5.10 A 1280×720 Back-Illuminated Stacked Temporal Contrast Event-Based Vision Sensor with 4.86µm Pixels, 1.066GEPS Readout, Programmable Event-Rate Controller and Compressive Data-Formatting Pipeline

Thomas Finateu[1], Atsumi Niwa[2], Daniel Matolin[1], Koya Tsuchimoto[2], Andrea Mascheroni[1], Etienne Reynaud [1], Pooria Mostafalu[3], Frederick Brady[3], Ludovic Chotard[1], Florian LeGoff[1], Hirotsugu Takahashi[2], Hayato Wakabayashi[2], Yusuke Oike[2], Christoph Posch[1]

[1]PROPHESEE, Paris, France
[2]Sony Semiconductor Solutions, Atsugi, Japan
[3]Sony Electronics, Rochester, NY

Event-based (EB) vision sensors pixel-individually detect temporal contrast exceeding a preset relative threshold [1,2] to follow the temporal evolution of relative light changes (contrast detection, CD) and to define sampling points for frame-free pixel-level measurement of absolute intensity (exposure measurement, EM) [3,4]. EB sensors gain popularity in high-speed low-power machine vision applications thanks to temporal precision of recorded data, inherent suppression of temporal redundancy resulting in reduced post-processing cost, and wide intra-scene dynamic range operation.

Information about temporal contrast (CD) is encoded in the form of "events": data packets containing the originating pixel's X,Y coordinate, time stamp, and contrast polarity. To maximally benefit from the ability of the individual pixel to sample visual information at high temporal precision, early time-stamping and high readout throughput are crucial to preserve event timings. Different readout schemes that trade off temporal precision in favor of simpler architectures and reduced bit rates at the output (frames-of-events) have been proposed [5,6]. This approach, however, is not well matched to the fundamental principle of asynchronous EB pixel arrays and limits the sensor's usability in high-speed applications like, e.g., fast time-coded structured light depth sensing, high-speed tracking and optical flow with kHz update rates, or visible-light communication (VLC). As shown below, other methods to reduce the bit/event number can be devised that do not impact the temporal precision of the event data.

EB vision sensors, due to complex in-pixel circuitry, have long suffered from large pixel sizes and low resolution; therefore, they particularly benefit from direct wafer bonding technology and the resulting possibility of stacking efficient phototransduction of BI CIS on top of high-density analog signal processing on CMOS to achieve competitive resolution and size. This paper presents a 3D stacked EB vision sensor that uses pixel-level Cu-Cu bonding interconnects to achieve 4.86µm pixel-pitch, yielding HD resolution below ½" optical format. In-pixel circuits communicate CD events via a stall-safe low-latency interface to an asynchronous row-selection tree. Events from active rows are immediately time-stamped and arranged for bit-efficient vector readout. A digital event signal processing (ESP) pipeline features a programmable event rate controller (ERC) and an event data formatter (EDF) for dynamic bit compression. A parallel output interface (EOI) with data packing feeds the event data directly to a processor or to an external MIPI or USB transceiver.

Figure 5.10.1 shows a pixel block diagram illustrating the CIS front-end with Cu-Cu interconnect to the CMOS asynchronous delta-modulation (ADM) based CD circuit and asynchronous readout interface and state-logic (ISL) block. Input latches for CON/COFF reduce power in case of slow comparator switching and prevent ringing around the switching point. Gated latches (K) robustify event registration by preventing late reqX generation and resulting event loss; events arriving after row-readout has started are latched for subsequent readout as soon as the current cycle is finished (ackY removed). Only pixels with events stored in the input latches, locally generate $CTRL_{ADM}$ to reinitialize their CD circuitry upon receiving ackY, thereby removing the necessity for column-wise ackX signals and simplifying readout control logic, column readout circuitry and pixel array signal routing.

Figure 5.10.2 shows a block diagram of the chip, illustrating pixel array readout and event data flow via the ESP pipeline to the output data interface. The active-row selection operation is pipelined; new row selection happens in parallel to processing previous row data. The asynchronous-to-synchronous interface is supervised by a time-out column at the far end of the pixel rows, ensuring correct timing for synchronizing the VecX data into clocked latches at the interface to the ESP pipeline and preventing event loss due to run time variations in the asynchronous data paths. Timestamps are attached to the events at this interface. This readout architecture preserves the event temporal precision out of the pixel array. Typical pixel activation to time-stamp delays of 60-to-70ns and about 120ns row-to-row selection timing are achieved.

Figure 5.10.3 shows a block diagram of ESP pipeline and output interface. ESP functions include address filtering, throughput regulation and data formatting. The digital readout (RO) block synchronizes and timestamps pixel event data row-wise and splits them into 32-event vectors. A LUT-based address filter removes selected events, e.g., from defective pixels. The ERC block allows to cap the output event rate to a predefined limit rate by dynamically applying data drop on peaks that exceed this value. The limit rate is programmable over a wide range between 5k events per second (EPS) and 1GEPS. ERC continuously monitors input and output rates of a FIFO buffer and, in closed-loop configuration, regulates the instantaneous FIFO output rate by removing events, following various drop strategies that combine spatial and temporal criteria. An ROI-based dropping scheme subdivides the pixel array into 40-by-23 32×32 pixel blocks that can be programmed to each exercising a different one of 64 selectable weighs on drop rate, allowing for application-specific optimization of ERC operation (e.g., preferably drop events from the sky in an automotive scene). The EDF block converts the event stream to vectorized data formats (EVT). Advanced EVT formats combine differential and vectorized encoding to dynamically optimizing the number of bits per event by exploiting spatial and temporal relations between events. In situations of highest throughput, 1.6b on average encodes the full temporal and spatial information of an event. The EOI manages off-chip data transmission. Endianness support for natively managing several event granularities is provided. In addition to 16b parallel 100MHz synchronous mode, the interface can be configured for packet mode to better adapt to USB/MIPI transceivers. Clock gating reduces EOI power consumption. ESP integrates automatic test pattern generators (ATPG) in different places along the pipeline to ease chip verification.

Figure 5.10.4 shows sensor operation in example applications. A ~1lux traffic scene recording demonstrates low-light contrast sensitivity in line with lab test results. The high temporal precision due to low-latency pixel and high-speed readout operation allows the sensor to decode temporally encoded structured light patterns in a 3D depth sensing application.

Figure 5.10.5 shows test results on CD contrast sensitivity, dynamic range, BG noise and pixel uniformity at a single sensor setting. Nominal contrast threshold (NCT) is ~16% log contrast. Lower NCT down to 11% can be achieved at different settings (insets). The CDP curve defines low- and high-light cutoff points (LLCO, HLCO) that limit the DR of CD. LLCO was measured at 0.04lux for c=40% contrast, HLCO has not been seen up to 100klux; also, no increase in BG rates at high light levels due to parasitic photocurrents is observed. Resulting DR is >124dB. BG rates peak at ~8.3 events/pixel/s at 0.6lx and go down to below 10^{-1} for illuminances above few tens of lux. FPN-like CTNU is at ~3% contrast.

A 1280×720 ½" temporal contrast EB vision sensor is designed and fabricated on a Cu-Cu bonded wafer stack of 90nm BI CIS on 40nm CMOS with 4.86µm pixels achieving >77% fill factor. The chip consumes 32mW (static) to 84mW at high activity (300MEPS). Readout with 1µs time stamping handles internal peaks of up to 2.5GEPS out of the pixel array and sustained 1.066GEPS at chip output. Bits-per-event are dynamically compressed down to 1.6b while maintaining full spatial and temporal information. Wide DR (>124dB) is achieved due to good low-light CIS performance (40mlx LLCO) and absence of leakage activity from parasitic photocurrents at high light. Step-response latencies around 200µs are typical at illuminance levels >10lux.

References:

[1] P. Lichtsteiner et al., "A 128×128 120db 30mW Asynchronous Vision Sensor that Responds to Relative Intensity Change," *ISSCC*, pp. 2060-2069, Feb. 2006.
[2] B. Son et al., "4.1 A 640×480 Dynamic Vision Sensor with a 9µm Pixel and 300Meps Address-Event Representation," *ISSCC*, pp. 66-67, Feb. 2017.
[3] C. Posch et al., "A QVGA 143dB Dynamic Range Asynchronous Address-Event PWM Dynamic Image Sensor with Lossless Pixel-Level Video Compression," *ISSCC*, pp. 400-401, Feb. 2010.
[4] J. Huang et al., " Live Demonstration: A 768 × 640 Pixels 200Meps Dynamic Vision Sensor," *IEEE ISCAS*, 2017.
[5] C. Li et al., T. Delbruck, "A 132 by 104 10µm-Pixel 250µW 1kefps Dynamic Vision Sensor with Pixel-Parallel Noise and Spatial Redundancy Suppression," *Int. Image Sensor Workshop*, 2019.
[6] H. E. Ryu, "Industrial DVS Design; Key Features and Applications," *Conf. on Computer Vision and Pattern Recognition*, 2019, Accessed Nov. 3, 2019, <http://rpg.ifi.uzh.ch/docs/CVPR19workshop/CVPRW19_Eric_Ryu_Samsung.pdf>.

 978-1-7281-3205-1/20/$31.00

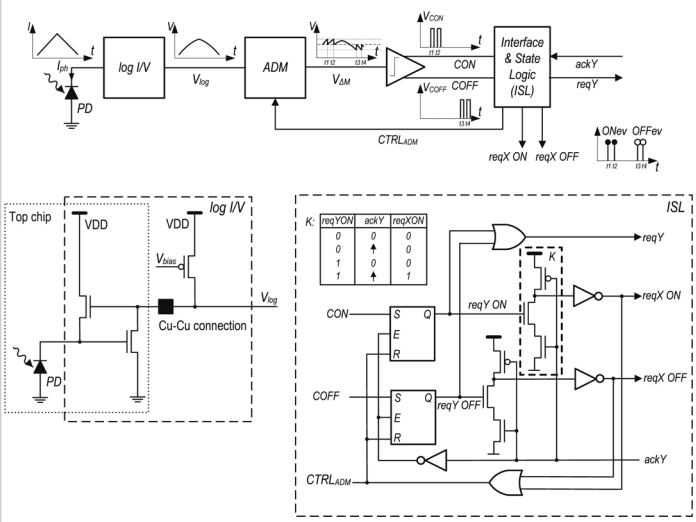


Figure 5.10.1: In-pixel circuitry: Logarithmic photoreceptor front-end, CD function block and readout interface.

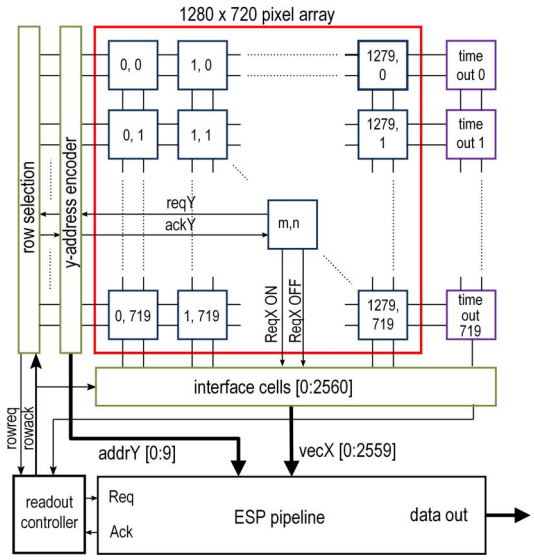


Figure 5.10.2: Chip block diagram.

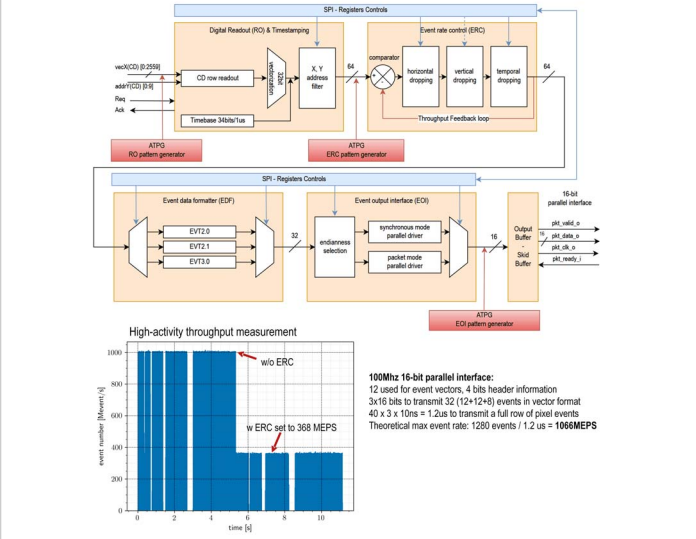


Figure 5.10.3: Event signal processing (ESP) digital pipeline with event output interface (EOI). Measurement of output event-rate.

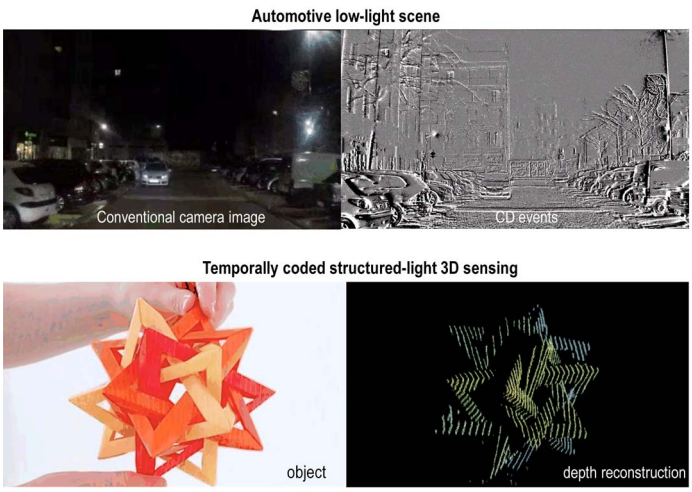
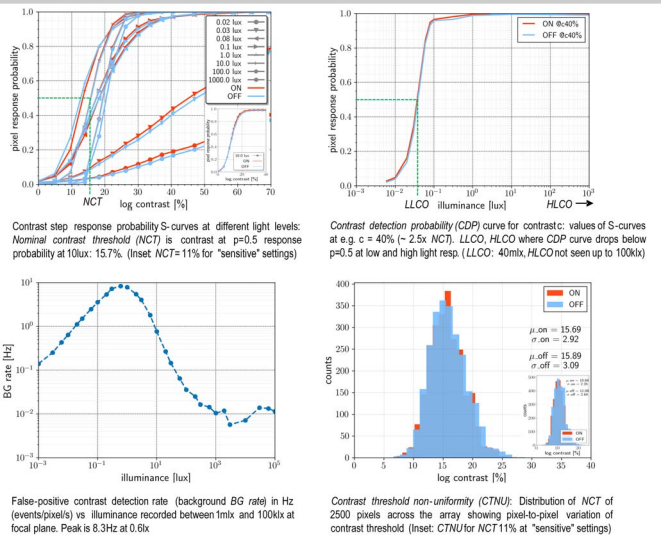


Figure 5.10.4: Sensor output event data illustrating low-light contrast sensitivity in an automotive scenario and fast 3D depth map reconstruction.



Contrast step response probability S-curves at different light levels: *Nominal contrast threshold (NCT)* is contrast at p=0.5 response probability at 10lux: 15.7%. (Inset: *NCT*= 11% for "sensitive" settings)

Contrast detection probability (CDP) curve for contrast c: values of S-curves at e.g. c = 40% (~ 2.5x *NCT*). *LLCO*, *HLCO* where *CDP* curve drops below p=0.5 at low and high light resp. (*LLCO*: 40mlx, *HLCO* not seen up to 100klx)

False-positive contrast detection rate (background *BG rate*) in Hz (events/pixel/s) vs illuminance recorded between 1mlx and 100klx at focal plane. Peak is 8.3Hz at 0.6lx

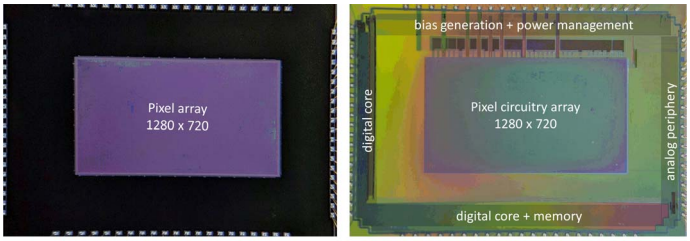
Contrast threshold non-uniformity (CTNU): Distribution of *NCT* of 2500 pixels across the array showing pixel-to-pixel variation of contrast threshold (Inset: *CTNU* for *NCT* 11% at "sensitive" settings)

Figure 5.10.5: Characterization results: Contrast sensitivity, low-light cutoff and DR, noise event rates, pixel uniformity.

		This work	[2] Samsung Gen2	[4] Celepixel Celex-IV	[5] Inivation DVS132
Technology		90nm BI CIS + 40nm CMOS	90nm CIS BSI	180nm CIS	65nm CMOS
Supply voltage		2.5, 1.1	2.8, 1.2	3.3	1.2
Resolution		1280 x 720	640 x 480	768 x 640	132 x 104
Pixel size (um^2)		4.86 x 4.86	9 x 9	18 x 18	10 x 10
Fill Factor		>77%	20%	9%	20%
Power (mW)	100kEPS[a]	32	27	-	0.25
	300MEPS	84	50	-	4.9[e]
Power/Pixel (nW)[b]		35	88	-	18
Energy/event (pJ)[b]		173	77	-	26
Max event rate (MEPS)		1066	300	200	180
Contrast Sensitivity NCT		11%	9% (19%[c])	-	-
Dynamic Range		>124dB	>80dB (90dB[d])	120dB	-

[a] EPS = events per second
[b] Definitions as detailed in Ref [5]
[c] Equivalent NCT definition (Ref [6])
[d] from Ref [6]
[e] at 180MEPS max rate

Figure 5.10.6: Comparison table.



micrograph of top part

micrograph of bottom part

Figure 5.10.7: Chip micrographs.

978-1-7281-3205-1/20/$31.00