

A high dynamic range image sensor with linear response based on asynchronous event detection

Juan A. Leñero-Bardallo, R. Carmona-Galán and Á. Rodríguez-Vázquez

Institute of Microelectronics of Seville (IMSE-CNM),
 Consejo Superior de Investigaciones Científicas y Universidad de Sevilla,
 C/ Américo Vespucio s/n, 41092, Seville, Spain

E-mails: {juanle, rcarmona, angel}@imse-cnm.csic.es

Abstract—This paper investigates the potential of an image sensor that combines event-based asynchronous outputs with conventional integration of photocurrents. Pixels voltages can be read out following a traditional approach with a source follower and analog-to-digital converter. Furthermore, pixels have circuitry to implement Pulse Density Modulation (PDM) sending out pulses with a frequency that is proportional to the photocurrent. Both read-out approaches operate simultaneously. Their information is combined to render high dynamic range images. In this paper, we explain the new vision sensor concept and we develop a theoretical analysis of the expected performance in standard AMS 0.18 μm HV technology. Moreover, we provide a description of the vision sensor architecture and its main blocks.

Keywords: Vision Sensor; AER; High Dynamic Range; Linear

I. INTRODUCTION

High dynamic range operation is important for a wide variety of applications like automotive applications, consumer products, quality medical imaging, and surveillance. There are several methods to capture high dynamic range images. The most extended approach is to use dual or multiple captures of several images with different integration times [1], [2] which usually require costly frame-memories for the different time exposures. They also need to choose carefully the ratio between different exposure times depending on the illumination conditions to maximize the dynamic range. Other authors use a compressive function to capture and codify high dynamic range images with a low number of bits [3], [4]. The drawback is that these sensors cannot capture all the illuminated areas with the same accuracy. Event-based vision sensors usually offer high dynamic range operation, [5]–[7]. Unfortunately, their event-based outputs are not compatible with frame-based displays and require some post-processing to convert their outputs into an equivalent frame-based composition.

We describe here a high dynamic vision sensor that combines an event-based asynchronous operation with a frame-based analog read-out to render high dynamic range images. Pixels photodiode voltages never saturate during the integration time achieving high dynamic range operation. They generate one event every time that the integrated charge reaches a limit. Then pixels self-reset and they start integrating charge again, until their outputs are read out following a

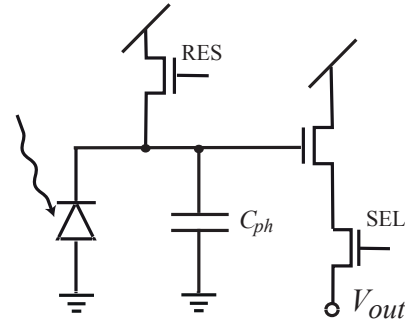


Fig. 1. Three-transistor active pixel sensor (APS) architecture.

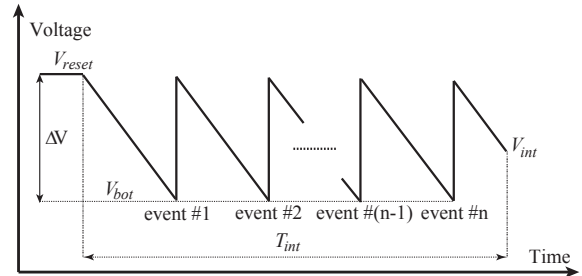


Fig. 2. High dynamic range operation new approach.

traditional approach, i.e. the voltage at the sensing capacitance read out via a voltage follower and digitized with an analog-to-digital converter. Finally, knowing the number of events (if any) associated to each pixel and its digital output, we will obtain a digital word associated to each pixel that represents its illumination value. This digital value is compatible with a conventional frame-based representation. The novelty of this new approach over prior implementations is the very high dynamic range of operation with linear outputs without compressing the information and without using multiple exposures times. The new sensor also allows to operate as a conventional imager with one exposure time with digitized outputs or as an asynchronous vision sensor with PDM outputs. Thus, it is possible to trade between dynamic range and speed and choose a frame based or a PDM output.

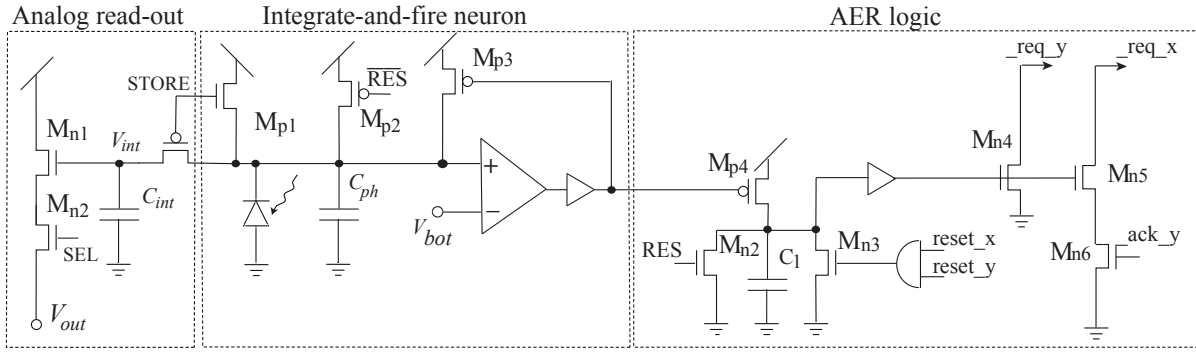


Fig. 3. Pixel schematics.

II. PRINCIPLE OF OPERATION

Fig. 1 displays a 3T-APS pixel. To sense the photocurrent, we need to read out the voltage at the photodiode capacitance C_{ph} , at the end of the integration period, T_{int} . Usually pixels have very different levels of illumination and it is not possible to sense all their photocurrent values. If the integration time is too long, pixels with high illumination will provide the same read-out voltage close to zero. If the integration time is shorter, we will be able to sense highly illuminated pixel outputs. On the contrary, pixels under low illumination will not have time to discharge the photodiode capacitance fast enough to measure their output voltage values.

The dynamic range of operation of a vision sensor is defined as the ratio between the lowest and the highest photocurrent that it can measure. It is usually expressed in decibels as:

$$DR = 20 \cdot \log_{10} \left(\frac{I_{ph_{max}}}{I_{ph_{min}}} \right) \quad (1)$$

If we represent the photocurrent values with binary words of N_{bits} , the dynamic range of the sensor is given by

$$DR = 20 \cdot \log_{10} (2^{N_{bits}}) \quad (2)$$

Hence, if we would like to render a 120dB-dynamic-range image, we need 20 bits. The dynamic range of the typical 3T-APS pixel is rarely beyond 50dB. Fig. 2 shows the new approach to maximize the dynamic range of one pixel. Pixels will never stop integrating charge, independently of the photocurrent value. Every time that the voltage at the photodiode's node reach the value V_{bot} , an event will be sent out of the chip with the coordinates of the pixel who has reached the voltage value V_{bot} . Immediately after, the pixel will self-reset and start integrating charge again. Finally, in the end of the integration period, T_{int} , all the pixels are read-out following a traditional approach, i.e. the voltage V_{int} is read out with a source follower and analog-to-digital converter. Thus the value of the photocurrent will be proportional to:

$$I_{ph} \propto (V_{reset} - V_{int}) + (V_{reset} - V_{bot}) \cdot (\#events) \quad (3)$$

If we know the number of events associated to each pixel (if any) and the value of V_{int} , it is possible to obtain a binary word with $N_s + N_b$ bits. The most significant N_s bits represent the number of events associated to each pixel. The less significant N_b bits represent the number of bits of analog-to-digital conversion. This approach does not need to read out the output voltage several times like multi-exposure cameras do. Pixels with high illumination will never saturate. There is a linear dependence between the output code associated to each pixel and its illumination. The only limitation is the maximum event rate that the arbitration system can handle, that will limit the frame rate.

III. PIXEL DESCRIPTION

Fig. 3 shows the pixel's topology. To implement the self-resetting approach of Fig. 2, pixels have to be reset immediately after sending an event out the chip when a programmable voltage threshold is reached. To implement this functionality, we use a integrate-and-fire neuron, that is made with a comparator that activates a digital signal that resets the voltage at $C = C_{ph} + C_{int}$ and every time the voltage value V_{bot} is reached and $STORE$ is active. To send out the chips events, pixels include some digital circuitry that has already been implemented in several designs and it is not within the scope of this work. For a deep understanding of the AER communication circuitry used in this work we recommend to look at Philipp's Häfliger PhD work [8]. Finally, on the right, the circuitry to read out V_{int} in the of the integration period is placed.

IV. SYSTEM DESIGN CONSIDERATIONS

The voltage at the photodiode's capacitance is a periodic signal which period is given approximately by:

$$T = \frac{C \cdot (V_{reset} - V_{bot})}{I_{ph}} + T_d \approx \frac{C \cdot \Delta V}{I_{ph}} \quad (4)$$

$T_d \approx 25\text{ns}$ is the controlled delay between the comparator and the digital buffer output of Fig. 3. Let us refer to the maximum photocurrent and the minimum photocurrent that can be sensed as $I_{ph_{max}}$ and $I_{ph_{min}}$. T_{min} will be the period that corresponds to the highest photocurrent that we can

measure. Under normal operation conditions, T_d will be much lower than T , as we will demonstrate at the end this section. Hence T_d influence can be neglected. If we assign N_s bits to codify the number of spikes, the integration time should be assigned to receive less than 2^{N_s} spikes per pixel within the time interval T_{int} , i.e.:

$$T_{int} < 2^{N_s} \cdot \frac{C \cdot \Delta V}{I_{ph_{max}}} = 2^{N_s} \cdot T_{min} \quad (5)$$

If we consider the less illuminated pixel that we can measure, to read-out its output voltage, we need to wait until its voltage drop is equal to one LSB of the analog-to-digital-converter:

$$T_{int} > T_{LSB} = \frac{C \cdot \Delta V}{I_{ph_{min}} \cdot 2^{N_b}} \quad (6)$$

Hence, for an optimum choice of the integration time, it must satisfy:

$$T_{int} = 2^{N_s} \cdot T_{min} = T_{LSB} \quad (7)$$

In our particular design, $N_s = 12$ bits and $N_b = 8$ bits, so the expected dynamic range is $DR = 20 \cdot \log_{10} \left(\frac{I_{ph_{max}}}{I_{ph_{min}}} \right) = 120$ dB. There is a trade-off between dynamic range and speed. In some applications, we could desire to make $T_{int} < T_{LSB}$ to increase the frame rate (FR). This is feasible, but DR will be reduced because that the value of the minimum detectable photocurrent, $I'_{ph_{min}}$ will be higher:

$$FR = \frac{1}{T_{int}} = \frac{I'_{ph_{min}} \cdot 2^{N_b}}{C \cdot \Delta V} \quad (8)$$

and DR would be:

$$DR = 20 \cdot \log_{10} \left(\frac{2^{N_b} \cdot I_{ph_{max}}}{FR \cdot C \cdot \Delta V} \right) \quad (9)$$

One practical limitation of our approach is that the arbitration system can handle a maximum output event rate MAX_{BR} that limits the number of times that pixels can spike during the integration time. So if we have an array with M rows and N columns, the sum of all the pixels spikes event rates during the integration time should be lower than MAX_{BR} . In the unlikely case that all the pixels are illuminated with the maximum illumination that we can sense, $I_{ph_{max}}$, we must satisfy:

$$M \cdot N \cdot \frac{1}{T_{min}} < MAX_{BR} \quad (10)$$

Since T_{min} depends on ΔV , we can control the global event rate by adjusting ΔV . In practical situations not all the pixels will be illuminated with the highest illumination. Let us define a parameter α that indicates that pixels percentage that exposed to the maximum illumination. Thus, to avoid loose information

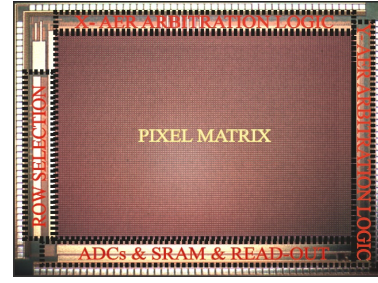


Fig. 4. Sensor's microphotograph. The main system blocks have been highlighted. Chip dimensions are 4.1mm \times 3.3mm.

due to the arbitration system limitations, we have to set a voltage variation ΔV that satisfies:

$$V_{DD} > \Delta V > \frac{\alpha \cdot M \cdot N \cdot I_{ph_{max}}}{C \cdot MAX_{BR}} \quad (11)$$

Obviously, there is a trade-off between speed, dynamic range and event rate (power consumption). We can lower the event rate by increasing ΔV , but the frame rate and the dynamic range will be lower. In our particular case, we have implemented a pixel matrix in the AMS 0.18 μ m HV standard fabrication process with the following parameters: $C = 40$ fF, $I_{ph_{max}} = 500$ pA, $M = 128$, $N = 96$, $MAX_{BR} = 15$ Meps, $\alpha = 0.2$, $FR = 25$ fps. Using Eq. 9 and Eq. 11, we obtain $\Delta V_{min} = 2V$ and $DR = 97$ dB. The design technology offers transistors with double oxide that can handle voltages of 5V. This gives more flexibility for the chip test. If we do not impose restrictions for the frame rate, it would be possible to obtain a dynamic range close to 120dB. Just choosing the value of the integration time according to Eq. 7, we obtain $T_{int} = 0.65$ s. Therefore, there is also a trade-off between dynamic range and frame rate as we discussed before. By increasing the number of bits N_s , we could go beyond 120dB.

The error ϵ introduced by delay between the comparator output and the digital buffer output, $T_d \approx 25$ ns, in Eq. 4 is in both cases in negligible because $T_{min} \gg T_d$: $\epsilon_{T_d} = \frac{T_d}{T_{min}} = T_d \cdot FR \cdot 2^{N_s} = 0.26\%$, with $FR=25$ fps.

V. SYSTEM ARCHITECTURE

The chip was fabricated in AMS 0.18 μ m HV technology. We are currently obtaining preliminary experimental results. Fig. 4 shows a chip photograph. Fig. 5 displays the system architecture. On the top and the right, we have placed the AER arbitration system. This arbitration system has been previously implemented in other designs [8]. On the left, we find digital circuitry that generates the control signals RES , SEL , and $STORE$ (see Fig. 3). On the bottom, all the circuitry necessary to read-out the pixels output voltages is placed. We have designed column-parallel ramp converters with 8 bits. The ramp is generated with a resistive divider and a set of switches that are activated sequentially. The ramp output voltage is driven by an analog folded cascode buffer. Each column has a comparator that activates a signal EOC when

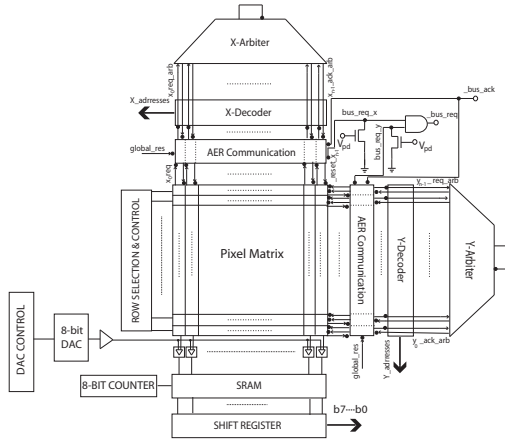


Fig. 5. System block diagram.

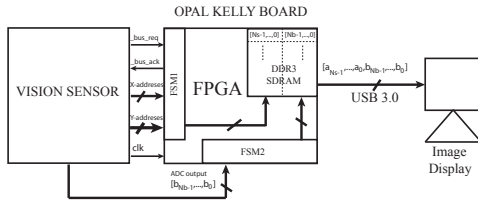


Fig. 6. Diagram showing the communication flow between the sensor, an Opal Kelly board, and a PC. The Opal Kelly board merges and stores the event and the ADC outputs associated to each pixel. Its outputs are digital words with $N_s + N_b$ bits that are transmitted to a PC.

the ramp reaches the pixel output voltage V_{out} . Finally, we have added a SRAM memory to latch all the digitized outputs for each row. It is connected to a shift register that sends out the chips all the bits of the analog-to-digital conversion. It has 8-parallel outputs operating with a frequency, $f_{clk} = 200\text{MHz}$.

Every pixel output will be represented with a digital word with $N_{bits} = N_s + N_b$. In order to combine the event and the ADCs outputs, we will use an external Opal Kelly K160T board to test the prototype. The board has a DDR3 SDRAM memory and a Kintex-7 FPGA that implements two FSMs to handle the asynchronous AER communication and store the ADC digital outputs (see Fig. 6). The first FSM will receive the signal $_bus_req$ and the address (X and Y coordinates) every time that an event is sent out. This information will be stored in the SDRAM memory. The events associated to each pixel will be the N_s most significant bits of the digital word. Every time that an event is received, we will increase the event number associated to the pixel by one. Finally, the FPGA will send the $_bus_ack$ signal when the event information has been processed. The second FSM will receive groups of N_b bits every system clock cycle. They correspond to the analog-to-digital conversion of each pixel output value. The state machine will store these bits in the SDRAM memory. They will correspond to the N_b less significant bits.

TABLE I
SENSOR SPECIFICATIONS

Technology	AMS 0.18 μm HV
Power Supply	1.8V/5V
Pixel Size	25 $\mu\text{m} \times 25\mu\text{m}$
Pixel Complexity	34 Transistors + 2 Capacitors
Fill Factor	10%
Dynamic Range	97dB@25fps, 120dB@1.5fps
Analog pixel Consumption	3 μA per column

VI. CONCLUSIONS

We have presented a new vision sensor concept with frame-based analog read-out and asynchronous event-based PDM read-out. Both flows are combined to render high dynamic range images. Pixels outputs are proportional to illumination, i.e. there is a linear dependence between the digital word assigned to each pixel and its illumination. Only one integration time is necessary to read out the image. Pixels exposed to high illumination values never saturate. Outputs are compatible with frame-based displays. The new system can combine the advantages of frame-based and event-based systems. The system has a trade-off between speed and dynamic range. System specifications are summarized on Table I.

VII. ACKNOWLEDGEMENTS

This research work has been supported by projects MONDEGO (TEC2012-38921-C02-02) MINECO (European Region Development Fund, ERDF/FEDER), IPT-2011-1625-430000 MINECO, and ONR grant N00014-14-1-0355.

REFERENCES

- [1] M. Mase, S. Kawahito, M. Sasaki, and S. Wakamori, "A wide dynamic range cmos image sensor with multiple exposure-time signal outputs and 12-bit column-parallel cyclic A/D converters," *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2787–2795, December 2005.
- [2] O. Yadid-Pecht and E. R. Fossum, "Wide intrascan dynamic range CMOS APS using dual sampling," *IEEE Trans. Electron Devices*, vol. 44, no. 10, pp. 1721–1723, October 1997.
- [3] S. Vargas-Sierra, G. Lián-Cembrano, and A. Rodríguez-Vázquez, "A 151dB high dynamic range CMOS image sensor chip architecture with tone mapping compression embedded in-pixel," *IEEE Sensors Journal*, pp. 1721–1723, July 2014, DOI: 10.1109/JSEN.2014.2340875.
- [4] A. Spivak, A. Belenky, A. Fish, and O. Yadid-Pecht, "Wide dynamic range CMOS image sensors -comparative performance analysis," *IEEE Transactions on Electron Devices*, vol. 56, no. 2, pp. 2446–2461, November 2009.
- [5] J. A. Leñero-Bardallo, T. Serrano-Gotarredona, and B. Linares-Barranco, "A five-decade dynamic-range ambient-light-independent calibrated signed-spatial-contrast aer retina with 0.1-ms latency and optional time-to-first-spike mode," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 57, no. 10, pp. 2632–2643, Oct 2010.
- [6] C. Posch, D. Matolin, and R. Wohlgenannt, "A QVGA 143dB dynamic range asynchronous address-event PWM dynamic image sensor with lossless pixel-level video compression," *IEEE Journal of Solid State Circuits*, vol. 46, no. 1, pp. 259–275, January 2010.
- [7] J. A. Leñero-Bardallo, T. Serrano-Gotarredona, and B. Linares-Barranco, "A 3.6 μs latency asynchronous frame-free event-driven dynamic-vision-sensor," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 6, pp. 1443–1455, June 2011.
- [8] P. Häfíger, "A spike based learning rule and its implementation in analog hardware," Ph.D. dissertation, ETH Zürich, Switzerland, 2000, <http://www.ifi.uio.no/~hafiger>.