ACE16k: A Programmable Focal Plane Vision Processor with 128 x 128 Resolution

Gustavo Liñán[*], Rafael Domínguez-Castro[*], Servando Espejo[*] and
Angel Rodríguez-Vázquez[*].

Abstract[**] **--- This paper presents a new generation 128x128 Focal Plane Analog Programmable Array Processor (FPAPAP), from a system level perspective. The design has recently sent to fabrication in a 0.35μm standard digital 1P-5M CMOS Technology. The chip has been designed to achieve the high-speed and moderate-accuracy constraints of most real time image processing applications. It has been designed to be easily embedded in conventional digital hosting systems: external data interchange and control are completely digital. The chip contains close to four millions transistors, 80% of them working in analog mode, and exhibits a relatively low power consumption (<4W, i.e. less than 1mW per transistor). Experimental results are expected for the date of paper presentation.**

1 Introduction.

Conventional vision systems 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 approach results in huge data rates which conventional computers can not analyze in real-time. Conventional computers and DSPs are able to manage such data rates for simple tasks like auto-focus, image stabilization, control of the luminance/chrominance, etc. However, the real-time execution of most spatial-temporal operations typical of front-end image processing tasks requires much more sophisticated digital processors. Consequently, conventional vision machines with real-time capabilities are bulky, expensive and extremely power-hungry. The contrast between the performance of artificial and "natural" vision systems is due, among other things, to the inherent *parallelism* of the latter.

Inspired by the efficiency of natural vision systems, universities and companies have focused their efforts on the development of new generations of devices, aiming to overcome the drawbacks of traditional ones by means of distributed parallel processing and concurrent signal acquisition and storage. One possible strategy is flip-chip bonding of separate sensing and processing devices. Another alternative is to "fuse" 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 [1][2]. Their development is expected to have a significant impact in quite diverse scenarios. However, industrial applications demand flexible and versatile chips, with programmable functions, distributed storage, and standard interfacing to conventional equipment. ACE16k, with its increased complexity and fully digital interface, is conceived to be one step closer to industrial applications than previous FPAPAPs prototypes.

2 System Description

2.1 System Architecture

ACE16k can be basically described as an array of 128x128 identical, locally interacting, analog processing units designed for high speed image processing tasks requiring moderate accuracy (around 8bits). The system contains a set of peripheral circuitries that, on one hand, allow a completely digital interface with the host, and on the other provide high algorithmic capability by means of conventional programming memories where the algorithms are stored.

Although ACE16k is an analog processor (computation is carried out in the analog domain), it can be operated in a fully digital environment. For this purpose, the prototype incorporates a bank of Digital-to-Analog (for input) and Analog-to-Digital (for output) converters at the images I/O port.

ACE16k is conceived to be used in two alternative ways. First, in applications where the images to be processed are directly acquired by the optical input module of the chip, and second, as a conventional image co-processor working in parallel with a digital hosting system that provides and receives the images in electrical form.

The architecture of the system is sketched in Fig. 1. The chip can be divided into five functional blocks.
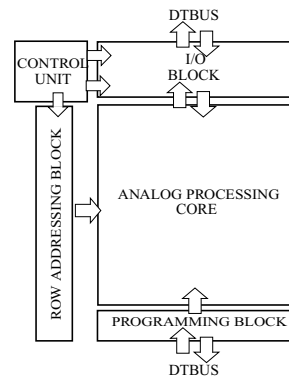
[*] Instituto de Microelectrónica de Sevilla, Centro Nacional de Microelectrónica, IMSE-CNM-CSIC. Avda. Reina Mercedes, s/n. Edif. CICA, 41012 Sevilla, SPAIN. E-mail: espejo@imse.cnm.es, Tel: +34-955-056666, Fax: +34-955-056686

Fig. 1:   Block Diagram

First, the analog processing core, which comprises the inner array of 128 x1 28 identical cells, a ring of border cells used to establish spatial boundary conditions for image processing, and several buffers driving analog and digital signals to the cell array. Second, the programming block, which contains two 64x32 SRAM digital memories used to store the algorithms to be executed by the chip, six 32x32 SRAM digital memories used to store, in 8bits format, different sets of the analog coefficients controlling the cell-to-cell interactions, some global bias signals, and some references used by the optical input module (precharging values, etc.). The programming memory contains also all the circuitry needed for external accesses to these memory blocks, and to transmit the programmed values to the analog processing core. This includes digital buffers for digital instructions, and digital-to-analog converters and analog buffers for weights (cell-to-cell interactions) and references.

The third to fifth blocks are dedicated to images I/O tasks. The global I/O control unit generates the signals required for I/O image accesses. This includes row and column addressing signals (automatically generated in a sequential way), and control of the Digital-to-Analog and Analog-to-Digital I/O converters bank.

The chip uses a 32 bits bidirectional data bus for external communication, and several address buses for the different blocks within the programming memory. The external interface follows very simple hand-shaking protocols.

ACE16k contains almost 4 millions transistors (80% of them operating in analog mode) and has been designed in a 0.35 μm CMOS 1P-5M layers, standard digital technology provided by ST Microelectronics. Table 1 shows the most relevant characteristic of the prototype.

Table 1: ACE16k Characteristics

Technology	STM-0.35 μm 5M-1P
Design Style	Full Custom (Analog Core) and Standard Cells (Digital I/O block)
Package	Ceramic QFP144
# of Cells	16384 (128 x 128 Array)
# of Transistors	3,748,170
# Transistors per cell	198
Cell Size	75.7 μm x 73.3 μm
Cell Density	~180 cells/mm^2
State Signal Swing	[0.6, 1.4]V (Programmable)
Weight Signal Swing	[2.15, 2.95]V (Programmable)
Time-Constant (lin. conv.)	~160ns
I/O Master Clock	32 MHz
Power Supply	3.3V +/- 10%
Power Consumption	< 4 Watts
# of Analog Instructions	32
# (Max.) Digital Instructions	64 x 64 Configurations
Die Size	11885.0 μm x 12230 μm

2.2 Programming Block

The programming block, described in Fig. 2, provides the algorithmic capability of ACE16k. It is basically a set of eight SRAM memory blocks with miscellaneous contents purpose, varying from digital vectors defining the algorithms to be executed (what we call "digital instructions"), to sets of cell-to-cell interaction weights and reference levels to be applied to the cell array (what we call "analog instructions").

The chip has two operating modes, namely the programming and the operation mode. During the programming mode, each of the 8 SRAM blocks can be independently accessed trough the data bus in order to be written (or read, just for testability purposes). On the other hand, in the operation mode, the contents of different groups of memory blocks are selected through different address buses, and transmitted is parallel to the cell array.

The programming block can be divided into three groups of memory blocks. Two of them (Operations Memory and Addresses Memory) are used to store digital instructions. Each of these blocks is designed to store 64 words of 32 bits (therefore fitting the data bus width). A digital instruction is defined as a 64bits digital vector that controls the configuration of the chip circuitry, comprising a word from the operations memory (32bit) and another one from the addresses memory (32bits). These two words can be addressed (transmitted to the processing core) independently through different address buses.

The third group (Weight and Analog References memory) is used to store cell-to-cell interaction weights and some references levels. This group consists of six identical SRAM blocks, each of them designed to store 32 words of 32 bits. Analog coefficients are defined by 8-bit words, and therefore each of these blocks stores 32 sets of 4 (digitalized) analog values. Thus, an analog instruction comprises 24 (i.e., 6 x 4) analog values that are transmitted in parallel to the processing core by means of a bank of 24 digital to analog converters. The selection of the analog instruction is done through an independent address bus.

For more homogeneous D-A conversion of the weight and reference signals, the resistor strings in the DACs are shorted (among those of equal conversion range) every few levels. In addition, for better homogeneity and also seeking sample-to-sample invariance, the
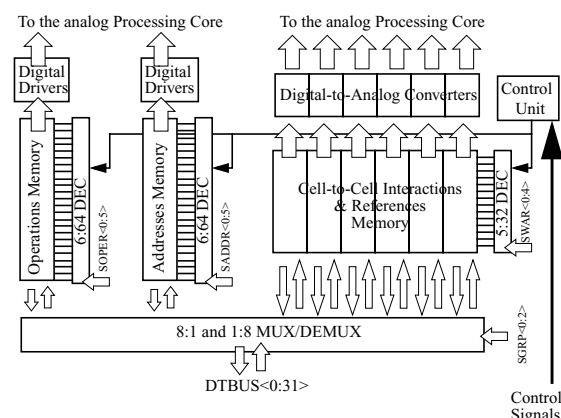

Fig. 2: Diagram of the Programming Block

analog buffers driving the processing core can be calibrated to eliminate their input offset.

In order to execute any task, the chip requires the application (selection, in operating mode) of both an analog and a digital instruction.

For external access to these memories (programming mode), an 8:1 MUX/DEMUX is used to multiplex the external 32-bits data bus to/from each of the 8 SRAM blocks. SRAM timing signals are generated from an internal Voltage Controlled Oscillator (VCO). A clock control block can optionally be used to modify the default VCO frequency (from about 0 to 250MHz) or to switch to an external reference. The external throughput will be above 32MHz in any case.

2.3 Analog Core

The analog Processing Core in ACE16k consists of an array of 128x128 locally interacting, identical processing units (cells) arranged in a rectangular grid. In addition, a ring of surrounding blocks is used to establish the proper spatial boundary conditions and to buffer the analog and digital instructions to the inner array. Each inner cell in the 128 x 128 array contains several analog functional blocks that provide the processing capabilities. These include:

- A configurable bank of analog multipliers that implement the cell-to-cell interactions with the 8 adjacent cells and with the cell itself. These multipliers are divided in four groups that can be driven by different, arbitrary images out of those stored in the ARAM (see below).
- A configurable nonlinear dynamic block, including linear and non-linear integrators, calibration circuitry, and an image arithmetic operator for weighted aggregation/substraction of images. The configuration of this block and of the analog multipliers bank defines the specific state equation of each cell.
- An Analog Random Access Memory (ARAM) with capacity for eight gray-scale pixel values with a resolution of 8bits.
- A Local Logic Unit, consisting of a programmable two-input one-output logic operator.
- An Optical Input module, consisting of different photosensor devices (Diffusion-Well Diode, Well-Substrate Diode and vertical PNP transistor) with configurable sensing schemes (integration, and logarithmic-compression type).
- An Address Event Downloading module, which allows the chip to download, sequentially, the location of active (black) pixels.
- A resistive grid module that allows continuous-time diffusion in a resistive-grid like manner.

2.4 I/O Interface

As compared to previous focal plane processor implementations ([1], [2], [3], [4]), and leaving aside the increase in the number of cells, the main improvement of ACE16k is the incorporation of a completely digital interface (not only for system control, but for digitalized gray-scale images I/O as well).

The chip incorporates 128 (one per column) Digital-to-Analog and Analog-to-Digital converters. DACs,
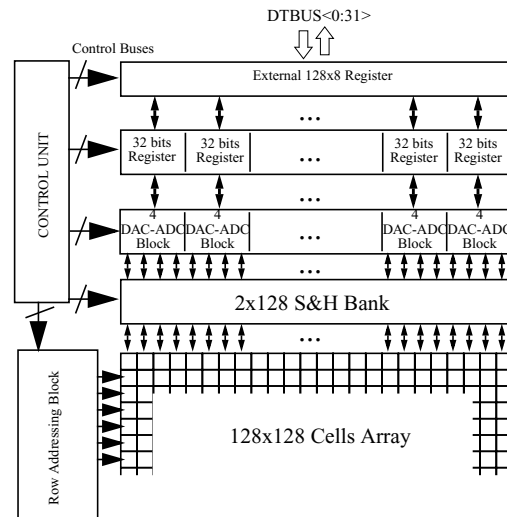

Fig. 3: I/O Block Diagram.

used for image input, are based on a resistor string and an analog multiplexer [5] while ADCs, for image output, follow a successive approximation approach [5]. These converter architectures provide a very good compromise in terms of area and power dissipation in this particular system. On one hand, the same DACs used for image input can be used as part of the successive approximation ADCs (the resistor levels are shifted up 1/2LSB to ensure a correct behavior of DAC-ADC loops). On the other hand, because the 128 converters work in parallel, a significant part of the digital circuitry needed to control the successive approximations can be shared in a common peripheral block, resulting in a substantial reduction in area and power dissipation. In fact, the ADC can be obtained from the DAC by adding just a comparator (used in each approximation step), a digital register (needed in any case for I/O processes) and some switches for reconfiguration. In addition, the use of the same resistor strings in the input and output converters ensures a correct DAC-ADC loop for every column. Resistor strings of all columns are shorted every few levels to provide homogeneous conversion among different columns. A final important detail towards the minimization of read-related fixed-pattern noise is comparators input offset. For this purpose, a self calibration process is automatically executed at the beginning of every data conversion.

At the system level, the I/O interface is based on a pipelined architecture. Data coming (going) from (to) the inner array goes through a double bank of 128 S&H circuits [6]. This double bank disposition allows a full row of data to be transferred to/from the processing core while, simultaneously, the next row is being converted at the DAC/ADC bank. On the other hand a 128 x 8 bits digital register is used for external data communication. This register, together with the 8 bit registers of the 128 ADCs, constitute a double digital row-wise memory needed to allow external data transferences to take place (through the 32-bits wide data bus) while the next (previous) row is being converted at the DAC/ADC bank. Fig. 3 shows a basic block diagram of the I/O block.

3 Global Throughput

The minimum time required for a D-A or A-D conversion at each converter in the image I/O D-A-D bank is 1μs. Within this time, a complete row is converted, since the 128 converters work in parallel. The time required for a full-row transference to/from the exterior of the chip is also 1μs (the external data bus, with a width of 4 bytes, operates at 32MHz).

Therefore the minimum time to load/download an image (128 rows) is 130μs (the two additional μs are needed because of the pipelined architecture).

The system is not designed for simultaneous image I/O and processing. Among other reasons, this avoids undesirable digital coupling with the analog processing circuitry. Therefore, the computation of a global processing throughput figure requires an allocation of image processing time. In most practical cases, an allocation of 140μs (similar time than a full-image input or output) is more than enough[*]. With this assumption, the time required to load, process, and download a 128x128 image is 400μs. This leads to an equivalent VGA[**] (640x480) processing speed of 100 Frames/second.

4 Previous Implementations

Table 2 presents a summary of some features associated to recent Analog Programmable Array of Processors (APAPs). Last row in this table refers to the presented prototype, ACE16k.

Speed is expressed in terms of equivalent operations per second. The equivalent multiply/add operations per second have been calculated assuming 10 discrete time-steps per time constant in the discrete-time emulation of the continuous-time algorithm. Assuming 10 interconnections per cell, with one multiplication and one addition required for each of them, yields 10 x 10 x 2 = 200 equivalent multiply/add operations per time constant and processor. Considering the 128 x 128 processors and a time constant of 0.8μs (minimum time constant for a continuous time dynamic evolution) yields a equivalent speed of about 4TeraOPS. Out of these chips, those reported in [2], [1], and also ACE16k have embedded distributed optical sensors, i.e. they are focal plane array processors. On the other hand, only ACE16k and the chip reported in [1] can handle grayscale inputs and outputs.

5 Summary

A new generation Focal Plane Analog Programmable Array Processor has been presented. The chip core consists of an array of 128x128 identical, locally interacting analog processing units. The systems is specially suited for real-time image processing applications. It has been designed in a standard, digital, 0.35μm CMOS 1P-5M technology provided by ST Microelectronics. The prototype contains close to 4 million transistors, 80% of them working in analog mode, with a power dissipation below 4W.

---

[*] Around 11 basic image processing tasks can be executed within 140μs.

[**] We divide the VGA frame into 6 x 4 chip-size images with sufficient overlap to avoid partition-related edge effects.

References

[1] G. Liñán, P. Foldesy, S. Espejo, R. Domínguez-Castro and A. Rodríguez-Vázquez. "A 0.5mm CMOS 10^6 Transistors Analog Programmable Array Processor for Real-Time Image Processing", *Proc. of the 25th European Solid-State Circuits Conference,* pp. 358-36, Duisburg-Germany, Sept. 1999.

[2] R. Domínguez-Castro, S. Espejo, A. Rodríguez-Vázquez, R. Carmona, P. Foldesy, A. Zarándy, P. Szolgay, T. Sziranyi and T. Roska, "A 0.8 μm 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.

[3] A. Paasio, A. Dawidziuk, K. Halonen and V. Porra, "Minimum Size 0.5μm CMOS Programmable 48 x 48 CNN Test Chip", *Proc. of the 1997 European Conference on Circuit Theory and Design*, pp. 154-156, Budapest, Hungary, September 1997.

[4] P. Kinget and M. Steyaert. "Analog VLSI Integration of Massive Parallel Processing Systems", Ed.Kluwer Academic Publishers, 1996.

[5] B. Razavi, *Principles of Data Conversion System Design.* IEEE Press, New York, 1995.

[6] Roubik Gregorian and Gabor C. Temes, *Analog MOS Integrated Circuits for Signal Processing.* John Wiley & Sons, New York, 1994.

[7] *Network Universal Machine"*, Ph.D. Thesis, Helsinki University of Technology, December 1998.

[8] S. Espejo: *"VLSI Design and Modeling of CNNs"*. Ph. Dissertation, University of Sevilla, March 1994.

[9] M.J.M Pelgrom, A.C.J. Duinmaijer and A.P.G. Welbers: *"Matching Properties of MOS Transistors"*. IEEE J. Solid-State Circuits, Vol. 24, pp 1433-1440, October 1989.

Table 2: Summary and comparison of recent chip implementations

Reference	Technology (CMOS (μm))	Array Size (cells)	Die Size (mm²)	Cells Density (cells/mm²)	Speed (XOPS)	Stored Program	Analog Resolution (eq. bits)	Optical Sensors	Electrical Input[a]	Electrical Output[c]	Embedded Images Memory	Digital External Control
[4][b]	0.7	20x20	25	17	12G		6-7		A	B		
[2]	0.8	20x22	30	28	0.13T	√	6-7	√	B	B		√
[3]	0.5	48x48	11.4	295	7.65T	√	2		B	B		√
[1][c]	0.5	64x64	87	81	0.40T	√	7-8	√	A+B	A+B	√	√
ACE 16k[d]	0.35	128 x128	130	180	4.1T	√	7-8	√	D	D	√	√

a. A=Analog, B=Binary (B&W), D=Digital.

b. The convolutors in this chip have vertical and horizontal interconnections, but not diagonals.

c. Some additional functionalities of this design include: local evolution enabling mask, global binary gates for fast binary output-images evaluation, cyclic spatial boundary conditions.

d. Preliminary data from simulations