ACE16k: The Third Generation of Mixed-Signal SIMD-CNN ACE Chips Toward VSoCs

Angel Rodríguez-Vázquez, *Fellow, IEEE*, Gustavo Liñán-Cembrano, L. Carranza, Elisenda Roca-Moreno, Ricardo Carmona-Galán, *Member, IEEE*, Francisco Jiménez-Garrido, Rafael Domínguez-Castro, and Servando Espejo Meana

Abstract—Today, with 0.18-μm technologies mature and stable enough for mixed-signal design with a large variety of CMOS compatible optical sensors available and with 0.09-μm technologies knocking at the door of designers, we can face the design of integrated *systems*, instead of just integrated circuits. In fact, significant progress has been made in the last few years toward the realization of vision systems on chips (VSoCs). Such VSoCs are eventually targeted to integrate within a semiconductor substrate the functions of optical sensing, image processing in space and time, high-level processing, and the control of actuators. The consecutive generations of ACE chips define a roadmap toward flexible VSoCs. These chips consist of arrays of mixed-signal processing elements (PEs) which operate in accordance with single instruction multiple data (SIMD) computing architectures and exhibit the functional features of CNN Universal Machines. They have been conceived to cover the *early* stages of the *visual processing* path in a *fully-parallel* manner, and hence more efficiently than DSP-based systems. Across the different generations, different improvements and modifications have been made looking to converge with the newest discoveries of neurobiologists regarding the behavior of natural retinas. This paper presents considerations pertaining to the design of a member of the third generation of ACE chips, namely to the so-called ACE16k chip. This chip, designed in a 0.35-μm standard CMOS technology, contains about 3.75 million transistors and exhibits peak computing figures of 330 GOPS, 3.6 GOPS/mm^2 and 82.5 GOPS/W. Each PE in the array contains a reconfigurable computing kernel capable of calculating linear convolutions on 3×3 neighborhoods in less than 1.5 μs, imagewise Boolean combinations in less than 200 ns, imagewise arithmetic operations in about 5 μs, and CNN-like temporal evolutions with a time constant of about 0.5 μs. Unfortunately, the many ideas underlying the design of this chip cannot be covered in a single paper; hence, this paper is focused on, first, placing the ACE16k in the ACE chip roadmap and, then, discussing the most significant modifications of ACE16K versus its predecessors in the family.

Index Terms—Analog programmable very large-scale integration (VLSI), early vision chips, silicon retinas.

I. INTRODUCTION

VISION involves extremely complex computational tasks [1]–[8]. So complex that, despite its huge set of applications and potential uses, no artificial vision system has been able to reach the level of efficiency of natural vision systems up to date. Indeed, performances of currently available artificial vision systems are far below those of the smallest insect, despite the usage of the most sophisticated latest generation computing devices. Is this paradox due to a lack of industrial or commercial interest? Clearly not, since the number of applications of artificial vision systems are enormous. Which can be hence the reason underlying the gap between natural and artificial vision systems?

Probably, the reason is that conventional signal processing architectures are not the best suited for vision. In these architectures, there exists a clear separation between signal *acquisition* and signal *processing*, with the role of analog processing being restrained to the front-end functions, namely transduction, signal conditioning and data encoding. The problem is that images contain a huge amount of data, many of them redundant, i.e., not carrying any information. Hence, does it make any sense to consume resources in handling, i.e., converting and processing, these data? Nature gives us some guesses about that. In natural vision systems, the *front-end* device, the *retina*, does not only acquire but also pre-processes the visual information [9], [10], such that the amount of data transmitted through the optic nerve to the brain gets compressed by a factor around 150.[1]

A similar compression of information occurs in any vision processing chain. As the signal climbs through consecutive levels in the processing path, its dimensionality shrinks whereas its abstraction increases. Thus, although using *serial* digital signal processing is advisable at the upper levels of the hierarchy, it might not be so adequate for early processing. Operating with images at the bottom level of the processing hierarchy implies intensive memory accesses and poses important constraints on the bandwidth of the communications between memory and processor. Also, having a chip to sense the visual information (*imager*) and another one to process it (*processor*), requires high-speed data conversions and transferences to achieve large frame rates. Using the conventional *Imager-Memory-DSP* architecture it is possible to reach 30 FPS, even for large resolution images. However, high-speed industrial applications requiring ultrafast frame rates[2] might turn unfeasible.

ACE chips render ultrafast operation feasible by using massively *parallel analog processing* at the early stages, as natural

[1]The human eye contains about 150 mill. photoreceptors whilst the optic nerve contains about 1 mill. fibers.

[2]In the order of 1000 FPS.

retinas do. Some reasons supporting this choice are [11]–[13] as follows.

1) The *accuracy* required for early processing is moderate or even low. Actually, the perceptual quality of the images does not drop significantly in the presence of perturbations (noise, spatial variances, nonlinearities, . . .); even if these perturbations are as large as 5% of the full scale[3].

2) The *speed* versus *power* efficiency of moderate-to-low resolution analog circuits is much larger than that of digital counterparts. This is relevant since very high speed is needed to achieve high frame rates for moderately large images.

3) The *area* efficiency of analog circuits for moderate-to-low resolution applications is better than that of digital counterparts.

The chip described in this paper represents the third generation of ACE chips and has been designed to overcome some limitations of its predecessors, particularly those of the the so-called ACE4k chip [5]. Major improvements of ACE16k include the following.

- **Incorporation of digital buses for grayscale data:** ACE16k embeds per-column data converters (arranged in analog-to-digital (A/D) and digital-to-analog (D/A) re-configurable pairs) for fully digital interfacing.

- **Exact control of the timing for input/output (I/O) access:** To that purpose, ACE16k does not include the possibility of individual cell selection; instead it incorporates an autonomous addressing scheme. Also, it employs a hand-shaking protocol to eliminate timing constraints.

- **Better internal organization of the processing cells:** ACE16k incorporates the so-called *ACE-BUS* to allow any functional block within the cell to communicate with any other.

- **Use of nonconventional logic blocks:** Particularly, the four local logic memories (LLMs) of ACE4k have been replaced by local analog memories (LAMs), and the local logic unit (LLU) has been designed to operate within reduced analog-compatible voltage ranges, instead of within complete digital voltage ones. Also, dynamic, instead of static, digital memories are used to store template masks. Finally, dedicated logic inverters with peak current limitation have been used instead of conventional ones.

- **Improvement of the optical interface:** ACE16k incorporates a re-configurable optical input module with the following features:
  - **User-defined photo-sensing device:** The user can select among a P-Diffusion/N-Well photo-diode, a N-Well/P-Substrate photo-diode or a P-N-P vertical photo-transistor.
  - **User-defined sensing scheme:** The user is allowed to select between normal linear integration modes or logarithmic compression sensing.

- **Incorporation of an address event detection scheme:** to simplify the extraction of information from black and white (B/W) images. The associated circuitry provides addresses (instead of images) corresponding to array locations where activity is detected. This scheme also embeds the functionality of the global gates—no address is provided if no active cells exist.

- **Improved power consumption management:** ACE16k has four times more cells than ACE4k, and much larger functional capabilities. However, it switches idle blocks off and uses scaleddown logic levels to keep the power consumption moderate—less than 180 μW per cell.

ACE16k has been designed in a digital CMOS 0.35-μm 5M-1P technology and contains more than 3.75 million transistors—85% of them working in analog mode. It can reach peak computing figures[4] of 330 GOPS, 3.6 GOPS/mm^2, and 82.5 GOPS/W. It provides and accepts 8-bit digitized images through a 32-bit data bus which works at 120 Mbytes/s

II. ACE16k IN ROAD MAP OF ACE MIXED-SIGNAL VISION CHIPS

ACE chips consist of an array of identical *processing elements* (PE) which execute the same instructions at the same time. Instructions are executed on data which are *locally* defined, i.e., at the PE level, while the sequence of instructions is controlled and timed by a digital controller which is shared by all the PEs. Typically, for implementation purposes, communications between PEs are restricted to the nearest neighbors. However, despite such an architectural limitation, ACE chips are able to implement most *early-vision* processing tasks [4]–[6], [13]. Adding the capability of sensing the visual information in a one-by-one pixel-to-PE correspondence makes these systems very well suited to implement the front-end stage of VSoCs. Obviously, processing images whose resolution is larger than the array size (necessarily limited due to the incorporation of programmable processing circuitry at pixel level) requires *windowing* and *time multiplexing*.

Regarding ACE chip architectures, different questions arise, which relate to:

1) functions to be incorporated within the PE;
2) complexity of the control unit;
3) interfacing with other hardware and/or equipment.

The answers to these questions are largely dependent of the intended application. However, due to size, design complexity, and fabrication costs of these chips, the design of special purpose devices is only advisable if a market niche absorbing mass production is ensured. Otherwise, the architecture of the PE must be flexible enough to guarantee the execution of the largest possible amount of vision algorithms under real-life illumination conditions. Thus, taking into account that most early vision processes consist of the application of *convolutions masks*, and the combination (either by Boolean operations in the case of B/W images, or by a local analog arithmetic operator) of their results in a bifurcated-flow algorithm, the following operators should be included at the PE level:

---

[3]The exact number is obviously application dependent.
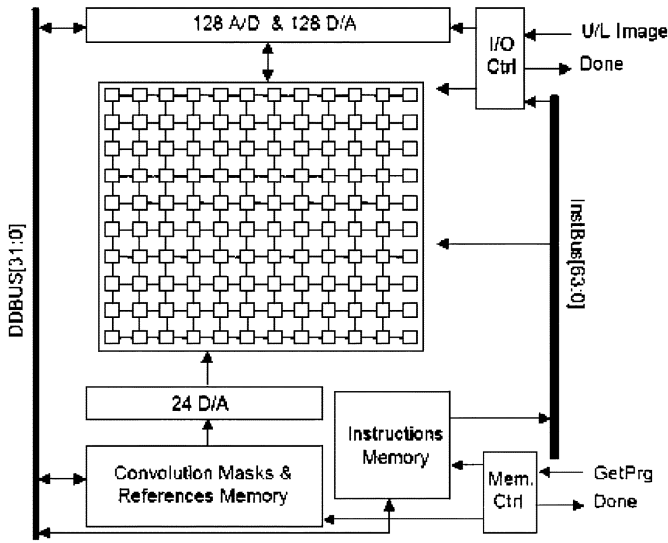
[4]These data correspond to experimental results.

Fig. 1.   Conceptual architecture of ACE16k.

1) multipliers and adders; for the convolution operation;
2) analog registers; to allow for the storage of previous results at the local level;
3) arithmetic operator and/or binary operator; to combine previously obtained results;
4) local masks; to allow for the conditional execution of certain operations at PE level depending on some locally defined value.
5) wide dynamic range optical input; to permit the light-sensing capability, and, hence, to avoid the bottleneck existing in data transmission from the sensory to the processing plane in conventional nonmassively parallel solutions.

To cope with the objective of covering the largest possible set of applications, all functions above must be programmable, including reliable setting of analog parameters, reconfiguration of topologies and control of internal data-flows. Regarding the control unit, its roles are:

1) controlling the sequence of operations to be executed on the array;
2) storing the machine code of the algorithms to be implemented;
3) storing the data which define the internal analog parameters of the array.
4) interfacing the external world using standard protocols;
5) performing high-level signal processing tasks.

Based on, first, the convenience of making the interfacing completely standard, and, second, the necessity to guarantee robustness in the control of the analog parameters, the control unit should be fully digital, with the obvious exception of the blocks which interchange information (both data and commands) with the array.

ACE chips have been designed with these guidelines in mind. Specifically, this is the case of ACE16k [6] whose conceptual architecture is depicted in Fig. 1. As already mentioned, ACE16k

represents the third generation of ACE chips. Fig. 2 depicts the evolution of these chips, where a bifurcation appears at the time when ACE16k was released. Such bifurcation is related to the different nature of the behaviors addressed by instances belonging to each of the branches. On the one hand, ACEXX chips are basically conceived to perform spatial image processing on temporal image flows. On the other hand, CACEXX chips are designed to emulate the spatial-temporal dynamic evolutions observed in mammalian retinas [14].

Table I summarizes some main features of the three different generations of ACEXX chips. It highlights a continuous improvement across time. ACE400, the first member of this family, was designed in 1996 using a standard 0.8-μm technology [4]. It was conceived to operate only on B/W image flows, and included reduced programming capability. Special attention was paid to the optical interface in order to achieve high speed capturing through the incorporation of Darlington-based photocurrent amplification.

Four years later, in 2000, the ACE4k chip was released [5]. Together with an increase by a factor of ten in spatial resolution, this chip incorporated much larger programming capabilities. Despite the increased complexity and its capability to handle grayscale images, this chip featured significantly larger PE density and lower power consumption while basically keeping the time constant unaltered. These ameliorations were basically the consequence of major architectural and circuital improvements, and marginally due to the scaling down of the fabrication technology—from 0.8 μm to 0.5 μm.

By the end of 2002, the first version ACE16k chip was made available from the foundry [6]. Improvements of ACE16k versus ACE4k have already been mentioned in the Introduction and are summarized in Table II. Details about the architectural and circuital tricks employed to achieve such significant enhancements can be found in [13]. In Section III, we basically discuss the modifications affecting the PE itself. Below, we give some hints regarding the programming memory and the I/O interface, whose circuit level details are presented in [6].

Regarding the programming memory of ACE16k, it is similar to that of ACE4k. However, three main differences exist.

- The instruction memory has been arranged into two blocks with 64 words of 32 bits each. This division aims to separate addresses from definition of operations—something like defining operations and operators separately. Thus, and thanks to the use of separate control buses, ACE16k has a programming memory of 64 × 64 words of 32 bits, instead of simply 64 words of 48 bits as with ACE4k. Such an increase in the memory gives the user the possibility of programming and testing more complex algorithms.
- ACE16k uses a memory control circuitry which includes a voltage-controlled oscillator to generate all the timing signals required for memory management.
- Finally, ACE16k uses hand-shaking protocols, instead of strobing signals, to control the access to the programming memory. This overly simplifies control.

A related major modification of ACE16k consists of the incorporation of self-calibration stages to the analog buffers
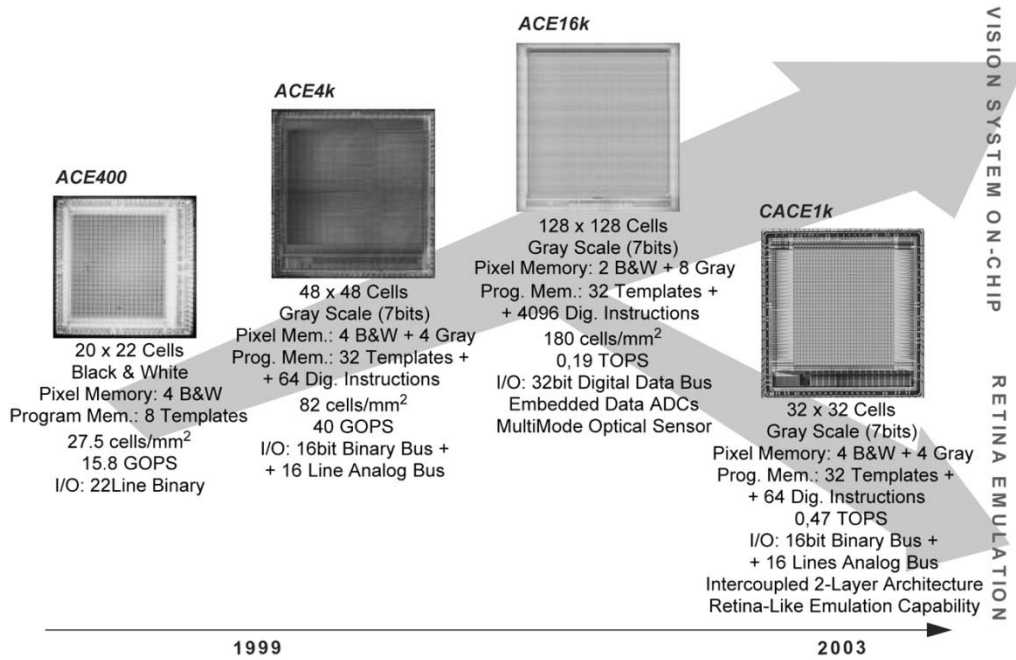
Fig. 2.   Historical roadmap of ACE chips.

TABLE I
OVERVIEW OF ACEXX CHIP FEATURES

Feature	ACE400	ACE4k	ACE16k
Technology @ Supply	0.8μm 2M-1P @ 5V	0.5μm 3M-1P @ 3.3V	0.35μm 5M-1P @ 3.3V
Design Style	Full Custom	Full Custom	Full Custom / Standard Cells
Signal Range	2V (Fully-Diff.)	0.6V – 1.4V	0.6V – 1.4V
Weight Range	2V (Fully-Diff.)	2. 15V – 2.95V	2. 15V – 2.95V
Analog Accuracy	7-bit	7.7-bit	8-bit
# Analog Instructions	8	32	32
# Digital Instructions	N/A	64	4096
# Memories per PE	4 BW	4 BW & 4 Grey	2 BW & 8 Grey
I/O Digital Speed	22 x 10 Mbit/s	16 x 10Mbit/s	120MByte/s
I/O Analog Speed	N/A	16 x 1MSamp/s	
Array Size	22 x 20	64 x 64	128 x 128
# Transistors on Chip	~200.000	~1,000,000	~3.75 mill.
PE Density (PE/mm^2)	27.5	82	180
Computing Power (GOPS)[a]	15.8	40	330
Speed / Area (GOPS/mm^2)	0.98	1	3.2
Speed / Power (GOP/ Joule)	25	39.5	100

a. For the ACE400 chip, speed figure refers to Boolean Operations per Second (BOPS) whereas for ACE4k and ACE16k it refers to 8-bit equiva-
    lent resolution operation, i.e. 8-bit additions or products.

which drive weights and analog references to the cell array. Although ACE16k uses the same distributed buffer strategy as ACE4k, the topology of the buffer includes extra circuitry for calibration purposes. Fig. 3 shows a simplified block diagram of the weight generation circuitry in ACE16k, including the RAM block in which coefficients are digitally stored, the 8-bit D/A converter (DAC), the two-level buffer structure, and the calibration circuitry.

Regarding I/O, ACE16k incorporates a fully digital port for image transferences. Fig. 3 shows a simplified block diagram of the I/O block in ACE16k. It includes a bank of 128 8-bit A/D converter (ADC) and DAC. Since the data bus is 32 bits wide, each word transmitted to/from the chip contains information about four adjacent cells—same row, consecutive columns. Then, and by just looking at the way of writing/reading images, the array can be divided into 32 identical blocks of four adjacent columns.

Data transference uses a two-stage pipeline architecture. In the input mode, data are sent to an input register of 8×128 bits (see Fig. 4). Once filled, this register is transmitted in parallel to an internal 8×128 register whose outputs (in blocks

TABLE II
COMPARING ACE16k VERSUS ACE4k

|	ACE4K	ACE16k	Scaling Factor
Technology	0.5μm 3M-1P	ST-0.35μm 5M-1P	200 % area scaling
Design Style	Full Custom (Analog Core) Standard Cells (Dig. I/O block)		
Signal Range	[0.6, 1.4]V (Programmable)		
Weight Range	[2.15, 2.95]V (Programmable)		
# Transistors per Cell	~180 (172 / 198)		
Analog Accuracy	~ 7 bits equivalent		
# Analog Instructions	32		
# Digital Instructions	64 Configurations	4096 Configurations (64 x 64)	64x plus programmable configurations
# Memories per Pixel	4 Gray + 4 B/N	8 Gray	4+ Gray Images on Chip
I/O Digital Speed	160 Mb/s@B/N	128 MBytes/s@Gray	640 % Digitized Gray Images
I/O Analog Speed	16 MS/s	128 MS/s	800 % Digitized Images
# Cells	4096 (64 x 64)	16384 (128 x 128 Array)	400 %
# Transistors	~1.000.000	3,748,170	375 %
Cell Density	~82 cells/mm^2	~180 cells/mm2	220 %
Speed (GOPS)	47	300	640 %
Speed / Area (GOPS/mm^2)	1	3.23	323 %
Speed / Power (GOP/J)	39.5	100	250 %
Power per Cell	~370 μWatts	~180 μWatts	49 %
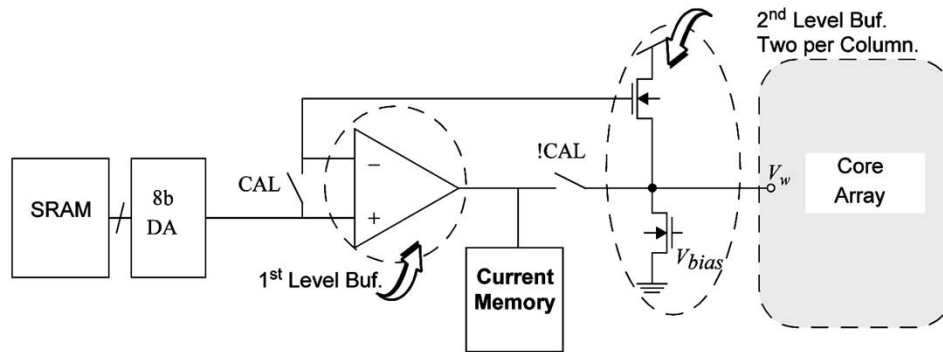
Fig. 3.   Distributed buffers in ACE16k.

of 8) are permanently connected to a bank of 128 DACs which operate in parallel. At the same time, the external register is again being filled with the information about the next row to be written—avoiding idle periods. At the end of the conversion, the first module of a double bank of 2 × 128 sample-and-hold (S&H) circuits acquires the converted data and sends it to the selected row of cells. While the first module of the bank of S&H sends the analog value to the array, the second module is this bank is capturing the next row of data which is being converted by the DACs.

During an output process, the first row is acquired by the first module of the S&H bank. In the next step, these data are held and converted while the second module of the S&H bank captures the second row. At the end of this step, the digital information (the result of the conversion of the first row) is sent to the external register where it is ready to be downloaded during the third step. In the third step, the content of the first row is read at the output of the external register, the content of the second row is being converted and the third row is being captured by the first module of the S&H bank.
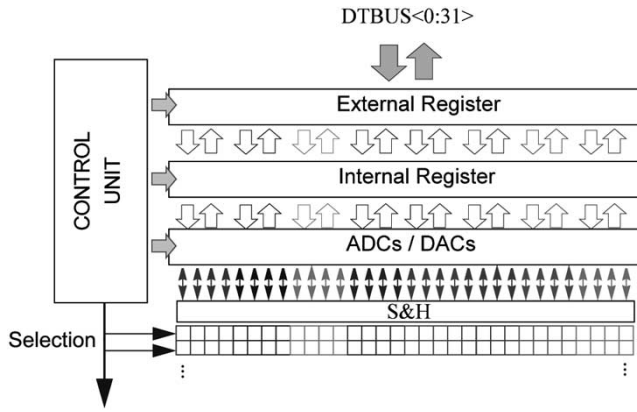
DTBUS<0:31>

Fig. 4.   I/O Block diagram.

III. ACE16K VERSUS ACE4K: MODIFICATIONS IN PE

A. ACE-BUS

The PE of ACE4k is designed to allow direct communication only between closely related blocks. However, ACE16k uses a new communication scheme, the *ACE-BUS*. The ACE-BUS is basically a node of the processing unit (PU), where every functional block connects its input and output ports. Communications between blocks always happen in the same way; first, one functional block (the data source) is configured in output mode, while a second one (the destination) is configured in input mode. It overly simplifies the definition of operations and data movements, and allows for rapid checking of conflictive switch configurations.

Fig. 5 shows the block diagram of the PU in ACE16k. Synapses take their inputs from the ACEBUS. They can be initialized by using either the LAM module contents, the result of LLU operations, the result of an optical acquisition, or the result of a passive diffusion realized by using one resistive grid embedded in the chip. The analog processing core steers the processed input current (the input current after eliminating all the offset contributions) to the ACE-BUS; this current can then be rooted either to the state capacitor or to any of the LAM modules.

B. Image-Processing Kernel

The synaptic analog multipliers are designed by using the same one-transistor technique as in ACE4k [13]. They are driven by voltages at both the signal and the scaling input and deliver a current at the output. The bank of multipliers, depicted at the conceptual level in Fig. 6, is driven by three different pixel values, $P_A, P_B,$ and P_C so that the current which flows into the PE is

$$I_{\text{tot}} = \mathbf{A} \bullet \mathbf{P_A} + b \cdot P_B + c \cdot P_C + z + I_{\text{off}} \qquad (1)$$

where the operator $(\bullet)$ denotes the convolution product of the template and the pixel value matrix, and I_{off} is the offset term generated by the one-transistor multipliers. This offset term is

eliminated by using a high-accuracy current memory [13], [15]. Fig. 7 shows a conceptual schematic of the PE input block including the S3I current memory used for offset cancellation, based on [15]. The resulting current

$$I_{\text{in}} = \mathbf{A} \bullet \mathbf{P_A} + b \cdot P_B + c \cdot P_C + z \qquad (2)$$

is either steered to the ACE-BUS, or to the input of a capacitive-input current comparator [16] whose output is connected to the ACE-BUS through an analog switch. Then, two situations may occur.

- A voltage codifying the sign of I_{in} (i.e., the sign of the outcome of the convolution operation) is delivered to the ACE-BUS

$$V_{\text{ACE-BUS}} = \begin{cases} X_{\text{min}} & \mathbf{A} \bullet \mathbf{P_A} + b \cdot P_B + c \cdot P_C + z > 0 \\ X_{\text{max}} & \mathbf{A} \bullet \mathbf{P_A} + b \cdot P_B + c \cdot P_C + z < 0. \end{cases} \qquad (3)$$

In this case, the output is a B/W pixel value.
- The analog current I_{in} is routed to one of the capacitors associated to the pixels and the output is a grayscale pixel value.

In any case, the specific pixel capacitor to which the output of the input block is routed is selected by the user through the activation of some bits in the digital instruction. By so doing, the evolution of the PU is described by a state equation whose actual expression depends on the selected integrating capacitor. Therefore, different kinds of processing kernels are available.

- Consider, for instance, that you want to execute a Sobel operator [8]. The convolution matrix is then defined in $\mathbf{A}$; the image is loaded into P_A; the following values are set: $c = z = 0$, and $b = -1$; and the signal current is routed to C_B. Hence, the equivalent state equation obtained for each PU is

$$C_B \frac{dP_B}{dt} = -P_B + \mathbf{A} \bullet \mathbf{P_A} \qquad (4)$$

whose steady state is $P_B = \mathbf{A} \bullet \mathbf{P_A}$, as corresponds to the desired convolution output.
- Consider now that the capacitor which receives the input current is C_A. Then, the cells are dynamically coupled and CNN spatio-temporal operations are realized.
- Consider finally that the current is routed to C_A; that all but the central entries of matrix $\mathbf{A}$ are null; and that this central entry is $A_{cc} = -1$. The steady-state solution is then

$$P_A = b \cdot P_B + c \cdot P_C + z \qquad (5)$$

which corresponds to the realization of grayscale image-wise arithmetic operations.

Although ACE16k also uses one-transistor synapses and, due to the similarities between the electrical parameters of 0.5-μm 0.35-μm technologies, the same voltage ranges as in ACE4k, the aspect of synapse transistors have been reduced from 2/20 to 1/20. This keeps the voltage drop across the metal line which
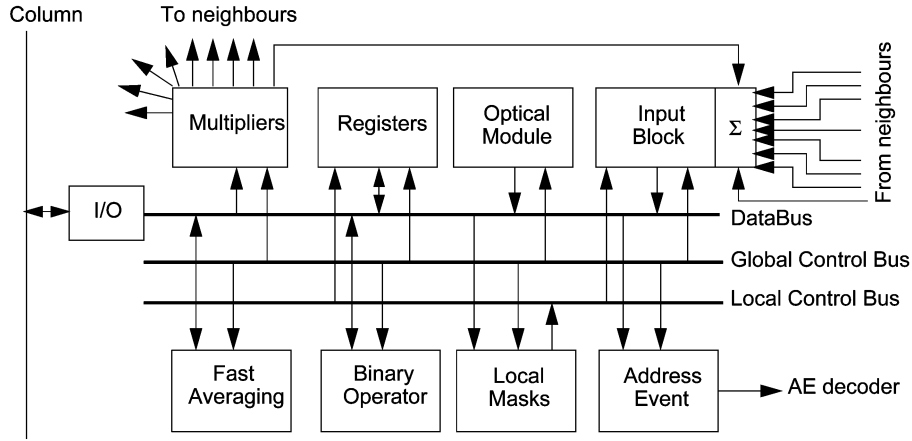
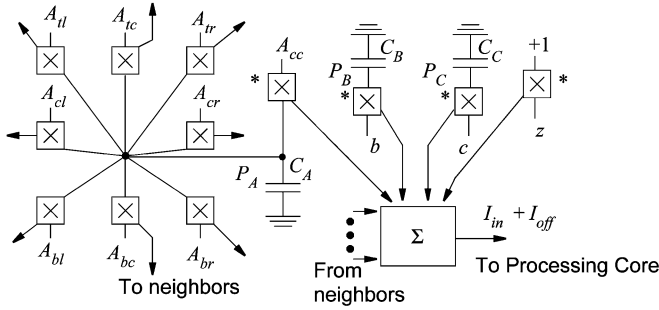Fig. 5.   Functional block diagram of the ACE16k PE.

Fig. 6.   Bank of multipliers in ACE16k.

drives the weight to the cell. After some calculations, the following expression can be obtained:

$$\frac{N_{\text{row}_{\text{ACE4K}}} \cdot \left(\frac{N_{\text{row}_{\text{ACE4K}}}}{2} - 1 \right)}{N_{\text{row}_{\text{ACE16K}}} \cdot \left(\frac{N_{\text{row}_{\text{ACE16K}}}}{2} - 1 \right)} \frac{L_{\text{cell}_{\text{ACE4K}}}}{W_{\text{met}_{\text{ACE4K}}}} S_{\text{syn}_{\text{ACE4K}}}$$

$$= \frac{L_{\text{cell}_{\text{ACE16K}}}}{W_{\text{met}_{\text{ACE16K}}}} S_{\text{syn}_{\text{ACE16K}}} \quad (6)$$

where S denotes the aspect of synapse transistors and W_{met} denotes the width of the metal layers driving the weights. Since, $N_{\text{row}_{\text{ACE4K}}} = 64$ and $N_{\text{row}_{\text{ACE16K}}} = 128$, (6) becomes

$$\frac{1}{4} \cdot \frac{L_{\text{cell}_{\text{ACE4K}}}}{W_{\text{met}_{\text{ACE4K}}}} S_{\text{syn}_{\text{ACE4K}}} \approx \frac{L_{\text{cell}_{\text{ACE16K}}}}{W_{\text{met}_{\text{ACE16K}}}} S_{\text{syn}_{\text{ACE16K}}}. \quad (7)$$

Since $L_{\text{cell}}/W_{\text{met}}$ is almost invariant from technology to technology (in the ideal case, both scale as the technology scaling factor does), the aspect ratio of the synapse transistor in ACE16k must be reduced by a factor of four in order to keep the same voltage drop as in ACE4k. However, because the number of multipliers is two times smaller in ACE16k, the aspect ratio is reduced only by a factor of two. The reason for reducing the width is that it does not practically affect the time constant. The counterpart is a degradation of matching which is attenuated by hardware.

C. Increasing the Cell Density

Lastly, the PE size is determined by the lines which carry the weights and control signals: their number, their width and the minimum separation between them. Obviously, having five metal layers (ACE16k@0.35-μm technology) instead of three (ACE4k@0.5-μm technology) gives some room for decreasing the cell size. However, the following hold.

- The top metal layer, metal 5, should be used only for power supply and ground. On the one hand, this layer has the maximum separation between adjacent lines. On the other hand, it has the greatest conductivity and hence the maximum current driving capability.
- ACE16k has a much larger number of PE-embedded functions than ACE4k (50 versus 35). Obviously this increases the number of control lines.

To meet the target of having cell densities larger than 150 cells $\times$ mm^{-2}, ACE16k employs an interaction pattern among cells different from that of ACE4k. As Fig. 6 shows, each PE contains 12 analog multipliers. Eight of them connect the cell to its neighbors; the other four provide additional inputs to the processing block. The multipliers marked with a star in Fig. 6 are double; they consist of the parallel aggregation of two multipliers. The purpose of this "double strength" is to increase the robustness in certain operations. From [17], it can be seen that in most cases, the central element of the template matrices is larger than the noncentral elements. At the electrical level it means that the corresponding multipliers have to be driven by quite different voltages, thus increasing mismatch-induced errors. By increasing the strength of central multipliers, the difference between weight voltages, and consequently the overall robustness, increases.

D. Digital Modules

The PE of ACE4k embeds conventional digital circuitry. This is not convenient because of the following.

- Level adapters are needed to transform the logic voltage levels, corresponding to full-scale swings, into levels compatible with the electrical operation of the PE analog circuitry.
- Protective measures must be taken to attenuate the impact of the large-power switching noise on the analog circuitry [18]. Last, this means greater area and penalizes cell density.
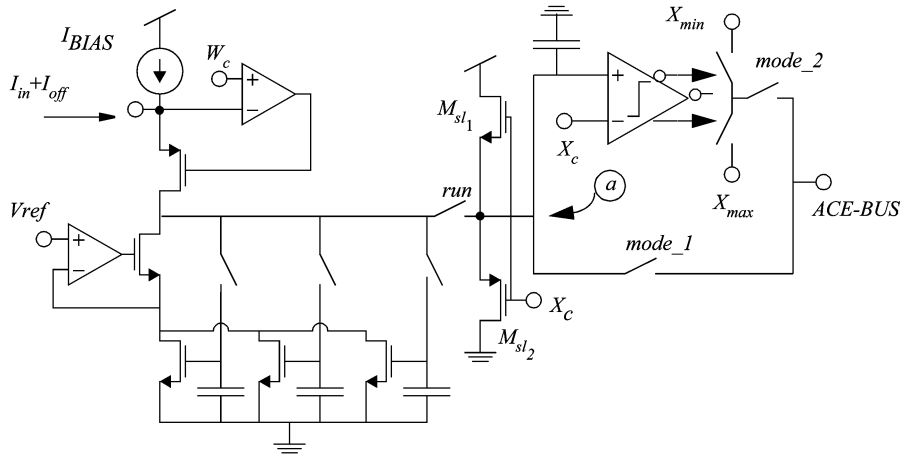
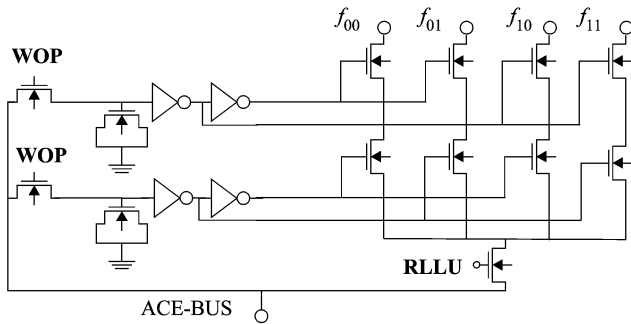Fig. 7.   Simplified schematic of the PE input block.

Fig. 8.   LLU in ACE16k.

Fig. 9.   Inverters and biasing circuitry in the cell in ACE16k.

In the case of ACE16k different measures have been taken to overcome these drawbacks, the following.

- The four LLMs of ACE4k have been replaced by LAMs. On the one hand, this eliminates the digital switching noise introduced by the LLMs. On the other hand, the impact on the silicon area is not very large because the readout amplifier is shared with the other LAMS. Finally, voltage level adapters between the LLMs and multipliers are not further needed.

    In addition to that, having eight instead of four LAM modules increases significantly the algorithmic capabilities of the chip.

- The LLU has been conceived to operate as an independent module which gets its inputs from the ACE-BUS and which also drives its output to the same ACE-BUS. This means a significant difference as compared to ACE4k. There, the LLU was intrinsically related to the LLM since its inputs were always taken from two fixed LLMs. In addition, although the LLU works as an intrinsically logic device, its inputs and outputs are provided via the ACE-BUS and have hence analog voltage levels.

Fig. 8 shows the LLU in ACE16k. Its two inputs, $OP0$ and $OP1$ are acquired from the ACE-BUS by using instruction bits **WOP0** and **WOP1**, while the result of the LLU operation is written to the ACE-BUS when the bit **RLLU** is activated. Logic inverters in the LLU (as well as any other inverter in the cell) are not conventional CMOS inverters but current-peak limited
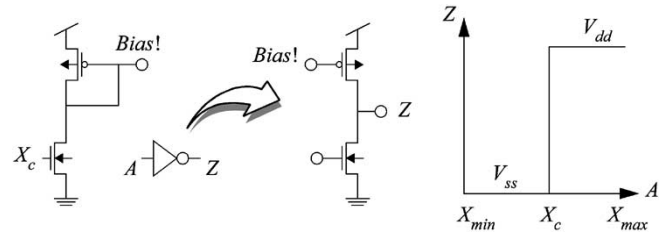
inverters. They have been designed using an NMOS transistor connected to a PMOS resistive load as depicted in Fig. 9. The resistive load is biased by a common biasing circuitry—shared by all the inverters in the cell. It establishes the quiescent point of the inverter around the middle of the voltage range for pixels.

E. Multimode Optical Sensor

Light sensing in ACE4k is realized by a parasitic diffusion-to-substrate diode of the LAM access switches. Thus, sensitivity is rather low, and cross-talk among the LAM modules arises. ACE16k incorporates a multimode optical sensor which has been conceived to be flexible enough to operate under very different illumination conditions. Fig. 10 shows its conceptual schematic, including three main blocks.

- The first one, a tri-state readout buffer, controls the communications between the sensor and other blocks in the PU. Sensor accesses are controlled by the global programming signal **ROPT**.
- The second one is devoted to transforming the photo-generated charges into a voltage. The user has the possibility of selecting the photo-transduction mechanism by means of signals **LOG1**, **LOG2**, **PCH**.
- The third block includes the optical sensor itself and two configuration switches $M_{s1,2}$ used to select one out of the three available photo-sensors. The selection of the sensor is carried out by signals **DW** and **WS**.

The optical sensor can be configured to operate in three different linear integration modes [Fig. 11(a)–(c)] and three different logarithmic compression modes [Fig. 11(d)–(f)]. In
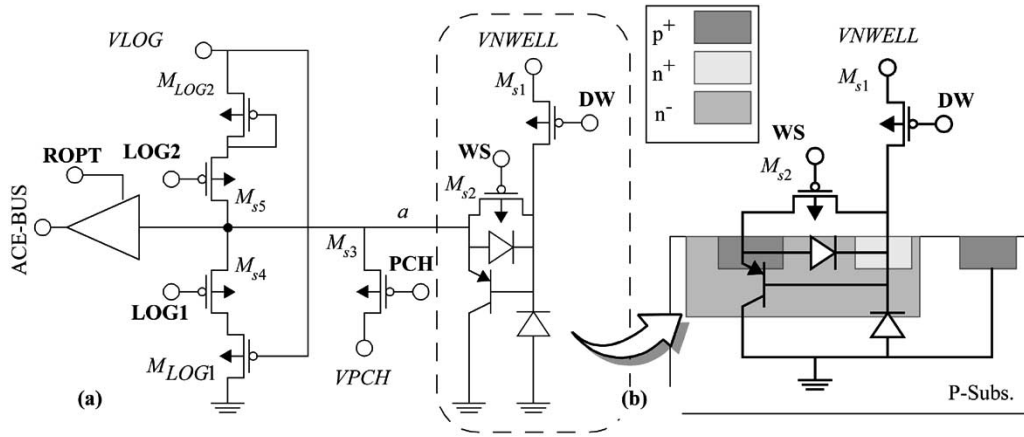
Fig. 10.   Multimode photosensor in ACE16k. (a) Equivalent schematic. (b) Cross section.
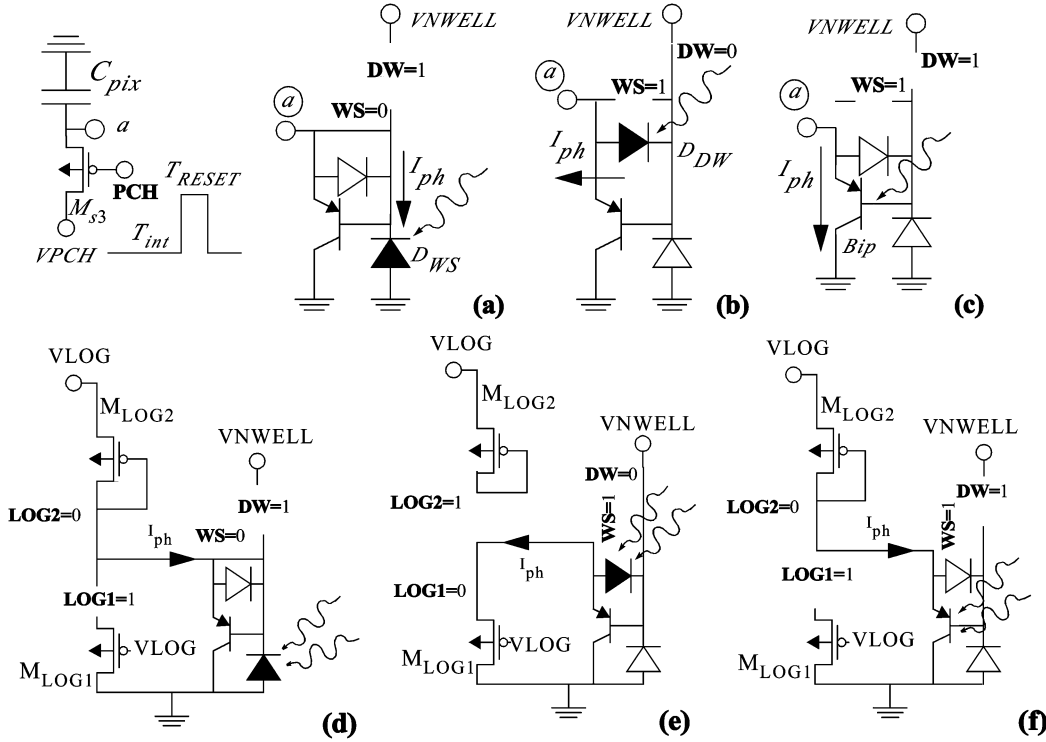

Fig. 11.   Available configurations for the optical sensor of ACE16k.

the integration modes, the sensing procedure is always carried out in the same way. First of all, $M_{s5,4}$ are turned off by making **LOG1 = LOG2 = 1**. Afterwards, switch M_{s3} precharges the internal node a to a user definable voltage *VPCH*. Finally, switch M_{s3} is turned off and the photo-generated current I_{ph} charges or discharges (depending on the selected photosensor) the pixel capacitor C_{pix}. Further details about the ACE16k sensor operation can be found elsewhere [13], [19].

Fig. 12 shows the global block diagram of the ACE16k-PU. There, the different building blocks can be identified. Control and configuration signals from the programming memory are in bold. A detailed description can be found in [13].

F. Cell Layout and Metal Distribution

The layout of the PU in ACE16k differs from that in ACE4k in various points.

- Metal 1 and metal 2 are used for internal routing, instead of just metal 1. This helps to increase cell density.
- As already mentioned, the last metal layer, metal 5, is employed for power and ground distribution. Therefore, power and ground lines can be as wide as almost half the cell height. This increases the quality of these signals: better uniformity across the array, less noise, lower probability of error during the fabrication, etc.

Fig. 12.   Functional block diagram of the cell.

- The existence of the ACE-BUS allows for a more organized layout. Generally speaking, the more involved the schematic, the more difficult the layout and the lower the cell density. Hence, the ACE-BUS concept contributes also to increase the cell density.
- A problem in the physical design of the ACE16k PU is related to the necessity to make a hole (in all the metal planes) to allow the light to reach the sensing area. This hole, located in the middle of the cell, just on top of the sensing area of the photosensor, reduces by three the number of available minimum width lines in each plane.
- As in ACE4k, digital instructions are sent to the cell by using a horizontal bus (metal 3) while weights are communicated by a vertical bus of metal 4 lines. Weights that

are connected to double-strength synapses are communicated through double-width lines.

Fig. 13 shows the layout and floor-planning of the PU in ACE16k. Cell size in ACE16k is 73.3×75.7 μm^2. It means a cell density of 180.2 cells $\cdot$ mm^{-2}. The width of the power and ground lines has been raised to almost 32 μm—about three times wider than in ACE4k.

IV. DISCUSSION

Applications of the ACE16k chip can be found in other papers of this special issue. Overall, these applications demonstrate that the chip is capable of operating with grayscale images at
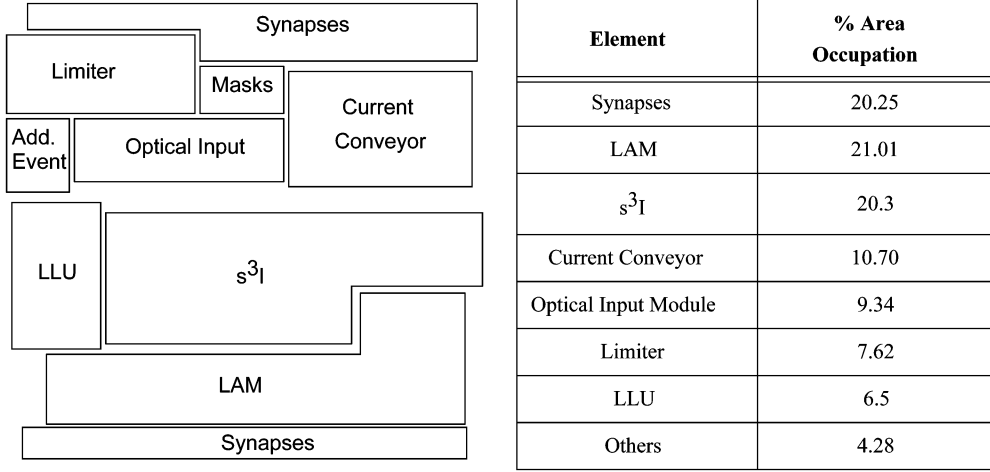
Element	% Area Occupation
Synapses	20.25
LAM	21.01
s^3I	20.3
Current Conveyor	10.70
Optical Input Module	9.34
Limiter	7.62
LLU	6.5
Others	4.28

Fig. 13.   Floor planning of the ACE16k PE and area occupation percentages.

TABLE III
FURTHER ACE16k VERSUS ACE4k COMPARISON.

Ref.	Tech. (micron)	Array	Pixel (format)	Memory	Speed(OPS)	Speed/Pow. (OP/Joule)	Photosensors
Further	0.5	64×64	Binary	4 Gray 4 Binary	47.5×10^9	39.5×10^9	Yes
This chip	0.35	128×128	Analog	8 Gray 4 Binary	330×10^9	82.5×10^9	Yes

frame-rates larger than 1000 FPS under room illumination conditions. This means a significant improvement as compared to other ACEXX chips which in turn have been shown to outperform other vision chips and architectures [13]. Further insight about the improvements yielded by ACE16k is provided by the data in Table III, where we have employed an equation which combines the number of operations (additions and products), the time constant of the process, and time constant units to keep settling errors below a given limit. In the particular case of linear convolutions

$$\text{OPS}_{\text{conv}} = \frac{(N_{\text{add}} + N_{\text{prod}}) \cdot N_c}{\tau_{\text{conv}} \cdot (n+1) \cdot Ln(2)} \tag{8}$$

where N_c is the number of elements in the array (128×128 in our case), N_{add} the number of additions (8 in 3×3 linear convolutions), N_{prod} the number of products (9 in $3 \times$ three linear convolutions), n the resolution for the settling error in an equivalent number of bits, and τ_{conv} the time constant of the process in (4), about 135 ns for the largest allowed b.

In summary, ACE chips, and specifically, the ACE16k prototype, are practical demonstration vehicles for the following statements.

- Sensory-processing concurrence is feasible with mixed-signal standard CMOS circuitry.
- Flexibility and programmability features can be incorporated by the smart synergy of analog and digital circuits.
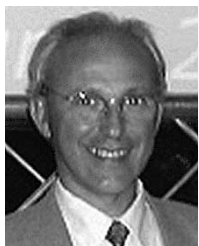
- Robustness can be achieved through proper analog design, and the use of calibration and error-correction techniques.
- Standard interfacing is a must which can be incorporated through embedded A-D and D-A converters.

These chips demonstrate that flexible analog early vision can be implemented in practice, and represent the first step toward the development of VSoCs. However, significant design challenges still have to be confronted to make true VSoCs capable of handling 10 000 Frames/s with moderate power consumption (below 1 W) and a large enough spatial resolution.

REFERENCES

[1] *The International Technology Roadmap for Semiconductors*, 2002. Update.
[2] CMOS Image Sensors at TSMC [Online]. Available: http://www.tsmc.com/english/technology/t0109.htm
[3] CMOS Image Sensors at UMC [Online]. Available: http://www.umc.com/english/process/m.asp
[4] 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 J. Solid-State Circuits*, vol. 32, pp. 1013–1026, July 1997.
[5] G. Liñán, S. Espejo, R. Domínguez-Castro, and A. Rodríguez-Vázquez, "ACE4k: An analog I/O 64×64 visual microprocessor chip with 7-bit analog accuracy," *Int. J. Circuit Theory Applicat.*, vol. 30, no. 2/3, pp. 89–116, 2002.
[6] G. Liñán-Cembrano, A. Rodríguez-Vázquez, L. Carranza, R. Domínguez-Castro, and S. Espejo-Meana, "A 1000 FPS@128×128 vision processor with 8-bit digitized I/O," in *Proc. 29th Eur. Solid-State Circuits Conf.*, Sept. 16–18, 2003, pp. 61–64.

[7] T. Roska and L. O. Chua, "The CNN universal machine: An analogic array computer," *IEEE Trans. Circuits Syst. II*, vol. 40, pp. 163–173, Mar. 1993.

[8] *Handbook of Computer Vision and Applications*, B. Jahne, H. Haubecker, and P. Geibler, Eds., Academic, London, U.K., 1999.

[9] R. H. Masland, "The fundamental plan of the retina," *Nature Neurosci.*, vol. 4, pp. 877–886, 2001.

[10] B. Roska and F. Werblin, "Vertical interactions across ten parallel, stacked representations in the mammalian retina," *Nature*, no. 410, pp. 583–587, Mar. 2001.

[11] P. Dudek, "A Programmable Focal-Plane Analogue Processor Array.," Ph.D. dissertation, University of Manchester Institute of Science and Technology, Manchester, U.K., May 2000.

[12] D. A. Martin, H. S. Lee, and I. Masaki, "A mixed signal array processor with early vision applications," *IEEE J. Solid State Circuits*, vol. 33, pp. 497–502, Mar. 1998.

[13] G. Liñán, "Design of low-power mixed-signal programmable vision Chips," Ph.D. dissertation, Univ. of Seville, Seville, Spain, Sept. 2002.

[14] R. Carmona, F. Jiménez-Garrido, R. Domínguez-Castro, S. Espejo, T. Roska, C. Rekeczky, and A. Rodríguez-Vázquez, "A bio-inspired two-layer mixed-signal flexible programmable chip for early vision," *IEEE Trans. Neural Networks*, vol. 14, pp. 1313–1336, Sept. 2003.

[15] J. B. Hughes and K. W. Moulding, "S^2I: A two-step approach to switched-currents," in *Proc. IEEE Int. Symp. Circuits and System*, May 1993, pp. 1235–1238.

[16] A. Rodríguez-Vázquez, R. Domínguez-Castro, F. Medeiro, and M. Delgado-Restituto, "High resolution CMOS current comparators: Design and applications to current-mode function generations," *Anal. Integr. Circuits Signal Processing*, vol. 7, no. 2, pp. 149–165, 1995.

[17] T. Roska, L. Kék, L. Nemes, Á. Zarándy, and M. Brendel, *CSL—CNN Software Library—Version 7.2*, Budapest, Hungary: Analogical and Neural Computing Laboratory, Computer and Automation Institute, Hungarian Academy of Sciences, 1998.

[18] A. Hastings, *The Art of Analog Layout*. Englewood Cliffs, NJ: Prentice-Hall, 2001.

[19] G. Liñán, A. Rodríguez-Vázquez, E. Roca, S. Espejo, and R. Domínguez-Castro, "A versatile sensor interface for programmable vision systems-on-chip," in *Proc. SPIE Conf. Electronic Imaging*, Santa Clara, CA, Jan. 2003.
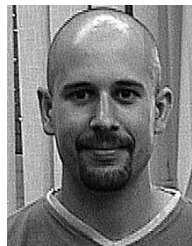
Gustavo Liñán-Cembrano received the Licenciado and Ph.D. degrees in physics, in the speciality of electronics, from the University of Seville, Seville, Spain, in June 1996 and Sept. 2002, respectively.

Since 2000, he has been an Assistant Professor in the Department of Electronics and Electromagnetism, School of Engineering, and the Faculty of Physics, University of Seville. His main areas of interest are the design and very large-scale integration implementation of massively parallel analog/mixed-signal image processors.

From 1997 to 1999, he was the recipient of a doctoral Grant of the Institute of Microelectronics of Seville, Centro Nacional de Microelectrónica (IMSE-CNM), Seville, Spain, funded by the Andalusian Government. He was the recipient of the "Best Paper Award 1999" from the *International Journal of Circuit Theory and Applications*. He was co-recipient of the "Most Original Project Award," of the "Salvà i Campillo Awards 2002," given by the Catalonian Association of Telecommunication Engineers.

L. Carranza received the B.S. degree in physics from the University of Seville, Seville, Spain, in 2000, and is working toward the Ph.D. degree in the Department of Electronics and Electromagnetism of the same university.

Since 2001, he has been with the Department of Analog Circuit Design, Institute of Microelectronics of Seville, Centro Nacional de Microelectrónica (IMSE-CNM), Seville, Spain. He research interests are in design of architectures and customized digital signal processors for vision system on chip.

Elisenda Roca-Moreno received the physics and the Ph.D. degrees from the University of Barcelona, Barecelona, Spain, in 1990 and 1995, respectively.

From November 1990 to April 1995, she worked at IMEC, Leuven, Belgium, working in the field of infrared detection aiming to obtain large arrays of CMOS compatible silicide Schottky diodes. Since 1995, she has been with Institute of Microelectronics of Seville, Centro Nacional de Microelectrónica (IMSE-CNM), Seville, Spain, where she holds the position of a Tenured Scientist. Her research interests are: design of CMOS analog focal-plane array processors for vision applications and CMOS imagers for the visible spectrum. She has been involved in several research projects with different institutions: Commission of the EU through the ESPRIT program, ESA, ONR-NICOP, etc. She has also co-authored more than 40 papers in international journals, books, and conference proceedings.

Angel Rodríguez-Vázquez (M'80–SM'95–F'96) received the Liceniado en físcia electrónica and the Ph.D. degrees from the University of Seville, Seville, Spain, in 1977 and 1983, respectively.

He is a Professor of Electronics in the Department of Electronics and Electromagnetism, University of Seville. He is also a Member of the Research Staff at the Institute of Microelectronics of Seville, Centro Nacional de Microelectrónica (IMSE-CNM), Seville, Spain, where he heads a research group on Analog and Mixed-Signal Integrated Circuits. His research interests are in the design of analog front-ends for mixed-signal circuits and systems-on-chip, telecom circuits, CMOS imagers and vision chips, sensory-processing-actuating systems-on-chip, and bio-inspired integrated circuits. On these topics, he has published seven books, 36 book chapters in other books, approximately 100 journal papers, and about 300 conference papers. He is also a member of the editorial staff of the *International Journal on Circuit Theory and Applications*, *Analog Integrated Circuits*, and *Signal Processing Journal*.

Dr. Rodríguez-Vázquez was co-recipient of the 1995 Guillemin–Cauer Award of the IEEE Circuits and Systems Society. In 1992, he received the Young Scientist Award of the Seville Academy of Science. In 1996, he was elected Fellow of the IEEE for contributions to the design and applications of analog/digital nonlinear ICs. He served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I: FUNDAMENTAL THEORY AND APPLICATIONS from 1993 to 1995. Currently, he is an Associate Editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II: EXPRESS BRIEFS. He was a Guest Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I: REGULAR PAPERS Special Issue on "Advances on Analog-to-Digital and Digital-to-Analog Converters."

Ricardo Carmona-Galán (M'99) received the degrees of Licenciado and Doctor (Ph.D.) in physics, in the speciality of electronics, from the University of Seville, Seville, Spain, in 1993 and 2002, respectively.

From 1994 to 1996, he worked at the National Center for Microelectronics, Seville, Spain, funded by an IBERDROLA S. A Grant. From July 1996 to June 1998, he was a Research Assistant in the Electronics Research Laboratory, Department of Electrical Engineering and Computer Sciences, University of California at Berkeley. He is currently with the Department of Analog Design, Institute of Microelectronics of Seville (IMSE), Centro Nacional de Microelectrónica (CNM-CSIC), Seville, Spain. Since October 1999, he is an Assistant Professor in the Department of Electronics and Electromagnetism, School of Engineering, University of Seville, where he teaches "Circuit Analysis and Synthesis" and "Circuit Synthesis Laboratory" for the degree of Telecommunication Engineer. His main areas of interest are linear and nonlinear analog and mixed-signal integrated circuits, in particular, the design and VLSI implementation of cellular neural networks and analog memory devices for real-time image processing and vision chips.

Francisco Jiménez-Garrido received the B.S. degree in physics and the B.S. degree in electronic engineering from the University of Seville, Seville, Spain, in 1998 and 2002, respectively, and is working toward the Ph.D. degree in the Department of Electronics and Electromagnetism of the same university.

Since 1999, he has been with the Department of Analog Circuit Design, Institute of Microelectronics of Seville, Centro Nacional de Microelectrónica (IMSE-CNM), University of Seville. He has research interests in linear and nonlinear analog and mixed-signal integrated circuits for image processing and communication devices.

Rafael Domínguez-Castro received the degree in electronic physics, the M.S. degree equivalent in microelectronics, and the Doctor en ciencias fisicas degree from the University of Seville, Seville, Spain, in 1987,1989, and 1993, respectively.

Since 1987, he has been with the Department of Electronics and Electromagnetism, University of Seville, where he is currently a Professor of Electronics. He is also a Member of the Research Staff, Institute of Microelectronics of Seville, Centro Nacional de Microelectrónica (IMSE-CNM-CSIC), University of Seville, where he is a Member of the Research Group on Analog and Mixed-Signal VLSI. His research interests are in the design of embedded analog interfaces for mixed-signal very large-scale integrated circuits, design of CMOS imagers and CMOS focal-plane array processors, and development on computer-aided design for automation of building blocks analog design, especially optimization and automatic sizing of basic building blocks for integrated circuits.

Dr. Domínguez-Castro is the co-recipient of the 1995 Guillemin–Cauer Award of the IEEE Circuits and Systems Society and the Best Paper award of the 1995 European Conference on Circuit Theory and Design.

Servando Espejo Meana received the licenciado en física degree, the M.S. degree equivalent in microelectronics, and the Doctor en ciencias físicas degree from the University of Seville, Seville, Spain, in 1987, 1989, and 1994, respectively.

He is currently a Professor of Electronics in the Department of Electronics and Electromagnetism, University of Seville, and also in the Department of Analog Circuit Design, Institute of Microelectronics of Seville, Centro Nacional de Microelectrónica (IMSE-CNM), University of Seville. From 1989 to 1991, he was an Intern at AT&T Bell Laboratories, Murray Hill, NJ, and an employee of AT&T Microelectronics, Madrid, Spain. His main areas of interest are linear and nonlinear analog and mixed-signal integrated circuits including neural networks electronic realizations and theory, vision chips, massively parallel analog array processing systems, chaotic circuits, and communication devices.

He is the co-recipient of the 1995 Guillemin–Cauer Award of the IEEE Circuits and Systems Society and the Best Paper award of the 1995 European Conference on Circuit Theory and Design.