

Review of CMOS Implementations of the CNN Universal Machine-Type Visual Microprocessors

T. Roska* and A. Rodríguez-Vázquez**

* Computer & Automation Institute – Hungarian Academy of Sciences
Kende-u. 13-17, Budapest, H-1111, HUNGARY

Phone: +36 1 209 5263, Fax: +36 1 209 5264, E-mail: roska@lutra.sztaki.hu

** Instituto de Microelectrónica de Sevilla – CNM-CSIC

Edificio CICA-CNM, C/Tarfia s/n, 41012- Sevilla, SPAIN

Phone: +34 95 5056679, Fax: +34 95 4231832, E-mail: angel@imse.cnm.es

ABSTRACT¹

While in most application areas digital processors can solve problems initially, in some fields their capabilities are very limited. A typical example is vision. Simple animals outperform super-computers in the realization of basic vision tasks. In order to overcome the limitations of these conventional systems, a fundamentally different array architecture is needed. This architecture is based on the new paradigm of *analogic cellular (CNN) computing* whose most advanced implementation is the so-called *CNN universal machine (CNN-UM)*. Its main components are: a) parallel architecture consisting of an array of locally-connected analog processors; b) a means of storing, locally, pixel-by-pixel, the intermediate computation results, and 3) stored on-chip programmability. When implemented as a mixed-signal VLSI chip, the CNN-UM is capable of image processing at rates of trillions of operations per second with very small size and low power consumption. On the other hand, when integrating the adaptive multi-sensor array in the CNN-UM, the resulting sensor+computer array offers unprecedented capabilities. This paper reviews the latest results on CNN-UM chips and systems, and outlines the envisaged roadmap for these computers.

1. Introduction

Conventional vision machines use a CCD camera for *parallel* acquisition of the input image, and *serial* transmission of a *digitized* version of the input data to a separate computer. This results in huge data rates which conventional computers are not capable of analyzing in real-time. For instance, a 3-color@512×512 camera delivers about $F \times 10^6$ bytes/second, where F is the frame rate. Conventional computers and DSPs are able to manage such a huge rate for auto-focus, image stabilization, control of the luminance/chrominance, etc. However, executing the spatial-temporal operations of image processing in *real-time*

requires much more sophisticated digital processors. Consequently, conventional vision machines with real-time capabilities are bulky, expensive and extremely power-hungry. This is in contrast to living beings, where even very tiny and power-efficient brains can analyze complex time-varying scenes in real-time. A prototype of this way of processing is manifested at the very front-end of the human vision system – the *retina* [9].

This contrast between the performance of artificial and “natural” vision systems is, among other things, due to the inherent *parallelism* of the processing realized by the latter. Such parallelism is observed already in the retina [8]. It contains *photoreceptor cells* of two different types – called *cones* (about 6 million in the whole retina) and *rods* (about 120 million) – which perform a logarithmic three-color imaging for around ten decades of light intensity range. It also contains *processing cells* – called horizontal, bipolar, amacrine and ganglion cells – to perform non-linear spatial-temporal processing operations on the incoming flow of images through a sequence of layers. Among many other tasks, such processing serves to extract important features from the raw sensory data and, thus, to reduce the amount of information transmitted for subsequent processing [3][9].

Inspired by the efficiency of natural vision systems, universities and companies have focused their efforts on the development of new generations of devices capable of overcoming the drawbacks of traditional ones through the incorporation of distributed parallel processing, and by making this processing act concurrently with the acquisition of the signal. One possible strategy to achieve that is through flip-chip bonding of separate sensing and processing devices; another possibility is to incorporate the sensory and the processing circuitry on the same semiconductor substrate. “Silicon retinas”, “smart-pixel chips” and “focal-plane array-processors” are members of this latter class of vision chips [5][4][6]. Their development is expected to have a significant impact in quite diverse scenarios. However, industrial applications demand chips capable of flexible operation, with programmable features and standard interfacing to conventional equipment. A powerful methodological framework for a systematic development of these types of chips is using the paradigm of

1. This work has been supported by the EU (IST-1999-19007) and the spanish CICYT (TIC99-0826).

analogic cellular (CNN) computing [1] and the Cellular Universal Machine (CNN-UM) processing architecture [7]. This paper reviews recent advances on system-level and chip-level results related to CNN-UMs, and outlines the envisaged roadmap for these computers.

2. New Directions in System Implementations

Right after the first digital microprocessor was made, Intel Corporation started to sell its associated development system, a tool to educate engineers how to use and program this new device. Likewise, a visual microprocessor development system has been devised to help software engineers and product designers to learn this new device and start developing new products [10].

The new version of our visual microprocessor development system is called ALADDIN: Analogic Application Development system for Dynamic Image processing and Navigation.

The main parts of ALADDIN system are shown at Fig. 1. We consider a PC based development system with cameras, video sources and multimedia accessories. Using this system, programs for the CNN-UM can be developed and tested in a dynamical visual environment.

The next issue is offering self contained Analogic Cellular Engine Boards (ACE Boards) with stored programmability. This means that walkman size units will be available integrated with or interfaced to sensors. Once a program is developed and tested at the ALADDIN site, we can download the program and use it immediately.

When a new version of an analogic cellular visual microprocessor is developed, only a small part of the ALADDIN system will be changed, the platform hosting the chip and a small part of the interface software. All the rest is the same, hence, programming efforts and reuse of hardware and software components is maintained.

A key element of the know-how is contained in the Analogic CNN Software Library. Templates (instructions), subroutines, and programs for well defined tasks are stored and distributed. Like during 60's when the first algorithms for digital microprocessors were developed, we are witnessing a similar process these years. Soon we will edit a new version of our Library, we will call, "Recipes in Alpha", containing hundreds of software modules tested on simulators and visual microprocessors as well.

3. Chip Implementations

During the last few years several CNN chips have been designed. Particularly, those having a size larger than 10×10 and whose operation have been actually demonstrated through experimental evidence are found in [11]-[16]. The attached table presents a summary of some features associated to these chips. Speed is expressed in terms of analog operations per second. The equivalent digital multiply/add operations per second can be calculated in such a way that 10 time step is supposed in a time constant. This is a default needed when the A template is full and analog input or output values are present. This means $10 \times 20 = 200$

equivalent multiply/add operations per time constant, so that calculating with 4096 cell processors and about 280ns time constant [16], the equivalent speed is about 3 TeraOPS.

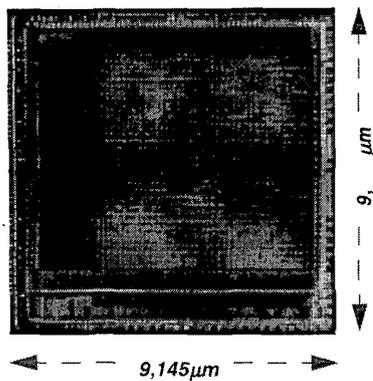
The data in this table reveals a trade-off between speed and accuracy – common to any analog integrated circuit. Out from these chips, those reported in [14] [16] have embedded distributed optical sensors; i.e. they are true focal plane array processors. On the other hand, only the latter is capable to operate with gray scale inputs and producing gray outputs, while at the same time having all functional features of CNN-UMs.

Relevant data pertaining to the chip in [16] are displayed in Fig. 2. Specially relevant are the low power consumption per unit cell and the large operation speed. This chip has also served as a vehicle to demonstrate the concept of true VLSI analog chips with robust, controlled and predictable response. From here the challenges are basically to increase the size and to improve the I/O [18]. Thus the a major next step will be the design of QCIF-resolution chip with embedded optical sensors in a $0.35\mu\text{m}$ or $0.18\mu\text{m}$ technology – a target that is scheduled to be reached during 2,001.

The integration of multiple sensors per pixel within the array computer probably defines the dominant medium- and long-term scenario for CNN-UM based systems [17]. The multiple sensors should be adaptive and capture different modalities, spectra, sensitivity and dynamics. Their control parameters should be set by underlying programmed calculations made by a CNN-UM. Hence, the multi-sensor image acquisition depends, pixel by pixel, on the actual changing scene to be analyzed.

4. References

- [1] L.O. Chua and T. Roska, "The CNN Paradigm", *IEEE Trans. Circuits & Systems-I*, Vol. 40, pp. 147-156, March 1993.
- [2] R. Domínguez-Castro, et al., "A $0.8\mu\text{m}$ CMOS 2-D Programmable Mixed-Signal Focal-Plane Array Processor with On-Chip Binary Imaging and Instructions Storage". *IEEE J. Solid-State Circuits*, Vo. 32, pp. 1013-1026, July 1997.
- [3] M.M. Gupta, G.K. Knopf (Eds.), *Neuro-Vision Systems, Principles and Applications*, IEEE Press, 1994. ISBN: 0-7803-1042-X
- [4] C. Koch, H. Li (Eds.), *Vision Chips, Implementing Vision Algorithms with Analog VLSI Circuits*, IEEE Press, 1995. ISBN: 0-8186-6492-4
- [5] A. Rodríguez-Vázquez, et al.: "Current-Mode Techniques for the Implementation of Continuous-Time and Discrete-Time Cellular Neural Networks", *IEEE Trans. Circuits and Systems II: Analog and Digital Signal Processing*, Vol. 40, pp. 132-146, March 1993.
- [6] B.J. Sheu, J. Choi, *Neural Information Processing and VLSI*, Kluwer Academic Publishers, 1995. ISBN: 0-7923-9547-6
- [7] T. Roska and L.O. Chua, "The CNN Universal Machine: An Analogic Array Computer", *IEEE Trans. Circuits & Systems-I*, Vol. 40, pp. 163-173, March 1993.
- [8] F. Werblin, T. Roska and L.O. Chua, "The Analogic Cellular Neural Network as a Bionic Eye", *Int J. of Circuit Theory and Applications*, Vol. 23, pp. 541-549, 1995.



Transistors	~1,000,000
Cell Density	~82 cells/mm ²
Accuracy	7.6 bits
I/O Digital Rate	10 MHz
I/O Analog Rate	1 MHz
Power Supply	1W(worst case)@3.3V
Templates Memorized	32
Coefficients Memorized	30 (10 A, 10 B, 10 Misc.)
Instructions Memorized	64 (3-byte-words)

Figure 2. Microphotograph and relevant numbers for the CNNUC3

	Tech.	Design Style	Size #Proc.	Size mm ²	Density cells/mm ²	Speed XPS ^a	XPS/cell	XPS/mm ²	XPS/mW	Functions
[11]	CMOS 1.0μm	Mixed-Signal	32 x 32	70	31	0.30T	0.30G	9.3G	---	Stored-Programmable 6-7/bit Analog Resolution Binary Outputs
[13]	CMOS 0.7μm	Analog	20 x 20	25	17	12.5G	31M	0.52G	82M	No Diagonal Interactions External-Programming 6-7/bit Analog Resolution Binary Outputs
[14]	CMOS 0.8μm	Mixed-Signal	20 x 22	30	28	0.13T	0.30G	8.25G	0.12G	Stored-Programmable 6-7/bit Analog Resolution Embedded Optical Sensors Binary Outputs
[12]	CMOS 0.5μm	Basically Digital	48 x 48	11.4	295	7.65T	3.76G	1.11T	25G	Stored-Programmable 2/bit Analog Resolution Binary Inputs and Outputs
[15]	CMOS 0.8μm	Analog	14 x 14	26	16	0.37T	1.89G	31G	1.24G	Stored-Programmable 4/bit Analog Resolution Analog Inputs and Outputs
[16]	CMOS 0.5μm	Mixed	64 x 64	87	81	0.40T	98M	7.93G	0.33G	Stored-Programmable 7-8/bit Analog Resolution Analog Inputs and Outputs Embedded Optical Sensors Embedded Ana. and Dig. Data RAM

a. XPS: Analog Operations Per Second, is an equivalent measurement indicating the number of analog arithmetic operations like addition, subtraction, multiplication and division. IPS: Instructions Per Second, is a typical measurement of a digital processor speed. Common instructions are bitwise addition, complement, shifting, etc.