

FAPS: A CMOS Sensor with multiple storage for fast imaging scientific applications

R. Turchetta, W. Gannon, A. Evans

Rutherford Appleton Laboratory, Chilton, Didcot, OX11 0QX,

United Kingdom, tel:+44.1235.446633, fax: +44.1235.445753, r.turchetta@rl.ac.uk

1 Introduction

This paper describes the design of a novel CMOS sensor with multiple storage capability.

In many applications the required frame rate is in excess of the available readout bandwidth, i.e. images must be acquired much faster than the maximum achievable readout speed. Examples can be found in high-speed video imaging for scientific and engineering applications [1]. Another example can be found in particle physics, where the aim is to detect charged particles instead of photons. Specifically, the driving application for this architecture is the design of a sensor for the future Linear Collider, an electron-positron accelerator whose construction is foreseen in the first half of the next decade.

We envisage the use of CMOS sensors for the vertex detector, one of the elements of the particle detector for the Linear Collider. The vertex detector is the closest detector to the collision point. Roughly, it has the shape of a cylinder, with the axis oriented as the axis of the colliding beams and its centre coinciding with the collision point. More specifically, it consists of several layers of sensors, situated at different distances from the interaction point. Each layer is comprised of several modules, or ladders, positioned in such a way as to assure high solid angle coverage (see for example [2]). At present the parameters for the vertex detectors are as indicated in table 1. Although not yet fixed, these parameters should be very close to the final numbers. It is worth noticing that the total number of pixels in the vertex detector is about 800 millions. It should also be noted that the

readout of the sensor should be done only on the edge at the end of the long side.

At least a couple of options for the design of the machine exist, but whatever the choice will be the need for a high rate of collisions between electrons and positrons sets stringent constraints on the speed of the sensors. In the so-called NLC/JLC design, collisions take place during 266 ns periods spaced by 8.3 ms. In the so-called TESLA design, electron/positron collisions take place almost continuously during 0.95 ms periods regularly spaced by 200 ms.

In the TESLA design, it is necessary to read each pixel several times during the collision periods to avoid an excessive accumulation of events in a single readout. Simulations show that on layer 1 each pixel should be read at least every 50 μ s, if not quicker. Since there are 1.625 million pixels in each sensor on layer 1, this would need an average readout speed of 325 Gpixel/sec, well in excess of the capability of present CMOS APSs. These requirements motivated the design described in the following sections. We call it Flexible Active Pixel Sensor, or FAPS, because, although specifically optimised for this particle physics application, this design could be of interest for any high-speed video imaging for scientific and engineering applications.

2 The Flexible Active Pixel Sensor (FAPS) architecture

The main concept behind the development of the FAPS is described in figure 1. On the left (figure 1a), is the pixel design of a standard APS. The unity gain buffer is dashed to indicate that it is only active during the readout. On the

right, is the FAPS architecture, which has two amplifiers integrated into the pixel. The first one, the write amplifier, has a gain A and can be powered at any time, in particular during the data taking. It can thus write the diode voltage into the bank of memory cells located within the pixel. The readout of the cells is performed later, by the read amplifier in the same way as in a standard APS. The detailed design of the FAPS, including the choice of the architecture for the two amplifiers or the size and number of the memory cells, depends on the application.

In order to prove on silicon the principle of the architecture, a test structure was designed in a CIS (CMOS Image Sensor) 0.25 μm process. This process is characterised by low leakage current, dual 2.5/3.3 V power supply for optimum analogue performances, up to 5 metal layers for routing (we only use 3), and precise metal-to-metal capacitors. The test structure contains several arrays with different pixel architectures, including standard 3 and 4 MOS pixels. Five arrays of 40x40 pixels at 20 μm pitch contain various designs of the FAPS architecture (figure 2). Each pixel contains 10 memory cells. The pitch is 20 μm , which is the baseline choice for the Linear Collider (see table 1). This application also requires 100 % efficiency for the detection of charged particles. This imposes the use of a N-well/P-epi diode as a detecting element [3], and of only NMOS transistors in the pixel design. This last constraint severely limits the choice of architectures that can be implemented in the pixel.

In the present test structure (figure 3), the write amplifier is a simple NMOS inverter with active, diode-mounted NMOS load. This architecture provides a DC gain A given by the ratios

$$A = \frac{g_{m,MIN}}{g_{m,MBI} + g_{mb,MBI}}$$

where g_m and g_{mb} indicates the transconductance and the bulk transconductance respectively.

The two transistors MW1 and MW2 act like switches controlled by the same WRITE pulse.

When WRITE goes high, the two switches are closed and the first (or 'write') amplifier is biased. When WRITE goes low, the two switches are open and no current flows in the amplifier. Two switches, instead of one are needed to effectively isolate the out_write node.

Each memory cell consists of an NMOS switch and a storage capacitor. Each SAMPLE switch is individually driven. When it goes high, the top plate of the storage capacitors is connected to the node out_write and the output voltage of the write amplifier can be written into the capacitors. Once the SAMPLE switch goes low, the capacitor is disconnected and holds the voltage. During the readout of the pixel, WRITE is low and READ is high. When SAMPLE goes high, the charge stored in the capacitors is transferred to the out_write node and generates the gate voltage of MIR.

The timing of the two phases, write and read, is shown in figures 4 and 5. In figure 4, the WRITE signal can follow either the dotted or full line. In this latter case, the write amplifier is powered only during the time needed to write a sample, thus reducing the overall power consumption.

During the writing phase, the voltage v_{sample} written onto the capacitance is given by

$$v_{\text{sample}} = \frac{Q_{\text{in}}}{C_{\text{in}}} * A$$

where Q_{in} is the charge at the input diode. It corresponds to a stored charge

$$Q_{\text{sample}} = C_{\text{mem}} * v_{\text{sample}} = A * \frac{C_{\text{mem}}}{C_{\text{in}}} * Q_{\text{in}} = G_Q * Q_{\text{in}}$$

where the charge-to-charge gain G_Q is

$$G_Q = A * \frac{C_{\text{mem}}}{C_{\text{in}}}$$

During the readout, the charge stored in the memory is dumped into the node out_write node, resulting in a voltage

$$V_{\text{out_write}} = \frac{Q_{\text{sample}}}{C_{\text{mem}} + C_{\text{stray}}} = \frac{C_{\text{mem}}}{C_{\text{mem}} + C_{\text{stray}}} * v_{\text{sample}} =$$

$$= \frac{G_Q}{C_{\text{mem}} + C_{\text{stray}}} * Q_{\text{in}}$$

and, if G_{SF} is the gain of the read source follower, the column output voltage is

$$V_{\text{out}} = \frac{G_Q * G_{\text{SF}}}{C_{\text{mem}} + C_{\text{stray}}} * Q_{\text{in}} =$$

$$= \left(1 - \frac{C_{\text{stray}}}{C_{\text{mem}} + C_{\text{stray}}} \right) * A * \frac{G_{\text{SF}}}{C_{\text{in}}} * Q_{\text{in}} \approx$$

$$\approx A * \frac{G_{\text{SF}}}{C_{\text{in}}} * Q_{\text{in}}$$

where the last equality holds in the case $C_{\text{mem}} \gg C_{\text{stray}}$. Choosing a relatively large size of C_{mem} is also beneficial in reducing the effect of any leakage on the storage capacitors, since the input referred effect of the leakage will be reduced by the factor G_Q . It will also reduce any mismatch effect. The leakage current on the storage capacitors will appear as a DC shift as well as introducing an additional shot noise component.

From the point of view of the DC behaviour, the output voltage $V_{\text{out_write}}$ of the write amplifier is limited on one side by $V_{\text{DSsat,MIN}}$ and on the other side by the largest V_T of any of the transistors MBI, MW1 and MW2. The read source follower introduces a further voltage shift, given by the V_T of transistor MIR.

The write amplifier is well approximated by a 1st-order low-pass filter, with a cut-off frequency given by

$$f_0 = \frac{1}{2\pi} \frac{g_{m,\text{MBI}} + g_{m,\text{b,MBI}}}{C_{\text{out_write}}}$$

where $C_{\text{out_write}}$ is the total capacitance at the node out_write, including any stray capacitance and the capacitance of the connected memory cells.

Considering the write phase, the temporal noise comes from the contribution of transistors MBI and MIN. Considering only the thermal noise,

the variance of the noise voltage sampled on the storage capacitors is

$$\langle v^2 \rangle_{\text{MIN}} = \gamma kT \frac{A}{C_{\text{out_write}}}$$

$$\langle v^2 \rangle_{\text{MBI}} = \gamma kT \frac{A}{C_{\text{out_write}}} \frac{g_{m,\text{MBI}} + g_{m,\text{b,MBI}}}{g_{m,\text{MIN}}}$$

respectively for transistors MIN and MBI.

The sampling leaves on the capacitor also a noise sample given by

$$\langle v^2 \rangle_{\text{MEM}} = kT \frac{1}{C_{\text{mem}}}$$

The reset noise due to the reset of the write amplifier can be eliminated by the correlated double sampling, however this will multiply the thermal noise as well as the sampling noise by $\sqrt{2}$. During the readout, the source follower will introduce an additional noise contribution (see [3] for example). Summing up all noise contributions the simulated noise is $15 e^-$ rms.

3 Floorplan of the test sensor and simulation results

The floorplan of the test sensor is shown in figure 6. The sensor was divided into 4 regions in the direction of the column readout, so that in each column only pixels with the same type of design are found. However, slight variations of the designs are found along each column. In the first 3 regions standard pixels with 3 and 4 MOS and pixels with charge amplification were designed and in the last regions, the FAPS pixels were designed. Row and columns Gray decoders provide the addressing of the selected pixels. The signals used to control the pixels are generated externally, so that the maximum flexibility of the timing and amplitude of the signals can be obtained. The column amplifiers are simple source followers and, in each column, two sets of hold capacitors and source followers are integrated so that the reset voltage can be read differentially with the signal voltage.

Figure 7 shows the simulation results on one pixel, after parasitic extraction on the layout. In this simulation, the voltage on the photodiode

was sampled every 100 ns, written onto the storage capacitors, and subsequently read out at 1 MHz. A charge amount of 1000 e⁻ was injected into the photodiode every 200 ns. As shown by the top, right-hand part of the figure, voltage steps appeared at the output every two samples as expected.

Figure 8 shows the output voltage as a function of the input signal. Good linearity is obtained up to and beyond 30,000 electrons. The main parameters of the FAPS pixels in the prototype sensors are found in table 2.

4 Conclusion

In this paper we presented the design of a novel pixel architecture for CMOS sensors. The main feature of this architecture is the ability to sample and store multiple samples in the pixels. This is achieved by integrating a write amplifier which can be biased in parallel into every pixel. A test sensor has been designed in a CIS 0.25 μm technology. The testing will start soon.

This sensor was specifically optimised for the detection of charged particles in a vertex detector at the future Linear Collider. In this case, multimillion pixel detectors need to be read out in less than 50 μs, a speed which is out of reach of present CMOS technology. We believe that the architecture presented here provides an efficient solution to this problem. This architecture can also be used for the detection of very fast phenomena using any type of ionising radiation, including visible light. The number of cells that can be integrated into a pixel is a function of the size of the pixel as well as of the required analogue performance. It also depends on the choice of a specific CMOS technology and we anticipate that, with the continuous shrinkage of minimum feature size, it will be possible to integrate a larger number of cells in a smaller area. This technology could provide an alternative choice to existing technologies for the recording of ultra-fast phenomena.

5 Acknowledgements

This work has been funded by the CCLRC Centre for Instrumentation CfI. A. Evans benefits from a CASE award from PPARC.

6 References

- [1] T. Goji Etoh et al., *An image sensor which captures 100 consecutive Frames at 1,000,000 frames/s*, IEEE Trans. on Electron Devices, vol. 50, no. 1, January 2003, 144-151
- [2] TESLA, *The Superconducting Electron-Positron Linear Collider with an Integrated X-Ray Laser Laboratory*, Technical Design Report, http://tesla.desy.de/new_pages/TDR_CD/start.html
- [3] R.Turchetta et al., *A Monolithic Active Pixel Sensor for Charged Particle Tracking and Imaging using Standard VLSI CMOS Technology*, in Nucl. Instr. And Meth A 458 (2001) 677-689

Layer	Distance from collision point mm	Ladders / layer	Sensor size LxW mm	Number of pixels per sensors
1	15	8	50x13	2500x650
2	26	8	125x22	6250x1100
3	37	12	125x22	6250x1100
4	48	16	125x22	6250x1100
5	60	20	125x22	6250x1100

Table 1. The pitch of the sensors is 20 μm.

Parameters	Value	Unit
Number of arrays	5	
Number of pixels per array	40x40 = 1600	
Number of memory cells per pixel	10	
Pitch	20	μm
Number of transistors per pixel	38	NMOS FET
Fill factor	100 %	For charged particle
Fill factor	16 %	For visible light
ENC	15	e ⁻ rms (simulated)
Full well capacity (linear)	30,000	e ⁻ (simulated)
Gain during write phase	32	μV/e ⁻ (simulated)
Total gain	20	μV/e ⁻ (simulated)

Table 2. Main parameters of the test sensor.

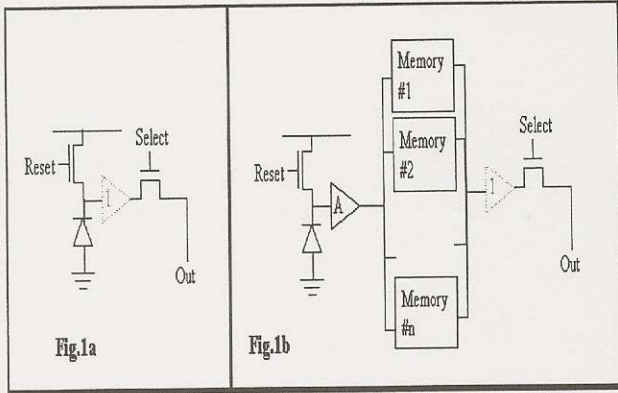


Figure 1. Fig 1a represents a standard APS pixel, while fig. 1b represents the FAPS concept described in this paper.

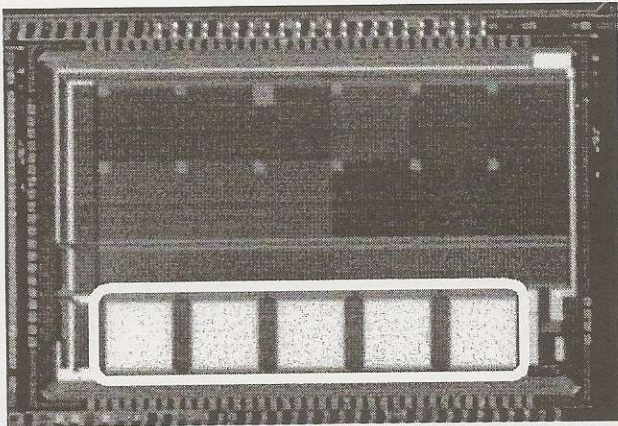


Figure 2. Photo of the test structures designed in a 0.25 μm CIS process. In the bottom part of the photo, the white shape highlights the areas where the five FAPS arrays are located.

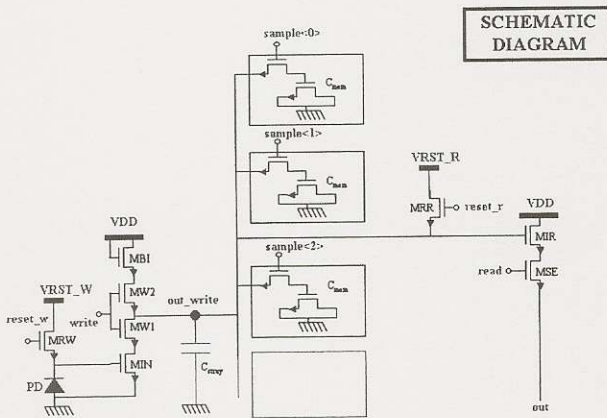


Figure 3. Schematic of the FAPS pixel implemented in the present test structure.

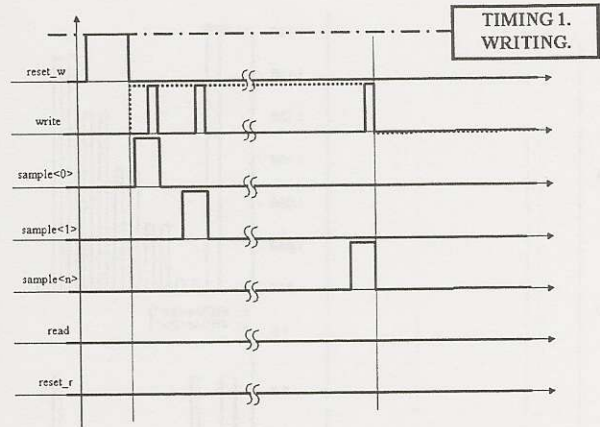


Figure 4. Timing of the writing phase.

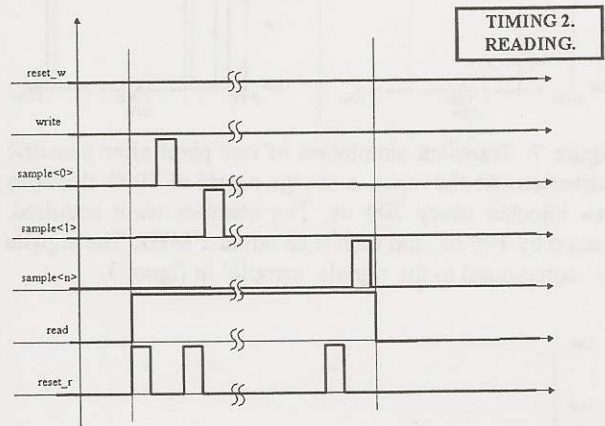


Figure 5. Timing of the reading phase.

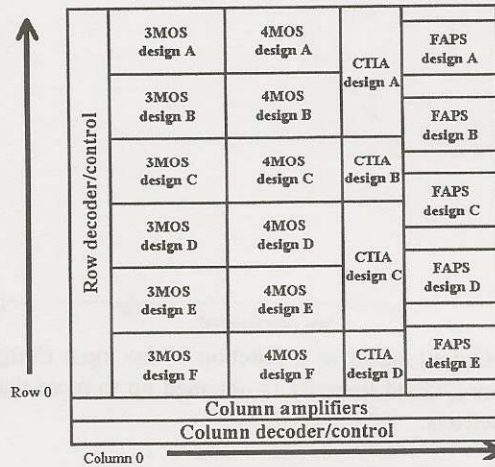


Figure 6. Floorplan of the test sensor.

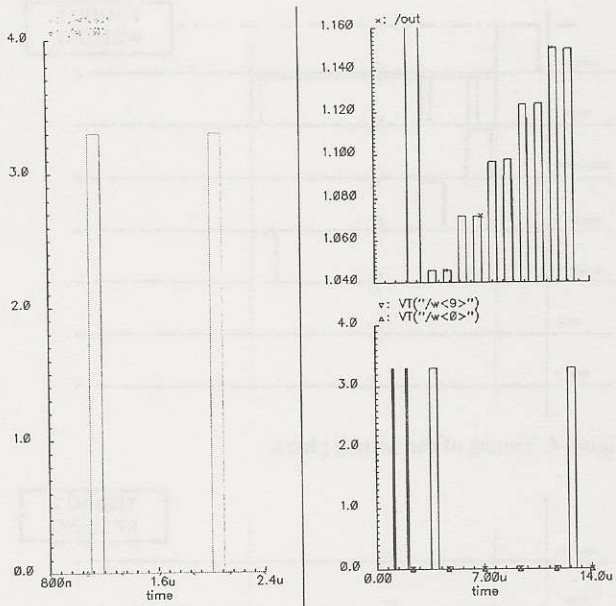


Figure 7. Transient simulation of one pixel after parasitic extraction. At the input, a charge signal of 1000 electrons was injected every 200 ns. Ten samples were acquired, spaced by 100 ns, and then read out at 1 MHz. The signals 'w' correspond to the signals 'sample' in figure 3.

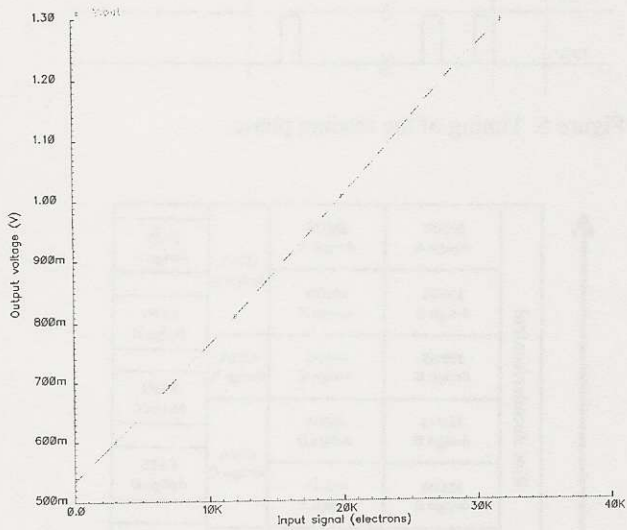


Figure 8. Output signal as a function of the input charge (in electrons). Good linearity is obtained up to more than 30,000 electrons.