

Global shutter CMOS vision sensors and event cameras for on-chip dynamic information

Marko Jaklin  | Daniel García-Lesta  | Paula López  | Victor M. Brea 

Centro Singular de Investigación en Tecnologías Intelixentes (CITIUS), Universidad de Santiago de Compostela, Santiago de Compostela, Spain

Correspondence

Marko Jaklin, Centro Singular de Investigación en Tecnologías Intelixentes (CITIUS), Universidad de Santiago de Compostela, Rúa de Jenaro de la Fuente Domínguez, Santiago de Compostela, A Coruña, Spain.
Email: marko.jaklin@usc.es

Funding information

H2020 Marie Skłodowska-Curie Actions; Xunta de Galicia; Marie Skłodowska-Curie grant, Grant/Award Number: 860370; Xunta de Galicia-Consellería de Cultura, Educación e Ordenación Universitaria Accreditation 2019–2022, Grant/Award Number: ED431G-2019/04; Reference Competitive Group Accreditation 2021–2024, Grant/Award Number: GRC2021/48; Spanish Ministry of Science, Innovation and Universities, Grant/Award Number: PID2021-128009OB-C32

Abstract

The on-chip extraction of dynamic information from a scene can be addressed with either frame-based CMOS vision, also called smart image sensors, or dynamic vision sensors, also known as event cameras. When implemented with a pinned photodiode (PPD) as a 4-transistor active pixel sensor (4T-APS), the former brings about the benefits of low temporal noise and dark current but without high dynamic range (HDR). The latter comes with the benefits of HDR and a fast event detection rate with low power consumption. The drawback is the background activity noise, which leads to additional hardware or algorithms to keep it low. This paper analyses the mismatch and noise of a global shutter 4T-APS implementation with local HDR through an overflow capacitor and correlated double sampling (CDS) to provide low noise events through frame differencing. The aim is to narrow the gap with dynamic vision sensors in terms of event rate and dynamic range. We show that our solution would be competitive with event cameras in scenarios with slow moving objects and a relatively wide dynamic range (85 dB).

KEYWORDS

CMOS vision sensors, dynamic vision sensors, event cameras, frame difference, high dynamic range, smart image sensors

1 | INTRODUCTION

Temporal and spatial redundancy of a static background scene do not call for continuous streaming or the continuous running of elaborate computer vision models such as object detection, tracking, or action recognition. Indeed, it is the entrance of an object into the field of view of the camera, or, in general, object motion, that triggers further computer vision processing and continuous streaming. Thus, it is clear that there are two main operation modes in video surveillance with different compute and power needs, which can be exploited to extend battery life toward stand-alone, always-on visual edge computing nodes.^{1–3} In this context, dynamic information extraction is key to separating said two operation modes.

Dynamic vision sensors (DVS) or event cameras like that in Figure 1 do it through time-stamped spikes of ON and OFF events that detect increasing and decreasing intensity changes in the scene above a certain threshold.^{4–7} These events are sent outside of the pixel array in address-event representation (ARE). They implement a logarithmic

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *International Journal of Circuit Theory and Applications* published by John Wiley & Sons Ltd.

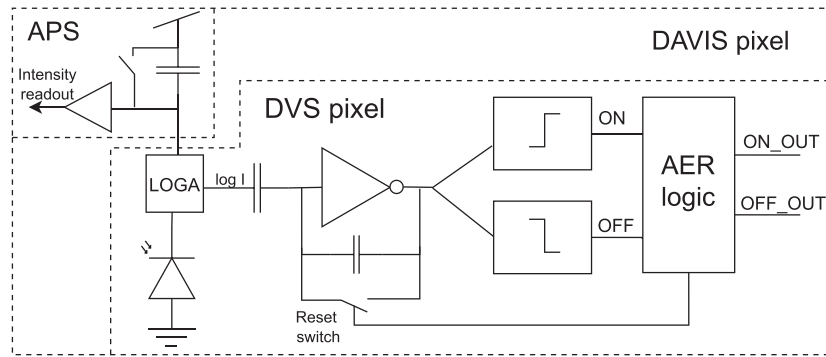


FIGURE 1 DAVIS solution to the dynamic vision sensor concept.¹⁷

photodiode that offers high dynamic range and allows for continuous operation of the pixel, leading to finer temporal resolution. The DVS is usually accompanied by a 3T-APS structure, leading to a Dynamic and Active Pixel Vision Sensor (DAVIS); see Figure 1.

Frame-based cameras can extract dynamic information through background subtraction or motion detection with smart image sensors implemented as either 3T- or 4T-APS sensors.^{1,8–10} Global shutter pixels working in integration mode with frame differencing functionality are another approach to generating events. In this case, frame differencing is performed through two consecutive frames with a fixed integration time and a global reset phase in between. Global reset keeps false event generation through leakage currents at bay, although at the cost of worse temporal resolution than that of event pixels.^{7,11} As a benefit, the pinned photodiode (PPD) in 4T-APS allows for lower dark current and reset noise with correlated double sampling (CDS) techniques. Apart from their worse temporal resolution, global shutter pixels with frame difference features lack HDR capability. The purpose of this design is to lower the gap in dynamic range between dynamic vision sensors and frame difference-based event cameras. Former frame differencing pixels either do not output events^{12–14} or do not implement any dynamic range extension methods.^{10,15} This paper describes a frame differencing-based event pixel with a lateral overflow capacitor that collects saturated photoelectrons as a dynamic range extension method.¹⁶ We give performance metrics of area, speed, and power consumption through post-layout simulations, as well as of false event generation through Monte Carlo and noise simulations.

2 | GLOBAL SHUTTER HDR 4T-APS PIXEL FOR EVENT GENERATION

Our global shutter HDR 4T-APS pixel provides events in the form $e_k = (x_k, t_k, p_k)$, with the event's position in the array x_k , the timestamp t_k , and the polarity of the intensity change p_k . A preliminary version of our pixel has been previously introduced in Jaklin et al.¹⁸ Apart from a less in-depth description of all the concepts and circuits, that paper lacks layouts, post-layout performance metrics, noise analyses, and a comparison with dynamic vision sensors.

2.1 | Pixel operation

The schematics and timing diagram of our pixel can be seen in Figure 2 and Figure 3 with three main operations: (i) image acquisition, (ii) HDR algorithm, and (iii) frame differencing. Figure 2 also includes simplified schematics for key phases of operations of the pixel by configuring switches along the data path. All these phases are explained throughout the next paragraphs below. The analog memory bank and the comparator are labeled AMB and CMP, respectively.

The image acquisition of our pixel features conventional and HDR integration modes. The conventional mode, named as $S1$ integrates electrons onto the floating diffusion node FD . The HDR mode, referred to as $S2$, adds the capacitance of the CS node to the floating diffusion node FD .¹⁶ This permits to store more electrons at $FD + CS$, extending the dynamic range at the price of a smaller conversion gain. The selection is automatic, that is, each pixel on its own determines whether to use $S1$ or $S2$. Figure 2B illustrates the decision on working in conventional or HDR mode with the comparison of signal $S1$ with a user programmed signal $V_{THS1/S2}$ with circuit CMP.

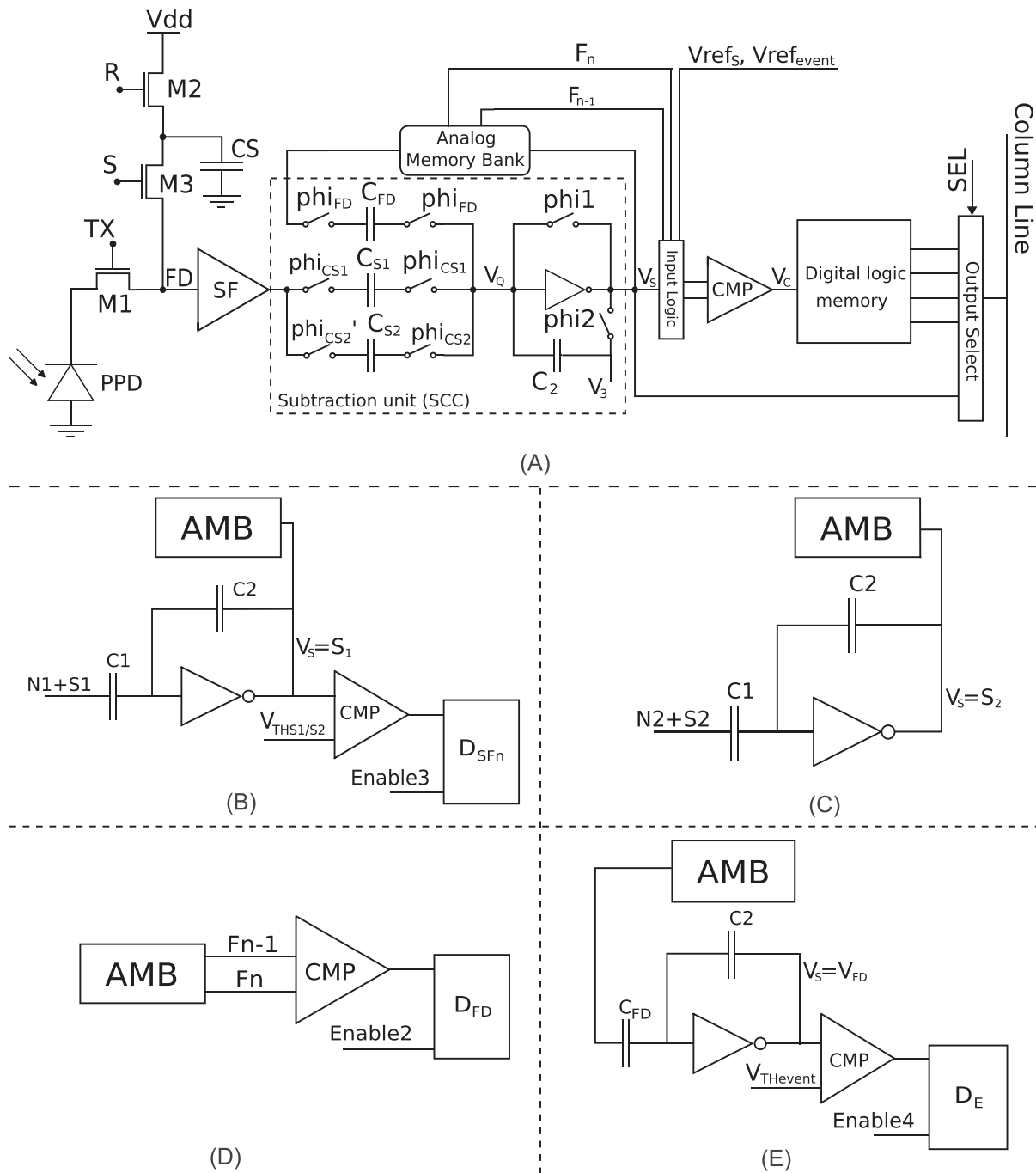


FIGURE 2 (a) Schematic of a pixel. (b) S1/S2 decision phase. (c) S2 acquisition phase. (d) Polarity decision phase. (e) Frame differencing and event detection phase.

The operation of our global shutter HDR 4T-APS pixel in more detail is as follows. First, right after reset at t_1 in Figure 3 and with signal S pulsed high, the noise level N_2 of the node $FD + CS$ is read and stored on the capacitor C_{S2} driven by the source follower (SF). During the integration time, signal S is pulsed high, keeping the M_3 transistor active and allowing for any saturated electrons that flow through M_1 when TX is low at this time to be stored on the overflow capacitor CS . Just before integration time ends, signal S is pulled low, isolating FD from CS and immediately after, at t_2 , the noise level N_1 associated with the electrons of signal S_1 of the now isolated FD node is read and stored on C_{S1} . Next, the switch TX is pulsed high, transferring the electrons from the PPD to the FD node, and thus, generating $N_1 + S_1$, and sending it to C_{S1} . The CDS is performed with the arrival of $N_1 + S_1$ at t_3 by setting signal ϕ_{i2} high in the subtraction unit, yielding S_1 on the node V_s . As said above, this is illustrated with Figure 2B. S_1 signal represents the

voltage value generated by nonsaturated electrons only, while the signal $S2$ which is calculated shortly after $S1$, represents the value generated by nonsaturated and saturated electrons. Similarly to $S1$, Figure 2C shows the circuit at the time of calculating $S2$. At $t4$, in order to set the correct amount of charge on the quiescent node, ϕ_{CS2} is pulsed high. Also at $t4$, TX is again pulsed high to deposit the electrons collected during the calculation of $S1$ signal. At this moment, the new integration cycle begins. And at $t5$ $S2$ is calculated on the node V_S . The $S1$ - $S2$ crossing is given by a user-defined threshold $V_{THS1/S2}$. If the accumulated voltage $S1$ exceeds said threshold, the analog memory stores signal $S2$, otherwise it stores $S1$. COMP signal enables the comparator and is also pulsed high at $t3$ to compare $S1$ to $V_{THS1/S2}$.

Frame differencing, executed after polarity decision through $F_n > F_{n-1}$ is performed with subtraction and comparator circuits, which are also used for the decision making on the switch from $S1$ to $S2$. F_n and F_{n-1} frames are compared to each other between $t5$ and $t6$ when signal COMP is pulsed high. Figures 2D and 2E show the polarity decision and frame differencing phases, respectively. At $t7$, the frame differencing is complete and the result is compared to a user-defined threshold voltage $V_{THevent}$. After the comparison the cycle begins anew with pulsing the reset signal R high. We apply circuit sharing techniques, as said circuits are the same for ON/OFF event generation and for CDS operations to mitigate the effect of mismatch and noise levels $N1$ and $N2$ corresponding to signals $S1$ and $S2$, respectively.

2.2 | Pixel circuits

All our circuits feature power gating in order to decrease power consumption. The supply voltage of the 4T-APS is $V_{dd} = 3.3$ V, searching for a wide dynamic range. This voltage supply is set to 1.8V for the rest of the circuitry, aiming for low power consumption.

2.2.1 | 4T-APS

Our 4T-APS sensing structure with a PPD comprises a programmable overflow metal insulator metal (MIM) capacitor CS set to 50, 100, and 150 fF for the HDR algorithm, which allows for different upper limits in the incident light. The aspect ratios W/L ($\mu\text{m}/\mu\text{m}$) in the transistors of our 4T-APS implementation are: 5.4/0.8, 0.35/0.35 and 0.35/0.5 for $M1$, $M2$ and $M3$, respectively. The source follower is biased with a current $I_{SFAPS} = 0.5$ μA with an aspect ratio of

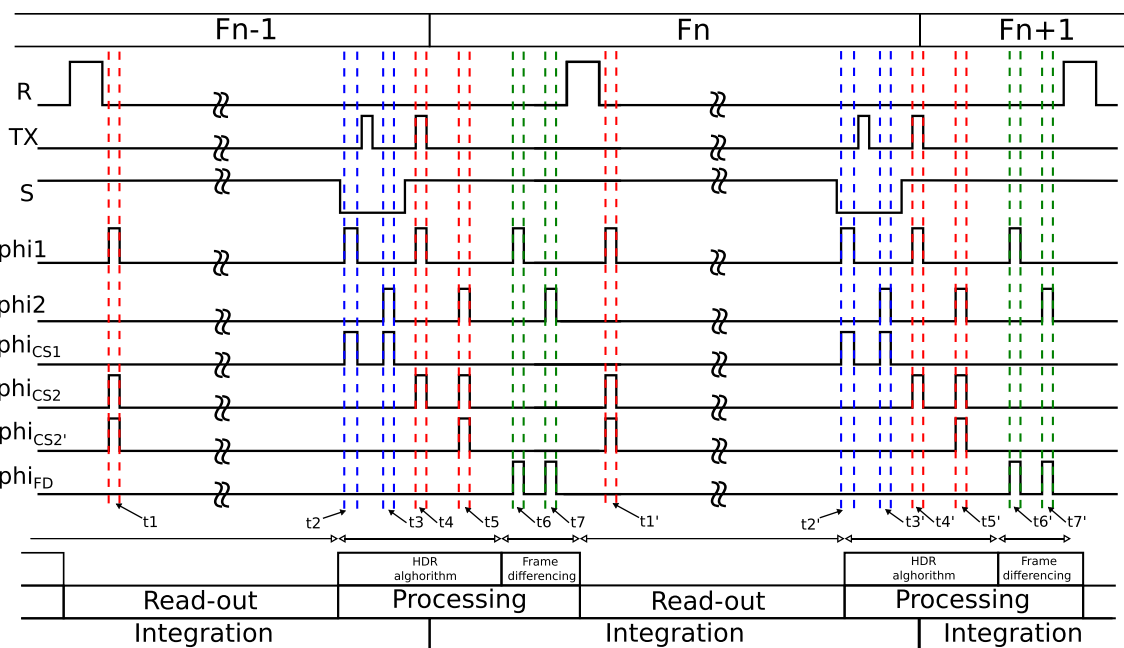


FIGURE 3 Timing diagram of our HDR 4T-APS for event generation. Control signals during two sequential frames, “ F_{n-1} ” and “ F_n ,” are shown. Red lines represent steps for $S2$ calculation, blue lines for $S1$, and green ones mean frame difference operation. [Colour figure can be viewed at wileyonlinelibrary.com]

0.22/0.9 ($\mu\text{m}/\mu\text{m}$), while the biasing transistor aspect ratio is 0.22/1.5, again in ($\mu\text{m}/\mu\text{m}$). The capacitance of the floating diffusion node FD has been estimated by post-layout simulations as 12 fF.

2.2.2 | Subtraction unit

The subtraction unit is a double cascode inverting amplifier in feedback mode with the capacitor C_2 and the reset switches phi_1 and phi_2 . This unit runs CDS for signals $S1$ and $S2$ and frame differencing between the current F_n and the previous frame F_{n-1} . This structure is the same as that of the dynamic vision sensor⁶ or other CMOS vision sensor solutions like.¹⁹

The sensitivity of the frame differencing and $S1/S2$ signals is given by the $C_{FD}/C_{S1}/C_{S2}$ to C_2 ratios. We have set C_2 as a programmable device to $1\times$, $2\times$ and $3\times$ gains. Capacitors $C_{FD}/C_{S1}/C_{S2}$ have been sized to 90 fF, while C_2 is scaled accordingly. Signal V_3 adds a user-defined offset, which in our case sets the comparator input transistors to saturation. The input transistors of the comparator are NMOS types with a threshold voltage of around 650 mV, and V_3 of minimum 700 mV is advised. All the switches have been designed as NMOS transistors with minimum dimensions. The double cascode inverting amplifier has an open-loop gain of 100 dB.

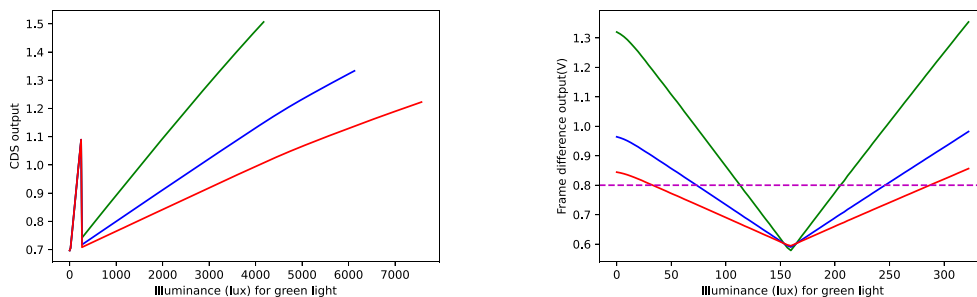
The sequence of operations of the different switches can be seen in Figure 3. The CDS operation for signals $S1$ and $S2$ to mitigate the effect of noise $N1$ and $N2$ is given by the formula:

$$S_i = \frac{C_{S_i}}{C_2} (N_i - (N_i + S_i)) + V_3 \quad (1)$$

with subindex i referring to either $S1/S2$ or $N1/N2$. This result is stored in the analog memory bank, which holds the values of the current frame F_n and the previous one F_{n-1} .

The programmability of the overflow capacitor CS offers flexibility in HDR mode. Figure 4A collects a simulation for monochromatic light at wavelength of 555 nm (green) for the three gains defined in our pixel. The Y-axis is the voltage V_S in Figure 2A, after running CDS. The pixel operates only with the floating diffusion capacitance, and thus with high conversion gain (signal $S1$), at low illumination levels. At a given illumination level, the pixel enters the HDR region, with the excess electrons being collected on the overflow capacitor CS . This is the region of signal $S2$, where the conversion gain decreases due to the capacitance of two shunted nodes, $FD + CS$. The higher the curve in the HDR region, the lower the CS capacitance. The three curves in Figure 4A correspond to our three cases of CS ; 50, 100 and 150 fF. The peak seen in Figure 4A occurs when pixel switches from $S1$ to $S2$ signal. It occurs when $S1 > V_{THS1/S2}$. In this case, $V_{THS1/S2}$ is set to 1.1 V.

The frame differencing is calculated as



(A) HDR extension up to the linear limit of $S2$. At this point $N1$ is saturated resulting in nonlinearity and false threshold $V_{THS1/S2}$ detection. (B) Frame difference gain programability.

FIGURE 4 Simulations showing HDR extension and gain programability. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/cim.3925)]

The comparator is implemented as a two-stage open-loop amplifier with a 5-transistor operational amplifier (OTA) architecture and a differential input driving an inverter. Noise and mismatch can cause incorrect polarity when two frames are close, so although it does not feature offset cancellation, it has been designed with large transistors in order to make it mismatch resilient. Post-layout simulations yield: (i) $\xi_s = 1.1$ mV as the static resolution of the comparator. This metric is achieved with a large input differential pair, sized to $W = 5$ μm , $L = 0.58$ μm .

2.2.5 | Digital memory block: Event generation and read-out

The digital memory block shown in Figure 6A contains four D-latches to (i) provide events; (ii) set the polarity of the event; and (iii) assess the signal uniformity of S1 or S2 between frames.

The behavior of the four D-latches is conveyed in Figure 6B. If D_e is set to the logical “1” and D_{SF_n} and $D_{SF_{n-1}}$ have the same logical values, either “0” or “1,” an event is issued. In this case, the local value at D_{FD} determines the polarity of the event. A logical “1” means that the voltage variation along the integration time at the current frame exceeds that of the previous frame, providing an ON event, and vice versa. The fact that two consecutive frames come from two different sensitivities (S1 and S2) is accounted for with D_{SF_n} and $D_{SF_{n-1}}$ issuing different logical states. In this situation, there is an ON event if the state from D_{SF_n} is a logical “1” and that from $D_{SF_{n-1}}$ is a logical “0” regardless of D_e and D_{FD} . There is an OFF event if the state from D_{SF_n} is a logical “0” and that from $D_{SF_{n-1}}$ is a logical “1” regardless of the states of D_e and D_{FD} . D_{SF_n} and $D_{SF_{n-1}}$ holding the same logical values means that the change from the previous to the current frame generates events.

Finally, our chip provides both raw images and events. This is managed by the “Output Select” block in Figure 2A. The raw image is read out with a unity gain buffer as an analog signal as a result of the CDS operation. The status of the D-latches through a 4-bit word is read out serially and sent for off-chip processing that, based on the table in Figure 6B, determines what kind of event occurs, either ON or OFF events. An example of an OFF event is shown in Figure 7. The input light in Figure 7A is an abrupt decrease in input light power in the S2 regime. Accordingly, the output corresponds to an OFF event with the code “1011.” Said output is a continuous-time signal on the output column line or bus of the pixel (see Figure 2), as shown by the blue curve in Figure 7B. This continuous-time signal on the output column line is the result of a sequential writing of the outputs of the four digital or D-latches, as labeled within vertical green lines in Figure 7B. Finally, the read-out speed is estimated at around 1000 efps by post-layout simulations.

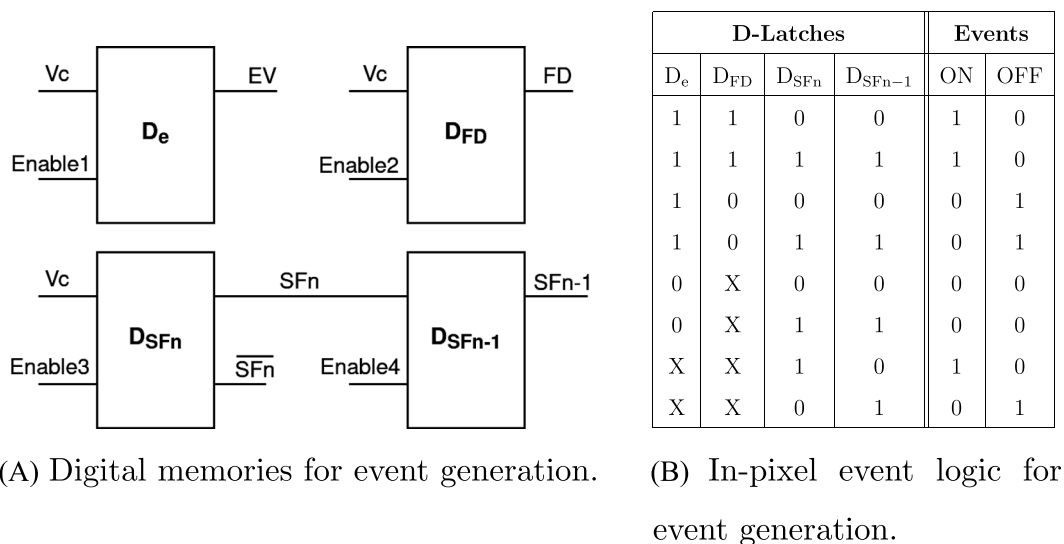


FIGURE 6 Digital logic structure and logic. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

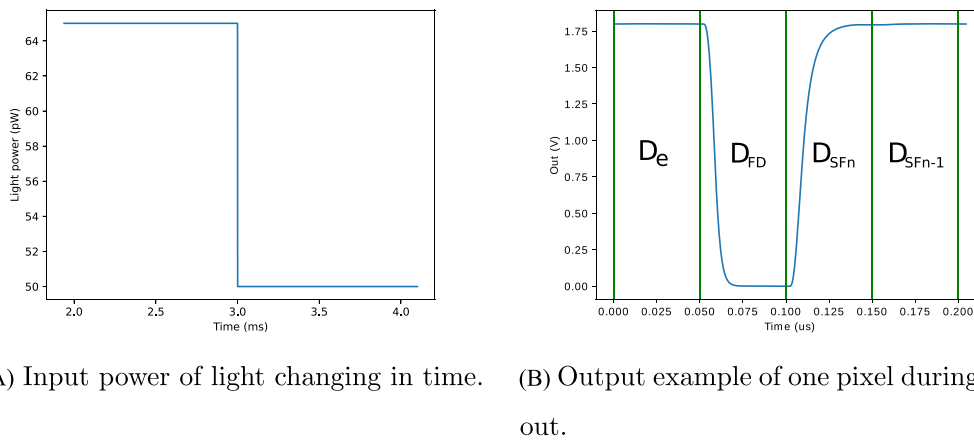


FIGURE 7 The two graphs show the relation between the input power of light during two sequential frames and the output of the second frame. The simulations provide an OFF event in S2 mode with the code “1011.” [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

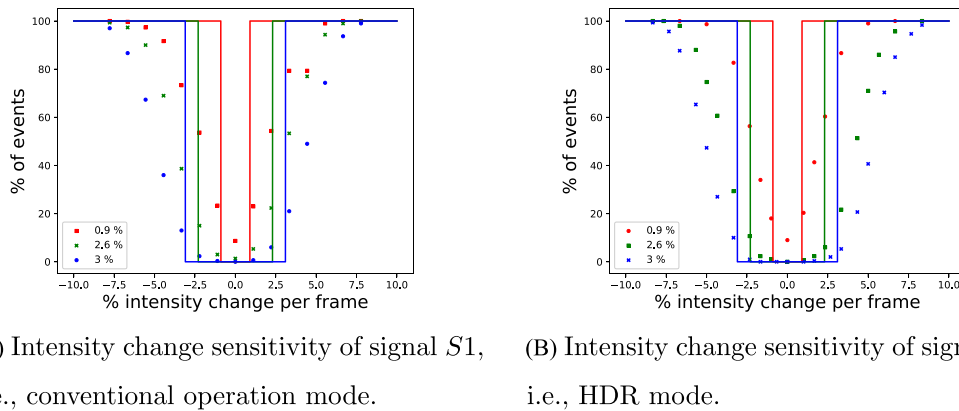


FIGURE 8 False event simulations due to mismatch effects in our HDR 4T-APS solution. Full lines represent transition in the absence of mismatch. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

3 | PERFORMANCE METRICS

3.1 | Spatial accuracy

The spatial uniformity of a pixel array in our global shutter HDR 4T-APS is given by the mismatch of every pixel along the data path. Our solution comprises the two integration modes $S1$ and $S2$, so we have run Monte Carlo simulations for both cases in order to assess the sensitivity to intensity changes of our approach.

Mismatches of individual circuits are listed below:

- $\sigma_{CMP} = 3.2$ mV
- $\sigma_{SCC} = 0.8$ mV
- $\sigma_{AMB_SF} = 16.2$ mV
- $\sigma_{FD_SF} = 16.3$ mV

with σ_{CMP} being the standard deviation of the offset of the comparator CMP, σ_{SCC} the standard deviation of the output of the subtraction unit SCC, and σ_{AMB_SF} and σ_{FD_SF} the standard deviation of the output of the source followers or buffers in the analog memory bank AMB and in the 4T-APS, respectively. As seen, the comparator CMP and the different source followers in our pixel are the major sources of mismatch in our pixel. All these contributions accumulate along the datapath of the pixel, contributing to false event generation, which can be seen in Figure 8A,B that collect the

effect of mismatch on event generation for $S1$ and $S2$. The X-axis is the intensity change in percentage between two consecutive frames, while the Y axis shows the percentage of ON and OFF events from Monte Carlo simulations for different user-defined threshold voltages, namely, $V_{THevent} = 7, 17$ and 23 mV, which correspond to percentage changes in the light intensity of 0.9%, 2.6% and 3%. An ideal scenario is that of a sudden jump from 0% to 100% of events for a given threshold, shown with continuous lines in Figure 8A,B. Mismatch and temporal noise in actual circuits cause a minimum threshold to generate events.

We have run Monte Carlo simulations to emulate a whole array of global shutter HDR 4T-APS pixels for event generation. Every dot in the plots of Figure 8A,B is a percentage of 300 nominal Monte Carlo simulations that we have thought of as a percentage of pixels in an array yielding events. Thus, the plots in Figure 7 can be seen as a simulation of pixel fixed pattern noise (PFPN). The percentage % of intensity change per frame along the X-axis is calculated with the formula $(P1 - P2)/FSO_{Si}$, where $P1$ and $P2$ mean the light input power of two consecutive frames, and FSO_{Si} is the full-scale output of either signal $S1$ or $S2$. The results expressed as percentage of events in Figure 8A,B are similar to one another, but it should be taken into account that the power of signals $S1$ and $S2$ differ. The light input power is kept constant during the first frame ($P1$) at 117 lux for $S1$ signal (low illumination) and 4675 lux for $S2$ signal (high illumination), while it is subject to increasing and decreasing variations during the second frame to generate the percentage of illumination changes along the X-axis. The input light is simulated to be a green monochromatic light with a wavelength $\lambda = 555$ nm, making the conversion from photometric to radiometric units more straightforward. The integration time has been set to 1 ms. Finally, Figure 8A,B shows that the 100% of correct cases is only met with around of 5% of intensity change for $S1$ (conventional mode) and $S2$ (HDR mode) signals, as can be seen by the red points. Temporal noise adds more inaccuracies. As apparent, the application dictates the user-defined threshold voltage $V_{THevent}$. The simulations include leakages caused by off-resistance of switches and the ones caused by the inverse pn junctions of the transistors. Leakages caused by the impact of light, that is, parasitic light sensitivity, are not included.

3.2 | Temporal noise

The temporal noise has to be added to the spatial noise to determine the noise floor in our implementation in order to calculate the dynamic range of our HDR 4T-APS pixel for frame differencing.

3.2.1 | Circuit noise

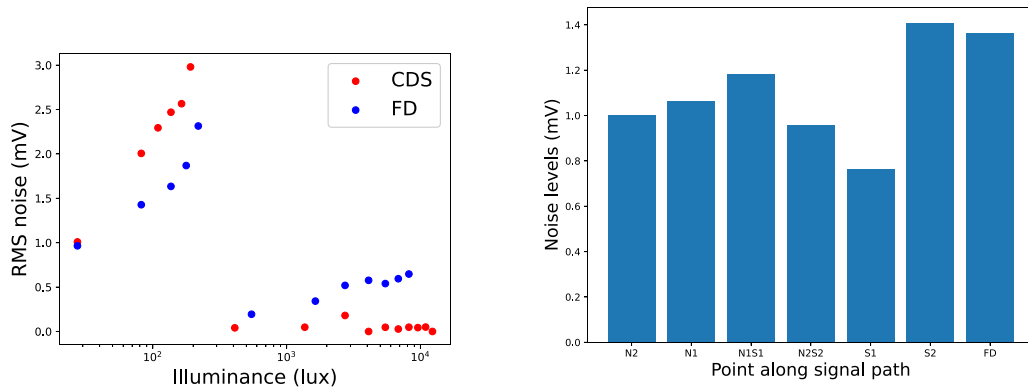
The effect of thermal noise in our implementation has been obtained by averaging 10,000 nominal transient noise simulations. $N1$ and $N1S1$ ($N1 + S1$ in Figure 3) are noise samples taken on the FD node of our circuit (Figure 2) at their corresponding time instants of the timing diagram of Figure 3, while $N2$ and $N2S2$ ($N2 + S2$ in Figure 3) are taken on the $FD + CS$ node. The subtraction of $N1S1$ and $N2S2$ from $N1, N2$ runs the CDS operation.

3.2.2 | Photon shot noise

Photon shot noise has also been added for a given input light, with the number of photons as $N_p = \frac{PT_e}{E_f}$, where P is the light input power with a given wavelength, in our case $\lambda = 555$ nm for an easy conversion from radiometric to photometric units, T_e is the integration time, and E_f is the energy of a single photon. From the number of photons, the Poisson distribution has been derived and added to our CAD simulator.

3.2.3 | Total noise

Figure 9A shows root mean squared (RMS) values of photon shot and circuit noise on the output node (V_S) of the subtraction circuit of Figure 2A, with and without CDS for a sudden transition from low illumination, where photon shot noise dominates, and the global shutter HDR 4T-APS works in the normal region with signal $S1$, to high illumination, where circuit noise prevails, and the pixel works within the HDR extension, with a lower conversion gain with signal $S2$. Circuit simulations do not account for parasitic light sensitivity and the effects of leakage currents caused by the



(A) RMS of total noise (circuit and photon shot noise). (B) RMS of temporal noise along the signal path.

FIGURE 9 Noise analysis of our HDR 4T-APS pixel for event generation. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/cia.3925)]

reset transistor of the subtraction unit.²¹ Nevertheless, we do not expect a large impact because the global shutter HDR 4T-APS works with reset between two consecutive frames, where integration times are usually in the order of ms, and, thus, short for the leakage currents to have a significant impact.⁸ Figure 9B shows the temporal noise along the data path of our HDR 4T-APS for event generation. The noise floor is estimated to be 0.5 mV_{rms} .

3.3 | Dynamic range

The estimation of dynamic range (DR) is given by (3). $E_{v_{sat2}}$ is the upper limit of S2 and after this limit signal N1 is saturated resulting in nonlinearity and false threshold $V_{THS1/S2}$ detection. $E_{v_{floornoise}}$ is the lowest illuminance that can be detected.

$$DR = 20 \cdot \log \left(\frac{E_{v_{sat2}}}{E_{v_{floornoise}}} \right) \quad (3)$$

The HDR extension raises the dynamic range from 53 dB up to 85 dB, which is lower than that of the original HDR pixel with overflow capacitor,¹⁶ which achieves 100 dB. This is due to two factors, the higher number of active elements in our circuit to generate events, which increases the noise floor, and the lower power supply voltage, 3.3 V versus 5 V.¹⁶

3.4 | Design considerations for CS capacitance

As stated in Akahane et al,¹⁶ it is important to maintain a high enough signal to noise ratio(SNR) at S1/S2 switching point. The signal to noise ratio of signal S2 at the S1/S2 switching point is given by equation (4) in which, $S2_{S1/S2}$ is the voltage level of the signal S2 at the switching point and $NS2_{S1/S2}$ is the noise of S2.

$$SN_{S2} = 20 \log \left(\frac{S2_{S1/S2}}{NS2_{S1/S2}} \right) \quad (4)$$

At the S1/S2 switching point signal S2 is

$$S2_{S1/S2} = \frac{C_{FD}}{C_{FD} + C_{CS}} \cdot S1_{S1/S2} \quad (5)$$

where $S1_{S1/S2}$ is the signal S1 at the S1/S2 switching point. C_{FD} is the capacitance of the FD node and C_{CS} is the capacitance of the CS capacitor. And the noise of S2 is given by

$$NS2 = \frac{qQ_{\mu}}{C_{FD} + C_{CS}} \quad (6)$$

where q is elementary charge and Q_{μ} number of residual charges still remaining after the CDS of S2.

The dynamic range is given by

$$DR = 20 \log \left(\frac{S2_{SAT} \cdot \frac{C_{FD} + C_{CS}}{C_{FD}}}{NS2_{S1/S2}} \right) \quad (7)$$

The sizes of C_{FD} and C_{CS} will affect SNR of S2 and DR of the pixel. From (4), we see that we increase the SNR by increasing the signal S2 at S1/S2 switching point. This can be done by changing the capacitance ratio $\frac{C_{FD}}{C_{FD} + C_{CS}}$. However, this will negatively affect the DR. Thus, a choice of the capacitance's is a matter of trade off between DR and SNR.

3.5 | Chip data and comparison with prior art

We have laid down a 64×64 pixel array in 180 nm CMOS technology. Our pixel is able to run at 1000 event frames per second (efps), measured as post-layout simulations. Our pixel pitch is $32.3 \mu\text{m}$. The photodiode size is $5.4 \times 5.4 \mu\text{m}^2$. The layout of the pixel with all its individual blocks labeled can be seen in Figure 10A. The layout of the complete chip is displayed on Figure 10B. The area of the chip is $2.81 \times 2.90 \text{ mm}^2$. The array is surrounded by row decoders on the left, row drivers on the right, column drivers on the top, and select circuits on the bottom side of the array.

Table 1 shows a comparison with prior art. Post-layout simulations show a higher power consumption in the range of 505–555 nW per pixel, depending on the incident light power. Our excess power consumption comes mainly from the in-pixel HDR algorithm. As an example, peak power consumption happens when input power generates S1 signal that is similar to the $V_{THS1/S2}$. A way to cut power consumption is the exchange of conventional source followers- used

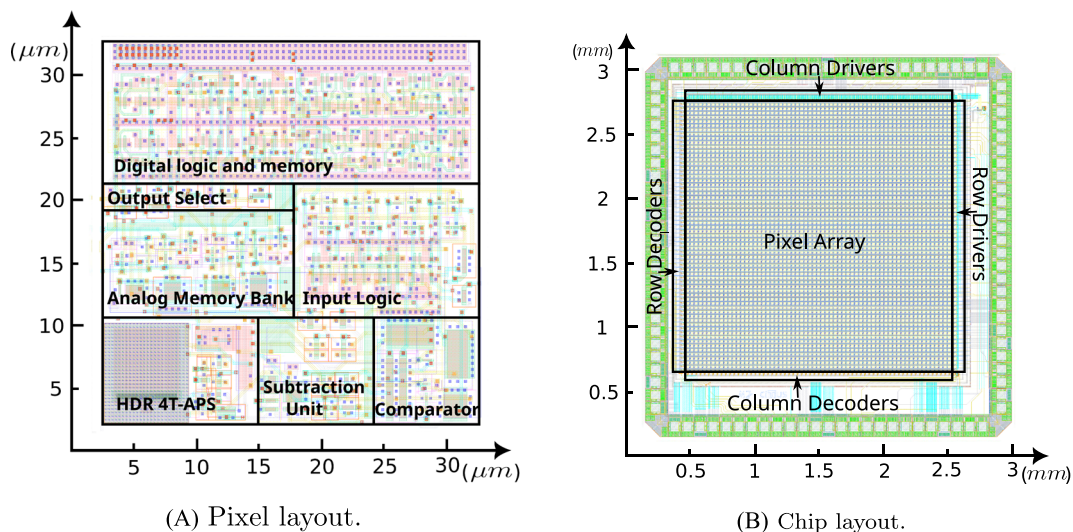


FIGURE 10 Pixel and chip layouts of our HDR 4T-APS for event generation. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/cia.3925)]

TABLE 1 Chip data and comparison with prior art.

	This work	ISSCC 2020 ^{a10}	TED ¹⁴	ISCAS 2020 ⁵	VLSI 2019 ⁶	ISSCC 2020 ^{b4}
Technology	CMOS	CMOS	FD-SOI	CMOS	CMOS	CMOS
Process	180 nm	180 nm	22 nm	65 nm	65 nm	90 nm
Resolution	64 × 64	64 × 64	-	1280 × 960	132 × 104	1280 × 720
Pixel Size	32.305 × 32.305 μm^2	15 × 15 μm^2	4 × 4 μm^2	4.95 × 4.95 μm^2	10 × 10 μm^2	4.86 × 4.86 μm^2
Fill Factor	2.8%	21%	76%	22%	20%	>77%
Supply	3.3 V/1.8 V analog 1.8 V digital	0.8 V	0.6 V	2.8 V/1.8 V analog 1.0 V digital	1.2 V	2.5/1.1 V
Max event rate/ Frame rate	1000 efps	510 fps	715 fps	1.3 Geps	180 Meps	1066 Meps
Power Consumption	505-555 nW	18.1 nW	-	122 nW	357 nW	35 nW
Dynamic range	85 dB	64.2 dB	-	-	-	>124 dB
Readout	Sequential	Sequential	Sequential	Sequential	Sequential	Asynchronous

in our implementation- for dynamic source followers (DSF). We have seen that the power consumption could be lowered in the range 32% to 45% with the adoption of DSFs, again according to the input power of the incident light.²² As seen in Table 1, this would make our chip in 180 nm competitive with that of reference,⁶ manufactured in 65 nm.

Compared to prior art- Table 1- our design suffers from a low fill-factor and an overhead pixel pitch caused mainly by the in-pixel HDR algorithm to deal with signals S_1 and S_2 , and by a less advanced technology node when compared to that of the solutions presented in the literature.^{5,14} Nevertheless, this can be partially alleviated on two-tier vertical technologies with the split of our pixel in a pattern of 2×2 photosites sharing common processing circuitry, similarly to previous event pixel sensors,^{5,23} and focal-plane processors by previous authors.^{8,19} The same methodology applied to our pixel by sharing the subtraction unit, the comparator and the logic unit would increase the fill-factor by 4 times, that is, from 2.8% up to approximately 11.2%. The remaining gap with the prior art in Table 1 would have to be tackled through a further effort in diminishing the area of the bulkiest blocks of the pixel, that is, digital memories and input logic, for example, through dynamic logic instead of the current static logic.

The dynamic range of our implementation falls below 120 dB, the state-of-the-art for event cameras. The dynamic range can be extended by increasing the size of the overflow capacitor CS . This would lead to a lower fill-factor, and to a decrease in the conversion gain, resulting in a trade-off between dynamic range and sensitivity. Also, as stated in Akahane et al,¹⁶ an increase in the power supply could be another method to extend the dynamic range of pixels with the overflow capacitance approach. Nevertheless, this solution cannot be adopted in a straightforward manner in our pixel due to its complexity with its many blocks and the interactions among them. This might lead to new modifications in our pixel, which might result in a worse fill-factor.

In summary, event generation through frame differencing with global shutter CMOS vision sensors with HDR through an overflow capacitor could be competitive in scenarios or applications with slow enough objects and a relatively high dynamic range while providing low power consumption. More demanding situations with fast moving objects and a very wide dynamic range would call for dynamic vision sensors.

4 | OUTLOOK AND CONCLUSION

This paper has delved into the two main CMOS options to run on-chip dynamic information from a scene, namely, global shutter pixels and dynamic vision sensors. We provide post-layout simulations of a global shutter 4T-APS with HDR extension for event generation by means of a local algorithm on an in-pixel overflow capacitor in order to cut the HDR edge of dynamic vision sensors on global shutter pixels. Our HDR mode extends the dynamic range from 53 to 85 dB, which, although it can be improved, is still far from the state-of-the-art value of 124 dB of dynamic vision

sensors. Additional circuits in global shutter pixels for HDR extension and event generation hamper the fill factor and increase the noise floor. Both figures of merit can be tackled through low power techniques like the inclusion of dynamic source followers instead of conventional ones, and the pixel split in several tiers with CMOS-3D technologies. Still, our solution provides low enough power consumption and it would be competitive in scenarios with slow enough moving objects and a relatively high dynamic range when compared to dynamic vision sensors.

ACKNOWLEDGEMENTS

This work has received funding from the projects: European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 860370; and the European Union (European Regional Development Fund): from the Xunta de Galicia-Consellería de Cultura, Educación e Ordenación Universitaria Accreditation 2019–2022 ED431G-2019/04 and Reference Competitive Group Accreditation 2021–2024, GRC2021/48. This work was also funded by Spanish Ministry of Science, Innovation and Universities under grant PID2021-128009OB-C32.

ORCID

Marko Jaklin  <https://orcid.org/0000-0002-6153-3412>

Daniel García-Lesta  <https://orcid.org/0000-0002-9895-9839>

Paula López  <https://orcid.org/0000-0002-8218-8945>

Victor M. Brea  <https://orcid.org/0000-0003-0078-0425>

REFERENCES

- Zhong X, et al. A fully dynamic multi-mode CMOS vision sensor with mixed-signal cooperative motion sensing and object segmentation for adaptive edge computing. *IEEE J Solid-State Circ.* 2020;55(6):1684-1697.
- Choi J, et al. Always-on CMOS image sensor for mobile and wearable devices. *IEEE J Solid-State Circ.* 2016;51(1):130-140.
- Park I, et al. A 640×640 fully dynamic CMOS image sensor for always-on operation. *IEEE J Solid-State Circ.* 2020;55(4):898-907.
- Finatue T, et al. 5.10 a 1280–720 back-illuminated stacked temporal contrast event-based vision sensor with 4.86 km pixels, 1.066geps readout, programmable event-rate controller and compressive data-formatting pipeline. In: 2020 IEEE International Solid-State Circuits Conference - (ISSCC); 2020:112-114.
- Suh Y, et al. A 1280–960 dynamic vision sensor with a 4.95- μm pixel pitch and motion artifact minimization. In: 2020 IEEE International Symposium on Circuits and Systems (ISCAS); 2020:1-5.
- Li C, et al. A 132 by 104 10 μm -pixel 250 μW 1kefps dynamic vision sensor with pixel-parallel noise and spatial redundancy suppression. In: 2019 Symposium on Vlsi Circuits; 2019:C216-C217.
- Lichtsteiner P, Posch C, Delbruck T. A 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor. *IEEE J Solid-State Circ.* 2008;43(2):566-576.
- García-Lesta D, et al. CMOS vision sensor for background subtraction. In: IEEE International Symposium on Circuits and Systems; 2020.
- Benetti M, et al. A low-power vision system with adaptive background subtraction and image segmentation for unusual event detection. *IEEE Trans Circ Syst I: Reg Pap.* 2018;65(11):3842-3853.
- Hsu T-H, et al. A 0.8V multimode vision sensor for motion and saliency detection with ping-pong pwm pixel. In: 2020 IEEE International Solid-State Circuits Conference - (ISSCC); 2020:110-112.
- Guo S, Delbruck T. Low cost and latency event camera background activity denoising. *IEEE Trans Pattern Anal Mach Intel.* 2023;45(1):785-795.
- Ma S-Y, Chen L-G. A single-chip CMOS APS camera with direct frame difference output. *IEEE J Solid-State Circ.* 1999;34(10):1415-1418.
- Milirud V, et al. A wide dynamic range CMOS active pixel sensor with frame difference. In: 2005 IEEE International Symposium on Circuits and Systems (ISCAS); 2005:588-591 Vol. 1.
- Liu L, et al. FD-SOI-based pixel with real-time frame difference for motion extraction and image preprocessing. *IEEE Trans Electron Dev.* 2023;70(2):594-599.
- Chi YM, et al. CMOS camera with in-pixel temporal change detection and ADC. *IEEE J Solid-State Circ.* 2007;42(10):2187-2196.
- Akahane N, et al. A sensitivity and linearity improvement of a 100 dB dynamic range CMOS image sensor using a lateral overflow integration capacitor. In: Digest of Technical Papers. 2005 Symposium on Vlsi Circuits, 2005; 2005:62-65.
- Gallego G, et al. Event-based vision: a survey. *IEEE Trans Pattern Anal Mach Intel.* 2022;44(1):154-180.
- Jaklin M, et al. HDR 4T-APS pixel for event generation by frame differencing. In: 2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS); 2021:921-924.
- Surez M, et al. Low-power CMOS vision sensor for Gaussian pyramid extraction. *IEEE J Solid-State Circ.* 2017;52(2):483-495.
- García-Lesta D, et al. In pixel analog memories for a pixel based background subtraction algorithm on CMOS vision sensors. *Int J Circ Theory Appl.* 2018;2018:1631-1647.
- Nozaki Y, Delbruck T. Temperature and parasitic photocurrent effects in dynamic vision sensors. *IEEE Trans Electron Dev.* 2017;64(8):3239-3245.

22. Jaklin M, et al. Low-power techniques on a CMOS vision sensor chip for event generation by frame differencing with high dynamic range. In: 2022 29th IEEE International Conference on Electronics, Circuits and Systems (ICECS); 2022:1-4.
23. Niwa A, et al. A 2.97 μm -pitch event-based vision sensor with shared pixel front-end circuitry and low-noise intensity readout mode. In: 2023 IEEE International Solid-State Circuits Conference (ISSCC) IEEE; 2023:4-6.

How to cite this article: Jaklin M, García-Lesta D, López P, Brea VM. Global shutter CMOS vision sensors and event cameras for on-chip dynamic information. *Int J Circ Theor Appl*. 2024;1-14. doi:[10.1002/cta.3925](https://doi.org/10.1002/cta.3925)