

A 0.8 μ m CMOS Programmable Analog-Array-Processing Vision-Chip with Local Logic and Image-Memory

S. Espejo, R. Carmona, R. Domínguez-Castro and A. Rodríguez-Vázquez

Centro Nacional de Microelectrónica-Universidad de Sevilla
Dept. of Analog Design
Edificio CICA, C/ Tarfia sn, 41012-Sevilla, SPAIN
Phone: +34 5 4239923, FAX: +34 5 4231832, email: espejo@cnm.us.es

ABSTRACT: An operational vision-chip prototype with a wide-range of potential applications in artificial-vision systems is presented. Its functionality includes concurrent image-transduction, programmable image-processing, image-storage, and algorithmic control over a network of 20 x 22 identical cells. The prototype has been designed and manufactured in 0.8 μ m CMOS standard technology, and has a total area of 30mm². Experimental results are reported.

1. Introduction

Conventional image-processing systems use a sensory plane for *parallel phototransduction*, and *serial transmission* of the complete input scene to a separate processing element. A major drawback of this architecture is the necessity to transmit the raw input data, out of which only a small percentage carries relevant information. Recently, a new class of vision systems overcome this drawback with the incorporation, at the phototransduction level, of simple *parallel* analog circuitry employed for image-features extraction, thus reducing the amount of transmitted data.

The resulting new class of front-end *vision-chips* [1] can be described as *arrays of interacting analog cells* with local sensing and processing functions. Their global sensing and processing capabilities are spatially-distributed, and determined by individual-cells' functionality and the strength of their interactions. Vision chips intend to emulate the parallel processing functions of the human retina [2], in which different layers appear to perform different well-defined tasks. One possibility is to employ a dedicated fixed-function chip for some particular task; for instance for edges or borders extraction, for motion detection, etc. Preferably, several functions can be realized by a single electrically-*programmable* chip. Furthermore, by incorporating *distributed image-memories* and a small amount of control circuitry, the resulting front-end chips can perform complex and diverse image-processing applications, through sequential and bifurcated-flow algorithms.

This paper presents a 0.8 μ m CMOS vision-chip prototype which follows these guidelines. The processing function is based on the paradigm of Cellular Neural Networks (CNN) -- a very suitable framework for *systematic* design of image processing chips [3]. The complete programmability of the interconnection strengths, its internal image-memories, and other additional features make the prototype a powerful front-end for the realization of simple and medium-complexity artificial vision tasks. Also, although the fundamental processing function is analog and continuous-time, as in other CNN chips [4], the interface of our prototype is completely digital, making it extremely easy to control with conventional computing systems. To the best of our knowledge, this chip is the first fully operational CNN vision-chip reported in literature which combines the capabilities of image-transduction, programmable image-processing and algorithmic control on a common silicon substrate.

2. System Architecture and Functionality

Architecture: The prototype contains 20 x 22 identical cells arranged in a rectangular grid, each of them with local *transduction, processing, control*, and *storage* capabilities. In addition, global control, interfacing and storage circuitry is placed on the surrounding of the cell array, yielding a total silicon area of 30mm², the maximum allowed by the foundry in MPW runs.

Transduction: Image acquisition relies on photogenerated currents at floating-base vertical BJTs, available on standard CMOS technologies. An automatic adaptive scheme is used to ensure appropriate contrast levels by shifting the observed scene to obtain a zero-mean distribution of pixel values. Optional external circuitry can be employed to adjust the mean of the distribution when needed, for instance for highly regular images with dominant background.

Processing: Image-processing is based on the Cellular Neural Network (CNN) computing paradigm [3]. CNNs can be described as artificial neural networks (ANNs) with neurons (or cells) spatially distributed on a *regular* grid (over the chip surface) and *locally* interconnected. Regularity serves to the purpose of scene discretization into a pixel matrix, while local connectivity, commonly restricted to adjacent neighbors, results in feasible local-routing requirements. An additional important property of image-processing oriented CNNs is *translational invariance*, meaning that the interconnection pattern among neighbor-cells does not change throughout the array, and can therefore be described in terms of *templates*: 3 x 3 matrices whose entries represent the strength of local interconnections. The resulting array of identical neurons can thus be controlled by a reduced number of synapse-programming analog lines, rendering feasible global-routing requirements.

CNNs input data consists of two images. One of the images is used as initial conditions of state variables and dynamically convoluted with the *feedback* template, while the other is assumed to be time-invariant and statically convoluted with the *control* template. These two templates, together with an off-set term and some spatial-boundary conditions complete the processing coefficients of the network, all of them programmable in our prototype. As in other ANNs, processing is analog, but input and output data is binary. This results in simple realizations of internal image memories and control circuitry.

Control and Storage: The basic functions of image acquisition and processing are complemented in this prototype with storage, algorithmic control, and programmable boolean operations among images. Four images can be stored on-chip. Individual pixels are physically located within corresponding cells, facilitating parallel data-transference tasks. Image-memories can be loaded from the image-sensor, the output of the neural network, or the output of the programmable boolean operator. Any stored image can be used as any of the two input images of the network, or as input to the boolean operator. Any image can also be loaded or downloaded, on a row by row basis, through an external I/O bidirectional bus.

Global control and interface: Eight complete sets of CNN coefficients can be stored on chip. Although their internal control is analog, they are discretized and stored in digital form, with a resolution of 7bits+sign, more than enough for the expected accuracy of the analog processing circuitry [5]. These digital values are internally transformed into analog synapse-control signals by adaptive loops comprising linear D/A converters and a synapse identical to those used within the cells. This strategy results in simple and robust external control, virtually independent of the detailed relationship between the control-signal and the actual weight programmed in the synapses [5]. The internal storage of images and processing coefficients, together with the possibility of using them in any order and any number of times, yields a highly flexible system usable for relatively complex and generic image-processing tasks including sequential and bifurcated-flow algorithms. Also, although the fundamental processing function is analog, the digital nature of the interface makes the prototype extremely easy to control with conventional computing systems.

3. Circuit Implementation

The major trend in the design has been the maximization of the number of cells in the array under the area limitation imposed by the foundry (30mm^2), while maintaining a reasonable degree of accuracy in the analog operations. The involved area-accuracy trade-off [6] has been addressed through intensive structural and parametric optimization [5]. The prototype contains 20 x 22 cells, each with an area of $180 \times 180 \mu\text{m}^2$; it means a cell density of 31 cells/ mm^2 .

Image Acquisition Circuitry: This is identical to that previously described by the authors in [7] for fixed-function CNN CMOS chips. Pixel sensors consist of two vertical BJTs arranged in Darlington configuration, as shown in Fig. 1a. The photogenerated current at the base-collector junction (n-well/p-substrate) of transistor Q_1 is amplified by a factor $(\beta_F+1)^2$ by Q_1 and Q_2 , yielding output current levels of about $0.8\mu\text{A}$ under an environmental laboratory lighting of $0.9\text{W}/\text{m}^2$. The acquired gray-scale image is converted to binary by comparison to the spatial average of the image, thus ensuring proper contrast adjustment over a wide range of illumination conditions. The averaging and comparison circuitry, illustrated in Fig. 1b, is included in every cell and globally interconnected through a common node SUM. A CMOS inverter transforms the shifted output current to digital levels, which can then be stored at internal memories. The area of the imaging circuitry (sensor+regulation) amounts up to 7% of the cell area.

CNN Processing Circuitry: Its fundamental building blocks are programmable synapses and non-linear integrators (integrators with saturation). Fig. 2a shows the structure selected for the former, consisting of a linear multiplier core with transistors operating in ohmic region [8], and two source-follower

buffers. Its selection is based on exhaustive analysis of the area-accuracy trade-off in alternative CMOS synapses [5]. Note that synapse inputs are voltages, which eases the intra-chip distribution of template coefficients, and the intra-cell distribution of the state-variable to all the cell synapses. On the other hand, the synapse output is a current, thus facilitating the summation of different contributions at integrators' input nodes. The nonlinear behavior of the output with respect to the differential weight signal $V_{wp}-V_{wn}$, mainly due to the resistively-loaded source-followers is not important, since it is taken into account by the weight-control adaptive circuitry. On the other hand, the behavior of the structure with respect to the state-variable signal $V_{xp}-V_{xn}$ is, as required, highly linear.

The differential nature of the synapse signals, and the necessity to reduce common-mode parasitics led us to the fully-differential integrator architecture of Fig. 2b, realized with two current-conveyors and two grounded capacitors. The latter are realized through the gate capacitance of the MOS transistors tied to the state-variable terminals of the neuron synapses, which are indeed the natural load for the current conveyors in an integrator configuration. Integrator saturation is achieved by a nonlinear resistor with a sharp voltage-limiting characteristic, realized with two diode-connected transistors as shown in Fig. 2c, and connected between the differential-rail signals.

Local Logic and Control Circuitry: The realization of the programmable boolean operator and the control circuitry is based on switches and conventional digital circuitry. The 4bit memory, based on charge storage, employs metal-1 shields over sensitive areas to avoid the adverse effect of light on reverse diode-currents, which could result in a significant reduction of storage-time. About 30% of the cell area is dedicated to these digital capabilities.

Interfacing Circuitry: About 50% of the prototype area is dedicated to miscellaneous circuitry placed at the periphery of the cell array. This includes the weight-control stages used to generate analog programming signals from their digitally stored values, boundary cells employed to establish spatial boundary conditions, biasing circuitry, and bonding pads. Also, some conventional digital blocks are needed for the control of the loading and downloading processes of images and CNN coefficients.

4. Experimental Results

The prototype, whose photograph is shown in Fig. 3, has been designed and manufactured in a standard, digitally oriented, two-metals, one-poly, n-well, 0.8 μ m CMOS technology available through the EUROCHIP consortium. Two additional smaller chips with analog and digital parts were fabricated for testing and characterization purposes. The total area of the prototype is 30mm², the time constant of the analog network is 0.25 μ s, and the operation speed of the digital circuitry is 10MHz.

Global Performance: The prototype has been globally tested and its functionality verified by the authors and by an independent research group [9]. It has been successfully used for such diverse functions as low-pass image filtering, corners and borders extraction, hole filling, motion detection, and many other CNN applications reported in literature. Task sequencing and algorithmic control has been also verified, in particular for texture-detection applications.

While the functionality of the prototype has been completely verified, the accuracy of the analog circuitry (around 6 bits) is slightly under the objective (7 bits). The source of this degradation has already been identified and could be easily avoided with minor changes in the design.

Basic Blocks Performance: most of the basic analog blocks have been exhaustively tested and characterized. Fig. 4a shows the response of the programmable synapse for different weight values. Integral linearity is better than 0.4% and total harmonic distortion of 0.2% within the required state-variable signal-range. Statistical characterization, based on 10 samples located on different chip-units, yield a standard deviation of the output current offset of 0.1% relative to the maximum output current; and for weight values it is 0.8% of its full-range. Fig. 4b illustrates the large signal V-I characteristic observed at the low-impedance input node of current conveyors. Statistical characterization over 20 samples shows a standard deviation for the input offset-voltage of 3mV. Finally, Fig. 4c contains the I-V characteristic of the voltage-limiter employed for integrators saturation. Saturated voltage levels exhibit a standard deviation of 2% of the full signal range.

5. Summary

We have briefly described the design of a fully programmable vision-chip with a wide range of potential applications. The processing function is based on the Cellular Neural Network paradigm, while the phototransduction relies on vertical BJTs available on standard CMOS technologies and includes an automatic contrast-centering circuitry. Additional features like internal image memories, algorithmic control, and programmable logic operators provide a high versatility for simple and medium complexity artificial-vision applications. Although the internal operation is fundamentally analog, the interface of the prototype is completely digital, making it directly controllable by conventional computing devices. The prototype has been manufactured in a standard 0.8 μ m CMOS technology and successfully tested.

6. References

- [1] C.Koch and H.Li (editors): *Vision Chips: Implementing Vision Algorithms using Analog VLSI Circuits*. New York, IEEE Press 1994.
- [2] M.M. Gupta and G.K. Knopf: "Neuro-Vision Systems: A Tutorial". *Neuro-Vision Systems: Principles and Applications*, New York, IEEE Press 1994.
- [3] L.O. Chua and L. Yang.: "Cellular Neural Networks: Theory and Applications". *IEEE Trans. Circuits and Systems*, Vol. 35, pp. 1257-1290, October 1988.
- [4] P. Kinget and M. Steyaert: "An Analog Parallel array Processor for Real-Time Sensor Signal Processing", *1996 Int. Solid State Circuits Conference*, paper 6.1
- [5] S. Espejo: *VLSI Design and Modeling of Cellular Neural Networks*. Ph.D. dissertation, University of Seville, 1994.
- [6] M.J.M Pelgrom, A.C.J. Duijnmaier and A.P.G. Welbers: "Matching Properties of MOS Transistors". *IEEE J. Solid-State Circuits*, Vol.24, pp 1433-1440, October 1989.
- [7] 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 Journal of Solid-State Circuits*, Vol. 29, pp. 895-905, August 1994.
- [8] B. Song: "CMOS RF Circuits for Data Communication Applications". *IEEE J. Solid-State Circuits*, Vol.21, pp 310-317, April 1986.
- [9] P. Foldesy, A. Zarandy, P. Szolgyay and T. Roska: *Measurement Results of the 20 x 22 CNUM Chip*, Analogic and Neural Computing Laboratory, Hungarian Academy of Sciences, February 1996.

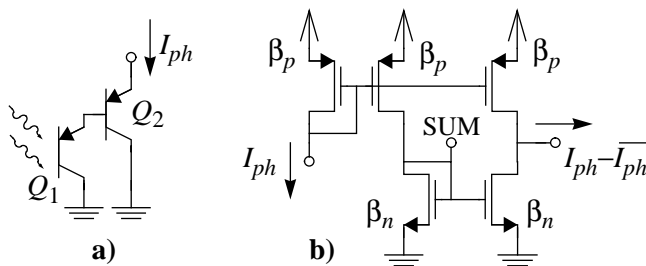


Figure 1: a) Optical transducer implementation, b) Automatic threshold adjustment circuitry.

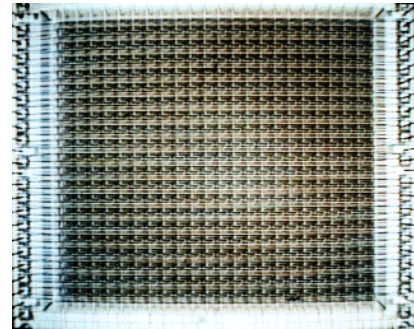


Figure 3: Prototype's photograph.

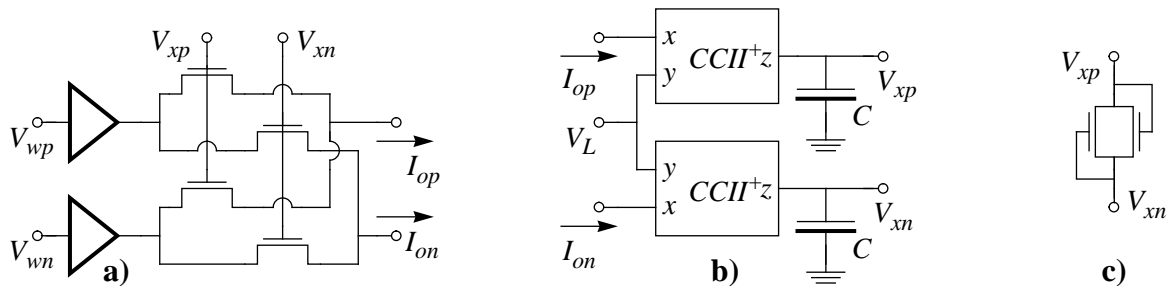


Figure 2: Main CNN processing blocks, a) programmable synapse, b) integrator, c) voltage limiter.

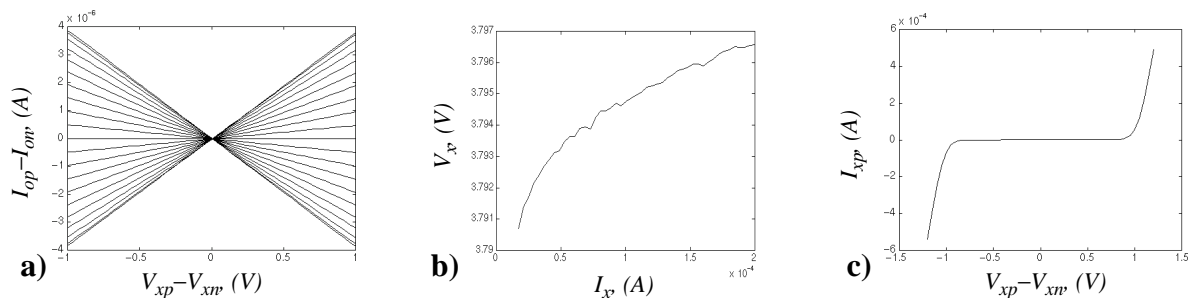


Figure 4: Measured response of a) the programmable synapse, b) the low-impedance input node of CCII, c) the voltage-limiter element.