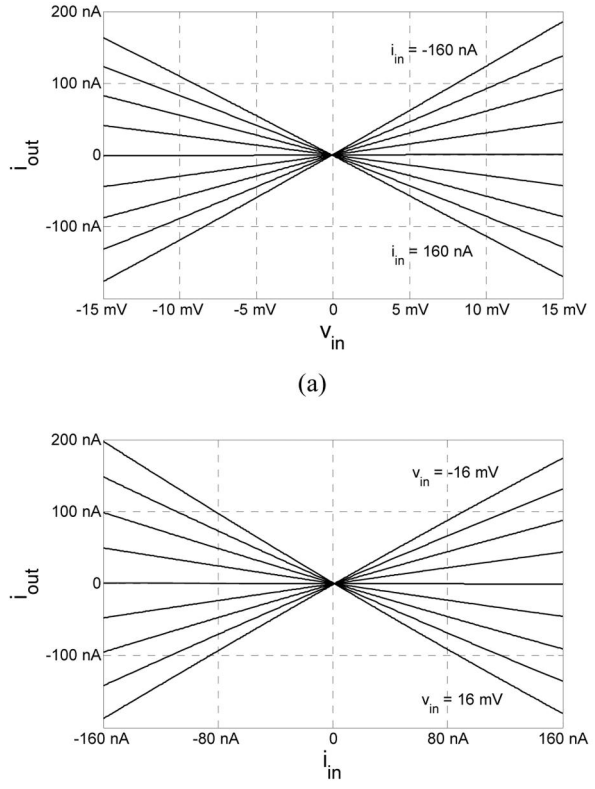


Figure 5. Simulated DC transfer characteristics of designed four-quadrant multiplier. (a) constant input current: i_{in} = -160 to 160 nA (step of 40 nA); (b) constant input voltage: v_{in} = -16 to 16 mV (step of 4 mV).

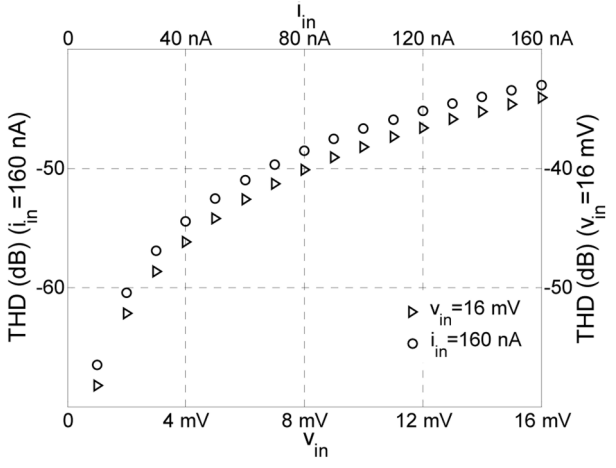


Figure 6. Simulated Total Harmonic Distortion of the designed four-quadrant multiplier for: i_{in} = 160 nA and 1 kHz sinusoidal v_{in}(circles); v_{in} = 16 mV and 1 kHz sinusoidal i_{in} (triangles).

C. Complete FSR Cell

Fig.7 illustrates the FSR cell proposed in this work [13]. Each synapse block accomplishes a multiplication task, as required in the FSR block diagram of Fig.1. In Fig.7, either the feedback (A_d^c) or the input (B_d^c) synaptic operators are represented by current signals, as i_{in} in the weight generator block, which corresponds to the left-most part of the multiplier core in Fig.2a. The weight generator converts each current signal representing a multiplication weight into a voltage signal, such as v_w, to be applied to the gates of transistors M_{3A} and M_{3B} of the synapse block in Fig.7, thus conveying the source transconductance of M_1.

Each input (u_d) or output ($x^c = y_{0,0}$) variable is embedded in a voltage signal such as v_{ish} in Fig.7, which is applied to the gates of transistors M_{5A} and M_{5D} of the synapse block. Indeed, in the case of the output (state) variable, v_{ish} is the result of shifting, through the level shifter block, the output voltage of the clamper block in Fig.7.

The output currents of all synapses (1 to N) are summed in node X and are integrated through the gate capacitances of transistor M_7 in the level shifter and transistors M_{8A} and M_{8B} in the clamper. Despite the integrator block in the diagram of Fig.1, the integration of the state variable in Fig.7 is not explicitly performed by a particular block. Whenever the clamper enters in operation, its output current (I_{clamp}) is also added to synapses output currents in order to maintain node X voltage inside input range. It should be noted that the node X is the clamper output node for current signals and is the clamper input node for voltage signals. Node X voltage, which is proportional to the integrated currents, is also the input of the level shifter. The shifted voltage (cell output v_o) is then applied to the synapses inputs of other cells or to a synapse input of the same cell in a feedback loop.

Due to the multiplier architecture modularity, block sharing allows significantly decreasing in circuit complexity, size and mismatching.

In order to accomplish the numerous synapses present in a high density CNN with minimal area, the weight generators, each one consisting of a current input and transistors M_1 and M_2 of the multiplier core in Fig.2(a), are shared by all FSR cells, assuming space invariant CNN. Moreover, the block that processes level shifting of the multiplier input voltage in Fig.2(a) (transistor M_7, and associated bias current source) is moved to the cell output in Fig.7. Since the cell output voltage is already level-shifted, the need of a level shifter for each synapse input is avoided. The blocks generating biasing voltages V_{ISHDC} and V_{WDC} are shared by all cells. V_{WDC} is associated to g_{ms1}, which is constant and proportional to I_B in this block.

We have adopted the clamping circuit presented by Hegt et al. in [14], as illustrated in Fig.7. In our FSR cell, the clamper has been dimensioned to maintain state voltage within the synapse input range (±15 mV), with a quiescent power dissipation of approximately 4.8μW.

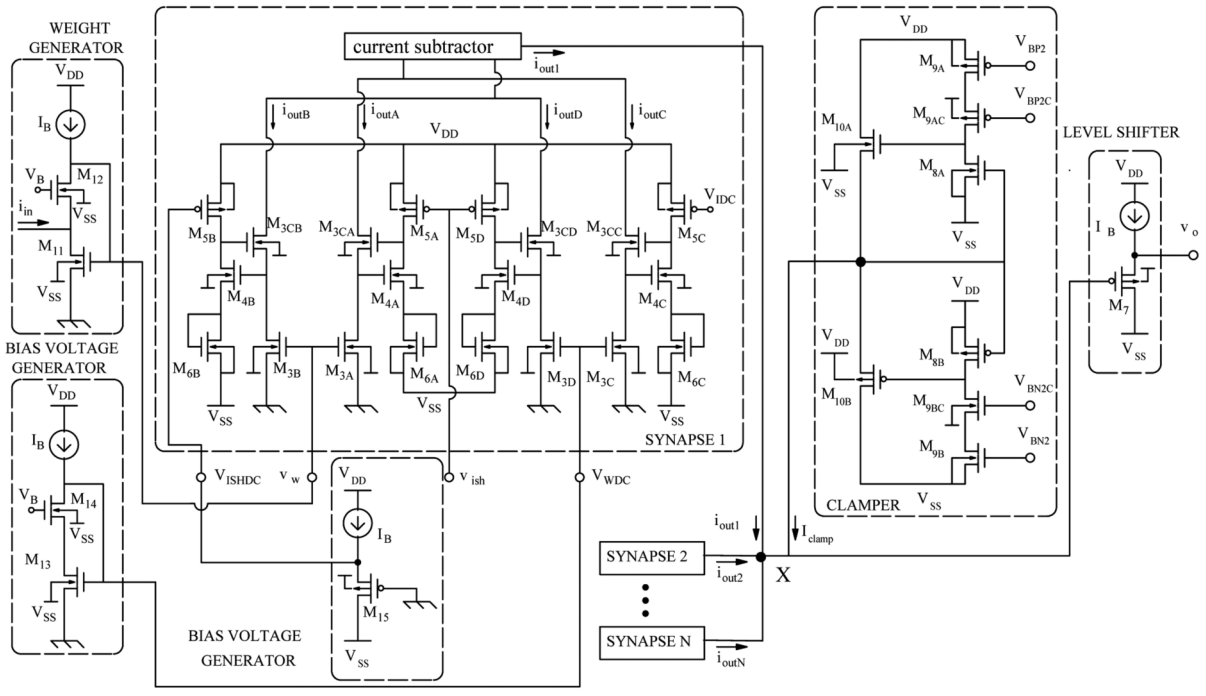


Figure 7. Circuit of FSR cell.

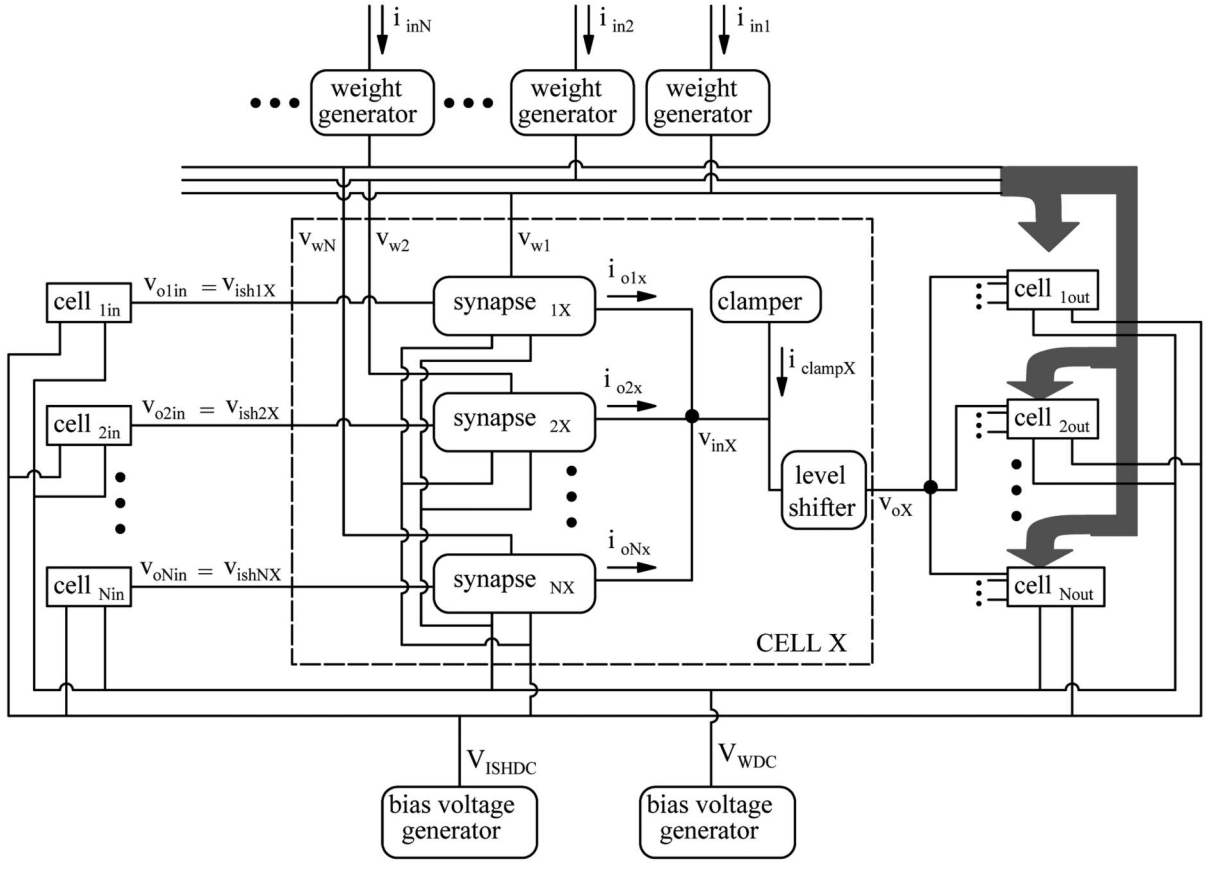


Figure 8. Block diagram of a space invariant neighborhood of the proposed FSR cells.

4. APPLICATIONS AND SIMULATION RESULTS

A. Connected Component Detector

The feasibility and performance of the proposed cell architecture has been analyzed through the simulation of a CCD (connected component detector) of sixteen cells in a row (on Mentor Graphics Tools using IBM-8RF DM design kit). For a given initial state of black and white pixels, the expected result from the CCD template is a row of alternate black and white pixels from right to left, whose number of black (white) pixels is equal to the number of connected blacks (whites) in the initial state, in case of the leftmost pixel is white (black).

Since CCD configuration of templates lies on the stability boundary of parameter space, this application is frequently referred as a ubiquitous CNN test circuit, being useful to check robustness of circuit implementation [15].

In order to synthesize the weights of CCD template [16] we have adopted current inputs of 30 nA, 30 nA and -30 nA in the left, self-feedback and right template elements, respectively. Constant state values have been assigned to left and right border cells, which influence the previously explained behavior of the CCD template. As shown in Fig.9, the correct final state values for nominal model parameters have been reached in less than 6 μs.

B. Border Extraction

The applicability of the proposed cell architecture to image processing in two dimensions is illustrated by the simulation of a border extractor (using Mentor Graphics Tools and IBM-8RF DM design kit). A 5x5 cell array has been designed for this purpose, using the feedback and input cloning templates of Fig.10, in which the actual values of synapses current inputs are presented. The threshold has been implemented through a synapse in which the voltage and current input values are -15 mV and 15 nA, respectively. In border cells, the value of -15 mV has been assigned to state and input variables.

Fig.11 shows the initial and final state voltages linearly scaled between [-15 mV, 15 mV] in a gray level representation for two different input images (equal to corresponding initial states).

In Fig.12 a three-dimensional representation of the normalized state voltages illustrates the temporal evolution and signal propagation of the network. Convergence to the final expected values is achieved in less than 31 μs.

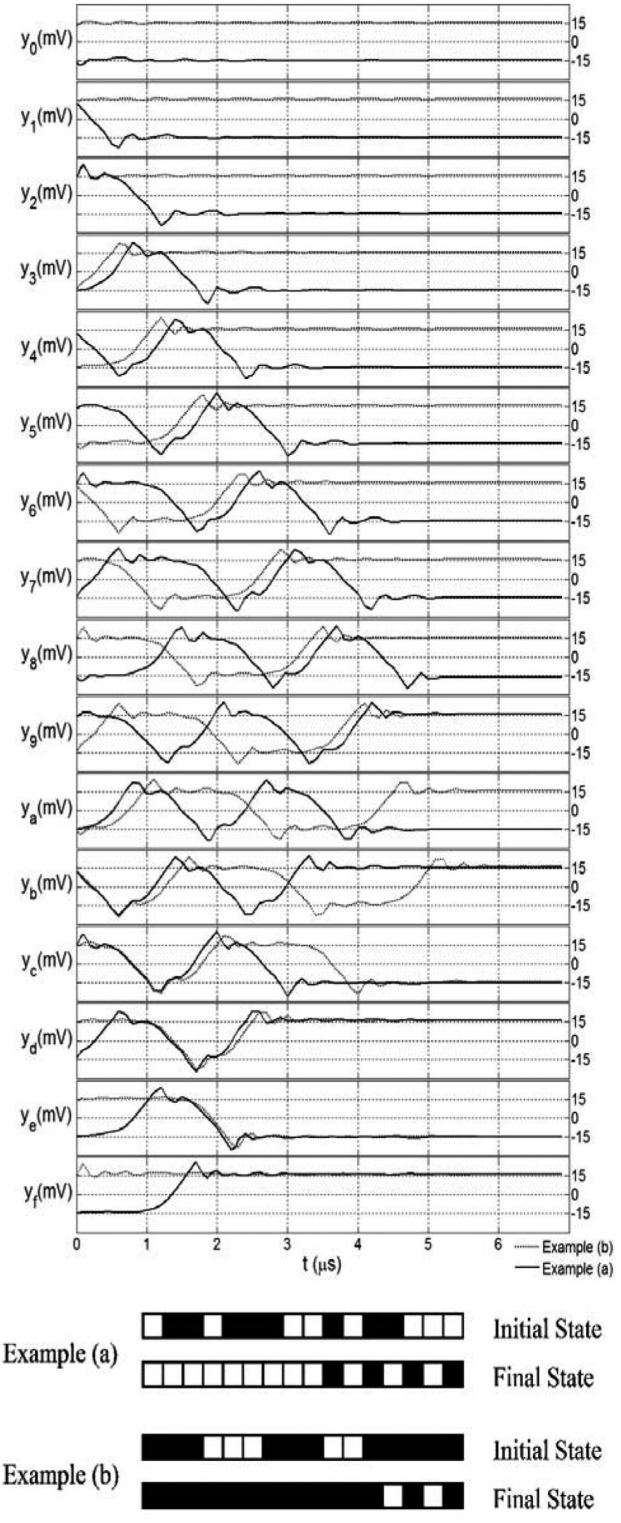
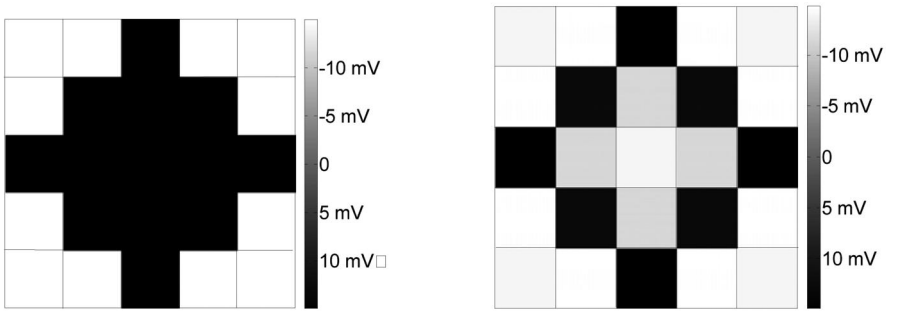
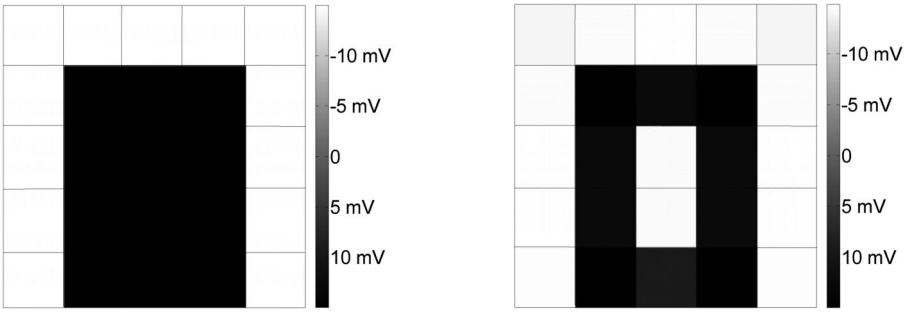


Figure 9. Transient response of a 16-cell CCD for initial state (a) in solid line and (b) in dashed line. Constant state voltage in left border cell is equal to -15 mV for both examples and in right border cells is equal to -15 mV for example (a) and to 15 mV for example (b).



(a)



(b)

Figure 11. Simulation results of the border extraction application for two input images example.

$\overline{A} =$

-2.5 nA	-2.5 nA	-2.5 nA
-2.5 nA	20 nA	-2.5 nA
-2.5 nA	-2.5 nA	-2.5 nA

$\overline{B} =$

0	0	0
0	10 nA	0
0	0	0

Figure 10. Input current for the synapses in border extraction templates.

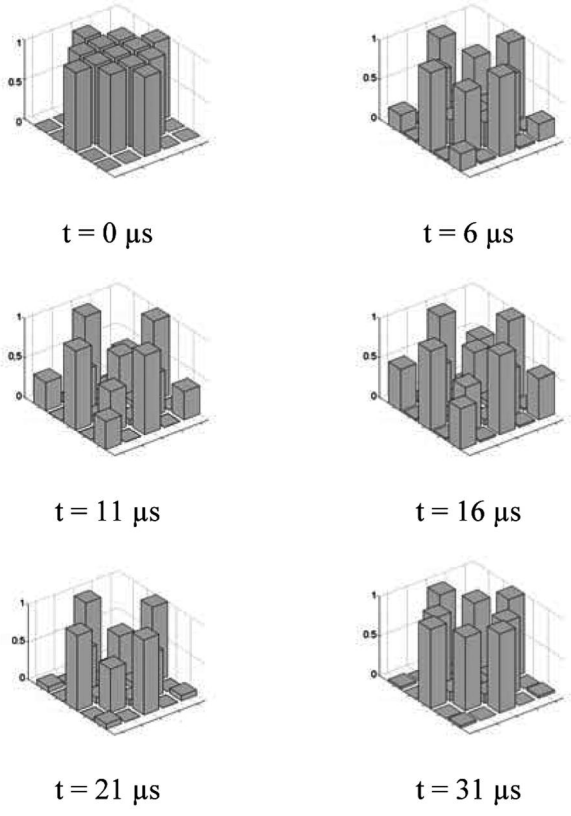


Figure12. Temporal evolution and signal propagation for input image example (a) of Fig.11.

5. CONCLUSIONS

This work presents a new proposal to the analog implementation of a FSR model-based CNN. Architectural premises have been used to reduce circuit complexity, leading to a compact network with low-power consumption, as verified in simulation results.

Transient simulation of a CCD template using the proposed CMOS architecture of FSR cell revealed a satisfactory performance regarding robustness and velocity. Two-dimensional image processing has been illustrated through the simulation of the border extraction function using a 5x5 array of the proposed FSR cell circuit. The expected final states have been achieved with enough resolution within a short period of time.

It should be emphasized that the use of current signals to synthesize template weights allows network programmability in a continuous sense, thus providing a high degree of flexibility to the analog implementation of complex image processing operations.

ACKNOWLEDGEMENTS

The authors would like to acknowledge FAPESB, NAMITEC and CNPQ for the financial support, and MOSIS for providing design kits.

APPENDIX

The relationship between v_{in} and v'_{in} in the core circuit of the proposed multiplier, illustrated in Fig. 3(a), results from the following assumptions: (i) transistors M_4, M_5, M_6 and M_7 operate in saturation; (ii) transistors M_4, M_5 and M_6 have equal inversion levels ($i_{f4} = i_{f5} = i_{f6}$). Applying these conditions to the expressions of the Advanced Compact MOSFET (ACM) model listed in Table III, leads to:

$$\frac{\left(v'_{in} - V_{SS} - V_{T0N}\right)}{n_N} - V_{G6} = \phi_t F(i_{f4}) \qquad \text{(A1)}$$

$$\frac{-V_{G5} - |V_{T0P}|}{n_P} = \phi_t F(i_{f5}) \qquad \text{(A2)}$$

$$\frac{V_{G6} - V_{T0N}}{n_N} = \phi_t F(i_{f6}) \qquad \text{(A3)}$$

$$\frac{\left(V_{DD} - v_{in} - |V_{T0P}|\right)}{n_P} + V_{G5} = \phi_t F(I_B / I_{S7}) \qquad \text{(A4)}$$

In (A1)–(A4), as well as in the expressions of Table III, i_f is the normalized forward saturation current (inversion level), V_G is the gate-to-bulk voltage, $V_{S(D)}$ is the source (drain)-to-bulk voltage, V_{T0} is the threshold voltage, φ_t is the thermal voltage, n is the slope factor, I_S is the specific current, μ is the carrier mobility, C'_{ox} is the oxide capacitance per unit area, W is the channel width, L is the channel length and V_{DSSAT} is the drain-source voltage on the onset of saturation. Index N(P) denotes channel N(P) device.

Solving (A1)-(A4) for v'_{in}, results,

$$k = -\frac{n_N\left(n_N + 1\right)}{n_P^2} \qquad \text{(A5)}$$

$$V_{IDC} = n_N\left(n_N + 1\right)\left(\frac{V_{DD}}{n_P^2} + \frac{V_{SS}}{n_N\left(n_N + 1\right)} + \frac{V_{T0N}}{n_N} - |V_{T0P}|\frac{n_P + 1}{n_P^2} - \frac{F(I_B / I_{S7})}{n_P}\right) \qquad \text{(A6)}$$

According to (A6), the value of V_{IDC} can be adjusted by properly selecting the inversion level of M_7.

Table III. Expressions of the Advanced Compact MOSFET (ACM) model [18].

Symbol	Expressions
I_D	$I_S i_f$ (in saturation)
I_S	$\mu C'_{ox}\, n \frac{\phi_t^2}{2}\frac{W}{L}$
g_{ms}	$\frac{2I_S}{\phi_t}\left(\sqrt{i_f + 1} - 1\right)$
$\frac{V_P - V_S}{\phi_t}$	$F(i_f) = \sqrt{i_f + 1} - 2 + ln\left(\sqrt{i_f + 1} - 1\right)$
V_P	$\frac{V_G - V_{T0}}{n}$
V_{DSSAT}	$\phi_t\left(\sqrt{i_f + 1} + 3\right)$

REFERENCES

[1] K. A. Zaghloul and K. Boahen, "Optic Nerve Signals in a Neuromorphic Chip I: Outer and Inner Retina Models," IEEE Transactions on Biomedical Engineering, vol. 51, no. 4, 2004, pp. 657–666.

[2] K. A. Zaghloul and K. Boahen, "Optic Nerve Signals in a Neuromorphic Chip II: Testing and Results," IEEE Transactions on Biomedical Engineering, vol. 51, no.4, 2004, pp. 667–675.

[3] R. C. Galán et al, "A bio-inspired two-layer mixed-signal flexible programmable chip for early vision," IEEE Transactions on Neural Networks, vol. 14, 2003, pp. 1313–1336.

[4] J. H. Liu, C. Y. Wang, and Y. Y. An, "A Survey of Neuromorphic Vision System: Biological Nervous Systems Realized on Silicon," in Proceedings of International Conference on Industrial Machine and Automation, 2009, IEEE CODE 10749897, pp. 154–157.

[5] L. Chua and L. Yang, "Cellular neural networks: theory," IEEE Transactions on Circuits and Systems, vol. 35, n. 10, 1988, pp.1257-1290.

[6] A. Rodríguez-Vázquez, S. Espejo, R. Dominguez-Castro, J. Huertas, and E. Sanchez-Sinencio, "Current-mode techniques for the implementation of continuous- and discrete-time cellular neural networks," IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, vol. 40, n. 3, 1993, pp.132–146.

[7] S. Espejo, R. Carmona, R. Dominguez-Castro, and A. Rodriguez-Vazquez, "A VLSI-oriented continuous-time CNN model", International Journal of Circuit Theory and Applications, vol. 24, n.3, 1996, pp. 341–356.

[8] J. Varrientos, E. Sanchez-Sinencio and J. Ramirez-Angulo, "A current-mode cellular neural network implementation," IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, vol.40, n.3, 1993, pp. 147–155.

[9] J. Cruz and L. O. Chua, "A CNN chip for connected component detection," IEEE Transactions on Circuits and Systems, vol. 38, n. 7, 1991, pp. 812 –817.

[10] Dominguez-Matas, C. M. et al., "Robust symmetric multiplication for programmable analog VLSI array processing," In Proceedings of 13th IEEE International Conference on Electronics, Circuits and Systems, 2006, IEEE CODE 9718653, pp. 1332–1335.

[11] E. P. Santana, R. C. S. Freire, and A. I. A. Cunha, "A CMOS analog four-quadrant multiplier for CNN synapse," accepted for publication in Proceeding of 8th IEEE International Caribbean Conference on Devices, Circuits and Systems, 2012.

[12] M. B. Machado et al, "CMOS multiplier based on the relationship between drain current and inversion charge," IET Circuits, Devices and Systems, vol. 3, no. 5, 2009, pp. 239–247.

[13] E. P. Santana, R. C. S. Freire, and A. I. A. Cunha, "A compact low-power CMOS a nalog FSR model-based CNN," accepted for publication in Proceedings of 3rd IEEE Latin America Symposium on Circuit and Systems, 2012.

[14] J. Hegt, D. Leenaerts, and R. Wilmans, "A novel compact architecture for a programmable full-range CNN in 0.5 m CMOS technology," in Proceedings of Fifth IEEE International Workshop on Cellular Neural Networks and Their Applications, 1998, IEEE CODE 5973180, pp. 288–293.

[15] Leon O. Chua and Tamás Roska, *Cellular Neural Networks and Visual Computing: Foundations and Applications*, Cambridge, 2002.

[16] T. Matsumoto, L.O. Chua and H. Suzuki, "CNN cloning template: connected component detector," IEEE Transactions on. Circuits and Systems, vol. 37, no. 5, pp. 633-635.

[17] The mentor graphics website. [Online]. Available: http://www.mentor.com/

[18] A. Cunha, M. Schneider, and C. Galup-Montoro, "An MOS transistor model for analog circuit design," IEEE Journal of Solid-State Circuits, vol. 33, no. 10, pp. 1510–1519.