



A 0.5 μ m CMOS CNN Analog Random Access Memory Chip for Massive Image Processing

R. Carmona¹, S. Espejo², R. Domínguez-Castro², A. Rodríguez-Vázquez²,
T. Roska³, T. Kozek¹, L. O. Chua¹

¹ Electronics Research Laboratory, University of California, Berkeley
258M Cory Hall, Berkeley, CA 94720, USA.

Phone: 1-510-6425311 Fax: 1-510-6438869 E-mail: rcarmona@fred.eecs.berkeley.edu

²Instituto de Microelectrónica de Sevilla-CNMI-Universidad de Sevilla.
Edificio CICA, C/Tarfia s/n, 41012-Sevilla, SPAIN.

³MTA-SZTAKI, Analogic & Neural Computing Laboratory. H-1111 Budapest, Hungary.

ABSTRACT: *An analog RAM has been designed to act as a cache memory for a CNN Universal Machine. Hence, all the non-standard chips are available for the CNN Chipset architecture. Time-multiplexed analog routines in the CNN processor require fast and efficient short-time signal storage in an analog buffer. This can be achieved by an extended sample and hold scheme able to address every sample to specific memory locations. Several arrays of capacitors are multiplexed sharing controlling circuitry and I/O buses. The design has the following key parameters: 637 analog memory cells/mm² with 0.4% accuracy, 100ns access time and 170ms storage time (within 1% error).*

1. Introduction

An analog random access memory (ARAM) chip has been designed and integrated in a 0.5 μ m n-well, single-poly, triple-metal CMOS technology. This circuit is the only missing important part of the chipset of the CNN universal machine [1] [2]. It works as a memory cache for the analog visual microprocessor. In order to take advantage of the high speed of operation of the CNN analog processing unit [3], a compatible, fast and reliable analog short-term storage circuit is needed. It will be interfaced either to the analog processor and/or the image acquisition devices. It is based in an extended sample and hold structure [4] in which signals are transmitted in the form of a voltage but stored as charges in an array of capacitors. An adequate and externally controllable timing scheme will be used to reduce switching feedthrough error improving accuracy. Random access to the memory array is available and can be serial or parallel via a fixed size multiplexed I/O bus.

2. The ARAM in the CNN Computer

Fig. 1 shows the role of the ARAM in the CNN chipset.

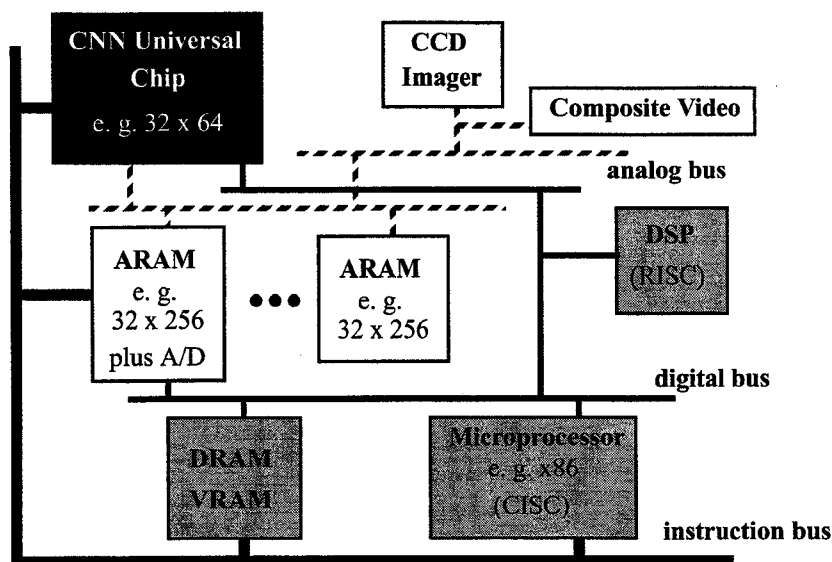


Figure 1: ARAM in the CNN chipset architecture.

It is connected to a CNN Universal Chip and to an imager or video source. It might also contain A/D converters (not discussed here) to be interfaced to some conventional digital memory devices for mid-term storage of information. The interface to a composite video signal source can be easily implemented, as shown in Fig. 2.

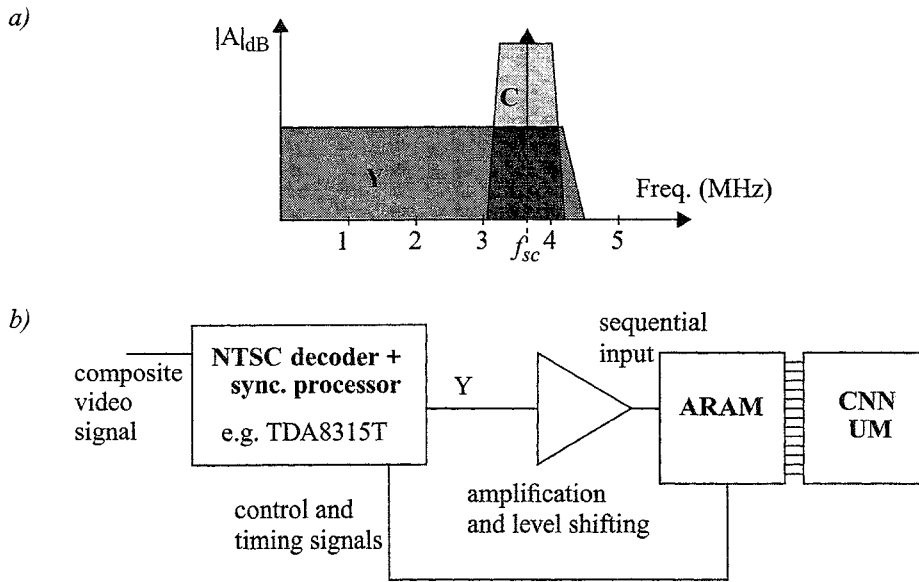


Figure 2: a) Envelope spectrum of a NTSC signal, showing the luminance (Y) and the chrominance (C) centered around the color subcarrier ($f_{sc} = 3.86\text{MHz}$), and b) video signal input to the CNN chipset.

This analog RAM chip consists of an array of 32×256 capacitors arranged in a collection of subsets. Each subset (Fig. 3) is an extended sample and hold structure that includes a group of capacitors, switches for the access and feedthrough cancellation and an OPAMP. Several trade-offs has been resolved in the design of the chip concerning speed, accuracy, and storage time. As a result 100ns access time, 7-8 bits of resolution (0.4% relative error) when comparing the recovered signal with the input, and 170-180ms storage time within 1% error, has been achieved.

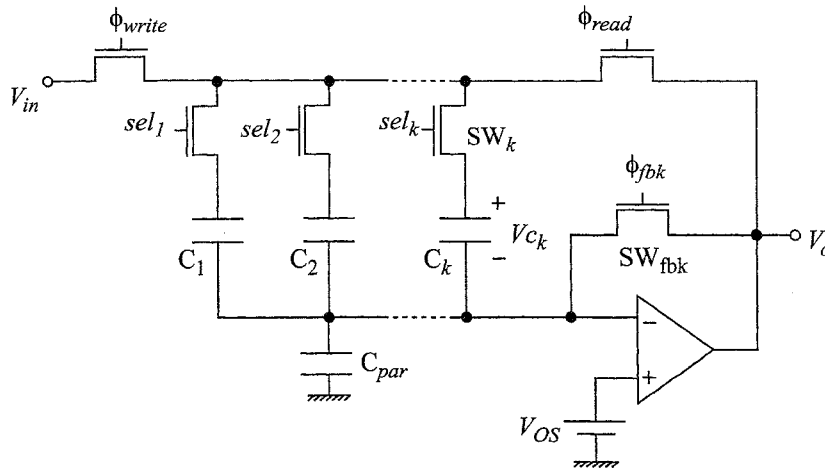


Figure 3: Capacitor subset with S&H circuit including parasitic capacitance and OPAMP offset.

Fig. 4 depicts the system architecture. Address selection and memory allocation is performed by externally controlled decoders generating the proper row and column selection signals. Updating and downloading the memory contents is realized either in serial mode or via a 16-lines analog bus. This I/O bus is multiplexed to the I/O lines of the 32 different rows of the array with the help of the digital address code. Furthermore, separated power supplies and ground connection for the analog and the CMOS logic circuitry has been used to minimize coupling between the analog stored voltages and the noisy digital signals through the substrate. Careful and conservative layout of the capacitor groups and substrate diffusion contacts and guard rings contributes, as well, to the isolation of the analog signals from the digital control lines. Test pads has been provided for calibration of the output buffer in order to enhance measurements performed onto the memory array output. After tuning the proper bias voltages the operation of the chip is completely controlled by digital signals.

3. Sample and Hold Circuit

3.1. Offset-free sample and hold circuit

The employed S&H circuit is an offset free scheme [6] that has been used in previous works [4] for the implementation of a scanning delay-line (Fig. 3). As compared to other circuits used in similar applications [5] it has the ability of non-destructive reading of the memory contents. Assume first that the opamp has infinite gain and that switch feedthrough is negligible. Then the voltage set at the input node (V_{in}), stored as a charge at the capacitor C_k , is reproduced at the output (V_o) and shows no trace of the opamp offset,

$$\left. \begin{aligned} V_{C_k}(n) &= V_{in}(n) - V_{OS} \\ V_o(n+1) &= V_{C_k}(n) + V_{OS} \end{aligned} \right\} V_o(n+1) = V_{in}(n) \quad (1)$$

Consider now the actual case of a finite opamp gain. The output voltage is then given by,

$$V_o \approx \frac{1}{1 + \frac{1}{A_o} \left(1 + \frac{C_{par}}{C_k}\right)} V_{in} + \frac{1}{A_o} \left(1 + \frac{C_{par}}{C_k}\right) V_{OS} \quad (2)$$

which shows the influence of the parasitic capacitance attenuated by the nominal capacitance and the opamp DC gain. Thus, providing a large enough gain for the OPAMP, the error introduced by parasitics can be highly reduced.

In the selected process (MOSIS offered HP 0.5 μ m CMOS n-well, single-poly, triple-metal) capacitors can be implemented by polysilicon over *n-type* diffusion (Fig. 5). These are embedded in a weakly-doped n-well in order to reduce the importance of the parasitic diode (*n-diff* to *p-substrate*).

3.2. Feedthrough error reduction strategy

Signals sel_k and ϕ_{fbk} regulate the access to the capacitor plates. Either opening switch SW_k , that controls the access to the upper capacitor plate, or the virtual ground feedback (via SW_{fbk}) we can isolate one plate of the storage capacitor preventing the other from strongly varying the amount of charge present on it. Only the coupling to the substrate via the parasitics can modify its stored charge. Hence, a slight difference between the falling edges of sel_k and ϕ_{fbk} results in a small and acceptable amount of voltage degradation due to feedthrough of the clock signal. Therefore, access to both plates of the capacitors is required to accomplish this S&H scheme. If the feedback switch (SW_{fbk}) is chosen to be disconnected first, the charge injected to the virtual ground node is shared by the whole capacitor group, due to the fact that they are all embedded into the same n-well. Then, the falling edge of signal ϕ_k will not considerably distort the stored voltage. It is important to use minimum size devices to implement SW_{fbk} .

3.3. Speed vs. accuracy and storage time: capacitor sizing

For the selection of the capacitor size some trade-offs must be resolved. Operating with low frequency signals, the switching error is the major source of inaccuracy. Although there is not a simple model for charge injection phenomena, error due to feedthrough of the switch control into the stored voltage via the parasitic capacitance of the switch can be expressed as a function of the nominal capacitor [6]:

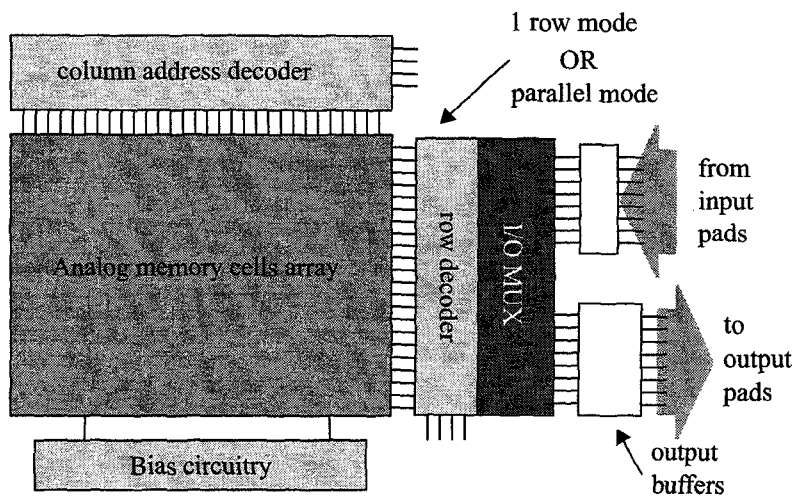


Figure 4: System architecture of the ARAM chip.

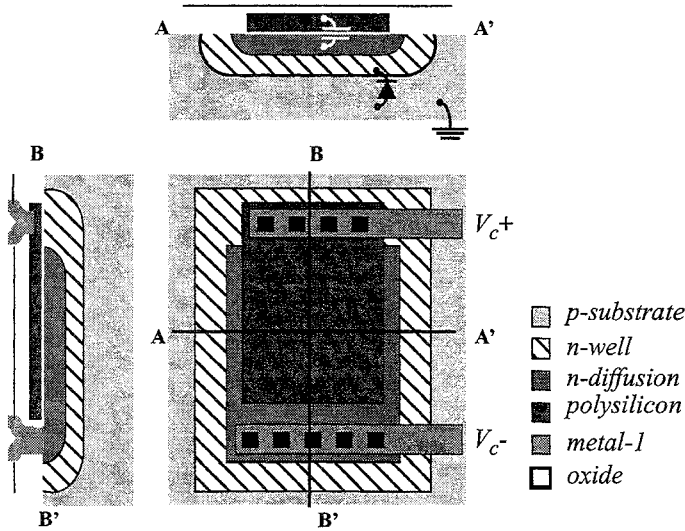


Figure 5: Polysilicon over n-diffusion capacitor.

$$\varepsilon_f(C) = \frac{C_{gs}}{C_{gs} + C} (V_{SS} - V_{mem} - V_{Tn}) \quad (3)$$

Moreover, for high frequency inputs the major source of error is due to the time constant of the switch-capacitor circuit. Hence, it can be written:

$$\varepsilon_\tau(C) = (V_{max} - V_{min}) e^{-\frac{t_a}{RC}} \quad (4)$$

where t_a is the desired access time (100ns). A capacitor of about 200fF allows these quantities to be approximately of the same magnitude. In addition, the storage time increases with the size of the storage capacitor, though not linearly because of an independent self discharge rate of the capacitive structure in this technology:

$$\Delta t = \frac{|\Delta V|C}{I_{ds-leak} + I_{junction-leak} + r_{self-discharge}C} \quad (5)$$

which implies a maximum storage time dependent on the desired accuracy level:

$$\Delta t_{max} = \frac{|\Delta V|}{r_{self-discharge}} \quad (6)$$

For a 1% accuracy (12mV), equation (6) yields 240ms maximum storage time, i.e. capacitances bigger than 3pF do not provide significantly longer storage time. For our particular selection (200fF) it results in 170ms. Several HSPICE simulations has been performed to confirm these predictions using vendor level-13 MOS models. Fig. 6 illustrates performance of a 64 analog registers set sampling a 156.25kHz sine-wave at 10MS/s (100ns access time) resulting in a 3.1mV standard deviation (0.26% of the 1200mV range) for the output-to-input error.

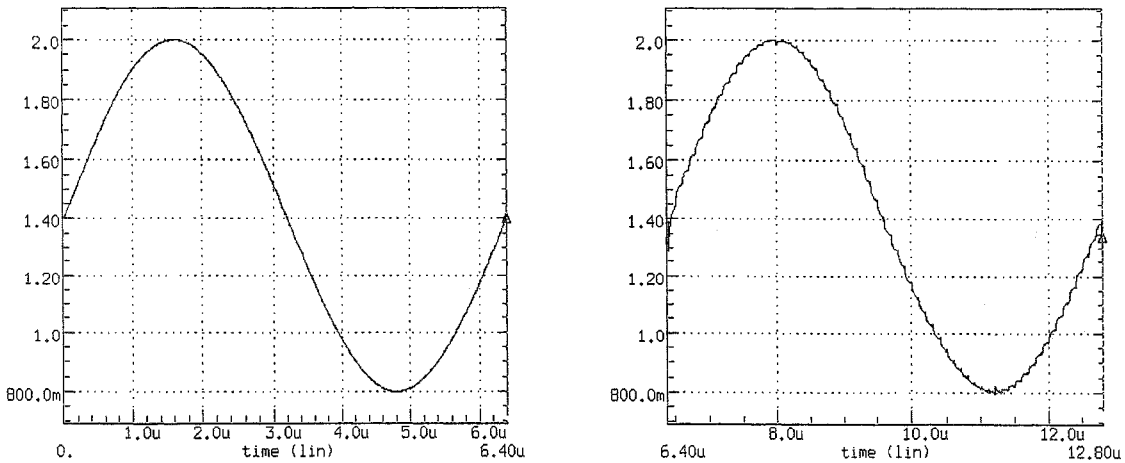


Figure 6: Input voltage waveform (156.25kHz sine-wave) and recovered signal from a HSPICE simulation.

4. Operation Examples

Simulation of the designed circuit with HSPICE give the results depicted in Fig. 7. These figures represent the input and output images of a row of the array, and the relative error computed from comparing the recovered signal with the initially stored one. Images are sampled and output at a rate of 10MS/s.

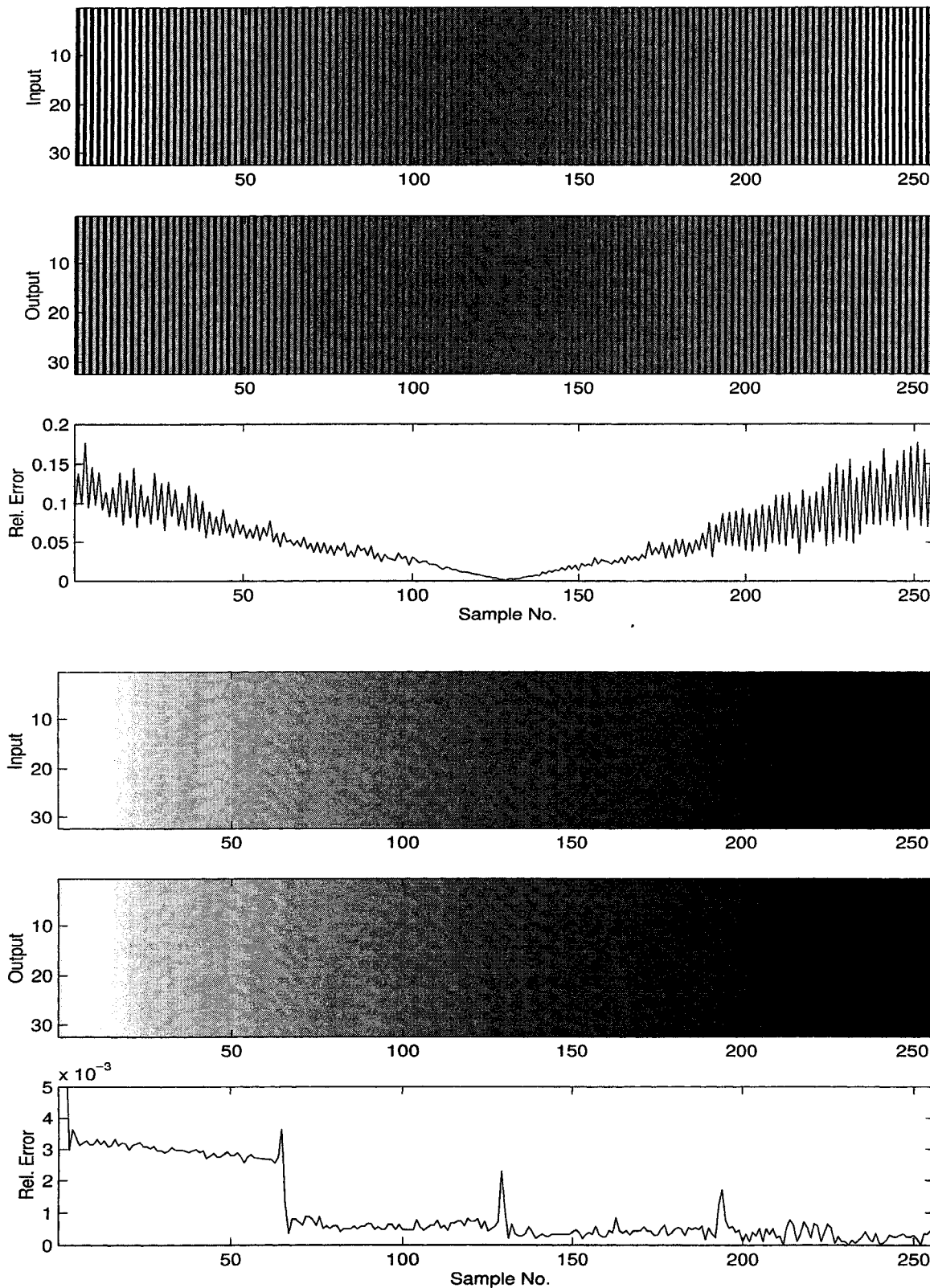


Figure 7: Test images processing.

5. ARAM prototype chip data

Power Supply	3.3V
Power dissipation	72.86mW @ 3.3V
Number of pixels	8192 (32 x 256)
Cell-array area	3733.2 μ m x 3446.4 μ m= 12.86mm ²
Cell-density	637 analog memory cells/mm ²
System area (w/o pads)	4130.1 μ m x 3886.2 μ m= 16.05mm ²
Die Area	4774 μ m x 4474 μ m= 21.36mm ²
Sampling Rate	10 Msamples per second
Electrical I/O rate	10MHz
Package	PGA-84M
Input range	[0.8, 2.0] V
Output swing	[0.8, 2.0] V
Storage time (1% accuracy)	150-180ms
Output to Input accuracy	7-8bits (0.4-0.7%)

Table 1: ARAM prototype chip data.

6. Acknowledgments

This work is supported by JSEP Grant No. FDF49620-97-1-0220-03/98 and by ONR Grant No. N00014-98-1-0052.

7. References

- [1] T. Roska and L. O. Chua: "The CNN Universal Machine: An Analogic Array Computer". IEEE Transactions on circuits and Systems-II: Analog and Digital Signal Processing, vol. 40, No. 3, pp. 163-173, March 1993.
- [2] T. Roska: "CNN Chip Set Architectures and the Visual Mouse". Proceedings of the 4th IEEE Int. Workshop on Cellular Neural Networks and their Applications, pp 487-492. Sevilla, Spain, June 1996.
- [3] R. Dominguez-Castro, S. Espejo, A. Rodríguez-Vázquez, R. Carmona, P. Foldesy, A. Zarándy, P. Szolgay, T. Sziranyi and T. Roska: "A 0.8 μ m CMOS Programmable Mixed-Signal Focal-Plane Array Processor with On-Chip Binary Imaging and Instructions Storage". IEEE Journal of Solid State Circuits, vol. 32, No. 7, pp. 1013-1026, July 1997.
- [4] K. Matsui, T. Matsura, S. Fukasawa, Y. Izawa, Y. Toba, N. Miyake and K. Nagasawa, "CMOS Video Filter Using Switched Capacitors 14-MHz Circuits". IEEE Journal of Solid-State Circuits, vol. 20, No. 6, pp. 1096-1102, December 1985.
- [5] K. A. Nishimura and P. R. Gray, "A Monolithic Analog Video Comb Filter in 1.2 μ m CMOS". IEEE Journal of Solid-State Circuits, vol. 28, No. 12, pp. 1331-1339, December 1993.
- [6] R. Gregorian and G. C. Temes: Analog MOS integrated circuits for signal processing. John Wiley & Sons, New York, 1994.