Activity-Driven, Event-Based Vision Sensors

Tobi Delbrück[1], Bernabe Linares-Barranco[2], Eugenio Culurciello[3], Christoph Posch[4],
[1]Inst. of Neuroinformatics,UNI-ETH Zürich, Switzerland, [2]Center for Microelectronics, Sevilla,
[3]Dept. of Electrical Engineering, Yale Univ., [4]Austria Inst. of Technology GmbH, Vienna

Abstract - The four chips [1-4] presented in the special session on "Activity-driven, event-based vision sensors" quickly output compressed digital data in the form of events. These sensors reduce redundancy and latency and increase dynamic range compared with conventional imagers. The digital sensor output is easily interfaced to conventional digital post processing, where it reduces the latency and cost of post processing compared to imagers. The asynchronous data could spawn a new area of DSP that breaks from conventional Nyquist rate signal processing. This paper reviews the rationale and history of this event-based approach, introduces sensor functionalities, and gives an overview of the papers in this session. The paper concludes with a brief discussion on open questions.

1. RATIONALE

The well-known leader in image sensor development Eric Fossum [6] stood up at the International Image Sensor Workshop in 2005 and defined "the perfect image sensor" as having "infinite resolution, dynamic range, and frame rate, together with zero pixel size and power consumption." This statement illustrates the divide between camera image producers and machine vision consumers: The output from this ideal sensor would also be infinitely expensive to process. Biology teaches that real-world vision relies on a sensor (the retina) that does local gain control and massive amounts of computation to produce at its output (the optic nerve), an asynchronous stream of digital data (spikes) which represents only the relevant information for vision. Work on event-based vision sensors extends Fossum's ideal by including one more metric: A "perfect vision sensor" should have perfect mutual information between the bits of the output and the vision problem to be solved. Because it depends so much on the problem, this definition is hard to pin down and must be proven by successful market application of the sensors.

The notion that binds work on event-based neuromorphic systems is a desire to emulate biology's use of asynchronous, exceedingly sparse, data-driven digital signaling as a core aspect of its computational architecture. Address-event representation (AER) systems naturally provide a way to incorporate demand-based computation, where data originating in one place drives computation in another. The high speed of silicon electronics allows communication of sparse, low frequency asynchronous digital address-events by sharing high speed digital buses. Simultaneous events from multiple sources share the bus, sometimes by arbitration [7] and in other designs by collision detection [8], or polling (e.g. [9]).

AER vision sensors aim to encapsulate "particles of visual information" in their output events (Fig. 1), so that each event carries more useful information than a gray level. The events should also be timely to take advantage of the asynchronous representation. At the same time, the pixel size should be affordable, the fill factor reasonable, and the pixel variability acceptable.

2. HISTORY

The first AER vision system was built by Mahowald and her colleagues in her outstanding 1992 PhD work [10]. The performance of early systems suffered because they had to simultaneously combine a new computational paradigm with tricky delay-insensitive asynchronous logic and massively parallel analog computation. After Mahowald's AER retina, only Boahen's group [11] published anything significant for the next 10 years. Aside from the pioneering work of Lazzaro on audition [12], work up to about 2005 was aimed at building complete neuromorphic systems, i.e. no conventional processor was wanted (e.g. [13-15]). In 2003 there was an impressive JSSC paper from Ruedi et al. [16] on a

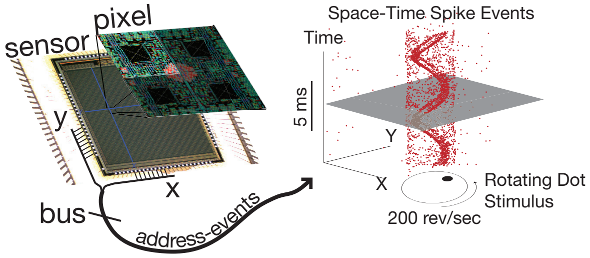

Fig. 1 AER vision sensor concept. This example from the dynamic vision sensor [5] shows how a high speed stimulus generates a sparse, asynchronous digital stream of address-events which rapidly signify changes in scene reflectance.

978-1-4244-5309-2/10/$26.00 ©2010 IEEE

simultaneous spatial contrast and local orientation vision sensor. This paper showed that it was possible by innovative circuit and architecture design to overcome the mismatch and low performance that plagued the prior work, and enabled the first industrial applications of AER sensor outputs [17]. The development of the dynamic vision sensor (DVS) in 2006 [5] opened up the field for dynamic vision problems.

3. TYPES OF AER VISION SENSORS

Although gray-level image sensors that use the AER protocol have been proposed (e.g. [18-23]), these simply use the AER protocol to transmit pixel intensity values, but have the drawback of costing expensive silicon pixel area without providing the benefit of either redundancy or latency reduction. Broadly divided, AER "silicon retinas" which do have these functionalities fall into the following classes:

- **Spatial contrast (SC) sensors** which reduce spatial redundancy based on intensity *ratios*, vs. **spatial difference (SD) sensors** which use intensity *differences*. SC is more useful with varying scene illumination while SD is cheaper to implement.
- **Temporal contrast (TC) sensors** which reduce temporal redundancy based on *relative* intensity change, vs. **temporal difference (TD) sensors** which use *absolute* intensity change. TC is more useful with non-uniform scene illumination but more expensive to implement than TD, especially with AE sensors (see next).

The exposure, readout, and pixel reset mechanisms can be lumped into two classes:

- **Frame Event (FE) sensors**, which use a synchronous exposure of all pixels and then schedule the event readout in order of presumed relevance, e.g. in order of SC or based on detected TD.
- **Asynchronous Event (AE) sensors**, which have autonomous pixels that continuously generate events based on a local decision about relevance, e.g. TC.

Contrast here means Weber contrast as opposed to Raleigh contrast, so that a uniform spatial or unchanging temporal input produces no events. Additional classifications and combinations are possible but are omitted here. Table 1 compares some specifications of published work, including the classifications from above. Three of the papers presented in this session are included in the table.

4. OVERVIEW OF PAPERS

The four papers in this special session show results from functional silicon and three of them are from complete systems.

Posch et al. [1] at the Austria Inst. of Technology, present a breakthrough in functionality for AER vision sensors, with the first asynchronous time-based image sensor (ATIS). It combines in each pixel the notions of TC detection [5] with PWM intensity encoding [24], using a new time-based correlated double sampling circuit [25], to output pixel gray level values only from pixels that change. Pixel gray levels can also be polled. Gray level is encoded by the time between the TC event and the PWM event from the same pixel. The resolution of 304x240 pixels in 180nm technology is high for this field which has typically been academically funded. The high dynamic range and temporal redundancy suppression, combined with the low latency of the temporal contrast events, will find application in real-time vision in robotics, networked surveillance, and high DR medium resolution video e.g. in scientific applications.

Linares-Barranco's group in Sevilla [2] present a second generation prototype [26] for AE sensing of signed SC (Weber contrast). Reported spatial sensors (either SD or SC), except for [16] (which is FE type) have been plagued by poor pixel matching. In this paper, new mechanisms for in-pixel calibration are demonstrated which allow for settings a fine threshold for minimum-detectable SC across the array. It continues the running theme from this lab in evolving the use of local calibration technology. Computed spatial contrast is coded as pixel event frequency (rate coding), but the chip also includes a global pixel reset mechanism for **TTFS** (Time-to-First-Spike) coding.

Culurciello's group at Yale [3] report a compact 20x20um^2 pixel with 40% fill factor. It has three operating modes—TD, SD and intensity readout—all at a low power consumption of 1mW, making it attractive for wireless sensor networks, like [27].

Delbruck and Berner's paper [4] is more preliminary and reports a single pixel design for low-contrast TC detection, which is aimed at applications in bio-luminescence. The test pixel can generate events for 0.3% contrast change, which is a factor of about 50 better than previous designs. This pixel, however, will need to be sped-up to work at realistic light levels.

5. DISCUSSION AND OUTLOOK

Of the many open questions, we list only the following:

1. Why has no one built a usable color AER vision sensor? Color is a basic sense for all animals.

Table 1 Comparison between AER vision sensor devices. Pixel size is given both in lambda (the scaling parameter) and um units. Power consumption is at chip and not board or system level. Best metrics are in bold. (Extended from [5].)

	Prior work						This session		
Year	2001	2003	2005	2006	2008	2009	2010	2010	2010
Source	Zaghloul, Boahen [30]	Rüedi et al. [16]	Mallik et al.[9, 32]	Lichtsteiner et al. [5, 33]	Massari et al. [27]	Ruedi et al.[31]	Posch et al. 2010 [34]	Linares-Barranco et al. [2]	Culurciello et al. [3]
Functionality	Asynchronous spatial and temporal contrast,	Frame-based spatial contrast and gradient direction, ordered output	Temporal frame-difference intensity change detection APS imager	Asynchronous temporal contrast dynamic vision sensor (DVS)	Binary spatial and temporal contrast	Digital log pixel + RISC proc.	Async. Time-based Image Sensor (ATIS)	Async. Weber Contrast (SC), with either rate or TTFS coding	Temporal intensity change or spatial difference can trigger readout
Type (Sec.3)	SC TD AE	SC FE	TD FE	TC AE	SD TD FE	SC, embedded	TC AE	SC AE	TD SD FE
Gray picture output			●			●	●	●	●
Pixel size um (lambda)	34x40 (170x200)	69x69 (276x276)	25x25 (**100x100**)	40x40 (200x200)	26x26.5 (130x130)	**14x14** (311x311)	30x30 (333x333)	80x80 (400x400)	16x21 (??)
Fill factor (%)	14%	9%	17%	8.1%	20%	20%	10%(TC)/20%(gray)	2.5%	**42%**
Fabrication process	0.35um 4M 2P	0.5um 3M 2P	0.5um 3M 2P	0.35um 4M 2P	0.35um 4M 2P	**180nm** 1P6M	**180nm** 4M 2P MIM	0.35um 4M 2P	**180nm** SiGe BiCMOS 7M
Pixel complexity T=MOS,C=cap	38T	>50T, 1C	**6T (NMOS) 2C**	26T(14 anal), 3C	45T	~80T, 1C	77T, 4C, 2PD	131T, 2C	**11T**
Array size	96x60	128x128	90x90	128x128	128x64	**320x240**	304x240	32x32	128x128
Die size mm^2	3.5x3.5	~10x10	3x3	6x6.3	11	5.2x8.4	9.9x8.2	2.5x2.6	??
Power consumption	62.7mW @ 3.3V	300mW @ 3.3V	30mW @ 5V (50 fps)	24mW @ 3.3V	**100uW@2V, 50fps**	**80mW** (11mW sensor)	50-175mW	0.66-6.6mW	<**1.4mW**@3V
Dynamic range	~50dB	120dB	51dB	120dB 2lux to >100 klux scene	100dB	**132dB** 6V/lux s 39dB SNR	**143dB** (static) 125dB@30FPS 56dB SNR	100dB 1lx to 100klx	2V/s/(uW/cm^2) @550nm. 1.14uV/e$^-$
PD dark current@25C	?	300fA	?	4fA (~10nA/cm^2)	NA	44mV/s	1.6nA/cm^2	NA	
Response latency, frames/sec (fps), events/sec (eps)	~10Meps	< 2ms 60 to 500 fps	< 5ms? 200 fps?	15µs @ 1 klux chip illumination 2Meps	Max 4000fps	30fps	**3.2us**@1klux 30Meps peak, 6Meps sustained	100us @50klx 66Meps	200-800fps dep. on mode 13Meps
FPN matching	1-2 decades	2% contrast	**0.5%** of full scale, 2.1% TD change	2.1% contrast	10% contrast	**0.8%**		0.87% contrast	

Attempts so far based on stacked junctions [28, 29] have had limited success owing to the weak color separation capability of this wavelength-dependent absorption. Conventional color filter technology has only recently become available from multi-project wafer services; but the cost (~$20k) for a prototype run still places casual experiments out of reach.

2. How can we build a pixel array that reduces redundancy across the multiple stimulus dimensions of spatial, temporal, and spectral contrast simultaneously? And do so with reasonable fill factor, pixel size, and matching? This may need to await 3d stacked wafer technology, require heterogeneous arrays [30], or be moved to be next to the pixel array instead of inside it [31].

3. Given a good AER data stream, how do we process the data? Leaving aside for now the longer range goal of processing in AER hardware (e.g. as in successors to [13]), here we can draw on the neuroscience and machine vision communities, but only if they are supplied with user-friendly implementations and open-source software APIs. We have tried to jump-start this process [35, 36], but much remains to be done in terms of standardization and community building.

4. Is frame-free the way to go? Going "frame-free" costs about 10 transistors per pixel. It requires a local decision and reset mechanism in the pixels and can rob the array of a synchronous reset signal which can be used to put all the pixels simultaneously in a known state, after which they can compare themselves to globally broadcast, possibly time-varying references, as is done in [16, 24, 31, 37]. But three big advantages of being frame-free are 1) the latency of response depends on the pixel, not the frame rate, 2) the pixels can run at their optimal rate for detecting signals, rather than being forced to a premature decision and 3) the use of bandwidth should scale to larger sizes more readily.

6. CONCLUSION

The field of AER vision sensor design demonstrates a trend over the past 5 years to show results that make it into applications such as surveillance (e.g. [38-41]) and high speed robotics [42, 43]. Eight papers on event based sensing have been accepted to the very competitive IEEE International Solid State Circuits Conference since 2003 (two on auditory sensors [44, 45] and six on vision sensors [9, 16, 27, 33, 46, 47]), showing that this approach is starting to impact mainstream electronics. We expect that the next few years will bring substantial progress in this approach to vision.

7. REFERENCES

[1] C. Posch, *et al.*, "High-DR Frame-Free PWM Imaging with asynchronous AER Intensity Encoding and Focal-Plane Temporal Redundancy Suppression," presented at the ISCAS, 2010.

[2] J. A. Leñero-Bardallo, *et al.*, "A Signed Spatial Contrast Event Spike Retina Chip," presented at the ISCAS, 2010.

[3] E. Culurciello and D. Kim, "A compact-pixel tri-mode image sensor," presented at the ISCAS 2010, 2010.

[4] T. Delbruck and R. Berner, "Temporal Contrast AER Pixel with 0.3%-Contrast Event Threshold," in *ISCAS 2010*, 2010.

[5] P. Lichtsteiner, *et al.*, "A 128×128 120dB 15us Latency Asynchronous Temporal Contrast Vision Sensor," *IEEE J. Solid State Circuits,* vol. 43, pp. 566-576, 2008.

[6] E. R. Fossum, "CMOS image sensors: electronic camera-on-a-chip," *Electron Devices, IEEE Transactions on,* vol. 44, pp. 1689-1698, 1997.

[7] K. A. Boahen, "A burst-mode word-serial address-event Link-III: Analysis and test results," *IEEE Transactions on Circuits and Systems I-Regular Papers,* vol. 51, pp. 1292-1300, Jul 2004.

[8] A. Mortara, "A pulsed communication/computation framework for analog VLSI perceptive systems," in *Neuromorphic Systems Engineering*, T. S. Lande, Ed., ed Norwell, MA: Kluwer Academic, 1998, pp. 217-228.

[9] U. Mallik, *et al.*, "Temporal change threshold detection imager," in *ISSCC Dig. of Tech. Papers*, San Francisco, 2005, pp. 362-363.

[10] M. A. Mahowald, "VLSI analogs of neuronal visual processing: a synthesis of form and function," PhD, Computation and Neural Systems, Caltech, Pasadena, California, 1992.

[11] K. Boahen, "A retinomorphic chip with parallel pathways: Encoding INCREASING, ON, DECREASING, and OFF visual signals," *Analog Integrated Circuits and Signal Processing,* vol. 30, pp. 121-135, Feb 2002.

[12] J. Lazzaro, *et al.*, "Silicon auditory processors as computer peripherals," *IEEE Trans.on Neural Networks,* vol. 4, pp. 523-528, 1993.

[13] R. Serrano-Gotarredona, *et al.*, "CAVIAR: A 45k Neuron, 5M Synapse, 12G Connects/s AER Hardware Sensory–Processing– Learning–Actuating System for High-Speed Visual Object Recognition and Tracking," *IEEE Trans. on Neural Networks,* vol. 20, pp. 1417-1438, 2009.

[14] T. Y. W. Choi, *et al.*, "Neuromorphic implementation of orientation hypercolumns," *IEEE Transactions on Circuits and Systems I-Regular Papers,* vol. 52, pp. 1049-1060, Jun 2005.

[15] E. Chicca, *et al.*, "A multi-chip pulse-based neuromorphic infrastructure and its application to a model of orientation selectivity," *IEEE Transactions on Circuits and Systems, part I (TCAS-I),* vol. 54, pp. 981-993, 2006.

[16] P. F. Ruedi, *et al.*, "A 128x128, pixel 120-dB dynamic-range vision-sensor chip for image contrast and orientation extraction," *IEEE Journal of Solid-State Circuits,* vol. 38, pp. 2325-2333, DEC 2003.

[17] E. Grenet, *et al.*, "High dynamic range vision sensor for automotive applications," *Proceedings of the SPIE, Multispectral and Hyperspectral Remote Sensing Instruments and Applications II, Edited by Larar, Allen M.; Suzuki, Makoto; Tong, Qingxi,* vol. 5663, pp. 246-253, 2005.

[18] J. G. Harris, "The changing roles of analog and digital signal processing in CMOS image sensors," in *Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASSP '02). IEEE International Conference on*, 2002, pp. IV-3976-IV-3979 vol.4.

[19] E. Culurciello, *et al.*, "A biomorphic digital image sensor," *IEEE Journal of Solid-State Circuits,* vol. 38, pp. 281-294, Feb 2003.

[20] E. Culurciello and R. Etiene-Cummings, "Second generation of high dynamic range, arbitrated digital imager," in *2004 International Symposium on Circuits and Systems (ISCAS 2004)*, Vancouver, Canada, 2004, pp. 828-831.

[21] M. Azadmehr, *et al.*, "A Foveated AER Imager Chip," in *2005 International Symposium on Circuits and Systems (ISCAS 2005)*, 2005, pp. 2751- 2754.

[22] C. Shoushun and A. Bermak, "Arbitrated Time-to-First Spike CMOS Image Sensor With On-Chip Histogram Equalization," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on,* vol. 15, pp. 346-357, 2007.

[23] G. Xiaochuan, *et al.*, "A Time-to-First-Spike CMOS Image Sensor," *Sensors Journal, IEEE,* vol. 7, pp. 1165-1175, 2007.

[24] X. Qi, *et al.*, "A Time-to-first-spike CMOS imager," in *2004 International Circuits and Systems Conference (ISCAS2004)*, Vancouver, Canada, 2004, pp. 824-827.

[25] D. Matolin, *et al.*, "True correlated double sampling and comparator design for time-based image sensors," in *Circuits and Systems, 2009. ISCAS 2009. IEEE International Symposium on*, 2009, pp. 1269-1272.

[26] J. Costas-Sanos, *et al.*, "A contrast retina with on-chip calibration for neuromorphic spike-based AER vision systems," *IEEE Trans. on Circuits and Systems-I,* vol. 54, pp. 1444-1458, 2007.

[27] N. Massari, *et al.*, "A 100uW 64×128-Pixel Contrast-Based Asynchronous Binary Vision Sensor for Wireless Sensor Networks," in *IEEE ISSCC Dig. of Tech. Papers*, 2008, pp. 588-638.

[28] J. A. M. Olsson and P. Hafliger, "Two color asynchronous event photo pixel," in *IEEE International Symposium on Circuits and Systems, 2008 (ISCAS 2008)*. 2008, pp. 2146 - 2149.

[29] R. Berner, *et al.*, "Self-timed vertacolor dichromatic vision sensor for low power face detection," in *ISCAS 2008*, Seattle, 2008, pp. 1032-1035.

[30] K. A. Zaghloul and K. Boahen, "Optic nerve signals in a neuromorphic chip II: Testing and results," *IEEE Transactions on Biomedical Engineering,* vol. 51, pp. 667-675, Apr 2004.

[31] P. F. Ruedi, *et al.*, "An SoC combining a 132dB QVGA pixel array and a 32b DSP/MCU processor for vision applications," in *IEEE ISSCC Dig. of Tech. Papers*, 2009, pp. 46-47,47a.

[32] Y. Chi, *et al.*, "CMOS camera with in-pixel temporal change detection and ADC," *IEEE JOURNAL OF SOLID-STATE CIRCUITS,* vol. 42, pp. 2187-2196, OCT 2007 2007.

[33] P. Lichtsteiner, *et al.*, "A 128×128 120dB 30mW Asynchronous Vision Sensor that Responds to Relative Intensity Change," in *ISSCC Dig. of Tech. Papers*, San Francisco, 2006, pp. 508-509 (27.9).

[34] C. Posch, *et al.*, "High DR, low data-rate imaging based on an asynchronous, self-triggered Address- Event PWM array with pixel-level temporal redundancy suppression," in *ISCAS 2010*, 2010.

[35] T. Delbruck. (2007, *jAER open source project*. Available: http://jaer.wiki.sourceforge.net

[36] T. Delbruck, "Frame-free dynamic digital vision," in *Proceedings of Intl. Symp. on Secure-Life Electronics, Advanced Electronics for Quality Life and Society*, Tokyo, 2008, pp. 21-26.

[37] M. Barbaro, *et al.*, "A 100 x 100 pixel silicon retina for gradient extraction with steering filter capabilities and temporal output coding," *IEEE Journal of Solid-State Circuits,* vol. 37, pp. 160-172, FEB 2002.

[38] M. Litzenberger, *et al.*, "Embedded Vision System for Real-Time Object Tracking using an Asynchronous Transient Vision Sensor," in *IEEE Digital Signal Processing Workshop 2006*, Grand Teton, Wyoming, 2006, pp. 173-178.

[39] A. Belbachir, *et al.*, "Estimation of Vehicle Speed Based on Asynchronous Data from a Silicon Retina Optical Sensor," in *IEEE Intelligent Transportation Systems Conference ITSC 2006*, Toronto, 2006, pp. 653-658.

[40] M. Litzenberger, *et al.*, "Vehicle Counting with an Embedded Traffic Data System using an Optical Transient Sensor," in *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*, 2007, pp. 36-40.

[41] Z. Fu, *et al.*, "An Address-Event Fall Detector for Assisted Living Applications," *IEEE Transactions on Biomedical Circuits and Systems,* vol. 2, pp. 88-96, 2008.

[42] T. Delbruck and P. Lichtsteiner, "Fast sensory motor control based on event-based hybrid neuromorphic-procedural system," in *ISCAS 2007*, New Orleans, 2007, pp. 845-848.

[43] J. Conradt, *et al.*, "An Embedded AER Dynamic Vision Sensor for Low-Latency Pole Balancing," in *5th IEEE Workshop on Embedded Computer Vision (in conjunction with ICCV 2009)*, Kyoto, Japan, 2009.

[44] R. Sarpeshkar, *et al.*, "An Analog Bionic Ear Processor with Zero-Crossing Detection," in *ISSCC Dig. of Tech. Papers*, 2005, pp. 78-79.

[45] B. Wen and K. Boahen, "A 360-channel speech preprocessor that emulates the cochlear amplifier," in *ISSCC Dig. of Tech. Papers*, 2006, pp. 556-557.

[46] C. Posch, *et al.*, "A Dual-Line Optical Transient Sensor with On-Chip Precision Time-Stamp Generation," in *Solid-State Circuits Conference, 2007. ISSCC 2007. Digest of Technical Papers. IEEE International*, 2007, pp. 500-618.

[47] C. Posch, *et al.*, "A QVGA 143dB DR Asynchronous Address-Event PWM Dynamic Vision and Image Sensor with Lossless Pixel-Level Video Compression and Time-Domain CDS," in *ISSCC Dig. of Tech. Papers*, 2010, p. in press.