A scalable low power imager architecture for compound-eye vision sensors

Farid Boussaid[1], *Senior Member, IEEE*, Chen Shoushun[2], and Amine Bermak[2], *Senior Member, IEEE.*
[1]*School of Electrical, Electronic and Computer Engineering, The University of Western Australia, Perth, Australia*
[2]*Department of Electrical and Electronic Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, SAR.*

***Abstract*—In this paper, we propose a scalable low power imager architecture for compound-eye vision sensors. The proposed hardware architecture is based on a time domain data representation as well as a biologically inspired read-out strategy using Address-Event-Representation (AER). The proposed AER approach to compound-eye imaging enables low power operation (10nA/pixel), efficient read-out, improved signal-to-noise ratio together with wide dynamic range. Moreover, the proposed AER-based VLSI architecture is scalable and well suited to the next generation of deep submicron silicon processes owing to decreased supply voltage, process variability and increased noise levels.**

I. Introduction

Miniaturized digital cameras will undoubtedly be an important feature of the next generation of consumer products. With the recent advances in microelectronic fabrication technology, it becomes possible today to fabricate ever thinner imaging systems [1]-[3]. However, a miniaturization of the optics would cause light to pass through a very small opening and to diffract or spread due to the interference of light waves. Diffraction would cause so much spreading that the image would be very blurry and essentially useless. To overcome diffraction effects, a fascinating approach is to design vision sensors mimicking the compound eyes found in tiny creatures such as insects [1]-[5]. A compound eye can comprise anywhere from a single visual unit (ommatidium) to over thirty thousand visual units, depending on the species [6]-[7]. Each facet or visual unit of the compound eye points in a slightly different direction, and images the scene onto its small number of photoreceptors. As a result, the compound-eye captures a mosaic of partially overlapping images. In contrast, the single-lens human eye images the scene onto many receptors, each of which captures only part of the image. Modern approaches to designing ultra-thin vision systems have adopted the concept of compound eyes [2]-[5]. However, research efforts have been so far focused on developing sensor designs that mimic the ultra-thin lens arrangement of a compound-eye. In this paper, the focus is towards VLSI hardware architectures enabling the development of high performance ultra-thin compound-eye vision sensors. We introduce a new approach to compound-eye imaging, based on a time domain data representation with a biologically inspired read-out strategy using Address-Event-Representation (AER) [8][9]. In a conventional digital CMOS camera, images are formed by scanning sequentially a photosensitive array [10]. Each pixel value is read-out after a fixed integration period, during which photo-generated electron-holes pairs are collected. The amount of signal generated by each pixel will depend on the amount of light that falls on the photosensitive cell as well as on the duration of the integration period [10]. This conventional read-out technique is not suitable for a compound-eye imaging system, which comprises a large number of pixels; this, because a systematic sequential scan of all pixels would lead to excessive power dissipation since the scanner is always active. Furthermore, this serial scanning of large arrays would dramatically reduce the frame rate, making real-time processing of the mosaic of captured images difficult to achieve.

To overcome these limitations, we propose a new approach to compound-eye imaging based on AER. In the proposed VLSI architecture, the read-out of information is initiated by the pixel itself, while access to the read-out bus is granted only once to each pixel after which it enters into a stand-by mode. This approach allows to greatly reduce dynamic power consumption and significantly enhances transmission bandwidth of conventional vision sensors. Wide dynamic range and improved signal-to-noise ratio are obtained as each pixel will set its own integration time, which is not dictated by a global timing circuit as is the case in conventional imager architectures. Finally, the proposed VLSI architecture integrates time encoding, to ensure a relative insensitivity to the ongoing aggressive reduction in supply voltage that is expected to continue for the next generation of deep submicron silicon processes. As a result, our proposed time encoding compound-eye sensor is scalable, allowing for further system miniaturization.

In the next section, the silicon implementation of a compound-eye vision sensor is briefly described. Section III presents the proposed pixel-driven strategy for fast and low power operation. Section IV describes the AER-based imager architecture for compound-eye vision sensors together with potential applications. Finally, concluding remarks are given in Section V.

II. Compound-Eye Sensor Design

Traditional camera systems are generally based on a single lens system, which provides a single view of the scene. The compound-eye that a dragonfly bears, on the other hand, can comprise a large number of visual unit systems, each looking in a different direction. Details of the design and analysis of compound-eye vision sensors have been presented by Sanders et al [7]. But, in spite of the efforts dedicated to implementing a compound-eye vision sensor, no practical device with acceptable imaging performance has yet been developed. However, the thin observation module by bound optics (TOMBO), recently proposed by Tanida el al [3]-[5] offers a potential avenue if pixel count is increased and image processing could be carried-out on-chip while still meeting power, speed and silicon area requirements. The TOMBO system consists of three components: a microlens array, a layer of separation walls to prevent crosstalk, and an imaging device which is currently a simple photo-detector array (Fig.1). In the following, we propose a VLSI hardware architecture for the photo-detector array to enable the full potential of TOMBO compound-eye imaging.

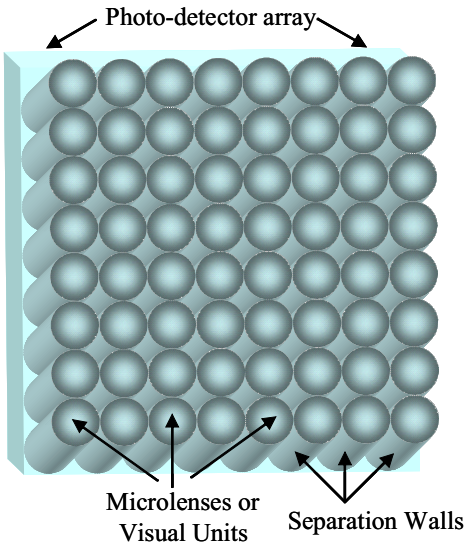

Fig. 1 TOMBO compound-eye architecture [3].

III. Pixel-Driven Operating Technique

To enable fast and low power operation, we propose a pixel-driven read-out strategy based the AER communication protocol, which is modeled after the transmission of neural information in biological systems [8][9][11]. AER is used to transmit "events" through a single communication channel. Events are generally in the form of a spike or a pulse. They are characterized by a location (address) and the time of occurrence. In the case of an image sensor, the communication channel will be an asynchronous digital bus. The address will identify one particular pixel of the array whereas the time of the event will here be defined as the time at which a pixel has reached a given threshold voltage. Each time an event occurs, a pixel generates a spike and communicates asynchronously with a peripheral arbiter, which will take the pixel address and place it on the bus. As a result, the asynchronous bus will carry a flow of pixel addresses. Access to the bus is allocated according to pixel demand. As a consequence, active pixels (i.e., pixels that have generated events) will be favored and granted access to the bus more frequently than less active pixels, which will in turn consume much less communication bandwidth. At the receiver end of the bus, address and time information are combined to retrieve the original data (e.g. pixel brightness value). The AER communication protocol makes efficient use of the available output bandwidth since read-out can be achieved at any time upon request. In terms of power consumption, AER is also more efficient than the conventional fixed time-slot (synchronous) allocation of resources; this because not all pixels are likely to require computation/communication resources at the same time, hence there is no waste of resources. The key element in the AER communication protocol is the event-driven pixel, which will be responsible for requesting access to the output bus, when a pixel has reached a predefined threshold voltage.

Fig. 2 shows the schematic of the proposed asynchronous event-driven pixel. It combines a photo-diode P_d, a reset circuit (*M1-M2*), a low power current-feedback event generator (*M3-M7*) [11] together with a handshaking circuitry (*M8-M14*) used to communicate the event to the outer array circuitry. The asynchronous event-driven pixel operates as follows. After a reset pulse (*Rst* high), the voltage V_N at the sensing node is reset to V_{Rst} and the integration phase begins with the collection of photo-generated electron-holes pairs. During the integration phase, the voltage V_N at the sensing node decreases as a function of the sensed photo-current I_{ph}. Here, an event occurs each time V_N reaches the threshold voltage V_{TH} of the inverter formed by (*M5-M7*), generating a spike or pulse at the output. The integrating time t_{int} required to generate this first event, encodes pixel brightness:

$$t_{\text{int}} = \frac{1}{I_{ph}} \times C_d (V_{Rst} - V_{TH}) \qquad (1)$$

Once an event has been generated, the pixel resets itself asynchronously (ASR active) and remains dormant until the next integration phase (falling edge of *Rst* pulse). With this approach, only a single transition or spike can occur per pixel and during a single integration period. In contrast, the output of an AER spiking pixel is typically a sequence of pulses or spikes, where the inter-spike interval encodes the brightness value of the pixel being read-out [12]. By dramatically reducing switching activity, the proposed approach leads to an average current of 10nA per pixel, which is 3 orders of magnitude lower than that of the conventional spiking pixel [12]. In addition, representing illumination information in the temporal domain provides a number of advantages. Firstly, this allows each pixel to have a large dynamic range because integration time is not dictated by a global scanning clock.

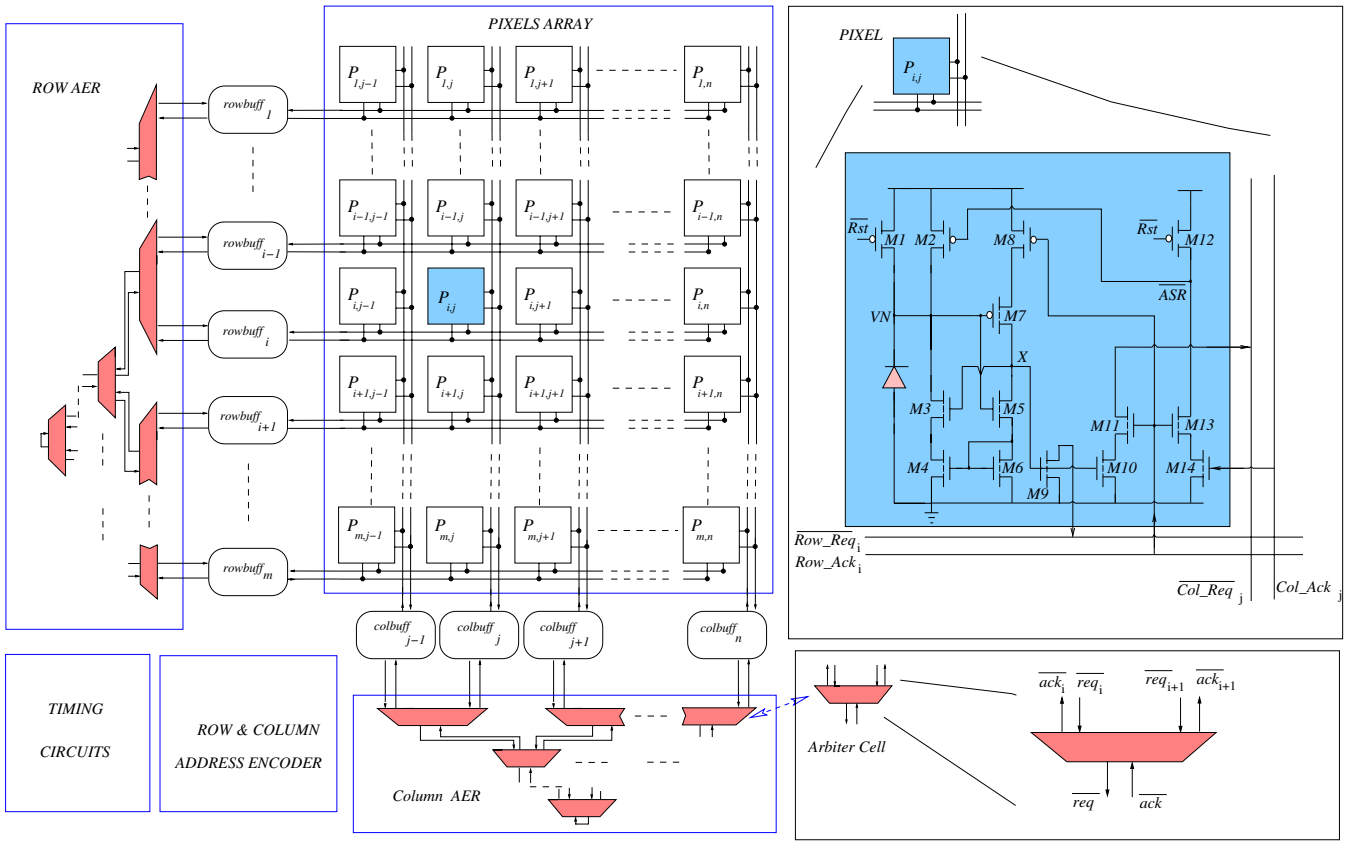

Fig. 2. AER-based imager architecture for compound-eye vision sensors (left); asynchronous event-driven pixel and its control signals (right).

Instead here, each pixel sets its own integration time t_{int}, which corresponds to the time required to bring the sensing node voltage from V_{Rst} down to the fixed voltage V_{TH}. The output charge that is read-out from any given pixel can be expressed as

$$Q(t_{\text{int}}) = \frac{1}{q}\left(I_{ph}t_{\text{int}} + Q_{shot} + Q_{reset} + Q_{read} + Q_{FPN}\right) \qquad (2)$$

when taking into account contributions from independent additive noise sources such as integrated shot noise, reset (kTC) noise, read-out circuitry noise which includes quantization noise, and offset/gain fixed pattern noise (FPN) which results from device mismatches [13]. If we assume that correlated-double sampling is performed, then we can neglect Q_{reset} and Q_{FPN} compared to Q_{shot} and express the signal-to-noise ratio SNR as

$$SNR(I_{ph}) = \frac{\left(I_{ph}.t_{\text{int}}\right)^2}{qI_{ph}t_{\text{int}} + q^2\sigma_{read}{}^2} \qquad (3)$$

where q is the electron charge, and $\sigma_{read}{}^2$ is the average power associated to the shot noise. In the case of our time encoding scheme, $I_{ph}.t_{int}$ is constant and equal to $(V_{Rst}-V_{TH})\times C_d$, where C_d is the capacitance at the sensing node. As a result, in the proposed asynchronous pixel, the SNR is not degraded at low illumination levels since each pixel integrates the same amount of charges regardless of the illumination conditions. An additional benefit of time encoding is that it provides a relative insensitivity to the ongoing aggressive reduction in supply voltage that is expected to continue for the next generation of deep submicron silicon processes. As a result, our proposed time encoding compound-eye sensor is scalable, allowing for further system miniaturization.

IV. AER-Based VLSI Architecture

The proposed AER-based VLSI architecture for compound-eye vision sensors is given in Fig 2. It comprises a pixel array, row and column buffers together with row and column AER circuitry, each organized in a tree structure. The handshaking in-pixel digital circuitry (M8-M14) has two functions: a) to generate a row request each time the pixel reaches a given threshold voltage and b) to reset the pixel when the request is acknowledged. It operates as follows. When a pixel produces a pulse, the row request Row_Req goes active and is presented to the row AER arbitration tree. The row arbiter may receive several bus requests at the same time. After arbitration, a single active row is selected, copied into the column buffer and finally acknowledged (Row_Ack). Next, column arbitration is carried-out within the buffers and not within the array, to eliminate delays associated to the large capacitances of the column lines. The column AER arbiter grants bus access only to those pixels within the buffered row that have generated an event and have requested access to the bus (Col_Req active). Each time a pixel is selected for read-out

its corresponding column is acknowledged by driving Col_Ack active. This causes the pixel to reset itself and remain dormant until the next integration phase (falling edge of Rst). Once the active pixels within the column buffers have been read-out, a new row is copied and a new round of column arbitration can start again. This pipeline processing between row and column arbiters significantly speeds-up the processing of the array. The proposed AER-based VLSI architecture was simulated and implemented using Alcatel 0.35µm CMOS technology, with a pixel size of 15×15µm and a fill-factor of 33%. Fig. 3 shows the control signals associated to the read-out sequence of a given pixel.

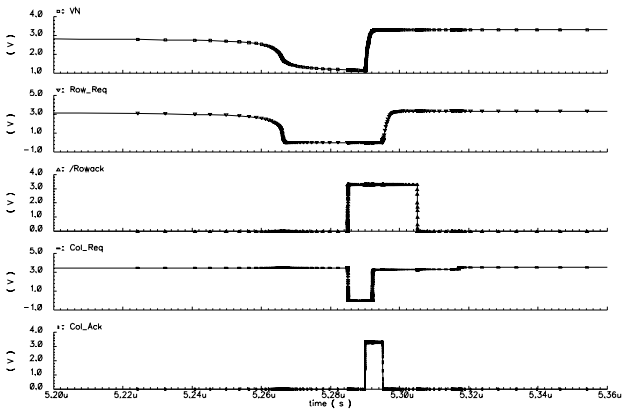

Fig. 3. Simulation results for the proposed asynchronous event-driven pixel: voltage V_N at the sensing node, row request Row_Req signal, Row acknowledgement Row_Ack signal, column request Col_Req signal, column acknowledgement Col_Ack signal, from top down respectively.

V. Conclusion

We have proposed an AER-based VLSI architecture for compound-eye vision sensors. The proposed scheme uses time encoding and provides simultaneously wide dynamic range, a dramatic reduction in dynamic power consumption (3 orders of magnitude), and efficient allocation of the transmission bandwidth to enable real-time on-chip processing. Moreover, the proposed time-encoding VLSI architecture ensures a relative insensitivity to the ongoing aggressive reduction in supply voltage that is expected to continue for the next generation of deep submicron silicon processes. As a result, the proposed compound-eye sensor architecture is scalable, allowing for further system miniaturization

Acknowledgment

The work described in this paper was supported in part by a grant from the Research Grant Council of HK SAR, China (Project No. HKUST6148/03E).

References

[1] R. Volkel, M. Eisner, and K. J. Weible, “Miniaturized imaging systems,” Microelectron. Eng. vol. 67–68, pp. 461–472, 2003.

[2] J. W. Duparré, P. Dannberg, P. Schreiber, A. Bräuer, and A. Tünnermann, “Artificial apposition compound eye fabricated by micro-optics technology,” Applied Optics, vol. 43, no. 22, pp. 4303–4310, 2004.

[3] J. Tanida et al, “Thin Observation Module by Bound Optics (TOMBO): Concept and experimental verification,” Applied Optics, vol. 40, no. 11, pp. 1806-1813, 2001.

[4] Tanida et al, “Color imaging with an integrated compound imaging system.” Optics Express: vol. 11, no. 18, pp. 2109–2117, 2003.

[5] Y. Kitamura et al, “Reconstruction of a high-resolution image on a compound-eye image-capturing system,” Applied Optics, vol. 43, no. 8, pp. 1719-1727, 2004.

[6] M. Land and D-E Nilsson, Animal Eyes. Oxford University Press, 2002.

[7] J. S. Sanders, and C. E. Halford, “Design and analysis of apposition compound eye optical sensors,” Optical Engineering, vol. 34, pp. 222–235, 1995.

[8] M. Sivilotti, “Wiring considerations in analog VLSI systems with applications to field programmable networks,” Ph.D. dissertation, California Institute of Technology, Pasadena, 1991.

[9] M. A. Mahowald, “VLSI analogs of neuronal visual processing: a synthesis of form and function,” Ph.D. dissertation, California Institute of Technology, Pasadena, 1992.

[10] E. R. Fossum, "CMOS image sensors: Electronic camera-on-a-chip", IEEE Trans. on Electron Devices, vol. 44, no.10, pp. 1689–1698, 1997.

[11] E. Culurciello, R. Etienne-Cummings, and K. A. Boahen, “A biomorphic digital image sensor,” IEEE Journal of Solid-State Circuits, vol. 38, pp. 281–294, 2003.

[12] J. Doge, G. Schonfelder, G.T. Streil, and A. Konig, “An HDR CMOS image sensor with spiking pixels, pixel-level ADC, and linear characteristics,” IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, vol. 49, no.2, pp. 155–158, 2002.

[13] S. Kavusi and A. El Gamal, “Quantitative Study of High Dynamic Range Image Sensor Architectures,” Proceedings of the SPIE Electronic Imaging 04 conference, vol. 5301, San Jose, CA, 2004.