A Programmable Vision Chip with High Speed Image Processing

Jérôme Dubois[a], Michel Paindavoine[b] and Dominique Ginhac[b]

[a]LCE CEA Saclay, 91190 Gif sur Yvette, Saclay, France;
[b]LE2I UMR CNRS 5158, 21078 Dijon Cedex, Dijon, France

ABSTRACT

A high speed Analog VLSI Image acquisition and pre-processing system is described in this paper. A 64×64 pixel retina is used to extract the magnitude and direction of spatial gradients from images. So, the sensor implements some low-level image processing in a massively parallel strategy in each pixel of the sensor. Spatial gradients, various convolutions as Sobel filter or Laplacian are described and implemented on the circuit. The retina implements in a massively parallel way, at pixel level, some various treatments based on a four-quadrants multipliers architecture. Each pixel includes a photodiode, an amplifier, two storage capacitors and an analog arithmetic unit. A maximal output frame rate of about 10 000 frames per second with only image acquisition and 2000 to 5000 frames per second with image processing is achieved in a 0.35 μm standard CMOS process. The retina provides address-event coded output on three asynchronous buses, one output is dedicated to the gradient and both other to the pixel values. A prototype based on this principle, has been designed. Simulation results from Mentor Graphics[TM]software and AustriaMicrosystem Design kit are presented.

Keywords: CMOS Image Sensor, Parallel architecture, High-speed image processing, Analog arithmetic unit.

1. INTRODUCTION

Today, improvements continue to be made in the growing digital imaging world with two main image sensor technologies: the charge coupled devices (CCD) and CMOS sensors. The continuous advances in CMOS technology for processors and DRAMs have made CMOS sensor arrays a viable alternative to the popular CCD sensors. New technologies provide the potential for integrating a significant amount of VLSI electronics onto a single chip, greatly reducing the cost, power consumption and size of the camera.[1–4] This advantage is especially important for implementing full image systems requiring significant processing such as digital cameras and computational sensors.[5] Most of the work on complex CMOS systems talks about the integration of sensors providing a processing unit at chip level or column level[6].[7] Indeed, pixel level processing is generally dismissed because pixel sizes are often too large to be of practical use. However, integrating a processing element at each pixel or group of neighbouring pixels seems to be promising. More significantly, employing a processing element per pixel offers the opportunity to achieve massively parallel computations. This benefits traditional high speed image capture[8910] and enables the implementation of several complex applications at standard video rates[11].[12]

Vision chips are designed based on the concept that analog VLSI systems with low precision are sufficient for implementing many low level vision algorithms. The precision in analog VLSI systems is affected by many factors, which are not usually controllable. As a result, if the algorithm does not account for these inaccuracies, the processing reliability may be severely affected. Vision chips implement a specific algorithm in a limited silicon area. They are always full custom designed which is a challenging task, known to be time consuming and error prone. One should consider issues from visual processing algorithms to low level circuit design problems, from photo transduction principles to high-level VLSI architectural issues.[13]

This paper describes the design and the implementation of a 64 × 64 active pixel sensor with per-pixel programmable processing element. Both the circuit design and layout are targeted for manufacturing in a

Send correspondence to J.D.

J.D.: E-mail: jerome1.dubois@cea.fr, Telephone: +33 (0)1 69 08 60 51

M.P.: E-mail: paindav@u-bourgogne.fr, Telephone: +33 (0)3 80 39 60 43

standard $0.35\,\mu$m double-poly quadruple-metal CMOS technology. The main objectives of our design are: (1) to evaluate the speed of the sensor, and, in particular, to reach a 10 000 frames/s rate, (2) to demonstrate a versatile and programmable processing unit at pixel level, (3) to provide a original platform dedicated to embedded image processing.

The rest of this paper is organized as follows: the section 2 describes the main characteristics of the overall architecture. The section 3 is dedicated to the description of the operational principle at pixel level in the sensor. In the following, the sections 4 and 5 respectively describe the details of the pixel design and the Analog Arithmetic Unit embedded in each pixel. Finally, some simulation results of high speed image acquisition with processing at pixel level are presented in the last section of this paper.

2. OVERALL ARCHITECTURE

The core circuit is, obviously, the bidimensional array of pixels. This array is organized into a cartesian arrangement of 64×64 pixels and contains 160 000 transistors on a $3.5\,\text{mm} \times 3.5\,\text{mm}$ die. The full layout of the retina is shown in Fig. 1 and the main chip characteristics are listed in Table 1. Each pixel contains a photodiode for the light-to-voltage transduction and 38 transistors integrating all the analog circuitry necessary to implement the algorithm described in section 3. This amount of electronics includes a preloading circuit, two 'Analog Memory, Amplifier and Multiplexer' ($[AM]^2$) and an 'Analog Arithmetic Unit' (A^2U) based on a four-quadrant multiplier architecture. The full pixel size is 35 μm $\times$ 35 μm with a 25 % fill factor.

The left part of the sensor is dedicated to a row decoder for adressing the successive rows of pixels. Below the chip core are the readout circuits with the three asynchronous buses described in section 4. The chip also contains test structures used for detailed characterization of the photodiodes and processing units. The test structures can be seen on the bottom left of the chip.
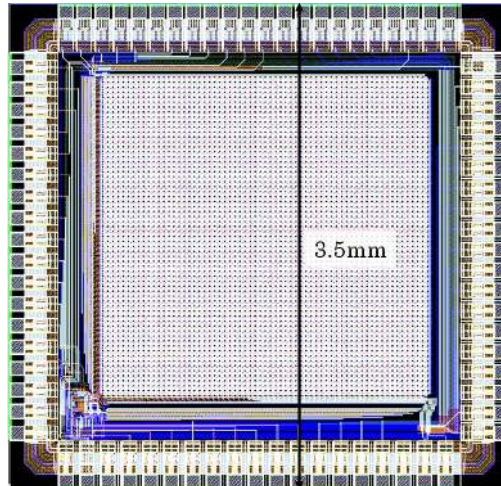

Figure 1. Layout of the Retina in a standard 0.35 μm CMOS technology

3. OPERATIONAL PRINCIPLE

3.1 Photodiode Structure

In a standard design, a pixel includes a photodiode (called *PD* in our chip) and a processing unit (implemented in the zone called *free surface*) as shown in Fig. 2(a). With the proposed approach, we focus on the optimization of the processing unit mapping. Each pixel integrates image processing based on neighborhood. So, each processing unit must easily have access to the nearest neighbors that's why an original structure was chosen as shown in Fig. 2(b). The major advantage of this structure is the possibility to limit the length of metal interconnection between adjacent pixels and the processing units, contributing to a better fill factor.

Table 1. Chip Characteristics

Technology	0.35μm 4-metal CMOS
Array size	64×64
Chip size	11 mm^2
Number of transistors	160 000
Number of transistors / pixel	38
Pixel size	35 μm $\times$ 35 μm
Sensor Fill Factor	25 %
Dynamics power consumption	250 mW
Supply voltage	3.3 V
Frame rate	10 000 fps

(a)   (b)
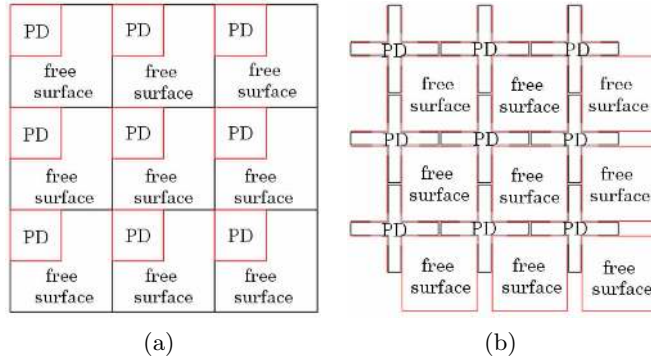
Figure 2. (a) Photosite classical structure, (b) Considered approach in our system

N-type photodiodes consist of a n$^+$-type diffusion in a p-type silicon substrate. During the preloaded period, the depletion region is formed in the neighborhood of photodiode cathode. However, considering the layout constraints, choosing a cross shape for the photodiode appears to be not realistic. Finally, a quasi-octagonal structure, shown in Fig. 3, was selected because of three main properties:

1. The reserved surface dedicated to the interconnections is about 12% lower compared to a square shape,

2. The depletion region is more efficient at the edges of the photodiode

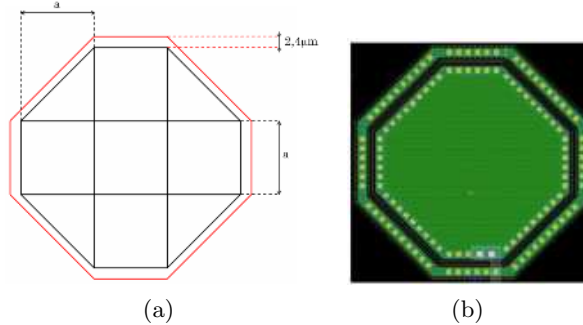3. This shape, based on 45° structures, is technologically realizable by the founder.

(a)   (b)

Figure 3. (a) Photodiode shape, (b) Photodiode layout

3.2 Spatial Gradients

The structure of our processing unit is perfectly adapted to the computation of spatial gradients.
The main idea for evaluating these gradients[14] in-situ is based on the definition of the first-order derivative of a 2-D function performed in the vector direction $\vec{\xi}$, which can be expressed as:

$$\frac{\partial V(x,y)}{\partial \vec{\xi}} = \frac{\partial V(x,y)}{\partial x'} \, cos(\beta) + \frac{\partial V(x,y)}{\partial y'} \, cos(\beta) \tag{1}$$

where β is the vector's angle. A discretization of the equation 1 at the pixel level, according to the Fig. 5, would be given by:

$$\frac{\partial V_5}{\partial \vec{\xi}} = (V_1 - V_3) \, cos(\beta) + (V_2 - V_4) \, sin(\beta) \tag{2}$$

where V_i, $i \in \{1;4\}$ is the luminance at the pixel i, *i.e.*, the photodiode output. In this way, the local derivative in the direction of vector $\vec{\xi}$ is continuously computed as a linear combination of two basis functions, the derivatives in the x' and y' directions.
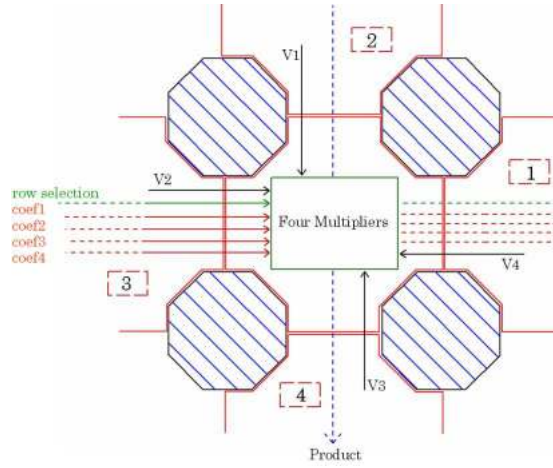

Figure 4. Implementation of four multipliers

Using a four-quadrant multiplier, shown in Fig. 4, the product of the derivatives by a circular function in cosines can be easily computed:

$$P = V_1 \, cos(\beta) + V_2 \, sin(\beta) - V_3 \, sin(\beta) - V_4 \, cos(\beta) \tag{3}$$

According to the Fig. 5, the processing element implemented at the pixel level carries out a linear combination of the four adjacent pixels by the four associated weights ($coef_i$, $i \in \{1;4\}$). To evaluate the equation 3, the following values have to be given to the coefficients:

$$\begin{pmatrix} coef1 & coef2 \\ coef3 & coef4 \end{pmatrix} = \begin{pmatrix} sin(\beta) & cos(\beta) \\ -sin(\beta) & -cos(\beta) \end{pmatrix} \tag{4}$$

3.3 Sobel operator

The structure of our architecture is also adapted to various algorithms based on convolutions using binary masks on a neighborhood of pixels. As example, the Sobel operator is described here. With our chip, the evaluation of the Sobel algorithm leads to the result directly centered on the photosensor and directed along the natural axes of the image according to the Fig. 5. In order to compute the mathematical operation, a 3×3 neighborhood
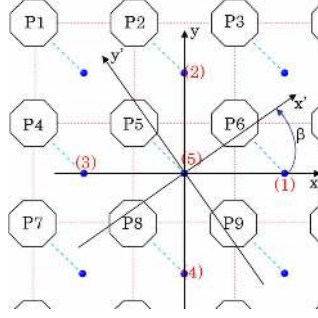
Figure 5. Array architecture

is applied on the whole image. To carry out the derived operation discretized in two dimensions, along the horizontal and vertical axes, it is necessary to build two 3×3 matrices called h1 and h2:

$$h_1 = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \quad h_2 = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \tag{5}$$

Within the four processing elements numbered from 1 to 4 (see Fig. 5), two 2x2 masks are locally applied. According to the equation 5, this allows the evaluation of the following series of operations:

$$h_1 : \begin{matrix} I_{11} = I_3 + 2I_6 + I_9 \\ I_{13} = I_1 + 2I_4 + I_7 \end{matrix} \quad h_2 : \begin{matrix} I_{22} = I_1 + 2I_2 + I_3 \\ I_{24} = I_7 + 2I_8 + I_9 \end{matrix} \tag{6}$$

with the I_{1k} and I_{2k} provided by the processing element (k). Then, from these trivial operations, the discrete amplitudes of the derivatives along the vertical axis ($I_{h1}=I_{11}-I_{13}$) and the horizontal axis ($I_{h2}=-I_{22}+I_{24}$) can be computed. All these operations can be carried out in two cycle's retina. We call cycle's retina, all operations of treatments carried out at the end of the screen of acquisition.

3.4 Second-order detector: Laplacian

Edge detection based on some second-order derivatives as the Laplacian can also be implemented on our architecture. Unlike spatial gradients previously described, the Laplacian is a scalar (see equation 7) and does not provide any indication about the edge direction.

$$\delta = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix} \tag{7}$$

From this 3x3 mask, the following operation can be extracted according to the principles used previously for the evaluation of the Sobel operator; $\delta = I_1 + I_2 + I_3 + I_4 - 4.I_5$. This operation can be carried out in only one cycle's retina.

4. ARCHITECTURE AND OPERATION

The proposed chip is based on a classical architecture widely used in the literature as shown on the Fig. 6. The CMOS image sensor consists of an array of pixels that are typically selected a row at a time by a row decoder. The pixels are read out to vertical column buses that connect the selected row of pixels to an output multiplexer. The chip includes three asynchronous output buses, the first one is dedicated to the image processing results whereas the two others provides parallel outputs for full high rate acquisition of raw images.

As a classical APS, all reset transistor gates are connected in parallel, so that the whole array is reset when is active the reset line. In order to supervise the integration period (*i.e.*, the time when incident light on each pixel generates a photocurrent that is integrated and stored as a voltage in the photodiode), the global output
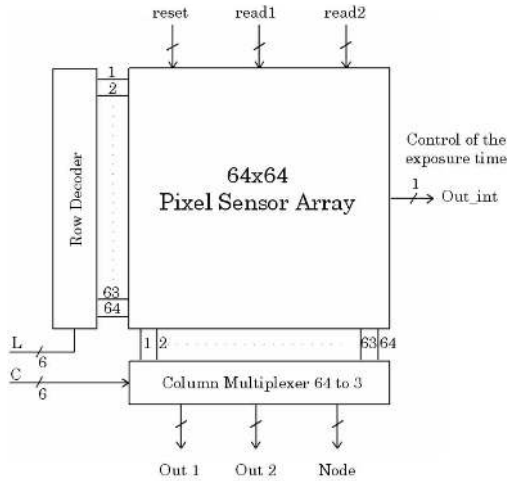
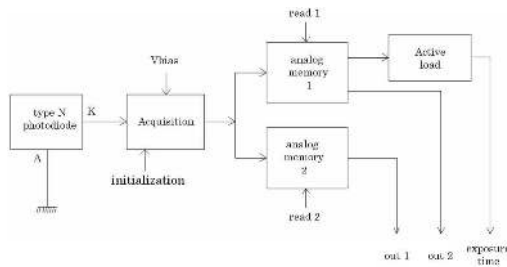Figure 6. Image sensor system architecture

Figure 7. Block diagram of reading structure

called Out_int provides the average incidental illumination of the whole matrix of pixels. So, if the average level of the image is too low, the exposure time may be increased. On the contrary, if the scene is too luminous, the integration period may be reduced.

To increase the algorithmic possibilities of the architecture, the acquisition of the light inside the photodiode and the readout of the stored value at pixel level are dissociated and can be independently executed. So, two specific circuits, including an analog memory, an amplifier and a multiplexer are implemented at pixel level as shown in Fig. 7.
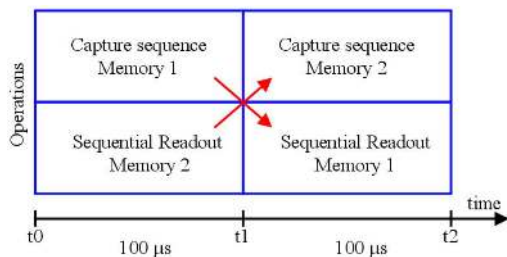

Figure 8. Parallelism between capture sequence and readout sequence

With these circuits called $[AM]^2$ (Analog Memory, Amplifier and Multiplexer), the capture sequence can be made in the first memory in parallel with a readout sequence and/or processing sequence of the previous image stored in the second memory (see Fig. 8). With this strategy, the frame rate can be increased without reducing the exposure time. Indeed, it is possible to double the speed of acquisition. Simulations of the chip show that frame rates up to 10 000 frames per second can be achieved with a brightness superior to 15 000 luxes.

The chip operates at a single 3.3 V power supply. In each pixel, as seen on Fig. 9, the photosensor is a NMOS photodiode associated with a PMOS transistor reset, which represents the first stage of the capture circuit. The
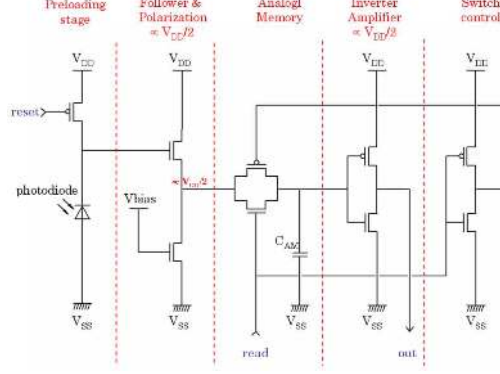
Figure 9. Frame capture circuit schematic with one analog memory

pixel array is held in a reset state until the 'reset' signal goes high. Then, the photodiode discharges according to incidental luminous flow. This signal is polarized around of $V_{DD}/2$, so the half power supply voltage. While the "read" signal remains high, the analog switch is open and the capacitor C_{AM} of the analog memory stores the pixel value. The C_{AM} capacitors are able to store, during the frame capture, the pixel values, either from the switch 1 or the switch 2. The following inverter, polarized on $V_{DD}/2$, serves as an amplifier of the stored value and provides a level of potential proportional to the incidental pixel illumination.
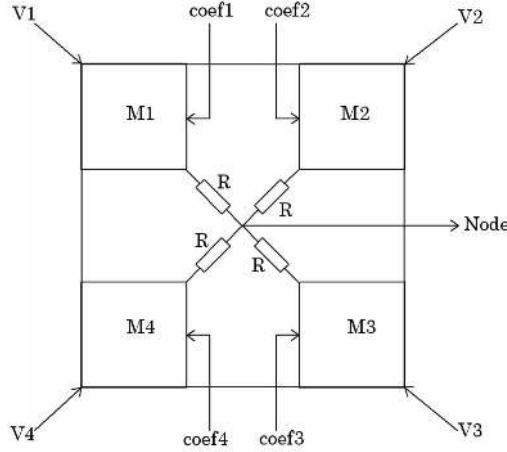

Figure 10. Architecture of the A^{2}U

5. ANALOG ARITHMETIC UNIT: A^{2}U

The analog arithmetic unit (A^{2}U) represents the central part of the pixel and includes four multipliers (M1, M2, M3 and M4), as illustrated on the Fig. 10 and 11. The four multipliers are all interconnected with a diode-connected load (*i.e.*, a NMOS transistor with gate connected to drain). The operation result at the 'node' point is a linear combination of the four adjacent pixels.

Fig. 12 and 13 show the simulation results of this multiplier structure with cosine signals as inputs, *i.e.*

$$coef_i = A.cos(2\pi f_1) \; with \; f_1 = 2.5kHz \tag{8}$$

$$V_i = B.cos(2\pi f_2) \; with \; f_2 = 20kHz \tag{9}$$

In this case, the output Node value can be written as following:

$$Node = \frac{A.B}{2}[cos(2\pi(f_2 - f_1)) + cos(2\pi(f_2 + f_1))] \tag{10}$$

MOS	W	L
P1	1.5	0.35
N1	0.4	0.35
N2	0.4	0.35
N3	0.4	0.35

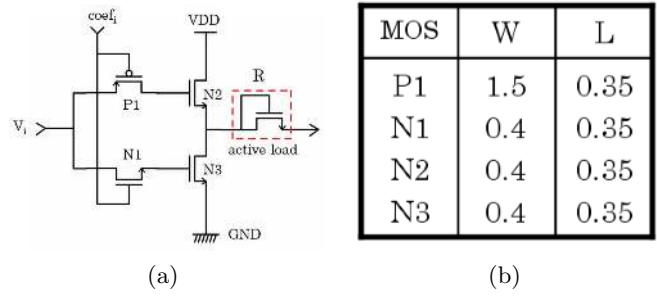(a)   (b)

Figure 11. (a) Multiplier schematic, (b) Transistor sizes in μm

The signal's spectrum, represented on Fig. 13 contains two frequencies (17.5 kHz and 22.5 kHz) around the carrier frequency. The residues which appear in the spectrum are known as 'inter modulation products'. They are mainly due to the nonlinearity of the structure (around 10 kHz and 30 kHz) and the defects in input pads insulation (at 40 kHz). However, the amplitude of these inter modulations products are significantly lower than the two main frequencies.
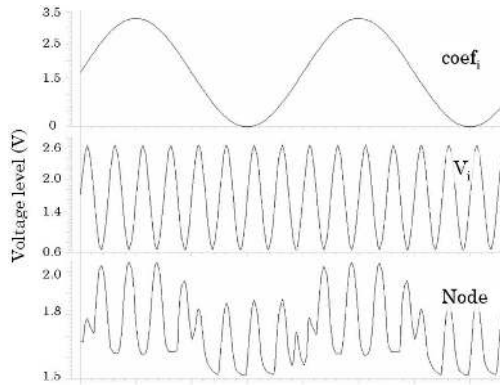

Figure 12. Multiplier of two cosine signals

Furthermore, in order to obtain the best linearity of the multiplier, the amplitude of the signal V_i is limited to a range of 0.6-2.6 V. In the full chip, the signal V_i corresponds to the voltage coming from the pixel and can be easily included in the range described before.
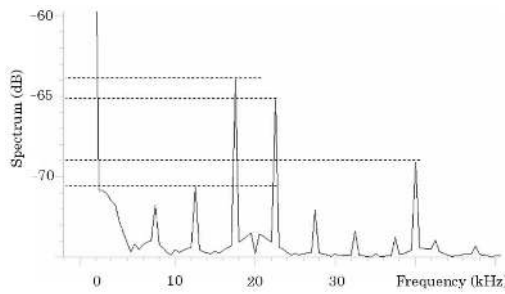

Figure 13. Output Node spectrum in dB

6. LAYOUT AND PRELIMINARY RESULTS

The layout of a 2×2 pixel block is shown in Fig. 14. This layout is symmetrically built in order to reduce fixed pattern noise among the four pixels and to ensure uniform spatial sampling. An experimental 64×64 pixel image sensor has been developed in a 0.35μm, 3.3 V, standard CMOS technology with poly-poly capacitors.
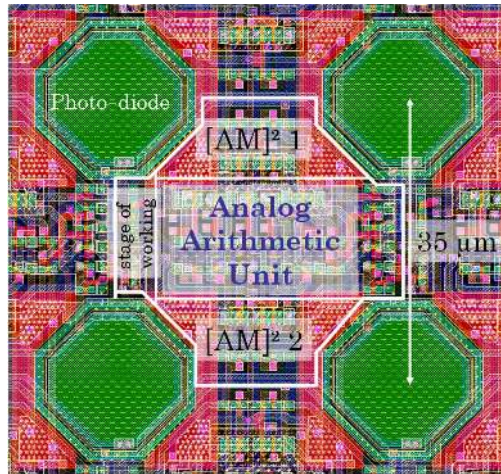
Figure 14. Layout of a pixel

This prototype has been sent to foundry at the beginning of 2006 and will be available at the end of the second quarter of the year. The Fig. 15 represents a simulation of the capture operation. Various levels of illumination are simulated by activating different readout signals (read 1 and read 2). The two outputs (output 1 and output 2) give the levels between GND and V_{DD}, corresponding to incidental illumination on the pixels. The calibration of the structure is ensured by the biasing (Vbias=1,35V). Moreover, in this simulation, the 'node' output is the result of the difference operation (out1-out2). The factors were fixed at the following values: coef1=-coef2=V_{DD} and coef3= coef4=V_{DD}/2. MOS transistors operate in sub-threshold region. There is no energy spent for transferring information from one level of processing to another level. According to the simulation's results, the voltage gain of the amplifier stage of the two [AM]^2 is Av=10 and the disparities on the output levels are about 4.3 %.
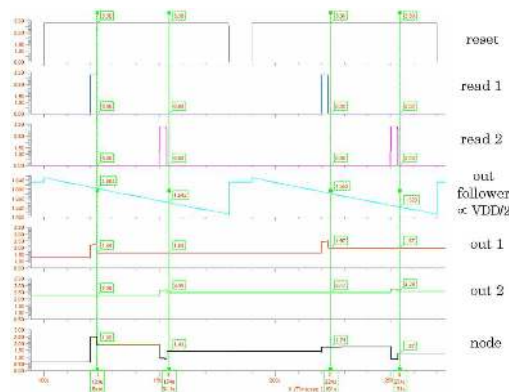

Figure 15. High speed sequence capture with basic processing

7. CONCLUSION

An experimental pixel sensor implemented in a standard digital CMOS $0.35\,\mu$m process was described. Each 35μm$\times35\mu$m pixel contains 38 transistors implementing a circuit with photo-current's integration, two [AM]^2 (Analog Memory, Amplifier and Multiplexer), and a A^{2}U (Analog Arithmetic Unit).

Chip simulations reveal that raw images acquisition at 10 000 frames per second can be easily achieved using the parallel A^{2}U implemented at pixel level. With basic image processing, the maximal frame rate slows to reach about 5000 fps.

The next step in our research will be the characterization of the real circuit as soon as the chip comes back from the foundry. Furthermore, we focus on the development of a fast analog to digital converter (ADC). The integration of this ADC on future chips will allow us to provide new and sophisticated vision systems on chip dedicated to digital embedded image processing at thousands of frames per second.

REFERENCES

[1] Fossum, E. R., "Active pixel sensors: Are CCDs dinosaurs?," *International Society for Optical Engineering (SPIE)* **1900**, 2–14 (1993).

[2] Fossum, E. R., "CMOS image sensor : Electronic camera on a CHIP," *IEEE Transactions on Electron Devices* **44**, 1689–1698 (October 1997).

[3] Litwiller, D., "Ccd vs. cmos: Facts and fiction," *Photonics Spectra* , 154–158 (January 2001).

[4] Seitz, P., "Solid-state image sensing," *Handbook of computer Vision and Applications* **1**, 165–222 (2000).

[5] El Gamal, A., Yang, D., and Fowler, B., "Pixel level processing – Why, What and How?," in [*Proceedings of the SPIE Electronic Imaging '99 conference*], **3650**, 2–13 (January 1999).

[6] Loinaz, M., Singh, K., Blanksby, A., Inglis, D., Azadet, K., and Ackland, B., "A 200mv 3.3v cmos color camera ic producing 352x288 24b video at 30 frames/s," *IEEE Journal of Solid-State Circuits* **33**, 2092–2103 (1998).

[7] Smith, S., Hurwitz, J., Torrie, M., Baxter, D., Holmes, A., Panaghiston, M., Henderson, R., Murrayn, A., Anderson, S., and Denyer, P., "A single-chip 306x244-pixel cmos ntsc video camera," in [*In ISSCC Digest of technical papers*], 170–171 (1998).

[8] Coaker, B., Xu, N., Latham, R., and Jones, F., "High-speed imaging of pulsed-field flashover of an alumina ceramic in vacuum," *IEEE Transactions on Dielectrics and Electrical Insulation* **2**, 210–217 (April 1995).

[9] Turko, B. and Fardo, M., "High-speed imaging with a tapped solid stat sensor," *IEEE Transactions on Nuclear Science* **37**, 320–325 (1990).

[10] Kleinfelder, S., Lim, S., Liu, X., and Gamal, A. E., "A 10 000 frames/s CMOS digital pixel sensor," *IEEE Journal of Solid-State Circuits* **36**, 2049–2059 (December 2001).

[11] Yang, D., El Gamal, A., Fowler, B., and Tian, H., "A 640 x 512 CMOS image sensor with ultra wide dynamix range floating-point pixel-level ADC," *IEEE Journal of Solid-State Circuits* **34**, 1821–1834 (December 1999).

[12] Lim, S. and El Gamal, A., "Integrating image capture and processing – beyond single chip digital camera," in [*Proceedings of the SPIE Electronic Imaging '2001 conference*], **4306** (january 2001).

[13] Moini, A., [*Vision Chips*], Kluwer Academic Publishers, Norwell, MA, USA (1999).

[14] Barbaro, M., Burgi, P., Mortara, A., Nussbaum, P., and Heitge, F., "A 100x100 pixel silicon retina for gradient extraction with steering filter capabilities and temporal output coding," *IEEE Journal of Solid-State Circuits* **37** (February 2002).