A 240x180 120dB 10mW 12us-latency sparse output vision sensor for mobile applications

Raphael Berner, Christian Brandli, Minhao Yang, Shih-Chii Liu and Tobi Delbruck

Inst. of Neuroinformatics, University of Zurich and ETH Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
tobi@ini.uzh.ch, +41 44 635 30 38

Abstract – **This paper proposes a CMOS vision sensor that combines event-driven asynchronous readout of temporal contrast with synchronous frame-based active pixel sensor (APS) readout of intensity. The image frames can be used for scene content analysis and the temporal contrast events can be used to track fast moving objects, to adjust the frame rate, or to guide a region of interest readout. Therefore the sensor is suitable for mobile applications because it allows low latency at low data rate and low system-level power consumption. Sharing the photodiode for both readout types allows a compact pixel design that is 60% smaller than a comparable design. The 240x180 sensor has a power consumption of 10mW. It is built in 0.18um technology with 18.5um pixels. The temporal contrast pathway has a minimum latency of 12us, a dynamic range of 120dB, 12% contrast detection threshold and 3.5% contrast matching. The APS readout has 55dB dynamic range with 1% FPN.**

I. Introduction

Conventional frame-based imagers are widely used in machine vision, but have drawbacks at the system level: The fixed frame rate limits the temporal resolution, the uniform pixel exposure time limits the dynamic range, and the repetitive sampling leads to redundant data and high power consumption. Asynchronous event-based dynamic vision sensors (**DVSs**) that report log intensity changes [1][2] circumvent all these drawbacks because their pixels asynchronously output address-events in response to scene reflectance changes. The DVS events represent novelty and automatically segment moving objects. The high dynamic range, sub-ms latency and sparse output of the DVS makes it useful in applications such as robotics [3][4] and real-time tracking [5]. However the DVS lacks information on static scene content, which is useful for object recognition and classification. The asynchronous time-based image sensor (ATIS, [6]) overcomes this limitation by adding an event-triggered PWM intensity readout to each DVS pixel. This PWM readout provides high dynamic range, but the PWM circuit effectively doubles the pixel area and requires an additional photodiode. It triples the amount of data to be transferred since every DVS event is followed by two PWM events encoding the intensity.

The proposed "apsDVS" sensor in this paper combines a DVS and an active pixel sensor at the pixel level (Fig. 1). It outputs image frames through the synchronous APS pathway and simultaneously outputs low latency temporal contrast events through the asynchronous DVS pathway using a shared photodiode. The shared photo-diode and small size of the APS circuit lead to an apsDVS pixel area that is 60% smaller than the ATIS pixel area in the same process.

Using the DVS pathway the sensor can be used to track and segment fast moving objects at low computational cost, while the APS output allows recognizing and classifying these objects. Because tracking is done using only DVS events, the frame rate of the APS output can be arbitrarily low. This makes the apsDVS sensor well suited for mobile applications with a tight power budget, because it allows low latency at low system-level power consumption.

II. Sensor Design

The temporal contrast computation on the pixel (DVS block in Fig. 1) is similar to [1]: the photo-current is logarithmically encoded by the photoreceptor output voltage Vpr. A source-follower buffer isolates the photoreceptor from the next stage. The self-timed switched-capacitor differencing amplifier amplifies the deviation from the last reset-level. This deviation Vdiff is continuously compared against two thresholds. As soon as either of the two comparator thresholds is crossed, an event is communicated to the periphery and the switched-capacitor amplifier is reset to store the new illumination level. An increase in light intensity leads to a so called ON event while a decrease produces an OFF event.

To communicate the DVS events, the sensor uses word-serial burst mode address event representation (AER) circuits evolved from [7]. If Vdiff crosses the threshold of either comparator, the pixel first requests in the row direction (RR). A non-greedy arbitration circuit [7] chooses among all requesting rows and acknowledges a single row (RA). In this row, all pixels that have crossed threshold assert the request

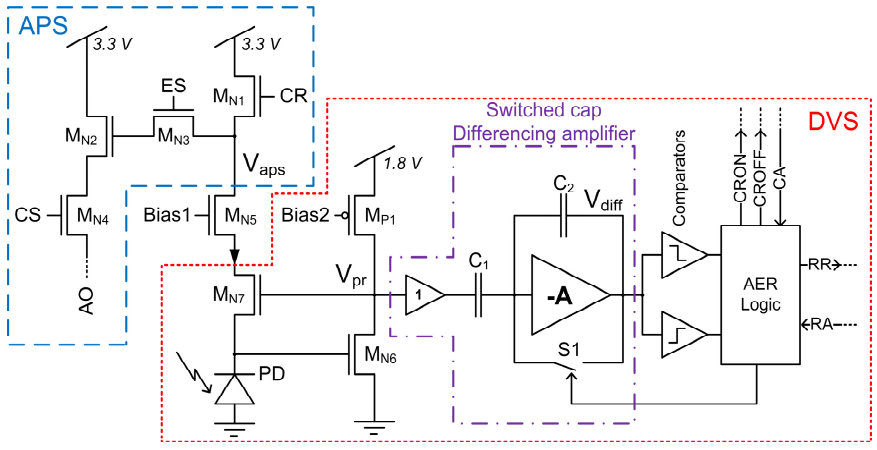

Fig. 1 apsDVS pixel schematic. Complete transistor level DVS schematics are in [1].

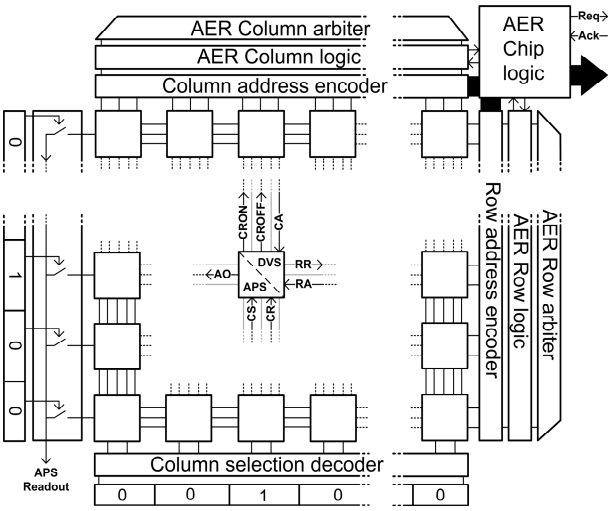

Fig. 2 apsDVS sensor readout architecture.

signal in the column direction (CRON or CROFF, depending on the polarity of the event). While the row address is transmitted off-chip, a small asynchronous state machine in each column latches the state of the request lines (requesting or not). A simplified arbitration tree chooses the leftmost requesting column and all addresses of the requesting columns are sequentially read out. All events in a column burst receive the same timestamp in microsecond resolution at the board level.

The pixel circuit allows simultaneous operation of the temporal contrast detection and the frame-based intensity readout without interfering with each other. Because the DVS photoreceptor (MN6/MN7/MP1) holds PD at virtual ground, the photocurrent is available at the drain of MN7, which allows the APS intensity readout with transistors MN1 to MN4. The cascode transistor MN5 protects the drain of MN7 from voltage transients due to the reset of Vaps. To prevent saturated pixels from affecting the DVS pathway, column reset CR is set to an intermediate voltage level instead of ground during integration. This ensures that MN7 is always in saturation.

The APS frame readout uses a single-tap rolling-shutter row-parallel analog output. After each column reset (CR) the reset levels are read out by selecting the column (CS). Later the signal level is read out by selecting the columns without resetting them. Correlated double sampling is performed in the digital domain by off-chip subtraction of signal and reset value. This readout scheme allows for increased dynamic range by multiple readouts during a single integration cycle. It has the disadvantage however that pixels integrate for up to one column readout time before the reset or readout level is sampled.

The 5 mm x 5mm chip (Fig. 6) is built in a standard (not CIS) UMC 0.18um 1P6M process with MiM capacitors. The pixel layout has a quad-mirrored form so that source contacts and bias lines can be shared, and the analog and digital parts of the pixel can be grouped together. The pixel pitch is 18.5um and the fill factor is 22%. An n-well is used as a photo-diode. The chip includes a configurable bias generator circuit [8] that generates PVT tolerant biases and that allows complete control of pixel and readout biases, which is essential for this complex pixel architecture.

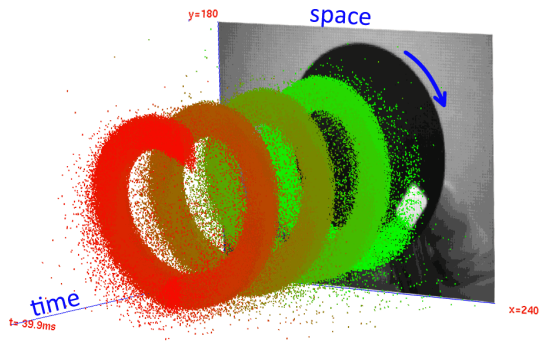

Fig. 3 Space-time 3D view of DVS events during 40ms of a white rectangle spinning on a black disk at 100Hz. Green events are older and red events are newer.

APS readout uses an external 10-bit ADC and control logic (implemented on a CPLD) for APS readout and DVS event address capture and time-stamping. This data is sent to a host controller through a high-speed Cypress FX2 USB interface. For a detailed analysis of the pixel function, a fully instrumented test pixel was added to the chip along with numerous digitally-selectable test points in the AER logic and bias generator circuits.

III. Experimental Results

The open-source software project jAER [9] is used to control, display, record, and process the output of the sensor.

Fig. 3 demonstrates the fast nature of the DVS output by plotting the events as points in 3D space-time: These events which are produced by a white rectangle spinning at 100Hz, form a helix when displayed in 3-D. In the background, a snapshot of the actual scene taken through the APS readout is displayed. The asynchronous nature of the temporal contrast output produces a stream of events that captures motion at speeds exceeding 40k pixels/s. The spinning rectangle produces a continuous stream of information at the position of its contours, allowing continuous and fast tracking with a high temporal resolution.

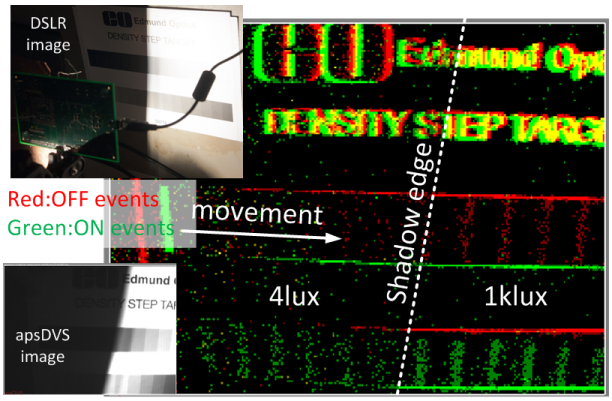

Fig. 4 A 60ms snapshot of accumulated DVS events from a scene of a moving Edmund density step chart. Insets show the DSLR image of the scene and the apsDVS APS frame.

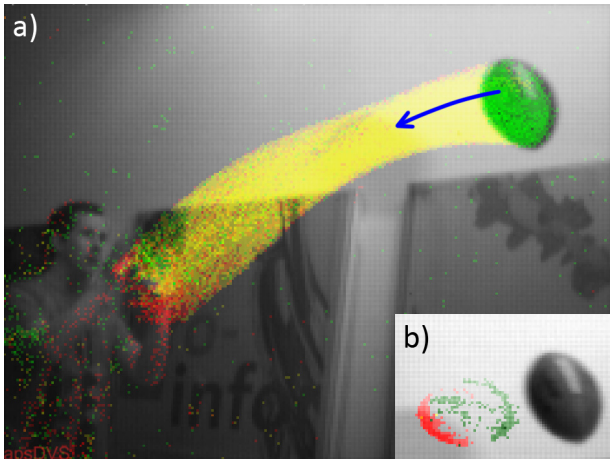

Fig. 5 (a) Combined 378ms snapshot showing an APS frame (gray) and the sparse OFF and ON DVS events (green/red) of a flying ball caught by a person. Almost all of the DVS events originate from the moving ball or the person. (b) A 5ms slice of DVS events of the ball in the air, taken 75ms after the APS snapshot which is also displayed here. The image of the ball is blurred by the APS exposure

Fig. 4 shows the high dynamic range of the DVS output in response to a moving Edmund density step chart on which a strong shadow of 250:1 contrast is cast. The DVS pathway still responds to 0.1 density contrast edges on both sides of the shadow, while the DSLR used to capture the inset photo saturates on the bright side. Although the APS dark current is high owing to the intervening transistors, the DVS circuit accesses the nwell photodiode directly, resulting in a DVS pathway dynamic range of more than 120dB, extending down to 0.1lux operation for high contrast inputs.

Fig. 5a shows the combined output of both readouts for a scene of an American football flying into the hands of a person. The sparse DVS output events (colored) mostly originate from the moving ball while most pixels remain silent. Fig. 5b demonstrates the temporal accuracy and resolution of the DVS events: a 5ms slice of DVS events (colored), acquired 75ms after the APS frame highlights the contours of the ball. The equivalent frame rate of the DVS is twice the pixel bandwidth (24kfps at 1.5klux scene illumination).

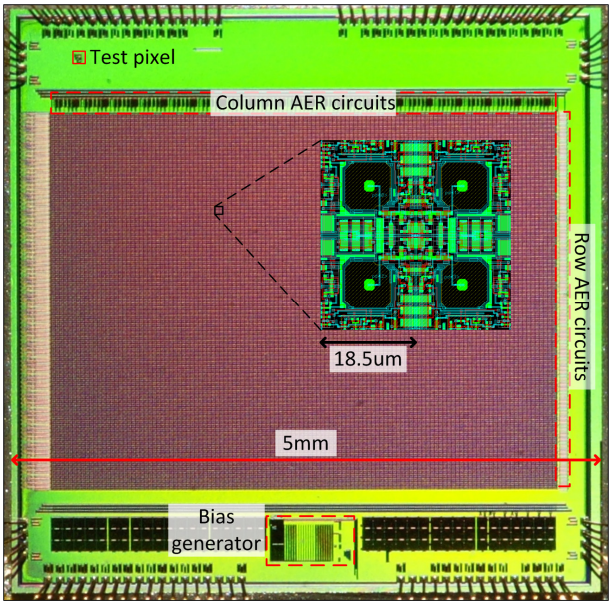

Fig. 6 Chip microphotograph and pixel layout of 2 by 2 pixels

Several shortcomings of the sensor are being addressed in a new design which is currently in fabrication. The rolling shutter readout is not ideal for some machine vision tasks because it produces motion artifacts. A minor checkerboard FPN observed in the APS readout is probably an artifact of an asymmetry of the quad-mirrored layout. The n-well photodiode is replaced by an optimized CMOS image sensor photodiode in the new design.

IV. Conclusion

With its compact frame-based image readout, the apsDVS enhances the computational advantages of the event-based asynchronous data-driven dynamic vision sensor. The sparse DVS events are well-suited for tracking and for automatic region of interest detection within the intensity frame on which more computationally-intensive conventional machine vision algorithms can then be applied. This dual readout sensor can thereby reduce the computational load and latency of vision systems. These advantages may outweigh the limited performance possibilities of the APS readout which currently cannot support in-pixel charge transfer and which contributes extra dark current.

Fig. 6 shows the chip photograph and pixel layout. Using the shared photo-diode, the intensity readout adds only 4 transistors and thus only about 10% additional area to the DVS pixel. Fig. 7 compares the specifications of the apsDVS with those of other event-driven vision sensors. This 240x180 sensor achieves a minimum DVS latency of 12us and a power consumption of 10mW. The DVS pathway has a dynamic range of 120dB with 12% contrast detection threshold and 3.5% contrast matching. The APS readout has a 55dB dynamic range with 1% FPN using readout of the reset level and a single integration value.

V. Acknowledgements

This research was supported by the European Union funded project SeeBetter (FP7-ICT-2009-6), the Swiss National Science Foundation through the NCCR Robotics, ETH Zurich, and the University of Zurich.

VI. References

[1] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128×128 120 dB 15 µs Latency Asynchronous Temporal Contrast Vision Sensor", ***IEEE J. Solid State Circuits***, vol.43, pp. 566-576, 2008.

[2] T. Serrano-Goterredona and B. Linares-Barranco, "A 128x128 1.5% Contrast Sensitivity 0.9% FPN 3us Latency 4mW Asynchronous Frame-Free Dynamic Vision Sensor Using Transimpedance Preamplifiers", ***IEEE J. Solid State Circuits***, vol. 48, Mar. 2013.

[3] J. Conradt, R. Berner, M. Cook, T. Delbruck, "An Embedded AER Dynamic Vision Sensor for Low-Latency Pole Balancing", in 5th IEEE Workshop on Embedded Computer Vision, ICCV, Kyoto, pp. 1-6, 2009.

[4] T. Delbruck and P. Lichtsteiner, "Fast sensory motor control based on event-based hybrid neuromorphic-procedural system",

	This work	Posch et al. [6]	Serrano-Gotarredona et al. [2]
Functionality	**async. temporal contrast+APS**	async. temporal contrast +level crossing intensity	async. temporal contrast
CMOS Technology	**0.18um 1P6M MIM**	0.18um 1P6M MIM	0.35um 2P4M
Chip size mm^2	**5 x 5**	9.9 x 8.2	4.9 x 4.9
Array size	**240 x 180**	304 x 240	128 x 128
Pixel size um^2	**18.5 x 18.5**	30 x 30	31 x 30
Fill factor	**22%**	20%,10%	10.5 %
Pixel complexity	**44 transistors, 2 MIM caps, 1 photodiode**	77 transistors, 3 caps, 2 photodiodes	N.A.
Supply voltage	**1.8V / 3.3V (pixels)**	3.3V analog, 1.8V digital	3.3V
Power consumption high activity low activity	 **13.5mW** **7.4mW**	 175mW 50mw	 4mW
Dynamic range	**120dB DVS (0.1 lx), 55dB APS**	Intensity 125dB, DVS N.A.	120dB (60dB intrascene)
Min. Contrast sensitivity	**12%**	13% @ 100lx, 30% @1klux	1.5%
FPN	**1 % APS, DVS 3.5% contrast mismatch**	<0.25% intensity, DVS N.A.	DVS 0.9%
Max. bandwidth	**30 MEvents/s, 25 fps**	N.A.	20 MEvents/s
Min. latency	**12us @ 1klux (mean of 20 pixels)**	<4us @ 1klux (fastest pixel)	3.2us @ 25klux (fastest pixel)

Fig. 7 Specifications compared to prior event-driven vision sensors

in *IEEE Intl. Symp. on Circuits and Systems (ISCAS)*, pp. 845-848, 2007.

[5] Z. Ni, et al., "Asynchronous Event-Based High Speed Vision for Microparticle Tracking", ***J. of Microscopy*****,** vol.245, pp 236-244, 2011.

[6] C. Posch, D. Matolin, and R. Wohlgenannt, "A QVGA 143 dB Dynamic Range Frame-Free PWM Image Sensor With Lossless Pixel-Level Video Compression and Time-Domain CDS", ***IEEE J. Solid State Circuits*****,** vol. 46, pp. 259-275, 2011.

[7] K. Boahen, "A Burst-Mode Word-Serial Address-Event Link-I: Transmitter Design", ***IEEE Trans. on Circuits and Systems I: Regular Papers*** , vol. 51, pp. 1269-1280, Jul. 2004.

[8] M. Yang, et al., "Addressable Current Reference Array with 170dB Dynamic Range", in *IEEE Intl. Symp. on Circuits and Systems (ISCAS)* pp. 3110-3113, May 2012.

[9] jAER Open Source Project. Available: http://jaerproject.net