Retinomorphic Event-Based Vision Sensors: Bioinspired Cameras With Spiking Output

Spatio–temporal cameras with spiking output can overcome multiple deficiencies of frame-based systems.

By Christoph Posch, *Senior Member IEEE*, Teresa Serrano-Gotarredona, *Member IEEE*, Bernabe Linares-Barranco, *Fellow IEEE*, and Tobi Delbruck, *Fellow IEEE*

ABSTRACT | State-of-the-art image sensors suffer from significant limitations imposed by their very principle of operation. These sensors acquire the visual information as a series of "snapshot" images, recorded at discrete points in time. Visual information gets time quantized at a predetermined frame rate which has no relation to the dynamics present in the scene. Furthermore, each recorded frame conveys the information from all pixels, regardless of whether this information, or a part of it, has changed since the last frame had been acquired. This acquisition method limits the temporal resolution, potentially missing important information, and leads to redundancy in the recorded image data, unnecessarily inflating data rate and volume. Biology is leading the way to a more efficient style of image acquisition. Biological vision systems are driven by events happening within the scene in view, and not, like image sensors, by artificially created timing and control signals. Translating the frameless paradigm of biological vision to artificial imaging systems implies that control over the acquisition of visual information is no longer being imposed externally to an array of pixels but the decision making is transferred to the single pixel that handles its own information individually. In this paper, recent developments in bioinspired, neuromorphic optical sensing and artificial vision are presented and discussed. It is suggested that bioinspired vision systems have the potential to outperform conventional, frame-based vision systems in many application fields and to establish new benchmarks in terms of redundancy suppression and data compression, dynamic range, temporal resolution, and power efficiency. Demanding vision tasks such as real-time 3-D mapping, complex multiobject tracking, or fast visual feedback loops for sensory-motor action, tasks that often pose severe, sometimes insurmountable, challenges to conventional artificial vision systems, are in reach using bioinspired vision sensing and processing techniques.

KEYWORDS | Address-event representation (AER); biomimetics; complementary metal–oxide–semiconductor (CMOS) image sensors; event-based vision; focal-plane processing; high dynamic range (HDR); neuromorphic electronics; neuromorphic engineering; silicon retina; time-domain correlated double sampling (TCDS); time-domain imaging; video compression

I. INTRODUCTION

Despite all the impressive progress made during the last decades in the fields of information technology, microelectronics, and computer science, artificial sensory and information processing systems are still much less effective in dealing with real-world tasks than their biological counterparts. Even small insects outperform the most powerful computers in routine functions involving, e.g., real-time sensory data processing, perception tasks, and motor control and are obviously capable of doing all this on an incredibly small energy budget. In stark contrast to human-engineered information processing and computation devices, biological neural systems rely on a large number of relatively simple, slow, and noisy processing elements and obtain performance and robustness

C. Posch is with the Institut de la Vision, Universite Pierre at Marie Curie, Paris 75012, France (e-mail: christoph.posch@inserm.fr).

T. Serrano-Gotarredona and **B. Linares-Barranco** are with the Instituto de Microelectrónica de Sevilla, 41092 Sevilla, Spain.

T. Delbruck is with the Institute of Neuroinformatics, University of Zurich and ETH Zurich, 8057 Zurich, Switzerland.

from a massively parallel principle of operation and a high level of redundancy where the failure of single elements usually does not induce any observable system performance degradation. Studying and understanding the computational principles of the brain and how they can be exploited to build intelligent artificial systems are fundamental for devising a new generation of neuromorphic systems, that, as the biological systems they model, are adaptive, fault tolerant and scalable, and process information using energy-efficient, asynchronous, event-driven methods.

A. Neuromorphic Engineering

Nature has been a source of inspiration for engineers since ancient times. In diverse fields such as aerodynamics, robotics, the engineering of surfaces and structures, or material sciences, approaches developed by nature through long evolutionary processes offer stunning solutions to engineering problems. Many synonymous terms like bionics, biomimetics, or bioinspired engineering have been used to name the flow of concepts from biology to engineering [1].

Also the idea of applying computational principles of biological neural systems to artificial information processing has existed for decades. An early work from the 1940s by McCulloch and Pitts introduced a neuron model and showed that it was able to perform computation [2]. Around the same time, Hebb developed the first models for learning and adaptation [3]. In 1952, Hodgkin and Huxley linked biological signal processing to electrical engineering in their famous paper entitled "A quantitative description of membrane current and its application to conduction and excitation in nerve" [4], in which they describe a circuit model of electrical current flow across a nerve membrane.

In the late 1980s, Mead at the California Institute of Technology (Caltech, Pasadena, CA, USA) introduced the "neuromorphic" concept to describe systems containing analog and asynchronous digital electronic circuits that mimic neural architectures present in biological nervous systems [5]–[7]. This concept revolutionized the frontier of computing and neurobiology to such an extent that a new engineering discipline emerged, whose goal is to design and build artificial neural systems, like computational arrays of synapse-connected artificial neurons, retinomorphic vision systems or auditory processors, using (micro)electrical components and circuits. This quickly expanding field is referred to as neuromorphic engineering.

The term "neuromorphic" has also been coined by Mead to name artificial systems that adopt the form of, or morph, neural systems. In a ground-breaking paper on neuromorphic electronic systems, published in 1990, in the Proceedings of the IEEE [6], Mead argues that the advantages of biological information processing can be attributed principally to the use of elementary physical phenomena as computational primitives, and to the representation of information by the relative values of analog signals, rather than by the absolute values of digital signals. He further argues that this approach requires adaptive techniques to correct for differences of nominally identical components, and that this adaptive capability naturally leads to systems that learn about their environment. Experimental results suggest that adaptive analog systems are 100 times more efficient in their use of silicon area, consume 10 000 times less power than comparable digital systems, and are much more robust to component degradation and failure than conventional systems [6].

Following further along these lines, it has been argued that these types of circuits can be used to develop a new generation of computing technologies based on the organizing principles of the biological nervous system [8], [9]. Indiveri and Furber argue that the characteristics of neuromorphic circuits offer an attractive alternative to conventional computing strategies, especially if one considers the advantages and potential problems of future advanced very large scale integration (VLSI) fabrication processes. By using massively parallel arrays of computing elements, exploiting redundancy to achieve fault tolerance, and emulating the neural style of computation, neuromorphic VLSI architectures can exploit to the fullest potential the features of advanced scaled VLSI processes and future emerging technologies, naturally coping with the problems that characterize them, such as device inhomogeneities and imperfections [10], [11].

B. Implementing Neuromorphic Systems

Neuromorphic devices today are usually implemented as VLSI integrated circuits or systems-on-chip (SoCs) on planar silicon, the mainstream technology used for fabricating the ubiquitous microchips that can be found in practically every modern electronically operated device.

The primary silicon primitive is the transistor. Interestingly, transistors share several physical and functional characteristics with biological neurons. For example, in the weak-inversion region of operation, the current through a metal–oxide–semiconductor (MOS) transistor exponentially relates to the voltages applied to its terminals. A similar dependency is observed between the active populations of ion channels as a function of the membrane potential of a biological neuron. Exploiting such physical similarities allows, e.g., constructing electronic circuits that implement models of voltage-controlled neurons and synapses and realize biological computational primitives such as phototransduction, multiplication, inhibition, correlation, thresholding, or winner-take-all selection [4], [6].

Representing a new paradigm for the processing of sensor signals, the greatest success of neuromorphic systems to date has been in the emulation of sensory signal acquisition and transduction, most notably in vision.

II. BIOINSPIRED VISION

A. Biological Retinas

The retina of vertebrates, e.g., humans, is a multilayered neural network lining the back hemisphere of the eyeball. The retina, initiating some 600 million years ago as an assembly of some light sensitive neural cells and further developed during a long evolutionary process, is the place where acquisition and first stage of processing of the visual information happens. As shown in Fig. 1, the retina is a complex structure with three primary layers: the photoreceptor layer, the outer plexiform layer, and the inner plexiform layer [13]–[16].

The photoreceptor layer consists of two classes of cells: cones and rods, which transform the incoming light into an electrical signal which affects neurotransmitter release in the photoreceptor output synapses. The photoreceptor cells in turn drive horizontal cells and bipolar cells in the outer plexiform layer.

The two major classes of bipolar cells, the ON bipolar cells and the OFF bipolar cells, separately code for bright spatio–temporal contrast and dark spatio–temporal contrast changes. They do this by comparing the photoreceptor signals to spatio–temporal averages computed by the laterally connected layer of horizontal cells, which form a resistive mesh.

The horizontal cells are connected to each other by conductive pores called gap junctions and are connected to bipolar cells and photoreceptors in complex triad synapses. Together with the input current produced at the photoreceptor synapses, this network computes spatio–temporal low-passed copies of the photoreceptor outputs. The horizontal cells feed back onto the photoreceptors to help set their operating points and also compute a spatio–temporally smoothed copy of the visual input.

The bipolar cells are effectively driven by differences between the photoreceptor and horizontal cell outputs. In the even more complex outer plexiform layer, the ON and OFF bipolar cells synapse onto many types of amacrine cells and many types of ON and OFF ganglion cells in the inner plexiform layer. The horizontal and amacrine cells mediate the signal transmission process between the photoreceptors and the bipolar cells, and the bipolar cells and the ganglion cells, respectively.

The bipolar and ganglion cells can be further divided into two different groups: cells with more sustained responses and cells with more transient responses. These cells carry information along at least two parallel pathways in the retina: the magno–cellular pathway, where cells are sensitive to temporal changes in the scene, and the parvo–cellular pathway where cells are sensitive to forms in the scene. This picture of a simple partition into sustained and transient pathways is too simple; in reality, there are many parallel pathways computing many views (probably at least 50 in the mammalian retina) of the visual input. In the following, a simplified view of biological vision that is feasible for silicon integrated circuit focal-plane implementation is presented.

The retina converts spatio–temporal information contained in the incident light from the visual scene into spike trains and patterns, output and conveyed to the visual cortex by retinal ganglion cells, whose axons form the fibers of the optic nerve. The information carried by these spikes is maximized by the retinal processing, encompassing highly evolved adaptive filtering and sampling mechanisms to improve coding efficiency [17], such as follows.

- Local automatic gain control at the photoreceptor and network levels: it eliminates the dependency on absolute lighting levels, and instead the receptors respond to changes in the incident light (also knows as temporal contrast). Local adaptation extends the retina's input dynamic range (DR) without increasing its output range.
- Bandpass spatio–temporal filtering: it limits spatial and temporal frequencies to an intermediate range, reducing redundancy by rejecting low frequencies and noise by rejecting high frequencies.
- Rectification in ON and OFF output cell types: it reduces spike-firing rates that would be required to signal both positive and negative signals on a single channel. ON/OFF encoding is used in bipolar cells as well as in ganglion cells, the retina's output cells.
- The varying distribution of different receptor types along with corresponding pathways across the retina (e.g., magno– and parvo–cellular pathways with more transient and more sustained response) combined with precise rapid eye movements elicit

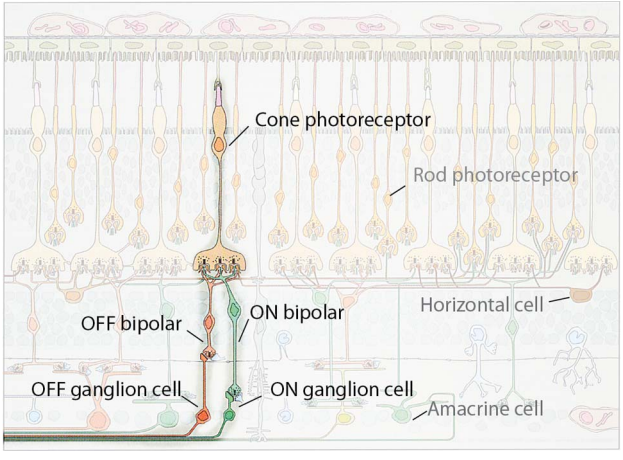

Fig. 1. ***Biological retina extracts multiple simultaneous "views" of the visual input by processing the photoreceptor outputs in several layers, using a multitude of cell types. Silicon retinas to date almost all simplify the biological functions to the highlighted cells, which extract rectified spatio–temporal contrast. (Adapted from [12] with permission.)***

the illusion of high spatial and temporal resolution everywhere, while, in reality, the retina samples coarsely in time in the center (however, at high spatial resolution) and coarsely in space in the periphery (however, at high temporal resolution).

In comparison to the human retina, a conventional image sensor sampling at the Nyquist rate requires transmitting more than 20 Gb/s to match the human eyes' photopic range (exceeding 100 dB), its spatial and temporal resolution, and its field of view. In contrast, by coding 2 b of information per spike [18], the optic nerve transmits just about 20 Mb/s to the visual cortex—a thousand times less.

Biological retinas have many desirable characteristics which are lacking in conventional image sensors but inspire and drive the design of retinomorphic vision devices. As discussed throughout this paper, many of these advantageous characteristics have been modeled and implemented on silicon. Local gain control, spatio–temporal filtering, and redundancy suppression encoding lead to unprecedented wide DR operation, video compression through redundancy suppression, and temporal resolution of pixel response far beyond that of most conventional, frame-based devices.

B. Limitations in Vision Engineering

In order to appreciate how biological approaches and neuromorphic engineering techniques could be beneficial for advancing artificial vision, it is inspiring to look at some shortcomings of conventional machine vision.

State-of-the-art image sensors suffer from limitations imposed by their frame-based operation. The sensors acquire the visual information as a series of "snapshots" recorded at discrete points in time, hence time quantized at a predetermined frame rate. Biology does not know the concept of a frame. In fact, comparing the performance of biological vision systems to the best state-of-the-art artificial vision devices, frames do not appear to be a very useful or efficient form of encoding visual information. This is even more obvious if one considers that the world, the source of the visual information we are interested in, works asynchronously and in continuous time. As a consequence, depending on the time scale of changes in the observed scene, a problem that is very similar to undersampling, known from other engineering fields, arises. Things happen between frames, and information gets lost. This may be tolerable for the recording of video for a human observer, but artificial vision systems in demanding applications such as, e.g., autonomous robot navigation, high-speed motor control, visual feedback loops, etc., may fail as a consequence of this shortcoming.

Nature suggests a different approach: Biological vision systems are driven and controlled by events happening within the scene in view, and not, like image sensors, by artificially created timing and control signals that have no relation whatsoever to the source of the visual information and its dynamics. Translating the frameless paradigm of biological vision to artificial imaging systems implies that control over the acquisition of visual information is no longer being imposed externally to an array of pixels but the decision making is transferred to the single pixel that handles its own information individually.

The second problem that is also a direct consequence of the frame-based acquisition of visual information is redundancy. Each recorded frame conveys the information from all pixels, regardless of whether this information, or a part of it, has changed since the last frame had been acquired. This method obviously leads, depending on the dynamic contents of the scene, to a more or less high degree of redundancy in the acquired image data. Acquisition and handling of these dispensable data consume valuable resources and translate into high transmission power dissipation, increased channel bandwidth requirements, increased memory size, and post-processing power demands.

Devising an engineering solution that follows the biological pixel-individual, frame-free approach to vision can potentially solve both problems.

C. Modeling the Retina in Silicon

The first silicon VLSI retina by Mahowald and Mead implemented a model of the photoreceptor cells, horizontal cells, and bipolar cells [19]. Each silicon photoreceptor mimics a cone cell and contains both a continuous-time photosensor and adaptive circuitry which adjusts its response to cope with changing light levels [20]. A network of MOS variable resistors mimics the horizontal cell layer, furnishing feedback based on the average amount of light striking nearby photoreceptors; the bipolar cell circuitry amplifies the difference between the signal from the photoreceptor and the local average and rectifies this amplified signal into ON and OFF outputs. The response of the resulting retinal circuit approximates the behavior of the human retina [4], [21].

Zaghloul and Boahen implemented simplified models of all five layers of the retina on a silicon chip starting in 2001 [8], [22]. This parvo–magno retina is a very different type of silicon retina that was focused on modeling of both outer and inner retinas including sustained (parvo) and transient (magno) types of cells, based on histological and physiological findings. It is an improvement over the first retina by Mahowald and Mead which models only the outer retina circuitry, that is, the cones, horizontal cells, and bipolar cells [19]. The parvo–magno design captures key adaptive features of biological retinas including light and contrast adaptation, and adaptive spatial and temporal filtering. By using small transistor log-domain circuits that are tightly coupled spatially by diffuser networks and single-transistor synapses, they were able to achieve 5760 phototransistors at a density of 722 per mm^2 and 3600 ganglion cells at a density of 461 per mm^2 in a 3.5×3.3-mm^2 silicon area, using a 0.25-μm complementary MOS (CMOS) process.

The outer retina's synaptic interactions realize spatio–temporal bandpass filtering and local gain control. The model of the inner retina realizes contrast gain control (the control of sensitivity to temporal contrast), through modulatory effects of wide-field amacrine cell activity. As temporal contrast increases, their modulatory activity increases, the net effect of which is to make the transient ganglion cells respond more quickly and more transiently. This silicon retina outputs spike trains that capture the behavior of ON- and OFF-center wide-field transient and narrow-field sustained ganglion cells, which provide 90% of the primate retina's optic nerve fibers. These are the four major types that project, via thalamus, to the primary visual cortex. Clearly, the parvo–magno retina has complex and interesting properties, but the extremely large mismatch between pixel response characteristics makes it quite difficult to use.

III. FROM BIOLOGICAL MODELS TO PRACTICAL VISION DEVICES

Many of the early developers of retinomorphic vision devices originated from the biological sciences community and saw their chips mainly as a means for demonstrating neurobiological models and theories, but did not relate their devices to real-world applications. Very few of the sensors so far have been used in practical applications, let alone in industry products. Many conceptually interesting pixel designs lack technical relevance because of, e.g., circuit complexity, large silicon area, low fill factors, or high noise levels, preventing realistic application. Furthermore, many of the early designs suffer from technical shortcomings of VLSI implementation and fabrication such as device mismatch, and did not yield practically usable devices.

Recently, an increasing amount of effort has been put into the development of practicable vision sensors based on biological principles [24], [25]. In this endeavor, the focus is less on a faithful reproduction of a biological model or a retina function, and more on devising engineering solutions to real-world vision problems.

Biologists have reported between 20 and up to 50 different types of cells in biological retinas [14], [23]. Their exact functionalities are still being investigated, however most experts agree on the following retina cell functions. There are: 1) cells sensitive to transients in luminance; 2) cells sensitive to sustained luminance (which is then, in turn, used to adjust other cells "operating point" depending on ambient light); 3) cells sensitive to direction of motion (which specialize to specific directions); and 4) cells sensitive to spatial contrast (which perform computations of the type ON-center OFF-surround). Besides this, there are cells sensitive to the wavelength of light (color), and in some animals, retinas have foveated topography. In the remainder of this paper, we will focus only on two of these aspects: sensitivity to light transients and sensitivity to absolute luminance.

A. Adapting to Technology: Address–Event Representation

Even though we observe striking parallels between VLSI hardware used to implement neuromorphic devices and neural wetware, some approaches taken by nature cannot be adopted in a feasible way. One prominent challenge posed is often referred to as the "wiring problem." Mainstream VLSI technology does not allow for the dense 3-D wiring observed everywhere in biological neural systems.[1]

In vision, the optic nerve connecting the retina to the visual cortex in the brain is formed by the axons of the about one million retinal ganglion cells, the spiking output cells of the retina. Translating this situation to an artificial vision system would imply that each pixel of an image sensor would have its own wire to convey its data out of the array. Given the restrictions posed by chip interconnect and packaging technologies, this is obviously not a feasible approach. However, VLSI technology does offer a workaround. Leveraging the five orders of magnitude or more of difference in bandwidth between a neuron (typically spiking at rates between 10 and 1000 Hz) and a digital bus enables engineers to replace thousands of dedicated point-to-point connections with a few metal wires and lots of switches, and to time multiplex the traffic over these wires using a packet-based or "event-based" data protocol called address–event representation (AER).

AER was originally proposed more than 20 years ago in Mead's Caltech research lab [26], [28], [29]. For over ten years, AER sensory systems were reported by only a handful of research groups, examples being Lazzaro *et al.* [30] and The Johns Hopkins University (Baltimore, MD, USA) [31] pioneering work on audition, or Boahen's early developments on retinas [32], [33]. However, during these years, some basic progress was made. A better understanding of asynchronous design [34], [35] leading to robust unarbitrated [36] and arbitrated [17] asynchronous event readout, combined with the availability of user-friendly field-programmable gate array (FPGA) external support for interfacing and new submicrometer technologies allowing complex pixels in reduced areas, heralded a new trend in AER bioinspired spiking sensor developments. Since 2003, many researchers have embraced this trend and AER has been widely used with retinomorphic vision sensors, in auditory systems and even for systems distributed over wireless networks [37].

In a point-to-point AER link, a transmitter chip (or module) includes, e.g., an array of neurons generating spikes. Each neuron is assigned an address, such as its x, y-coordinate within the array. Neurons generate spikes at

[1]Although neuromorphic engineers, inspired by cortical architectures, have looked at the increasingly available 3-D integrated circuit fabrication techniques and have built several experimental systems, 3-D VLSI yet remains a niche solution for neuromorphic devices [26], [27].

low frequency (10–1000 Hz), and these are arbitrated and put on an interchip (or intermodule) high-speed asynchronous AER bus, implementing a time-multiplexing strategy where all computing elements (pixels, neurons, etc.) share the same physical bus to transmit their pulses, together with the (implicit) timing information. In this asynchronous protocol, temporal information is self-encoded in the timing of the events, and it is explicitly added to the address in the form of a timestamp only when processing takes place in "non-AER" processing units, such as FPGAs or digital processors. Fig. 2(a) illustrates an AER interface servicing an array of spiking pixels. The ON-chip AER periphery contains address encoders, bus arbiters, and handshake circuitry, implementing a four-phase handshake with the data receiver [Fig. 2(b)] [17].

The AER bus is a multibit (either parallel, serial, or mixed) bus which transmits the addresses of the emitting neurons. Typical delays for transmitting address events between AER chips range from about 30 ns [38] to 1 μs [48] per event for parallel AER, and have been reported down to 16 ns per event for serial AER [39].[2]

Addresses are received, read, and decoded by the receiver chip (or module) and sent to the corresponding destination neuron or neurons. The use of AER splitters and mergers [40] allows extension to one-to-many, many-to-one, or many-to-many AER links. Inserting AER mappers [40] allows coordinate transformations (rotations, translations, etc.) to be performed while address events travel between modules. Current research is looking at how large numbers of AER convolutional modules can be combined through independent and multiple AER links to build high-speed object and texture recognition systems [41].

B. Retinomorphic Vision Devices

Today, practically all bioinspired vision sensors with spiking output use the AER protocol to communicate their data. These devices come in a variety of different types.

One class of bioinspired image sensors relies on AER to transmit pixel intensity values that are encoded in the relative timing or in the instantaneous rates of spike-like events, generated by pixel circuits in response to incident light levels [25]. These sensors do not achieve redundancy suppression or latency reduction, yet have interesting properties such as intrinsic wide DR operation and support for pixel-level analog-to-digital (AD) conversion.

The family of sensors, which are the most advanced and productized bioinspired vision devices today and that are the main focus of this paper, however follow the natural, event-driven, frame-free approach, capturing and being driven by transient events in the visual scene. These sensors' output is compressed at the sensor level, without the need of external processors, optimizing data transfer, storage, and processing, hence increasing power efficiency and compactness of the vision system. The dynamic vision sensor (DVS) is the first practically usable device of this class and has triggered a plethora of research in event-based vision. The asynchronous time-based image sensor (ATIS) continues the maturing and push to real-world applicability of bioinspired vision and derives its unique characteristics from combining the advantages of event-driven acquisition and time-domain spiking encoding of image information. The DAVIS sensor proposes a hybrid between frame-based and frame-free sensor technology.

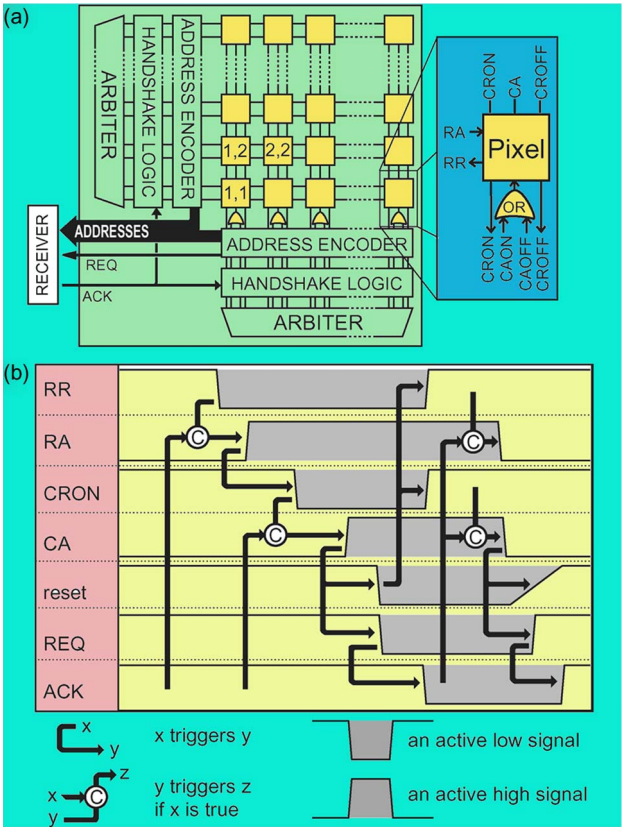

Fig. 2. *(a) Block level schematic of a pixel array embedded in AER communication periphery. (b) Timing diagram of a communication cycle for one ON event: RR—row request, RA—row acknowledge, CRON—column request ON polarity, CA—column acknowledge, reset—pixel reset, REQ—external request signal, ACK—external acknowledge signal. Delays are nondeterministic internal propagation delays except between REQ and ACK which is determined by the external receiver. (Adapted from [48].)*

[2]The speed advantage of the reported interchip serial AER interface is due to the use of low-voltage differential signaling (LVDS) transmission over pairs of impedance matched microstrip lines that can be driven near the physical limit of several gigabits per second. For example, for 32-b events using an 8-b/10-b encoding (which results in 40 b per event), an event is transmitted in only 16 ns [39]. Comparable transmission speeds are difficult to achieve with parallel AER connections as multiple parallel bit lines would need to be jitter- and skew-free down to the level of a few picoseconds.

1) Encoding Images in the Time Domain: First, let us have a closer look at one species of "event-based" cameras that use AER to encode and transmit pixel illuminance data. As in biology, these devices encode illuminance in the time domain, i.e., in the timing or rate of spike "events." Yet these devices are not "event driven" in the sense that their pixels autonomously react to visual events in the scene.

From an engineering point of view, time-domain encoding of visual information has technical merits as it optimizes the phototransduction individually for each pixel by abstaining from imposing a fixed integration time for all pixels in an array. Exceptionally high DR and improved signal-to-noise ratio (SNR) as compared to conventional imaging techniques are immediate benefits of this approach [25]. In particular, DR is no longer limited by the power-supply rails as in standard CMOS active pixel sensors, thus providing relative immunity to the aggressive supply-voltage scaling of modern CMOS technologies.

The so-called "octopus retina" [42] encodes and communicates individual pixel intensities in the instantaneous frequency (or interspike intervals) of AER events emitted concurrently by each pixel. In contrast to conventional, serially scanned arrays that allocate an equal portion of the bandwidth to all pixels independently of activity, this biologically inspired readout method offers activity-driven, pixel-parallel readout. In the octopus sensor, brighter pixels are favored because their integration threshold is reached faster than darker pixels, thus their AER events are communicated at a higher frequency. Consequently, brighter pixels request the output bus more often than darker ones and are updated more frequently. The rather large fixed-pattern noise (FPN) makes the sensor hard to sell for conventional imaging. The octopus sensor has the advantage of a smaller pixel size compared to other AER retinas, but has the disadvantage that the bus bandwidth is allocated according to the local scene luminance. Because there is no reset mechanism and 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 AER communication bus. Another drawback of this approach is the complexity of the digital frame grabber required to count the spikes produced by the array. The buffer must either count events over some time interval, or hold the latest spike time and use this to compute the intensity value from the interspike interval to the current spike time. This, however, leads to a noisier image. The octopus retina would probably be most useful for tracking small and bright light sources.

A biologically inspired approach based on relative spike timing is implemented in the so-called "time-to-first spike" (TTFS) imager [43]–[45]. In this encoding method, the system does not require the storage of large number of spikes since every pixel generates only one spike per frame. This coding method was also suggested as a scheme used by neurons in the visual system to code information [46]. The global threshold for generating a spike in each pixel can be reduced over the frame time so that dark pixels will still emit a spike in a reasonable amount of time. A disadvantage of the TTFS sensors is that uniform parts of the scene all try to emit their events at the same time, overwhelming the AER bus. This problem can be mitigated by serial reset of the, e.g., rows of pixels, but, of course, then the problem can arise that a particular image still causes simultaneous emission of events.

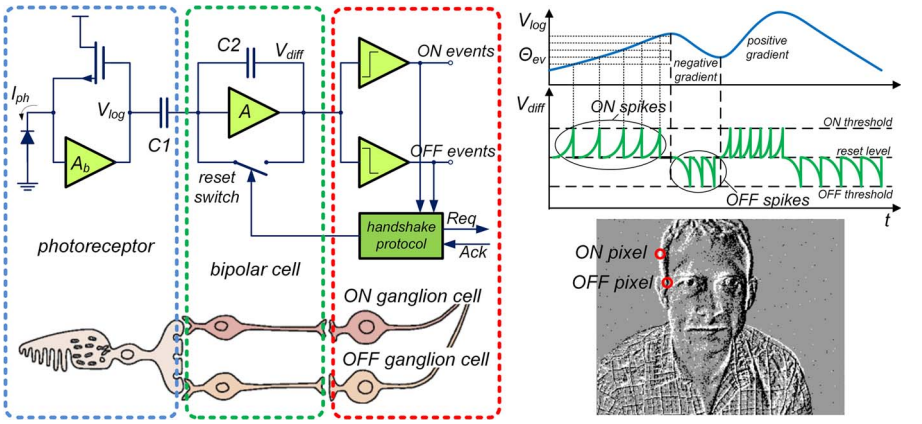

Fig. 3. ***Three-layer model of a human retina and corresponding DVS pixel circuitry (left). Typical signal waveforms of the pixel circuit are shown top right. The upper trace represents a voltage waveform at the node V_{log} tracking the photocurrent through the photoreceptor. The bipolar cell circuit responds with spike events (V_{diff}) of different polarity to positive and negative changes of the photocurrent, while being monitored by the ganglion cell circuit that also transports the spikes to the next processing stage; the amount of log-intensity change is encoded in the number of events, the rate of change in interevent intervals. The bottom right image shows the response of an array of DVS pixels to a natural scene (person moving in the field of view of the sensor). Events have been collected for some tens of milliseconds and are displayed as an event map image with ON (going brighter) and OFF (going darker) events drawn as white and black dots.***

2) Pixel-Autonomous Detection of Temporal Contrast—DVS: Practically all conventional frame-based image sensors completely neglect the dynamic information immanent to natural scenes and perceived in nature by the magno–cellular transient pathway. In an attempt to realize a practical transient vision device based on the functioning of the magno–cellular pathway, the DVS was developed [47]–[49]. This type of vision sensor is sensitive to the scene dynamics and directly responds to changes, i.e., temporal contrast, pixel individually, and near real time. The gain in terms of temporal resolution with respect to standard frame-based image sensors is dramatic. But also other performance parameters like the intrascene DR greatly profit from the biological approach. This type of sensor is very well suited for a plethora of machine vision applications involving high-speed motion detection and analysis, object tracking, and shape recognition.

The DVS pixel models a simplified three-layer retina (Fig. 3), implementing an abstraction of the photoreceptor-bipolar-ganglion cell information flow. Single pixels are spatially decoupled but take into account the temporal development of the local light intensity. The DVS pixel autonomously responds to relative changes in intensity at microsecond temporal resolution over six decades of illumination [49].

These properties are a direct consequence of abandoning the frame principle and modeling three key properties of biological vision: the sparse, event-based output; the representation of relative luminance change (thus directly encoding scene reflectance change); and the rectification of positive and negative signals into separate output channels (ON/OFF). The major consequence of this bioinspired approach and most distinctive feature with respect to standard imaging is that the control over the acquisition of the visual information is no longer being imposed to the sensor in the form of external timing signals such as shutter or frame clock, but the decision making is transferred to the single pixel that handles its own visual information individually and autonomously.

Fig. 4 illustrates the DVS principle of operation and demonstrates the high-speed, high-temporal-resolution operation of event response. Each DVS pixel produces a continuous-time spatio–temporal representation of the visual dynamics in its field of view by detecting relative changes (i.e., temporal contrast) in illuminance. In this example, a DVS pixel array is observing a light dot on an analog oscilloscope screen moving in a spiral pattern which is repeated at a 500-Hz rate [Fig. 4(a)]. The address events describe the motion of the dot at microsecond temporal resolution in space and time [Fig. 4(b)]. Fig. 3 shows the basic block diagram of a typical DVS pixel circuit [49], [52], [61], [64]. The first stage transduces photocurrent to a voltage proportional to the logarithm of light

$$V_{\log} = V_{\mathrm{DC}} + A_v U_T \ln I_{ph}$$

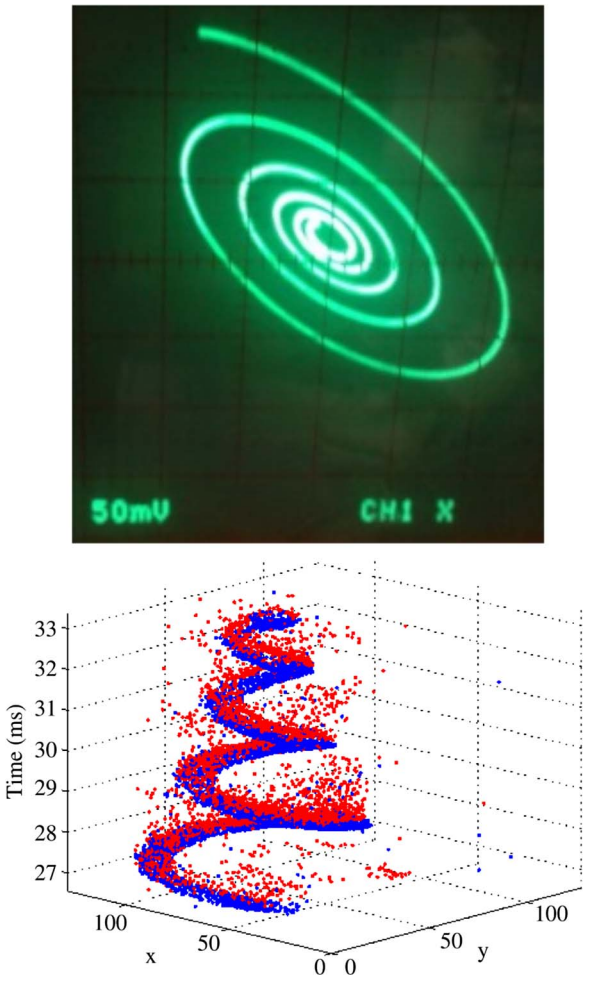

Fig. 4. ***Illustration of DVS output. (a) A DVS is observing a 500-Hz spiral on an analog oscilloscope. (b) DVS output is a continuous sequence of address events (x, y) in time. Each address event signals that the pixel at that coordinate experienced a change of light at that instant. Red and blue events represent a positive (increase) or negative (decrease) change of light.***

where U_T is thermal voltage, A_v is a voltage gain factor, and V_{DC} is a light independent direct current (dc) voltage level with high interpixel mismatch. The second stage amplifies $V_{\log}$ by C_1/C_2 resetting the charge integrated at C_2 every time $V_{\mathrm{diff}} = (C_1/C_2)V_{\log}$ varies a fixed threshold $\pm V_{\mathrm{th}}$ set by the comparators. This also eliminates the dc component at $V_{\log}$. The result is that each pixel generates a signed asynchronous output "event" every time its light changes by $\ln I_{ph}(t_2) - \ln I_{ph}(t_1) = \pm\theta_{ev}$, with

$$\theta_{ev} = \frac{C_2 V_{\mathrm{th}}}{C_1 U_T A_v}.$$

Consequently, pixel information is obtained not synchronously at fixed time steps δt (as in conventional video

sensors), but asynchronously, driven by data at fixed relative light increments $\theta_{ev} = |\ln(I_{ph}(t_2)/I_{ph}(t_1))|$, as shown in Fig. 3. The event output of the pixel encodes the temporal development of the local illuminance, however, it does not at any time contain information on the absolute intensity seen by the pixel, commonly referred to as gray level. Hence, DVSs do not acquire image information in the conventional sense.

Parameter θ_{ev} represents the minimum detectable temporal contrast or "contrast sensitivity" of a DVS pixel

$$\theta_{ev} = \left|\ln\left(\frac{I_{ph}(t_2)}{I_{ph}(t_1)}\right)\right| \approx \left|\frac{\Delta I_{ph}}{I_{ph}}\right|$$

in practice limited by the noise of the photoreceptor front–end and its bandwidth. In practical designs, noise-equivalent contrasts (NECs) with related contrast sensitivities of around 1% have been achieved [50], [52], [53].

The demand for compact pixel size imposes the use of simple, area-efficient two-transistor comparators with offset voltage mismatch standard deviations around 10–20 mV, resulting in minimum practical values for V_{th} of between 50 and 100 mV. In order to approach the contrast sensitivity levels achievable with single pixels, the effect of mismatch must be minimized by maximizing overall the voltage gain $A_T = A_v C_1/C_2$. Several approaches to increasing this gain factor have been proposed, including increasing the voltage gain A_v of the front–end [50], adding a preamp stage [51], [52] or using a two-stage capacitive feedback amplifier configuration [54].

Fig. 5(a) shows the original phototransduction stage [49], [64] with $A_v = n_n$, where n_n is the subthreshold slope factor of NMOS transistor M_{n1}. An overall voltage gain was obtained by setting $C_1/C_2 = 20$.

An alternative photoreceptor front–end circuit with additional preamplification stage is shown in Fig. 5(b). Increased front–end voltage gain combined with a reduced capacitor ratio improves overall voltage gain to about 60 (a factor 3 increase) while saving capacitor area [50]. Fig. 5(c) illustrates an improved transimpedance amplifier-based technique to increase contrast sensitivity by about a factor 10 while minimizing additional mismatch. To achieve this, gain A_v is now $n_n N$, where N, the number of stacked diode-connected transistors of a preamplification stage, is a mismatch-free factor. Voltage headroom limits the practical number of stacked transistors to $N = 4$ and the circuit reduces intrascene DR [52]. A two-stage version of the capacitive feedback amplifier also yielded an overall gain increase by about a factor 10, as reported in [54].

A recent DVS pixel design [55] is aimed at color vision [color DVS (cDVS)]. The cDVS simultaneously detects relative intensity and absolute wavelength change events using a single buried double junction (BDJ) photodiode. Measurements show that the cDVS color change pathway can detect light wavelength changes as small as 15 nm [55], although basic limitations on the color separation capabilities and dark current of parasitic BDJs make this approach challenging.

DVS relative change events and gray-level image frames are two highly orthogonal representations of a visual scene. The former contains the information on local relative changes, hence encodes all dynamic contents, yet there is no information about absolute light levels or static parts of the scene. The latter is a snapshot that carries an absolute intensity map at a given point in time, however has no information about any motion; hence, if scene dynamics are to be captured, one needs to acquire many of those frames. In principle, there is no way to recreate DVS change events from image frames nor can gray-level images be recreated from DVS events.

3) Pixel-Individual Image Acquisition—ATIS: Combining relative change detection with absolute exposure measurement at the pixel level leads to a sensor with very rich

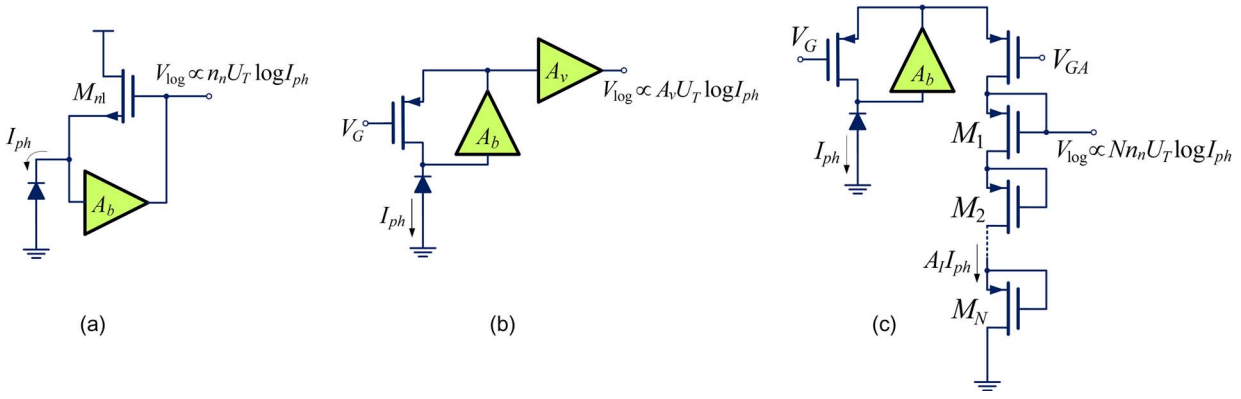

Fig. 5. *(a) DVS original photocurrent transduction circuit [49], (b) alternate transduction with mismatch sensitive preamplification [50], and (c) transimpedance transduction with mismatch insensitive preamplification [52].*

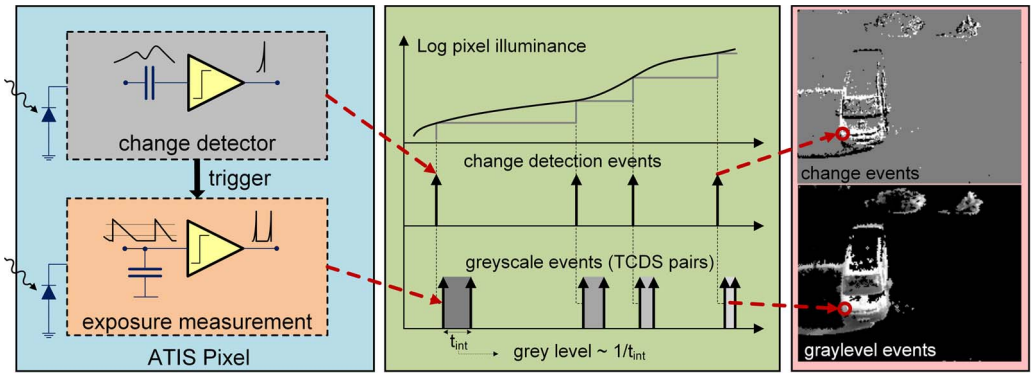

Fig. 6. *The ATIS pixel containing a change detector circuit and a conditional exposure block. An exposure measurement is executed when triggered by a change detection. As a result, two types of asynchronous AER events, encoding change and exposure information, are generated and transmitted separately. On the right, change detection events recorded during a short time window are displayed; associated gray-level updates at the corresponding pixel positions are shown below.*

visual information output that provides the means to overcome some of the persistent limitations faced in today's vision engineering.

Besides limited temporal resolution, data redundancy is the other major drawback of conventional frame-based imaging. Each frame contains data from all pixels, regardless of whether the information has changed since the last frame had been acquired. Consequently, pixel values from unchanged parts in the scene get recorded and transmitted over and over, even though they do not contain any (new) information. This serious inefficiency of the standard paradigm of image acquisition has been tolerated for decades and no viable remedy has been found until recently.

Ideally, only information that is relevant—because unknown—should be acquired, transmitted, and processed. Approaches such as sensor-level and even pixel-parallel frame differencing have been proposed [56]–[59], however all frame differencing imagers still rely on acquisition and processing of full frames of image data and are not able to self-consistently suppress temporal redundancy and provide real-time compressed video output. Furthermore, even when processing and difference quantization is done at the pixel level, the temporal resolution of the acquisition of the scene dynamics, as in all frame-based imaging devices, is still limited to the achievable frame rate and is time quantized to this frame rate. The main obstacle for sensor-driven video compression lies in the necessity to combine a pixel identifier and the corresponding grayscale value and implement conditional readout using the available array scanning readout techniques.

As with many surprisingly simple solutions to persistent problems, a combination of approaches developed independently in seemingly unrelated fields led to a breakthrough. Developing a detector readout chip for a high-energy physics experiment in the 1990s [60], [61], we proposed a circuit that detects an "event" (the event of an elementary particle passing through a particle detector) and encodes information on electrical charge related to this event asynchronously in the time domain (i.e., in the time between two pulse edges). Later, we discovered that AER, discussed in Section III-A, would provide the means to conveying asynchronous time-domain information off a large array of pixels and include the spatial information about the location of the source of the information in the array (the pixel x, y address). Finally, with an array of DVS pixel circuits [47]–[49], we now dispose of the information where in a visual scene and when (at very high temporal precision) something has changed and new information is available to be acquired. Combining these techniques, it becomes possible—for the first time—to acquire image information not frame-wise but conditionally only from parts in the scene where there is new information, and so to overthrow the seemingly inviolable paradigm of frame-based image acquisition [62].

The ATIS [63], [64] introduces fully autonomous pixels that combine a DVS change detector and a conditional exposure measurement circuit. The change detector independently and asynchronously initiates the measurement of a new exposure/grayscale value only if—and immediately after—a brightness change of a certain magnitude has been detected in the field of view of the respective pixel (Fig. 6). The exposure measurement circuit acquires absolute instantaneous pixel illuminance by converting the integrated photocharge into the timing of asynchronous pulse edges (see Figs. 7 and 8).

Letting each exposure measurement be triggered by a change detection, the ATIS pixel does not rely on external timing signals and autonomously requests readout access only when it has a change event or a new grayscale timing pulse to communicate. At the readout periphery, change and grayscale events are arbitrated, furnished with the pixel's array address by an address encoder and sent out on an asynchronous bit-parallel AER bus [17], [65]. Fig. 6 shows a functional diagram of the ATIS pixel.

Time-based encoding of image data: Instead of encoding pixel values in electrical quantities—voltage, current, or charge—image data can also be encoded in the time domain [25]. The basic principle of time-domain image sensors is illustrated in Fig. 7(a). To initiate a photomeasurement cycle, the photodiode node is reset to a defined voltage level $V_{\mathrm{pix}}, 0$ by applying a short pulse signal V_{reset} to a reset switch. Subsequently, the photodiode is discharged by the photo-generated current. The integration, and hence the photomeasurement, is finished not after a fixed exposure time as in standard imagers, but when the integration voltage reaches a given threshold level. The exposure information is encoded in the time between the V_{out} pulse and the reset with the incident light intensity being inversely proportional to the integration time t_{int}.

To eliminate comparator offset FPN, switching delay errors, and kTC noise, a time-domain differential mode can be realized. As shown in Fig. 7(b), an illuminance-proportional integration time t_{int} is established by measuring the time difference for the voltage V_{pix} to drop from a first reference voltage V_{refH} to a second reference voltage V_{refL}, such implementing a time-domain correlated double sampling (TCDS) light-to-time transduction [66]. The TCDS in-pixel circuitry uses a single comparator and pixel-level asynchronous digital logic to automatically control the switching between two reference voltages within the same integration cycle.

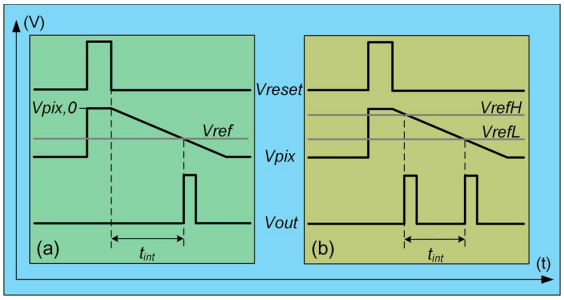

Fig. 7. ***(a) Principle of time-based imaging. (b) Time-based imaging with TCDS.***

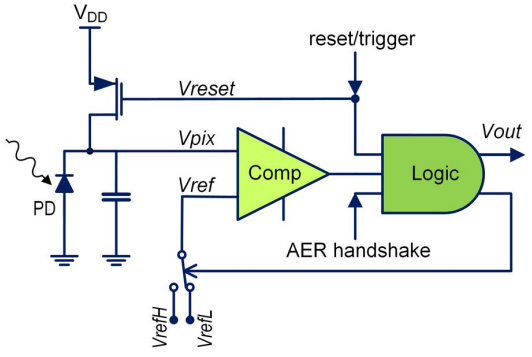

Fig. 8. ***ATIS pixel exposure measurement circuit.***

The benefits of ATIS image data acquisition are manifold.

- Sensor-level video compression: The temporal redundancy suppression of the change-detector-controlled operation ideally yields lossless focal-plane video compression with compression factors depending only on scene dynamics. Theoretically approaching infinity for static scenes, in practice, due to background noise-triggered events, the achievable compression factor reaches 1000 for bright low-activity scenes. Typical dynamic scenes yield compression ratios between 20 and several hundred. Fig. 9 shows a typical surveillance scene generating a 2500–50 000 events/s @ 18 bit/event continuous-time video stream.
- Fine temporal resolution: ATIS avoids the unnatural time quantization of frame-based image data acquisition. Continuous-time operation results in a temporal resolution of the acquired scene dynamics (depending on light levels) orders of magnitude better than standard imagers [e.g., 1000 frames per second (fps) equivalent at $>$ 100 lux].
- Wide DR and improved SNR: Time-domain encoding of the intensity information automatically optimizes the integration time separately for each pixel instead of imposing a fixed integration time for the entire array, resulting in exceptionally high DR and improved SNR: An intrascene DR of 143 dB (static) and 125 dB at 30 fps equivalent temporal resolution, and an SNR of $>$ 56 dB have been measured. In contrast to conventional image sensors, SNR of time-domain encoding image sensors is largely independent of light levels [64].

4) Hybrid DVS Plus APS Sensor: DAVIS: Another recent approach to combine dynamic and static information into a single pixel is the so-called dynamic and active pixel vision sensor (DAVIS) [67], shown in Fig. 10. This pixel combines conventional frame-based sampling of intensity with asynchronous detection of log intensity changes. It relies on the fact that a logarithmic photoreceptor can continuously measure the photocurrent without consuming it, so the photocurrent can also be integrated over time to produce a voltage signal. The DAVIS has the advantages of sharing the same photodiode with the DVS circuit and a small readout circuit that only adds a few transistors to the pixel, increasing the DVS pixel area by about 5%. It allows to capture conventional images, which are compatible with years of research in machine vision. Of course, this also brings back the disadvantages of providing a redundant, sampled intensity output with linear encoding of intensity. It remains to be seen if the DVS output can be used to efficiently trigger frame captures. If so, then perhaps the DAVIS can bring together conventional machine vision and bioinspired, event-based approaches.

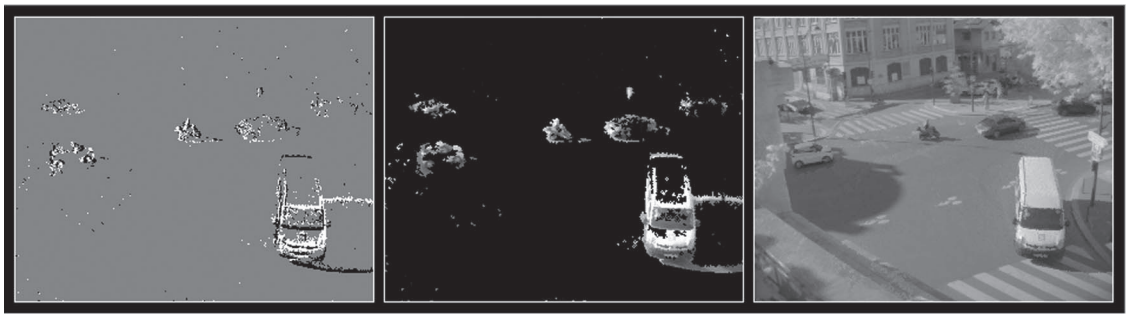

Fig. 9. *Typical surveillance scene data acquired with an ATIS, generating a 2500–100 000 events/s @ 18 bit/event continuous-time video stream. The actual event rate depends on instantaneous scene activity. Comparing corresponding bit rates—45 000 to 1 800 000 b/s—to the raw data rate of a QVGA 8-b grayscale sensor at 30 fps of 18 Mb/s demonstrates lossless video compression with compression factors between 10 and 400 for this example scene. Positive (ON) and negative (OFF) change events collected during 30 ms and displayed as black and white pixels (left); gray levels measured by the pixels that have registered a change event during the past 30 ms, for better illustration shown on an empty background. The effect of objects triggering exposure measurement in the pixel they hit while moving across the focal plane becomes apparent (middle); same point in time into the video sequence with full background data displayed, showing compressed video output from the ATIS.*

IV. DISCUSSION AND CONCLUSION

This paper focused on retinomorphic vision sensor designs. There are many areas of possible improvement and innovation in the design of silicon retinas and the processing of their outputs.

The CMOS image sensor (CIS) industry seems still to be mired in the "megapixel race," where a main aim is to offer more pixels for less money without sacrificing basic image capturing capability. Bioinspired vision sensors also face this problem, because pixels that are too large are hard to sell in mass production.

Event-based silicon retinas offer either spatial or temporal processing and none of them offer powerful spatial redundancy reduction. No one has thus far built a high-performance color silicon retina, although color is a basic feature of biological vision in all diurnal animals.

The rather poor quantum efficiencies and fill factors of silicon retinas are also a consequence of using standard CMOS technologies. One solution is the use of integrated microlenses, which concentrate light onto the photodiode. However, standard microlenses offered in CIS processes are optimized for pixels smaller than 5 um, so a CIS technology offering large micolenses is required. Another possible improvement could come from backside illumination (BSI). Normally, a vision sensor is illuminated from the top (front) of the wafer. However, for tiny-pixel CMOS imagers frontside illumination (FSI) is a big problem, because the photodiode sits at the bottom of a tunnel through all the overlaying metal and insulator layers, making it difficult to capture light, particularly at the edges of the sensor when using a wide angle lens. This problem led the development of BSI, where the wafer is thinned down to less than 20 μm and is illuminated from the back

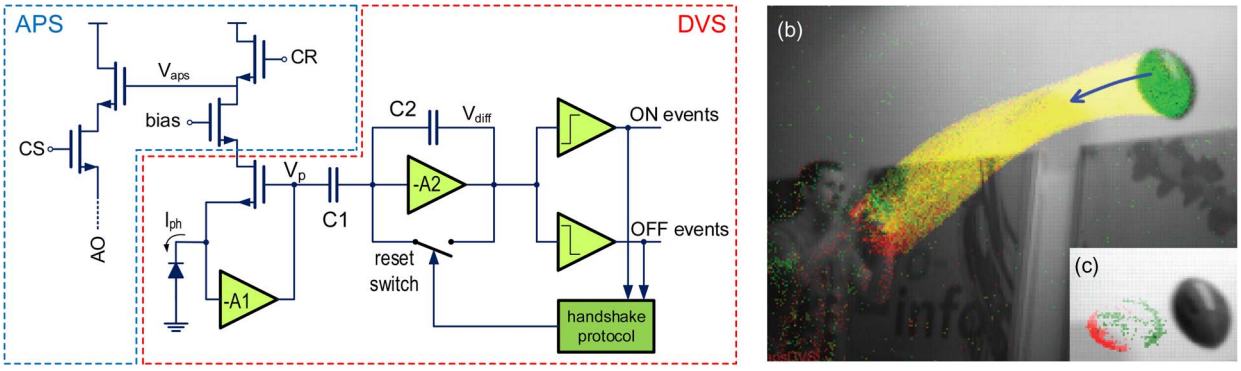

Fig. 10. *The DAVIS vision sensor. (a) The pixel circuit combines conventional APS with a DVS circuit. (b) Snapshot from 240 × 180 sensor showing a captured APS frame in grayscale with the DVS events in color. The ball was flying toward the person. (c) 5 ms of output right after the frame capture of the ball.*

rather than the front. Now all the silicon area can receive light and, if properly designed, most of the photocharge is collected by the photodiode. Intense development of BSI image sensors by industry may shortly make this technology more accessible for prototyping. New problems can arise, such as unwanted parasitic photocurrents in junctions other than the photodiode. These currents can disturb the pixel operation, particularly when the pixel stores a charge on a capacitor like a global-shutter CMOS imager or the DVS pixel.

The retina implementations described here show the variety of approaches taken toward building high-quality AER vision sensors. These sensors can now be used for solving practical machine vision problems. In order to take full advantage of the characteristics of these silicon retinas, development of customized, event-based processing techniques is required. Image processing for frame-based vision is highly developed as the field has been expanding for many decades. As a consequence, open source processing packages such as OpenCV are available. Event-based sensing and processing is a quite incipient field. Although event-based sensors have been reported for some time now, higher level processing is still very premature. Nonetheless, there are some initial promising results already reported. For example, Pérez-Carrasco *et al.* [68] have shown 1-ms symbol recognition latency including both sensing and processing under indoor light illumination, based on event-driven convolutional neural networks. Similarly, O'Connor *et al.* [69] have shown handwritten character recognition using event-driven stochastic encoding learning and processing. No complete package such as OpenCV for event-driven vision processing exists to date, but there is a growing open-source Java-based environment for controlling AER-type hardware as well as performing basic processing like optic flow, spatial filtering, ego-motion suppression, etc. [70].

Future developments in bioinspired vision will focus on continued sensor improvements, together with developments of algorithms and hardware architectures for processing the sensor outputs. These approaches should further improve performance in artificial vision systems, e.g., for wide DR high-speed imaging, while at the same time decreasing computational cost and latency by taking advantage of the sparsity and the precise timing of the event-based visual information encoding. ■

REFERENCES

[1] Wikipedia, "Bionics." [Online]. Available: http://en.wikipedia.org/wiki/Bionics

[2] W. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biol.*, no. 5, pp. 115–133, 1943.

[3] D. Hebb, *The Organization of Behavior*. New York, NY, USA: Wiley, 1949.

[4] A. Hodgkin and A. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *J. Physiol.*, vol. 117, pp. 500–544, 1952.

[5] C. Mead, *Analog VLSI and Neural Systems*. Reading, MA, USA: Addison-Wesley, 1989.

[6] C. Mead, "Neuromorphic electronic systems," *Proc. IEEE*, vol. 78, no. 10, pp. 1629–1636, Oct. 1990.

[7] M. A. C. Maher, S. P. Deweerth, M. A. Mahowald, and C. A. Mead, "Implementing neural architectures using analog VLSI circuits," *IEEE Trans. Circuits Syst.*, vol. 36, no. 5, pp. 643–652, May 1989.

[8] K. A. Boahen, "Neuromorphic microchips," *Sci. Amer.*, vol. 292, pp. 56–63, May 2005.

[9] R. Sarpeshkar, "Brain power-borrowing from biology makes for low power computing-bionic ear," *IEEE Spectrum*, vol. 43, no. 5, pp. 24–29, May 2006.

[10] G. Indiveri, "Synaptic plasticity and spike-based computation in VLSI networks of integrate-and-fire neurons," *Neural Inf. Process.—Lett. Rev.*, vol. 11, no. 4–5, pp. 135–146, Apr.–Jun. 2007.

[11] S. Furber and S. Temple, "Neural systems engineering," *J. Roy. Soc. Interface*, vol. 2007, no. 4, pp. 193–206, 2007.

[12] R. W. Rodieck, *The First Steps in Seeing*. Sunderland, MA, USA: Sinauer Associates, 1998.

[13] S. W. Kuffler, "Discharge patterns and functional organization of mammalian retina," *J. Neurophysiol.*, vol. 16, pp. 37–68, 1953.

[14] R. Masland, "The fundamental plan of the retina," *Nature Neurosci.*, vol. 4, pp. 877–886, 2001.

[15] R. W. Rodieck, "The primate retina," *Comput. Primate Biol.*, vol. 4, pp. 203–278, 1998.

[16] F. S. Werblin and J. E. Dowling, "Organization of the retina of the mudpuppy Necturus maculosus: II. Intracellular recording," *J. Neurophysiol.*, vol. 32, pp. 339–355, 1969.

[17] K. 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.

[18] D. K. Warland, P. Reinagel, and M. Meister, "Decoding visual information from a population of retinal ganglion cells," *J. Neurophys.*, vol. 78, pp. 2336–2350, 1997.

[19] M. A. Mahowald and C. A. Mead, "The silicon retina," *Sci. Amer.*, vol. 264, no. 5, pp. 76–82, May 1991.

[20] T. Delbruck and C. Mead, "Adaptive photoreceptor circuit with wide dynamic range," in *Proc. IEEE Int. Symp. Circuits Syst.*, 1994, vol. 4, pp. 339–342.

[21] M. Mahowald, *An Analog VLSI System for Stereoscopic Vision*. Boston, MA, USA: Kluwer Academic, 1994.

[22] K. A. Zaghloul and K. Boahen, "A silicon retina that reproduces signals in the optic nerve," *J. Neural Eng.*, vol. 3, pp. 257–267, 2006.

[23] H. Wassle, "Parallel processing in the mammalian retina," *Nature Rev. Neurosci.*, vol. 5, pp. 747–757, 2004.

[24] S. C. Liu and T. Delbruck, "Neuromorphic sensory systems," *Curr. Opinion Neurobiol.*, vol. 20, pp. 288–295, 2010.

[25] D. Chen, D. Matolin, A. Bermak, and C. Posch, "Pulse modulation imaging—Review and performance analysis," *IEEE Trans. Biomed. Circuits Syst.*, vol. 5, no. 1, pp. 64–82, Feb. 2011.

[26] H. Kurino *et al.*, "Smart vision chip fabricated using three dimensional integration technology," in *Advances in Neural Information Processing Systems 13*, T. Leen, T. Dietterich, and V. Tresp, Eds. Cambridge, MA, USA: MIT Press, pp. 720–726, 2000.

[27] E. Culurciello and A. Andreou, "Capacitive coupling of data and power for 3D silicon-on-insulator VLSI," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2005, pp. 4142–4145.

[28] M. Sivilotti, "Wiring considerations in analog VLSI systems with application to field-programmable networks," Ph.D. dissertation, Comput. Neural Syst., California Inst. Technol. (Caltech), Pasadena, CA, USA, 1991.

[29] M. A. Mahowald, "VLSI analogs of neuronal visual processing: A synthesis of form and function," Ph.D. dissertation, Comput. Neural Syst., California Inst. Technol. (Caltech), Pasadena, CA, USA, 1992.

[30] J. Lazzaro, J. Wawrzynek, M. Mahowald, M. Silvilotti, and D. Gillespie, "Silicon auditory processors as computer peripherals," *IEEE Trans. Neural Netw.*, vol. 4, no. 3, pp. 523–528, May 1993.

[31] G. Cauwenberghs, N. Kumar, W. Himmelbauer, and A. G. Andreou, "An analog VLSI chip with asynchronous interface for auditory feature extraction," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 45, no. 5, pp. 600–606, May 1998.

[32] K. Boahen, "Retinomorphic chips that see quadruple images," in *Proc. Int. Conf. Microelectron. Neural Fuzzy Bio-Inspired Syst.*, 1999, pp. 12–20.

[33] K. Boahen, "A retinomorphic chip with parallel pathways: Encoding increasing, on, decreasing, and off visual signals," *Int. J. Analog Integr. Circuits Signal Process.*, vol. 30, pp. 121–135, 2002.

[34] J. Sparsø and S. B. Furber, *Principles of Asynchronous Circuit Design: A Systems*

Perspective. Dordrecht, The Netherlands: Kluwer Academic, 2001.

[35] A. J. Martin and M. Nyström, "Asynchronous techniques for system-on-chip design," *Proc. IEEE*, vol. 94, no. 6, pp. 1089–1120, Jun. 2006.

[36] A. Mortara, E. A. Vittoz, and P. Venier, "A communication scheme for analog VLSI perceptive systems," *IEEE J. Solid-State Circuits*, vol. 30, no. 6, pp. 660–669, Jun. 1995.

[37] T. Teixeira, A. G. Andreou, and E. Culurciello, "Event-based imaging with active illumination in sensor networks," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2005, pp. 644–647.

[38] J. Costas-Santos, T. Serrano-Gotarredona, R. Serrano-Gotarredona, and B. Linares-Barranco, "A 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.

[39] T. Iakymchuk *et al.*, "An AER handshake-less modular infrastructure PCB with x8 2.5Gbps LVDS serial links," in *Proc. IEEE Int. Symp. Circuits Syst.*, Melbourne, Australia, 2014, pp. 1556–1559.

[40] R. Serrano-Gotarredona *et al.*, "CAVIAR: A 45 k neuron, 5M synapse, 12G connects/s AER hardware sensory-processing-learning-actuating system for high-speed visual object recognition and tracking," *IEEE Trans. Neural Netw.*, vol. 20, no. 9, pp. 1417–1438, Sep. 2009.

[41] J. A. Pérez-Carrasco *et al.*, "Fast vision through frame-less event-based sensing and convolutional processing. Application to texture recognition," *IEEE Trans. Neural Netw.*, vol. 21, no. 4, pp. 609–620, Apr. 2010.

[42] E. Culurciello, R. Etienne-Cummings, and K. Boahen, "A biomorphic digital image sensor," *IEEE J. Solid State Circuits*, vol. 38, no. 2, pp. 281–294, Feb. 2003.

[43] Q. Luo and J. Harris, "A time-based CMOS image sensor," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2004, vol. IV, pp. 840–843.

[44] X. Qi, X. Guo, and J. Harris, "A time-to-first spike CMOS imager," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2004, vol. 4DOI: 10.1109/ISCAS.2004.1329131.

[45] C. Shoushun and A. Bermak, "Arbitrated time-to-first spike CMOS image sensor with on-chip histogram equalization," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 15, no. 3, pp. 346–357, Mar. 2007.

[46] S. Thorpe, D. Fize, and C. Marlot, "Speed of processing in the human visual system," *Nature*, vol. 381, pp. 520–522, 1996.

[47] P. Lichtsteiner and T. Delbruck, "A 64 × 64 AER logarithmic temporal derivative silicon retina," *Res. Microelectron. Electron.*, vol. 2, pp. 202–205, Jul. 25–28, 2005.

[48] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128 × 128 120 dB 15μs latency asynchronous temporal contrast vision sensor," *IEEE J. Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, Feb. 2008.

[49] P. Lichtsteiner, C. Posch, and R. Delbruck, "A 128× 128 120 dB 30 mW asynchronous vision sensor that responds to relative intensity change," in *Dig. Tech. Papers IEEE Int. Solid-State Circuits Conf.*, Feb. 6–9, 2006, pp. 2060–2069.

[50] T. Delbruck and R. Berner, "Temporal contrast AER pixel with 0.3%-contrast event threshold," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2010, pp. 2442–2445.

[51] J. A. Leñero-Bardallo, T. Serrano-Gotarredona , and B. Linares-Barranco, "A 3.6 us latency asynchronous frame-free event-driven dynamic-vision-sensor," *IEEE J. Solid-State Circuits*, vol. 46, no. 6, pp. 1443–1455, Jun. 2011.

[52] T. Serrano-Gotarredona and B. Linares-Barranco, "A 128 × 128 1.5% contrast sensitivity 0.9% FPN 3μs latency 4 mW asynchronous frame-free dynamic vision sensor using transimpedance preamplifiers," *IEEE J. Solid-State Circuits*, vol. 48, no. 3, pp. 827–838, Mar. 2013.

[53] C. Posch and D. Matolin, "Sensitivity and uniformity of a 0.18 μm CMOS temporal contrast pixel array," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 15–18, 2011, pp. 1572–1575.

[54] C. Posch, D. Matolin, and R. Wohlgenannt, "A two-stage capacitive-feedback differencing amplifier for temporal contrast IR sensors," *Analog Integr. Circuits Signal Process. J.*, vol. 64, no. 1, pp. 45–54, 2010.

[55] R. Berner and T. Delbruck, "Event-based pixel sensitive to changes of color and brightness," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 58, no. 7, pp. 1581–1590, Jul. 2011.

[56] K. Aizawa *et al.*, "Computational image sensor for on sensor compression," *IEEE Trans. Electron Devices*, vol. 44, no. 10, pp. 1724–1730, Oct. 1997.

[57] V. Gruev and R. Etienne-Cummings, "A pipelined temporal difference imager," *IEEE J. Solid-State Circuits*, vol. 39, no. 3, pp. 538–543, Mar. 2004.

[58] J. Yuan, H. Y. Chan, S. W. Fung, and B. Liu, "An activity-triggered 95.3 dB DR 75.6 dB THD CMOS imaging sensor with digital calibration," *IEEE J. Solid-State Circuits,* vol. 44, no. 10, pp. 2834–2843, Oct. 2009.

[59] Y. M. Chi *et al.*, "CMOS camera with in-pixel temporal change detection and ADC," *IEEE J. Solid-State Circuits*, vol. 42, no. 10, pp. 2187–2196, Oct. 2007.

[60] C. Posch, S. Ahlen, E. Hazen, and J. Oliver, "CMOS front-end for the MDT sub-detector in the ATLAS Muon Spectrometer-development and performance," in *Proc. 7th Workshop Electron. LHC Exp.*, 2001, CERN-2001-005/CERN-LHCC-2001-034.

[61] C. Posch, E. Hazen, and J. Oliver, "MDT-ASD, CMOS front-end for ATLAS MDT," CERN, ATL-MUON-2002-003, 2002.

[62] C. Posch, D. Matolin, and R. Wohlgenannt, "An asynchronous time-based image sensor," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 18–21, 2008, pp. 2130–2133.

[63] C. Posch, D. Matolin, and R. Wohlgenannt, "A QVGA 143 dB dynamic range asynchronous address-event PWM dynamic image sensor with lossless pixel-level video compression," in *Proc. IEEE Conf. Solid-State Circuits*, Feb. 2010, pp. 400–401.

[64] 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," *J. Solid-State Circuits*, vol. 46, no. 1, pp. 259–275, Jan. 2011.

[65] K. 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.

[66] D. Matolin, C. Posch, and R. Wohlgenannt, "Correlated double sampling and comparator design for time-based image sensors," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 18–21, 2009, pp. 1269–1272.

[67] R. Berner, C. Brandli, M. Yang, S.-C. Liu, and T. Delbruck, "A 240 × 180 10 mW 12 us latency sparse-output vision sensor for mobile applications," in *Proc. Symp. Very Large Scale Integr. (VLSI)*, Kyoto, Japan, 2013, pp. C186–C187.

[68] J. A. Pérez-Carrasco *et al.*, "Mapping from frame-driven to frame-free event-driven vision systems by low-rate rate-coding and coincidence processing. Application to feed-forward ConvNets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2706–2719, Nov. 2013.

[69] P. O'Connor, D. Neil, S. C. Liu, T. Delbruck, and M. Pfeiffer, "Real-time classification and sensor fusion with a spiking deep belief network," *Front. Neurosci*, vol. 7, 2013, DOI: 10.3389/fnins.2013.00178.

[70] jAER, "jAER Open Source Project." [Online]. Available: http://jaerproject.org/