

A 128×128 120 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor

Patrick Lichtsteiner, *Member, IEEE*, Christoph Posch, *Member, IEEE*, and Tobi Delbruck, *Senior Member, IEEE*

Abstract—This paper describes a 128×128 pixel CMOS vision sensor. Each pixel independently and in continuous time quantizes local relative intensity changes to generate spike events. These events appear at the output of the sensor as an asynchronous stream of digital pixel addresses. These address-events signify scene reflectance change and have sub-millisecond timing precision. The output data rate depends on the dynamic content of the scene and is typically orders of magnitude lower than those of conventional frame-based imagers. By combining an active continuous-time front-end logarithmic photoreceptor with a self-timed switched-capacitor differencing circuit, the sensor achieves an array mismatch of 2.1% in relative intensity event threshold and a pixel bandwidth of 3 kHz under 1 klux scene illumination. Dynamic range is > 120 dB and chip power consumption is 23 mW. Event latency shows weak light dependency with a minimum of 15 μs at > 1 klux pixel illumination. The sensor is built in a 0.35 μm 4M2P process. It has $40 \times 40 \mu\text{m}^2$ pixels with 9.4% fill factor. By providing high pixel bandwidth, wide dynamic range, and precisely timed sparse digital output, this silicon retina provides an attractive combination of characteristics for low-latency dynamic vision under uncontrolled illumination with low post-processing requirements.

Index Terms—Address-event representation (AER), asynchronous vision sensor, high-speed imaging, image sensors, machine vision, neural network hardware, neuromorphic circuit, robot vision systems, visual system, wide dynamic range imaging.

I. INTRODUCTION

THE notion of a “frame” of video data has become so embedded in machine vision that it is usually taken for granted. This is natural given that frame-based devices have been dominant from the days of drum scanners and videcon tubes to today’s CCDs and CMOS imagers. There are undeniable advantages to frame-based imagers: They have small simple pixels, leading to high resolution, large fill factor, and low imager cost. The output format is well understood and is the basis for many years of research in machine vision.

On the other hand, frame-based architectures carry hidden costs because they are based on a series of snapshots taken at a constant rate. The pixels are sampled repetitively even if their values are unchanged. Short-latency vision problems require

high frame rate and produce massive output data (e.g., > 1 GB/s from 352×288 pixels at 10 kFPS in [1]). Pixel bandwidth is limited to half of the frame rate, and reducing the output to a manageable rate by using region-of-interest readout usually requires complex control strategies. Dynamic range is typically limited by the identical pixel gain, the finite pixel capacity for integrated photocharge, and the identical integration time. For machine vision in uncontrolled environments with natural lighting, limited dynamic range and bandwidth can compromise performance.

In this paper, we elaborate on [2] to describe a vision sensor whose pixels respond asynchronously to relative changes in intensity. The sensor output is an asynchronous stream of pixel address-events (AEs) that directly encode scene reflectance changes, thus reducing data redundancy while preserving precise timing information. These properties are achieved by abandoning the frame principle and modeling three key properties of biological vision: its sparse, event-based output, its representation of relative luminance change (thus directly encoding scene reflectance change), and its rectification of positive and negative signals into separate output channels. The proposed device improves on prior frame-based temporal difference detection imagers (e.g., [3]) by asynchronously responding to temporal contrast rather than absolute illumination, and on prior event-based imagers because they either do not reduce redundancy at all [4], reduce only spatial redundancy [5], have large fixed pattern noise (FPN), slow response, and limited dynamic range [6], or have low contrast sensitivity [7]. The prototype sensor has already been used successfully for various applications: high-speed robotic target tracking [8], traffic data acquisition [9], [10], and in internal work for tracking particle motion in fluid dynamics, tracking the wings of fruitflies, eye-tracking, and stereo vision based on temporal correlation.

The rest of this paper is organized as follows. After a review of the asynchronous communication protocol and of prior work, Section II describes the vision sensor design. Section III shows characterization results. Section IV concludes the paper.

A. Address-Event Representation

The basic idea of an asynchronous vision sensor is that the output is in the form of address-events (AEs, encoding the x, y -address of the pixel in the array) that are generated locally by the pixels. Pixels individually quantize the analog vision signal, usually after local gain control and spatial-temporal redundancy reduction. The output is thus in the form of an *address-event representation* (AER). This architecture arises from a merging of biology—where there are many parallel nerve fibers carrying continuous-time digital impulses—with

Manuscript received November 17, 2006; revised August 31, 2007. This work was supported by the University and ETH Zurich, EU FP5 project CAVIAR (IST-2001-34124), and by Austria Research Centers—Seibersdorf Research.

P. Lichtsteiner and T. Delbruck are with the Institute of Neuroinformatics, UNI-ETH Zurich, CH-8057 Zurich, Switzerland (e-mail: patrick@ini.phys.ethz.ch; tobi@ini.phys.ethz.ch).

C. Posch is with Austria Research Centers—Seibersdorf Research (ARC-sr), Vienna, Austria (e-mail: christoph.posch@arcs.ac.at).

Digital Object Identifier 10.1109/JSSC.2007.914337

TABLE I
COMPARISON OF TMPDIFF128 WITH OTHER DEVICES. PIXEL SIZE IS GIVEN BOTH IN LAMBDA (THE SCALING PARAMETER) AND μm UNITS.
POWER CONSUMPTION IS AT CHIP LEVEL, NOT BOARD OR SYSTEM LEVEL. MISMATCH IS COMPLEX;
THE SINGLE METRIC REPORTED HERE IS NOT FULLY DESCRIPTIVE IN SOME CASES

	TEMPDIFF128	Rüedi et al.[5]	Zaghloul, Boahen [13]	Kleinfelder et al.[1]	Mallik et al.[3]
<i>Functionality</i>	Asynchronous temporal contrast	Frame-based spatial contrast and gradient direction, ordered output	Asynchronous spatial and temporal contrast	In-pixel ADC APS imager	Temporal change detection APS imager
<i>Pixel size um (lambda)</i> <i>Fill factor (%)</i>	40x40 (200x200) 8.1% (PD area 151μm²)	69x69 (276x276) 9%	34x40 (170x200) 14%	9.4x9.4 (104x104) 15%	25x25 (100x100) 17%
<i>Fabrication process</i>	4M 2P 0.35um	3M 2P 0.5um	4M 2P 0.35um	5M 2P 0.18um	3M 2P 0.5um
<i>Pixel complexity</i>	26 transistors (14 analog), 3 capacitors	> 50 transistors 1 capacitor	38 transistors	37 transistors	6 transistors, NMOS 2 capacitors
<i>Array size</i>	128x128	128x128	96x60	352x288	90x90
<i>Die size mm²</i>	6x6.3	~ 10x10	3.5x3.5	5x5	3x3
<i>Interface</i>	15-bit word-parallel AER	8-bit bus, 16 x 24-bit FIFO, Non-arbitrated with collision detection	8-bit word-serial AER	64-bit (8-pixel) bus, 167 MHz	Serial, with event FIFO
<i>Power consumption</i>	24mW @ 3.3V 1.5mA core 0.3mA logic 5.5mA biases	300mW @ 3.3V	62.7mW @ 3.3V	50mW @ 3.3V (10kfps)	30mW @ 5V (50 fps)
<i>Dynamic range</i>	120dB 2 lux to > 100 klux scene illumination with f/1.2 lens	120dB	~50dB	~45dB	51dB
<i>Photodiode dark current at room temperature</i>	4fA (~10nA/cm²) Nwell photodiode	300fA	?	10nA/cm ²	?
<i>Response latency</i> <i>Frames/se or bandwidth</i>	15μs @ 1 klux chip illumination ~1M events/sec	< 2ms 60 to 500 fps	? ~10M events/sec	100us 10k fps	< 5ms? 200 fps?
<i>FPN, matching</i>	2.1% contrast	2% contrast	1-2 decades	0.8% of full scale before off-chip digital CDS	0.5% of APS full scale 2.1% change

silicon technology, which has the capability of building high-speed asynchronous digital buses [11].

B. Prior Work

The field of AER sensors is largely unexplored. Table I quantitatively compares our device with some existing AER vision sensors. The main obstacles to advancement have been unfamiliarity with asynchronous logic and very poor uniformity of pixel response characteristics. Industry is unfamiliar with non-frame-based vision sensors and understandably wary of large pixels with relatively small fill factors.

The first AER vision sensor was built by Mahowald and Mead [12]. This silicon retina incorporated adaptive photoreceptors, a spatial smoothing network, and self-timed communication. It was a demonstration device that was unusable for any real world task.

Zaghloul and Boahen [13] incorporated both sustained and transient types of cells with adaptive spatial and temporal filtering. This design comes closest to capturing key adaptive features of biological retinas. It is achieved by the use of small-transistor log-domain circuits that are tightly coupled spatially by diffuser networks. However, this circuit design style led to large mismatch: the pixel firing rates vary by a standard deviation of 1–2 decades and more than half the pixels do not spike at all for stimuli with 50% contrast. In addition, the use

of a passive phototransistor current-gain mechanism limits the dynamic range to about 2.5 decades and leads to a small bandwidth, particularly at low illumination. This chip was intended as a model of biology more than as a practical device.

The group at CSEM Neuchatel [5] presented a device that is closest to being dual in functionality to the one reported here in that its output encodes spatial rather than temporal contrast: After a global frame integration period, this device transmits events in the order of high-to-low spatial contrast. Thus, readout can be aborted early if limited processing time is available without losing information about high-contrast features. Each contrast event is followed by another event that encodes gradient orientation. This device has low 2% contrast mismatch and a large 6 decade dynamic range. They are presently in commercial development for automotive applications [14]. The main limitation of this architecture is that it does not reduce temporal redundancy (compute temporal derivatives), and its temporal resolution is limited to the frame rate.

Etienne-Cumming's group reported a temporal change threshold detection imager [3], which modifies the traditional active pixel sensor (APS) CMOS pixel so that it can detect a quantized absolute change in illumination. This synchronous device stores the addresses of pixels that signal change in a FIFO, making a new type of synchronous AER sensor. It has the big advantage that it offers a normal APS mode with small NMOS-only pixels, but the disadvantages of limited 2.5 decade

dynamic range and absolute—rather than relative—illumination-change threshold, meaning that the single threshold is only useful when the scene illumination is very uniform. It is also frame based, so the event times are quantized to the limited global sample rate.

Culurciello and Andreou [15] reported several imaging sensors that use AER to communicate the pixel intensity, either by inter-event interval or mean frequency. They have the advantage of relatively small pixel size, but the big disadvantage that the bus bandwidth is allocated according to the local scene luminance. Because there is no reset mechanism and because the event interval directly encodes intensity, a dark pixel can take a long time to emit an event, and a single highlight in the scene can saturate the bus.

Other recent developments include the time-to-first-spike (TTFS) imager [16] and the time-based imager [17] from Harris's group, a foveated AER vision sensor [18] from Häfliger's group a spatial-contrast AER retina [19] with in-pixel digitally programmed offset current calibration from Linares-Barranco's group, and a double line sensor based on the pixel reported here [20].

Kramer *et al.* [7], [21] reported the predecessors to the chip described here. The problem with these devices that led to the present development is mismatch in the transistor feedback elements, which makes it difficult to set a low contrast threshold across a large array. In addition, the leakage current in the feedback element results in a significantly non-zero corner frequency, i.e., the devices could not be adjusted to respond to very slow changes.

II. VISION SENSOR DESIGN

This section will describe the vision sensor design, starting with the pixel and then more briefly describing the rest of the chip design.

A. Pixel Design

The objective for this pixel design was to achieve low mismatch, wide dynamic range, and low latency in a reasonable pixel area. We met these challenges with a fast logarithmic photoreceptor circuit, a differencing circuit that amplifies changes with high precision, and cheap two-transistor comparators. Fig. 1(a) shows how these three components are connected.

The photoreceptor circuit has the desirable properties that it automatically controls individual pixel gain (by its logarithmic response) while at the same time responding quickly to changes in illumination. The drawback of this photoreceptor circuit is that transistor threshold variation causes substantial DC mismatch between pixels, necessitating calibration when this output is used directly [22], [23].

The DC mismatch is removed by balancing the output of the differencing circuit to a reset level after the generation of an event. The gain of the change amplification is determined by the well-matched capacitor ratio C_1/C_2 . The effect of inevitable comparator mismatch is reduced by the precise gain of the differencing circuit.

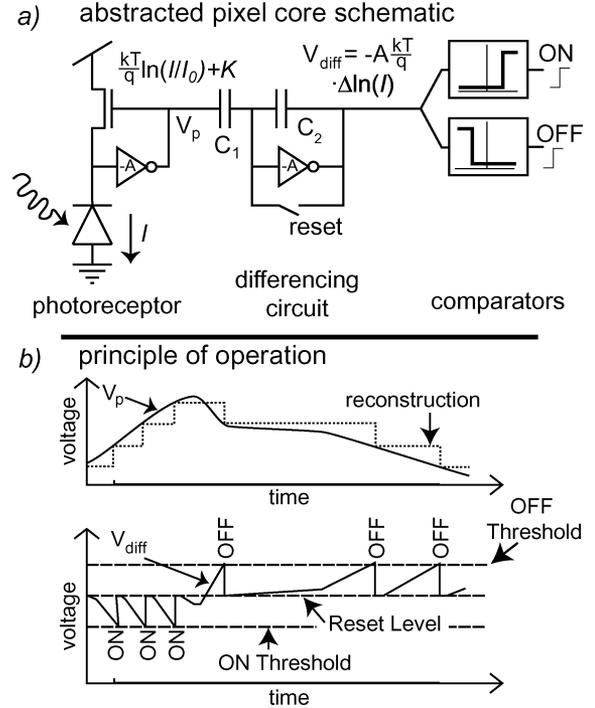


Fig. 1. (a) Abstracted pixel schematic. (b) Principle of operation. In (a), the inverters are symbols for single-ended inverting amplifiers.

Because the differencing circuit removes DC and due to the logarithmic conversion in the photoreceptor, the pixel is sensitive to temporal contrast $TCON$, which we define as

$$TCON = \frac{1}{I(t)} \frac{dI(t)}{dt} = \frac{d(\ln(I(t)))}{dt} \quad (1)$$

where I is the photocurrent. (The units of I do not affect $d(\log I)$). Fig. 2(b) illustrates the principle of operation of the pixel. In the rest of this section, we will consider in detail the operation of these component parts of the pixel circuit (Fig. 2).

The photoreceptor circuit comprises a photodiode whose photocurrent is sourced by a saturated NMOS transistor M_{fb} . The gate of M_{fb} is connected to the output of an inverting amplifier (M_{pr} , M_{cas} , M_n) whose input is connected to the photodiode. This well-known transimpedance configuration (see, e.g., [24]) converts the photocurrent logarithmically into a voltage and also holds the photodiode clamped at a virtual ground. The bandwidth of the photoreceptor is extended by the factor of the loop gain in comparison to a passive logarithmic photoreceptor circuit. This extended bandwidth is beneficial for high-speed applications, especially in low lighting conditions. Additionally, this photoreceptor circuit includes the option of adaptive biasing. Using a fraction of the low-pass-filtered sum of the photocurrents of all pixels to directly generate the bias voltage for M_{pr} [25] can reduce power consumption and maintain a constant resonance (constant quality factor Q) of the photoreceptor.

The photoreceptor output V_p is buffered with a source follower to V_{sf} to isolate the sensitive photoreceptor from the rapid transients in the differencing circuit. The source follower drives the capacitive input of the differencing circuit. The following

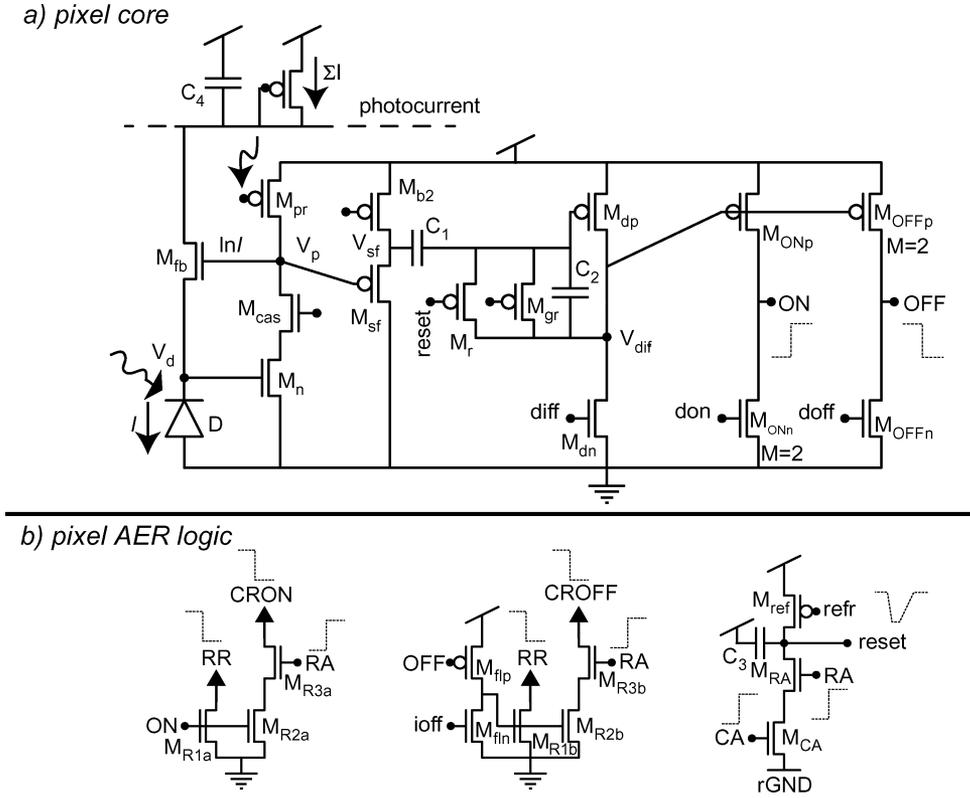


Fig. 2. Complete pixel circuit. (a) Transistor-level pixel schematic corresponding to the abstract schematic in Fig. 1(a). (b) Asynchronous logic circuits of the pixel. Transistor W/L ($\mu\text{m}/\mu\text{m}$) and capacitor values are as follows: M_{fb} 2/2, M_{pr} 1.6/5.6, M_{cas} 2/1.2, M_n 2/1.2, M_{b2} 1.2/1.2, M_r/M_{gr} 0.4/0.35, $M_{d*}/M_{ON*}/M_{OFF*}$ 1.5/3.2, M_{ref} 1.2/2.2, M_{fn} 1.2/2.4, other M are 0.4/0.6. $C_1 = 467$ fF, $C_2 = 24$ fF, $C_3 = 32$ fF. Using nominal bias currents, the gain of the photoreceptor feedforward amplifier using the cascode is about 500 and the open loop gain of the differencing amplifier and comparators is about 400.

capacitive-feedback inverting amplifier is balanced with a reset switch that shorts its input and output together, resulting in a reset voltage level.

A direct relation between temporal contrast $TCON$ and V_{diff} is given by

$$\begin{aligned} \Delta V_{diff} &= -A \cdot \Delta V_{sf} = -A \cdot \kappa_{sf} \cdot \Delta V_p \\ &= -A \frac{U_T \kappa_{sf}}{\kappa_{fb}} \ln \left(\frac{I(t + \Delta t)}{I(t)} \right) \\ &= -A \frac{U_T \kappa_{sf}}{\kappa_{fb}} \Delta \ln(I) \\ &= -A \frac{U_T \kappa_{sf}}{\kappa_{fb}} \int_t^{t+\Delta t} TCON(t') dt' \end{aligned} \quad (2)$$

where $A = C_1/C_2$ is the differencing circuit gain, U_T is the thermal voltage, and κ_X is the subthreshold slope factor of transistor M_X .

The comparators (M_{ONn} , M_{ONp} , M_{OFFn} , M_{OFFp}) compare the output of the inverting amplifier against global thresholds that are offset from the reset voltage to detect increasing and decreasing changes. If the input of a comparator overcomes its threshold, an ON or OFF event is generated.

Replacing in (2) ΔV_{diff} by comparator input thresholds don and $doff$ and solving for $\Delta \ln I$ yields the threshold positive and

negative temporal contrasts θ_{on} and θ_{off} that trigger ON or OFF events

$$\begin{aligned} \theta_{on} &= \Delta \ln(I)_{\min, ON} \\ &= \frac{\kappa_{fb}}{\kappa_{sf} \kappa_{ONp} U_T A} \\ &\quad \cdot (\kappa_{ONn} (don - diff) + U_T \ln 2) \end{aligned} \quad (3)$$

$$\begin{aligned} \theta_{off} &= \Delta \ln(I)_{\min, OFF} \\ &= \frac{\kappa_{fb}}{\kappa_{sf} \kappa_{OFFp} U_T A} \\ &\quad \times (\kappa_{OFFn} (doff - diff) - U_T \ln 2), \end{aligned} \quad (4)$$

where $don - diff$ is the ON threshold and $doff - diff$ is the OFF threshold; note that these equations take into account that M_{ONn} is $2M_{dn}$ and M_{OFFp} is $2M_{dp}$.

The threshold temporal contrast θ has dimensions of $\ln(\text{intensity})$ and is hereafter called *contrast threshold*. For smoothly varying temporal contrasts, the rate of generated ON and OFF events can be approximated with

$$f(t) = \text{Event Rate}(t) \approx \frac{TCON(t)}{\theta} = \frac{1}{\theta} \frac{d}{dt} \ln(I). \quad (5)$$

The ON and OFF events are communicated to the periphery by the circuits in Fig. 2(b) that implement the 4-phase AE handshaking with the peripheral AE circuits shown in Fig. 3(a). The row and column ON and OFF request signals (RR, CRON, CROFF) are generated individually, while the acknowledge

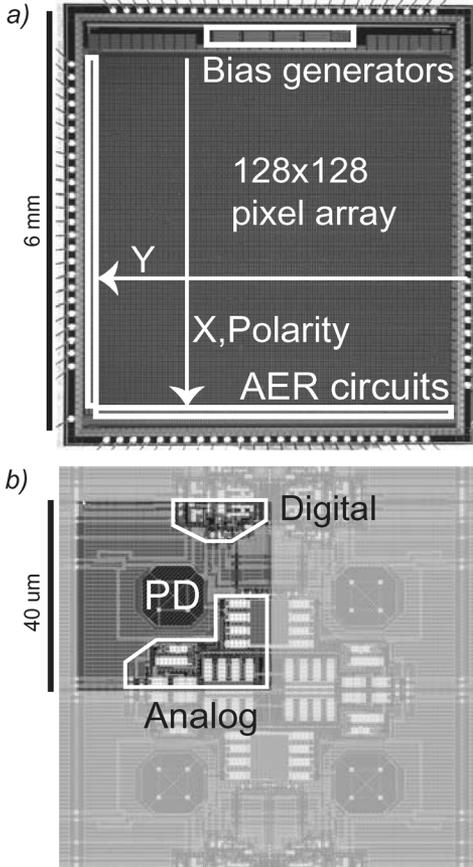


Fig. 4. (a) Die photo of the $0.35 \mu\text{m}$ 4M2P process chip. (b) Pixel layout is quad-mirror-symmetric with photodiode (PD) and analog and digital parts of the pixel. Most of the rest of the pixel is occupied by capacitance.

metal cut over the PD overlaps the n-well edge slightly to protect nFETs from parasitic photocurrent.

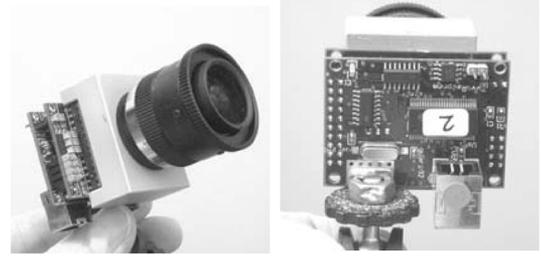
E. Interfacing to External Devices

The vision sensor can be directly interfaced to other AER components that use the same word-parallel protocol (this sensor is part of the CAVIAR multi-chip AER vision system [31]) or can be readily adapted to other AER protocols using simple commodity logic circuits. Our latest implementation (Fig. 5) streams time-stamped address-events to a host PC over a high-speed USB2.0 interface based on the Cypress FX2LP. On the host side, there is considerable complexity in acquiring, rendering, and processing the non-uniformly-distributed, asynchronous retina events in real time on a hardware single-threaded platform like most PCs. We developed an infrastructure consisting of several hundred Java classes in order to capture retina events, monitor them in real time, control the on-chip bias generators, and process the retina events for applications [32].

III. CHARACTERIZATION

Here we discuss characterization of the most important aspects of device operation: uniformity, dynamic range, pixel bandwidth, latency, and latency jitter.

a) Vision sensor system



b) Vision sensor USB interface

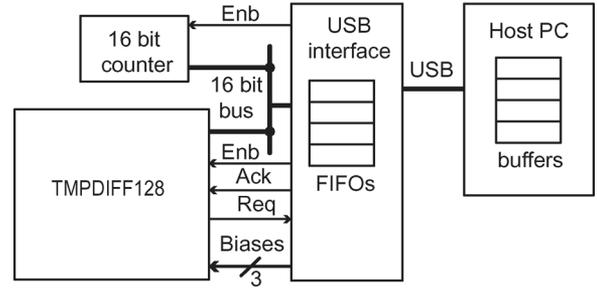


Fig. 5. Present implementation of the TMPDIFF128 camera system with USB2.0 interface. (a) Vision sensor system. (b) Schematic view of the USB hardware and software interface. The vision sensor (TMPDIFF128) sends AEs to the USB interface, which also captures time-stamps from a free-running counter running at 100 kHz that shares the same 16-bit bus. These time-stamped events are buffered by the USB FIFOs to be sent to the Host PC. The PC also buffers the data in USB driver FIFOs, “unwraps” the 16-bit time-stamps to 32-bit values, and offers this data to other threads for further processing. The same USB chip also uses a serial interface to control the vision sensor biases. Flash memory on the USB chip stores persistent bias values.

A. Uniformity of Response

For standard CMOS image sensors, the FPN characterizes the uniformity of response. For this vision sensor, the equivalent measure is the pixel-to-pixel variation σ_θ in the contrast threshold θ , which was introduced in Section II-A. θ depends on the settings of the comparator thresholds and σ_θ is due to pixel-to-pixel mismatch. We define contrast threshold mismatch σ_θ as follows:

$$\sigma_\theta = \text{standard deviation of threshold } \theta. \quad (6)$$

The dominant source of mismatch is expected to be found in the relative mismatch between differencing circuit reset level and comparator thresholds because: 1) device mismatch for transistors is in the order of 30% while capacitor mismatch is only in the order of 1%; 2) the amplifiers are simple two transistor devices without offset compensation; 3) the front-end steady-state mismatch is eliminated by differencing; and 4) gain mismatch (κ mismatch) in the front-end is expected to be in the order of 1%.

To measure the variation in event threshold, we use a black bar with linear gradient edges (reducing effects of the refractory period) which are moved at constant projected speed of about 1 pixel/10 ms through the visual field of the sensor. To quantify the pixel mismatch we counted events over a sequence of 40 stimulus presentations. Fig. 6 shows a histogram of events per pixel per stimulus edge for six different threshold settings.

We can measure the threshold mismatch from the width of these distributions combined with the known stimulus contrast

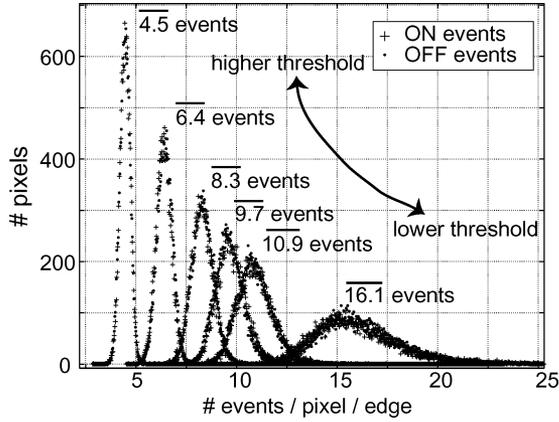


Fig. 6. Distributions in the number of events recorded per pass of the bar for 40 repetitions of the 15:1 contrast bar sweeping over the array, e.g., for the highest threshold setting there are an average of 4.5 ON and 4.5 OFF events per ON and OFF edge.

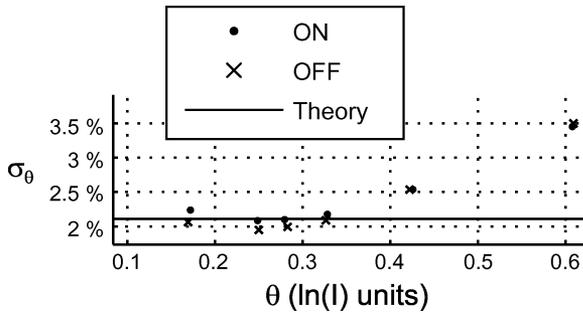


Fig. 7. Standard deviation of measured contrast threshold in % change of illumination plotted as a function of contrast threshold in $\ln I$ units. The line shows predicted threshold mismatch for 10 mV relative comparator mismatch.

of 15:1. Assuming an event threshold $\theta = \Delta \ln(I)$, (with the ON and OFF thresholds, here assumed identical), and a threshold variation σ_θ , we can compute (7) in that an edge of log contrast $C = \ln(I_{\text{bright}}/I_{\text{dark}})$ will make $N \pm \sigma_N$ events:

$$N \pm \sigma_N = \frac{C}{\theta \pm \sigma_\theta} \approx \frac{C}{\theta} \left(1 \mp \frac{\sigma_\theta}{\theta} \right). \quad (7)$$

From (7), we can compute expressions (8) for θ and σ_θ :

$$\begin{aligned} \theta &= \frac{C}{N} \\ \sigma_\theta &= \frac{\sigma_N}{N} \theta. \end{aligned} \quad (8)$$

C is measured from the stimulus; N and σ_N are measured from the histograms. Fig. 7 plots σ_θ versus θ . The solid line shows the analytically predicted mismatch [(3) and (4)] of 2.1% assuming a 10 mV relative mismatch in the comparator thresholds. For low θ (10%–40% illumination change), σ_θ lies between 2% and 2.5% (mean of 2.1%). For higher θ , σ_θ increases, suggesting that gain mismatch starts to become important.

Temporal contrast resolution is reduced for large temporal contrasts (e.g., fast moving, strong headlights of a car in nighttime) because of the refractory period. Fast, high-contrast stimuli produce fewer events than the idealized model given above.

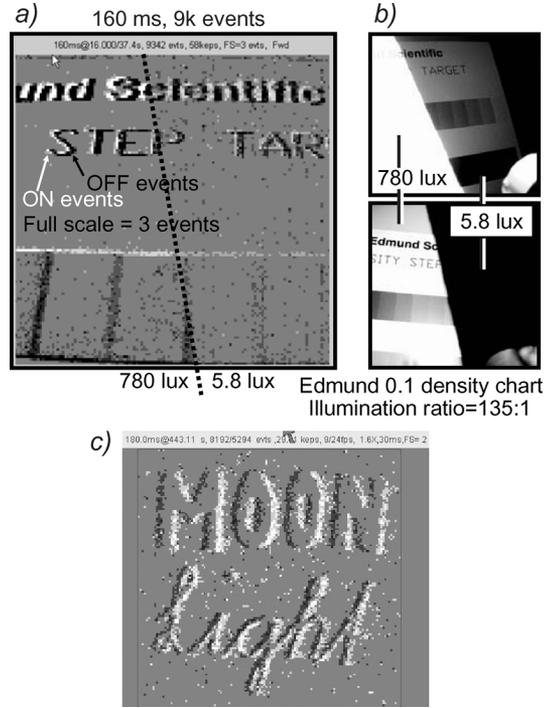


Fig. 8. Illustration of vision sensor dynamic range capabilities. (a) Histogrammed output from the vision sensor viewing an Edmund density step chart with illumination ratio of 135:1 (a shadow was cast to create this illumination step). (b) The same scene as photographed by a Nikon 995 digital camera to expose the two halves of the scene. (Adapted from [2] and [32]). (c) Moving black text on white background under 3/4 moon (<0.1 lux) illumination (180 ms, 8000 events).

B. Dynamic Range

The vision sensor wide dynamic range [illustrated in Fig. 8(a)–(c)] arises from the logarithmic compression in the front-end photoreceptor circuit and the local event-based quantization. We define the dynamic range as the ratio of maximum to minimum scene illumination at which events can be generated by high contrast stimuli. Photodiode dark current of 4 fA at room temperature (inferred from the global photodiode node, Fig. 2) limits the lower end of the range. Events are generated reliably and reproducibly down to less than 0.1 lux scene illumination using a fast $f/1.2$ lens [Fig. 8(c)]. At this illumination level, the signal (photocurrent induced by photons from the scene) is only a small fraction of the noise (background dark current). Operation at this low signal-to-noise ratio is only possible because the low threshold mismatch allows setting a low threshold. The sensor also operates up to bright sunlight scene illumination of 100 klux; thus, the achieved dynamic range amounts to at least 6 decades, or 120 dB. The full dynamic range can appear within a scene and will still be resolved by the sensor. The vision sensor is fully usable for typical scene contrast under nighttime street lighting of a few lux. The dynamic range is halved approximately every 8°C increase in temperature. Using a low-leakage imager process would increase the dynamic range by a factor of ~ 100 .

C. Pixel Bandwidth

At low illumination, the photoreceptor bandwidth is proportional to photocurrent because the bandwidth is determined by

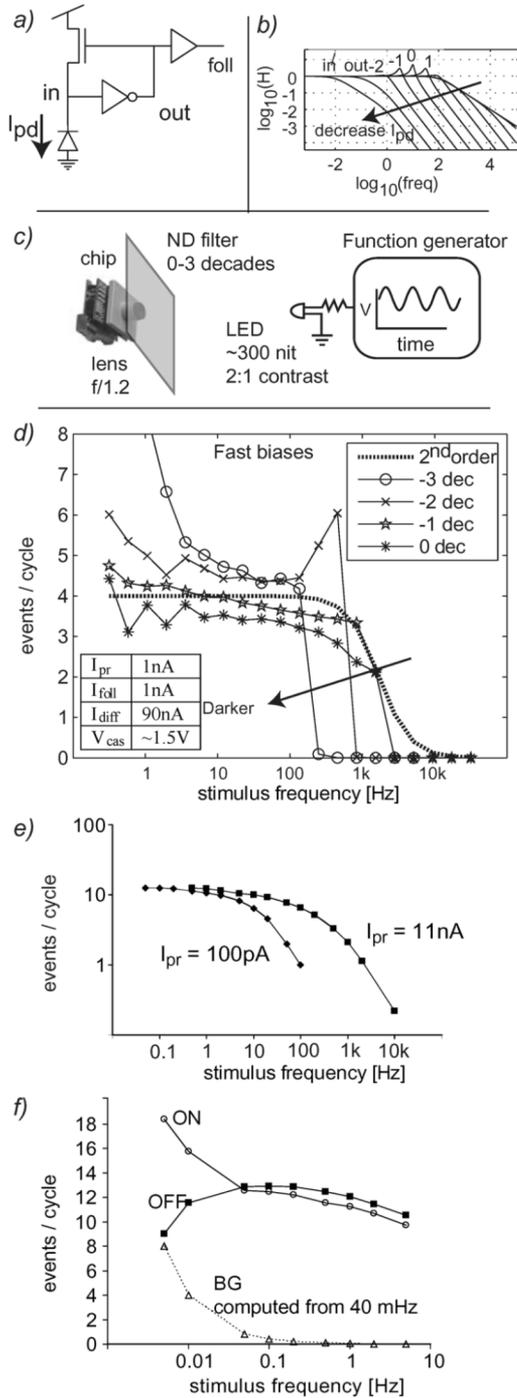


Fig. 9. Event transfer function from a single pixel in response to sinusoidal LED stimulation. Chip bias settings were held constant for all measurements and were optimized for maximum bandwidth with stability over the entire illumination range. (a) Model circuit. (b) Theoretical linear transfer functions based on (a) in which τ_{in} is varied over 8 decades while τ_{out} is held constant; these curves show the same response characteristics as the measured results. (c) Measurement setup. (d) Measured responses; curves are labeled with decade attenuation from bare LED with an unattenuated luminance of about 300 nit using a 6 mm $f/1.2$ lens. Data were collected from a single pixel over the duration of 10 s; then the number of events was divided by the stimulation frequency times the collection duration, leading to an average number of events per stimulation cycle. The inset shows programmed bias pixel bias currents. (e) Single pixel frequency response for two photoreceptor amplifier M_{pr} bias currents to demonstrate control of bandwidth. (f) Events produced at very low stimulus frequencies. The BG curve is computed from measured 40 mHz background activity rate.

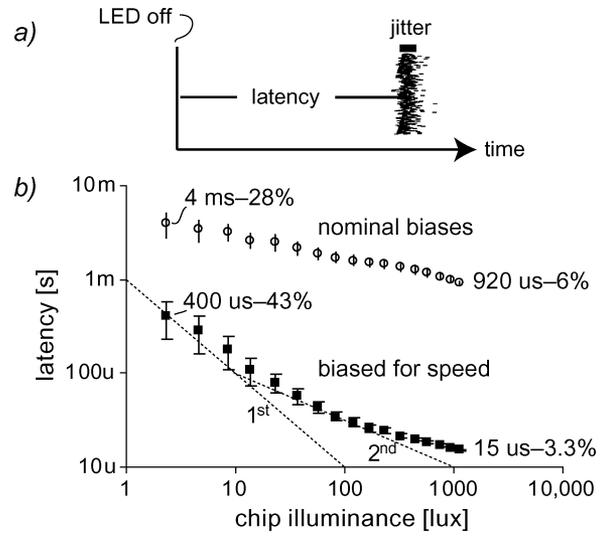


Fig. 10. Sensor latency and latency jitter (error bars) versus illumination in response to a 30% step increase of single pixel illumination. (a) Measurement of repeated single event responses to the step; jitter is shown by the error bars. (b) Results with two bias settings as a function of pixel illuminance.

the RC time constant formed by the photodiode parasitic capacitance and the resistance of the feedback transistor source [24]. Because this subthreshold conductance is proportional to photocurrent, the bandwidth is proportional to photocurrent at low intensities. At higher photocurrents, the feedback amplifier pole contributes to a second-order resonant response that can be modeled by the circuit shown in Fig. 9(a), resulting in transfer functions that vary with photocurrent as shown in Fig. 9(b). In the present implementation, the bandwidth is increased by using active feedback by a factor of about 15 compared to a passive logarithmic photoreceptor.

The bias current values we typically use for a wide range of lighting limit the maximum event frequency response to about 3 kHz. Fig. 9(c) shows the setup for measuring this transfer function, and Fig. 9(d) shows the measured temporal “event transfer function” for four different DC illumination levels. Each curve shows the average number of events (ON and OFF combined) generated per complete sinusoidal cycle of LED modulation. At low illumination (−3 dec) the transfer function shows the characteristic of a first-order low-pass filter. At higher illumination (−2 dec) there is resonant peaking. At the highest illumination levels (−1 and 0 dec), the measured single pixel bandwidth is about 3 kHz. Bandwidth can be increased more, but only if instability is tolerated for low illumination.

Fig. 9(e) shows that we can use the amplifier bias current to adjust bandwidth. Increasing the current in M_{pr} by a factor of about 100 (from 100 pA to 11 nA) increases bandwidth by only about 50×, probably because the amplifier starts to enter moderate-inversion operation.

The reset switch junction leakage produces background ON events. These only become significant at extremely low frequencies. Fig. 9(f) shows events per stimulus cycle for very low stimulus frequencies between 5 mHz and 10 Hz. The pixel under test has an average background ON-event rate of 40 mHz (one background event every 25 seconds) at room temperature. The BG

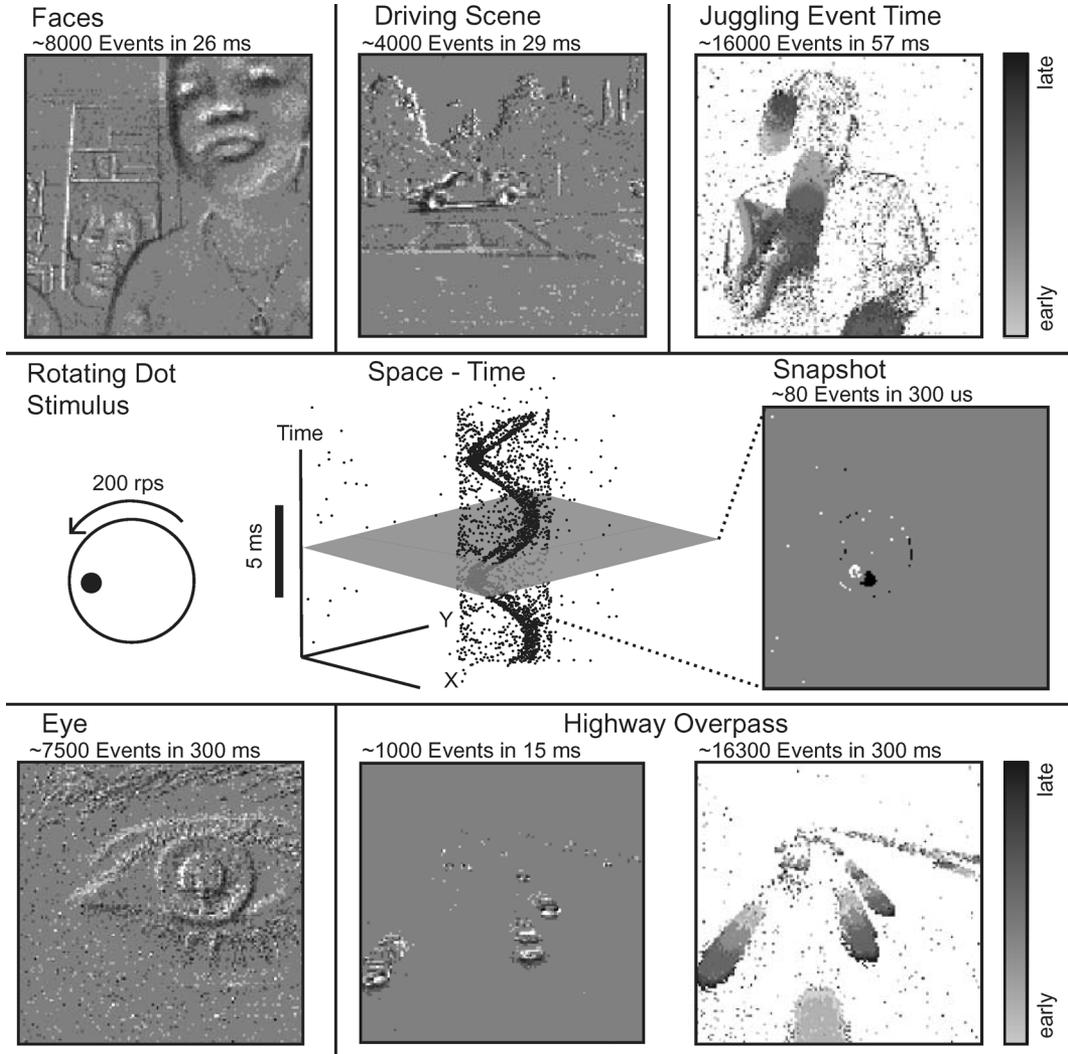


Fig. 11. Shows images taken under natural lighting conditions with either object or camera motion. These are rendered as contrast (gray scale represents reconstructed gray scale change), grayscale-time (event time as shown as grayscale, black = young, gray = old, white = no event), or 3-D space time.

curve shows how this background activity would affect measurement of very low frequencies. In the case of complete absence of reset switch junction leakage, the usable frequency range appears to practically extend down to DC. The uncorrelated background events are easily filtered away at the application level.

D. Latency and Latency Jitter

A basic prediction from Section III-C is that the latency should increase when we decrease the illumination level, and the increase should be proportional to reciprocal illumination. The latency was measured using a low-contrast (30%) periodic 10 Hz step stimulus at variable DC luminance (Fig. 10). The thresholds were set to produce exactly one event of each polarity per stimulus cycle. The overall latency is plotted versus stimulus chip illuminance for two sets of measurements, one at the nominal biases, the other at higher current levels for the M_{pr} photoreceptor and M_{b2} source follower biases (Fig. 2). The plots show the measured latency and the 1-sigma response jitter. The dashed lines show a reciprocal (1st) and

reciprocal-square-root (2nd) relationship between latency and illumination.

The most interesting aspects of this data are the following: 1) the minimum latency is only 15 μ s, representing an effective single pixel bandwidth of about 66 kHz; 2) the latency is a soft function of photocurrent; only at very low illuminance is the latency reciprocal with illuminance, and with nominal biases the latency changes only a factor of 4 over 3 decades of photocurrent; 3) at the nominal biases that we generally use, the latency is still only 4 ms at the lowest illuminance of a few lux; and 4) the jitter in the step response is a small fraction of the latency regardless of illuminance. The latency only changes very slowly with illuminance for the nominal biases because other mechanisms besides photoreceptor bandwidth limit the latency. In summary, this vision sensor's low latency makes it attractive for real-time control systems.

E. Example Data

Fig. 11 shows example image data from the vision sensor. We rendered the dynamic properties by using grayscale or 3-D

to show the time axis. The “Faces” image was collected indoors at night with illumination from a 15W fluorescent desk lamp. The “Driving Scene” was collected outdoors under daylight from a position on a car dashboard. The “Juggling Event Time” image shows the event times as grayscale while one of the authors juggles three balls under indoor daylight illumination. The “Rotating Dot” panel shows the events generated by a black dot drawn on a white disk rotating at 200 revolutions per second under indoor fluorescent office illumination of 300 lux. The events are rendered both in space–time over ~ 10 ms and as a briefer snapshot image spanning 300 μs. The “Eye” image shows events from a moving eye under indoor illumination. The “Highway Overpass” images show events produced by cars on a highway viewed from an overpass in late afternoon lighting, on the left displayed as ON and OFF events and on the right as relative time during the snapshot. This data was collected with similar digital bias settings from a variety of individual chips.

IV. SUMMARY AND DISCUSSION

The main achievement of this work is the implementation of a high-quality frame-free transient vision sensor that represents a concrete step towards solving vision problems in the event-based, data-driven, redundancy-reducing style of computation that underlies the power of biological vision. This sensor responds to relative changes in intensity, discarding most illuminant information, leaving precisely timed information about object and image motion. This information is useful for dynamic vision problems.

The pixel design uses a novel combination of continuous and discrete time operation, where the timing is self-generated. The use of self-timed switched capacitor architecture leads to well-matched pixel response properties and fast, wide dynamic range operation. We developed new techniques for characterizing this sensor, including metrics for matching, for pixel bandwidth, and for pixel latency. We characterized the sensor for these metrics over wide illumination range.

Table I shows the key performance metrics and compares them with other work. The vision sensor achieves wide dynamic range (> 120 dB), low latency (15 μs), low power consumption (23 mW), and low mismatch (2.1% contrast). The vision sensor also integrates a programmable bias generator that allows temperature-independent and process-independent operation. We use this programmability for dynamic control of operating parameters.

The main areas that could benefit from improvement are as follows. The AER bus bandwidth limits high-speed imaging for “busy” scenes. The chip should be built in a low-leakage imager process. The bias generator should be modified to reduce its power consumption and to allow for smaller bias currents in order to limit bandwidth when desired. An integrated means for measuring average scene brightness would be beneficial for automatic bias control. It remains to be seen how much lack of any DC response hinders application.

Applications areas for this vision sensor include high-speed low-bandwidth imaging, surveillance and traffic monitoring

under uncontrolled lighting conditions, wireless sensor networks, industrial vision for manufacturing or inspection, autonomous navigation systems (e.g., lane finding, flying vehicles), human interface devices (e.g., eye-trackers), and visual prosthetics. The processing of this vision sensor’s output for vision is beyond the scope of this paper [8]–[10], [32].

ACKNOWLEDGMENT

The authors thank S. Mitra, G. Indiveri, and K. Boahen for AER circuit layout, and M. Litzberger and S. C. Liu for discussions. They also thank B. Linares-Barranco, P. F. Ruedi, E. Culurciello, P. Häfliger, and the anonymous reviewers for constructive feedback on a draft of this paper. This type of vision sensor was first conceived by the late J. Kramer. The authors have benefited greatly from the support and environment of the Institute of Neuroinformatics.

REFERENCES

- [1] S. Kleinfelder, S. Lim, X. Q. Liu, and A. El Gamal, “A 10000 frames/s CMOS digital pixel sensor,” *IEEE J. Solid-State Circuits*, vol. 36, no. 12, pp. 2049–2059, Dec. 2001.
- [2] P. Lichtsteiner, C. Posch, and T. Delbruck, “A 128x128 120 dB 30 mW asynchronous vision sensor that responds to relative intensity change,” in *IEEE ISSCC 2006 Dig. Tech. Papers*, San Francisco, CA, 2006, pp. 508–509.
- [3] U. Mallik, M. Clapp, E. Choi, G. Cauwenberghs, and R. Etienne-Cummings, “Temporal change threshold detection imager,” in *IEEE ISSCC 2005 Dig. Tech. Papers*, San Francisco, CA, 2005, pp. 362–363.
- [4] E. Culurciello and R. Etienne-Cummings, “Second generation of high dynamic range, arbitrated digital imager,” in *Proc. ISCAS 2004*, Vancouver, BC, Canada, May 2004, vol. 4, pp. 828–831.
- [5] P. F. Ruedi, P. Heim, F. Kaess, E. Grenet, F. Heitger, P. Y. Burgi, S. Gyger, and P. Nussbaum, “A 128x128 pixel 120-dB dynamic-range vision-sensor chip for image contrast and orientation extraction,” *IEEE J. Solid-State Circuits*, vol. 38, no. 12, pp. 2325–2333, Dec. 2003.
- [6] K. A. Zaghloul and K. Boahen, “Optic nerve signals in a neuromorphic chip II: Testing and results,” *IEEE Trans. Biomed. Eng.*, vol. 51, no. 4, pp. 667–675, Apr. 2004.
- [7] P. Lichtsteiner, T. Delbruck, and J. Kramer, “Improved ON/OFF temporally differentiating address-event imager,” in *Proc. 11th IEEE Int. Conf. Electronics, Circuits, and Systems (ICECS2004)*, Tel Aviv, Israel, Dec. 2004, pp. 211–214.
- [8] T. Delbruck and P. Lichtsteiner, “Fast sensory motor control based on event-based hybrid neuromorphic-procedural system,” in *Proc. IEEE ISCAS 2007*, New Orleans, LA, May 2007, pp. 845–849.
- [9] M. Litzberger, B. Kohn, A. N. Belbachir, N. Donath, G. Gritsch, H. Garn, C. Posch, and S. Schraml, “Estimation of vehicle speed based on asynchronous data from a silicon retina optical sensor,” in *Proc. 2006 IEEE Intelligent Transportation Systems Conf. (ITSC’06)*, Toronto, ON, Canada, Sep. 2006, pp. 653–658.
- [10] M. Litzberger, C. Posch, D. Bauer, A. Belbachir, P. Schön, B. Kohn, and H. Garn, “Embedded vision system for real-time object tracking using an asynchronous transient vision sensor,” in *Proc. 12th Digital Signal Processing Workshop, 4th Signal Processing Education Workshop*, Grand Teton National Park, WY, Sep. 2006, pp. 173–178.
- [11] K. Boahen, “Neuromorphic microchips,” *Sci. Amer.*, vol. 292, pp. 56–63, 2005.
- [12] M. Mahowald, *An Analog VLSI System for Stereoscopic Vision*. Boston, MA: Kluwer, 1994.
- [13] K. A. Zaghloul and K. Boahen, “Optic nerve signals in a neuromorphic chip: Part I and II,” *IEEE Trans. Biomed. Eng.*, vol. 51, no. 4, pp. 657–675, Apr. 2004.
- [14] E. Grenet, S. Gyger, P. Heim, F. Heitger, F. Kaess, P. Nussbaum, and P.-F. Ruedi, “High dynamic range vision sensor for automotive applications,” in *Proc. SPIE*, 2005, vol. 5663, pp. 246–253.
- [15] E. Culurciello and A. G. Andreou, “CMOS image sensors for sensor networks,” *Analog Integr. Circuits Signal Process.*, vol. 49, pp. 39–51, 2006.

- [16] X. Qi, X. Guo, and J. Harris, "A time-to-first-spike CMOS imager," in *Proc. IEEE ISCAS 2004*, Vancouver, BC, Canada, May 2004, pp. 824–827.
- [17] Q. Luo and J. G. Harris, "A time-based CMOS image sensor," in *Proc. IEEE ISCAS 2004*, Vancouver, BC, Canada, May 2004, pp. 840–843.
- [18] M. Azadmehr, J. P. Abrahamsen, and P. Häfliger, "A foveated AER imager chip [address event representation]," in *Proc. IEEE ISCAS 2005*, Kobe, Japan, May 2005, vol. 3, pp. 2751–2754.
- [19] J. Costas-Santos, T. Serrano-Gotarredona, R. Serrano-Gotarredona, and B. Linares-Barranco, "A spatial contrast retina with on-chip calibration for neuromorphic spike-based AER vision systems," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 54, no. 7, pp. 1444–1458, Jul. 2007.
- [20] C. Posch, M. Hofstätter, D. Matolin, G. Vanstraelen, P. Schön, N. Donath, and M. Litzberger, "A dual-line optical transient sensor with on-chip precision time-stamp generation," in *IEEE ISSCC 2007 Dig. Tech. Papers*, San Francisco, CA, 2007, pp. 500–501.
- [21] J. Kramer, "An on/off transient imager with event-driven, asynchronous read-out," in *Proc. IEEE ISCAS 2002*, Phoenix, AZ, May 2002, vol. 2, pp. 165–168.
- [22] M. Loose, K. Meier, and J. Schemmel, "A self-calibrating single-chip CMOS camera with logarithmic response," *IEEE J. Solid-State Circuits*, vol. 36, no. 4, pp. 586–596, Apr. 2001.
- [23] S. Kavadias, B. Dierickx, D. Scheffer, A. Alaerts, D. Uwaerts, and J. Bogaerts, "A logarithmic response CMOS image sensor with on-chip calibration," *IEEE J. Solid-State Circuits*, vol. 35, no. 8, pp. 1146–1152, Aug. 2000.
- [24] T. Delbruck and C. A. Mead, "Analog VLSI adaptive logarithmic wide-dynamic-range photoreceptor," in *Proc. IEEE ISCAS 1994*, London, U.K., May 1994, pp. 339–342.
- [25] T. Delbruck and D. Oberhoff, "Self-biasing low power adaptive photoreceptor," in *Proc. IEEE ISCAS 2004*, Vancouver, BC, Canada, 2004, vol. 4, pp. 844–847.
- [26] K. A. Boahen, "Point-to-point connectivity between neuromorphic chips using address events," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 47, no. 5, pp. 416–434, May 2000.
- [27] K. A. Boahen, "A burst-mode word-serial address-event link—I: Transmitter design," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 51, no. 7, pp. 1269–1280, Jul. 2004.
- [28] ChipGen Silicon Compiler. [Online]. Available: <https://www.stanford.edu/group/brainsinsilicon/Downloads.htm>
- [29] P. Lichtsteiner and T. Delbruck, "A 64x64 AER logarithmic temporal derivative silicon retina," Ph.D. dissertation in microelectronics and electronics, Lausanne, Switzerland, 2005.
- [30] T. Delbruck and P. Lichtsteiner, "Fully programmable bias current generator with 24 bit resolution per bias," in *Proc. IEEE ISCAS 2006*, Island of Kos, Greece, 2006, pp. 2849–2852.
- [31] R. Serrano-Gotarredona, M. Oster, P. Lichtsteiner, A. Linares-Barranco, R. Paz-Vicente, F. Gómez-Rodríguez, H. K. Riis, T. Delbrück, S. C. Liu, S. Zahnd, A. M. Whatley, R. Douglas, P. Häfliger, G. Jimenez-Moreno, A. Civit, T. Serrano-Gotarredona, A. Acosta-Jiménez, and B. Linares-Barranco, "AER building blocks for multi-layer multi-chip neuromorphic vision systems," in *Advances in Neural Information Processing Systems 18, Proc. 2005 Conf.*, Vancouver, BC, Canada, Dec. 2005, pp. 1217–1224.
- [32] jAER Open Source Project. 2007 [Online]. Available: <http://jaer.wiki.sourceforge.net>



Patrick Lichtsteiner (M'05) received the Diploma (equivalent to M.S.) degree in physics and the Ph.D. degree from the Swiss Federal Institute of Technology, Zurich, Switzerland, in 2002 and 2006, respectively.

He is currently a Postdoctoral Fellow at the Institute of Neuroinformatics, University of Zurich and Swiss Federal Institute of Technology, Zurich, Switzerland. His research interests include CMOS imaging, neuromorphic vision sensors and high-speed vision.

Dr. Lichtsteiner and his colleagues have received four prizes for IEEE conference papers, including the 2006 ISSCC Jan Van Vessel Outstanding Paper Award.



Christoph Posch (M'07) received the M.Sc. and Ph.D. degrees in electronics engineering from Vienna University of Technology, Vienna, Austria, in 1995 and 1999, respectively.

From 1996 to 1999, he worked at CERN, the European Laboratory for Particle Physics in Geneva, Switzerland, on analog CMOS and BiCMOS IC design for semiconductor particle detector readout and control. From 1999 to 2004, he was with Boston University, Boston, MA, where his focus was on analog/mixed-signal integrated circuit design for

high-energy physics instrumentation. In 2004, he joined the Smart Sensors Group at the Austrian Research Centers (ARC) in Vienna, where he was promoted to Principle Scientist in 2007. His current research interests include development and design of neuromorphic CMOS image sensors and bio-inspired signal processing.

Dr. Posch was co-recipient of the Jan Van Vessel Award for Outstanding European Paper at the IEEE ISSCC 2006. He is a member of the Sensory Systems Technical Committee of the IEEE Circuits and Systems Society.



Tobin Delbruck (M'89–SM'06) studied physics and applied mathematics as an undergraduate and received the Ph.D. degree from the California Institute of Technology (Caltech), Pasadena, in computation and neural systems in 1993.

He is a group leader at the Institute of Neuroinformatics (INI), part of ETH Zurich and the University of Zurich, Switzerland, and a Visiting Scientist at Caltech. His main interest is in developing neuromorphic electronics, particularly vision sensor chips. He co-invented the standard neuromorphic

adaptive photoreceptor circuit and bump circuit. He worked for several years for Arithmos, Synaptics, National Semiconductor, and Foveon, where he was one of the founding employees. In 1998, he moved to Switzerland to join INI. In 2002, he was lead developer of the tactile luminous floor used in INI's exhibit "Ada: Playful Intelligent Space". He holds eight patents, and has over 30 refereed papers in journals and conferences, four book chapters, and one book.

Dr. Delbruck has received six IEEE awards, including the 2006 ISSCC Jan Van Vessel Outstanding European Paper Award.