A 90dB single-shot HDR, 0.5MP global-shutter image sensor with NIR QE enhancement, 20mW power consumption and smart event detection modes

A. Xhakoni, A. Fekri, P. Francis, S. Melsen, S. Janardhan, B. Hashani, A. Popa, K. Ruythooren, P. De Wit, R. Lafaille, K. Kawamura, A. Jaderi, K. Kimura, K. Ben Ali, R. Brolly, A. Andrade

ams OSRAM, Borsbeeksebrug 36, Antwerp, Belgium

adi.xhakoni@ams-osram.com, +32 3 2868902

1   Introduction

Consumer applications such as AR/VR require global shutter image sensors due to image artifact reduction and due to their capability to sync with a pulsed illuminator. Other key requirements are low system power, (achieved through high QE at NIR and sensor low power operation) and miniaturized silicon size. Furthermore, HDR operation is especially required for world facing cameras.

2   Sensor architecture

To satisfy the needs of consumer applications, we designed a 0.5MP image sensor with 2.79um voltage domain GS pixels with a packaged sensor footprint of only 2.3mm x 2.8mm, with architecture depicted in Fig 1. To meet the challenging requirements of footprint and power, we made extensive use of digital-friendly readout circuits starting with the ramp ADC, which contains minimal analog circuitry. Given the scaled technology node (40nm) of the logic layer, the power of the readout scaled dramatically compared to our previous designs in older technology node. Certain building blocks which were the most power hungry in our previous designs (e.g. ADC counter) experienced an order of magnitude of power consumption improvement. In addition, advanced power down techniques were used for optimizing current consumption by turning on specific blocks only when needed. The above techniques allowed for an allow ultra-low-power operation of the sensor (e.g. 18.5mW at 10bit, 30fps, 0.5MP).

To avoid potential column FPN issues, we decided to keep the ADC pitch same as the pixel pitch. This posed a challenge in the row driver design to keep the sensor footprint within specifications. Improvements in the row driver architecture and the use of 7 metal layers (made possible by the stacked technology) allowed achieving an exceptionally narrow row driver width of only 40um. The row driver was split in 2 parts (left and right of array) to half its width. The improvements in architecture, layout, and the usage of scaled technology node allowed for a sensor footprint which is mostly dominated by the pixel array rather than readout silicon.

3   HDR operation

The HDR technique validated in this sensor is shown in Fig 2. It uses charge-overflow-on-FD technique, with critical improvements over SOA. Overflow operation is done on FD, at a fraction (T1) of low-light exposure time (T0), with DR extended by T0/T1 ratio, with T0 exploiting the entire full well of PD and T1 exploiting the entire full well of FD. The technique makes use of the timing diagram of Fig 2 and works as follows. A) Long exposure T0 starts when TX toggles low. B) TX toggles high to a mid-level. This facilitates charge overflow from PD to FD (hence to reset) in case of medium or high light. C) Short exposure T1 starts when reset switch is turned off. At this point, any overflowing charge is collected on FD node. Since T1 is much shorter than T0 (e.g. 15x shorter), FD leakage affecting T1 and dark current affecting T0 are negligible. D) T1 exposure ends when TX toggles back to the

lowest level. T1 exposure is stored via S1 and S2 in C2. E) Finally, T0 exposure ends when TX toggles high. T0 signal is then stored on C1 via the S1 switch.

Despite storing 2 signal levels in the 2 in-pixel capacitors, CDS is still provided for the low light signal and DDS is provided for high light signal, as follows. Fig 3 shows 3 cases of light conditions (low, medium and high light). In case of low light condition, no overflow occurs. Therefore, T1 exposure stored on C2 is the correlated reset level for long exposure T0, allowing CDS hence low noise (~5e-) and low FPN (~3e-). In case of medium light level, overflow starts happening just after T1 starts (Fig 3, middle plot). To reconstruct the HDR image both CDS (C2-C1 signals) and DDS (C2 signal – dark reference) are needed during mid light. The dark reference to allow for DDS is achieved during row readout, by accessing the reset level of the pixel after reading the C2 and C1 signals. In case of high light (Fig 3, right plot) T0 exposure signal starts overflowing before T1 starts. This means that in case of high light, only T1 signal is used for the HDR reconstruction. Therefore DDS is used, CDS is discarded.

With the technique above, only 2 in-pixel capacitors are necessary for the 4 signals needed for HDR (low light + reset, high light + reset), as compared to the 4 caps needed by e.g. [2]. This method employing only 2 capacitors for 4 signals can be used with many other HDR techniques, including conventional LOFIC, dual exposure, etc. Compared to alternative HDR techniques, the proposed technique doesn't need FPN calibration, and pipeline readout is possible, improving low light performance. Fig 4 shows ~90dB DR with ~27dB at dip point.

4   Event detection mode

The sensor includes an event detection mode for low power operation. As shown in Fig 5, the sensor splits the pixel array into a programmable number of tiles composed of binned and/or subsampled pixels and runs at 1fps with low power (<3mW). A reconfigurable on-chip algorithm processes the variation of signal intensity in each tile and decides if an interesting event has happened. If yes, the sensor switches to a user-defined operation mode (e.g. 120fps, 10bit). This mode is quite useful for sparing system power in a multitude of applications including computing, doorbells, AR/VR, etc.

5   Conclusion

With a 2.3mm x 2.8mm packaged size (Fig. 6), the presented sensor is one of the smallest GS sensors ever reported. Despite its size, it packs 0.5MP, Mipi interface, ultra-low power operation, smart event detection modes and proves single shot HDR concept with pipeline operation. The specialty of the proposed HDR technique is that it reuses the same pixel storage capacitor for both reset level of low light condition and signal level of high light condition. Combined with high QE (~92%@vis, ~38%@940nm) it well compares against great works from the past [1],[2] and it is an ideal candidate for consumer AR/VR devices.
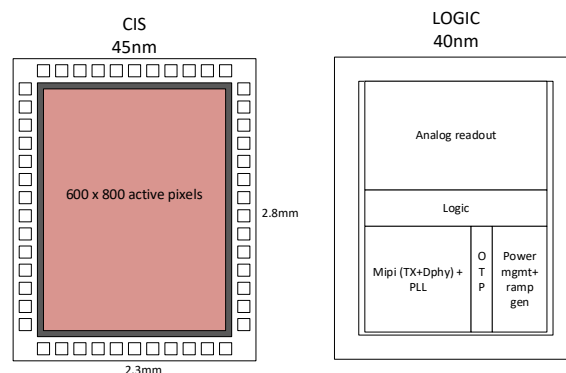

Fig. 1. CIS silicon (left) stacked on top of logic silicon (right) via hybrid bonding interconnect
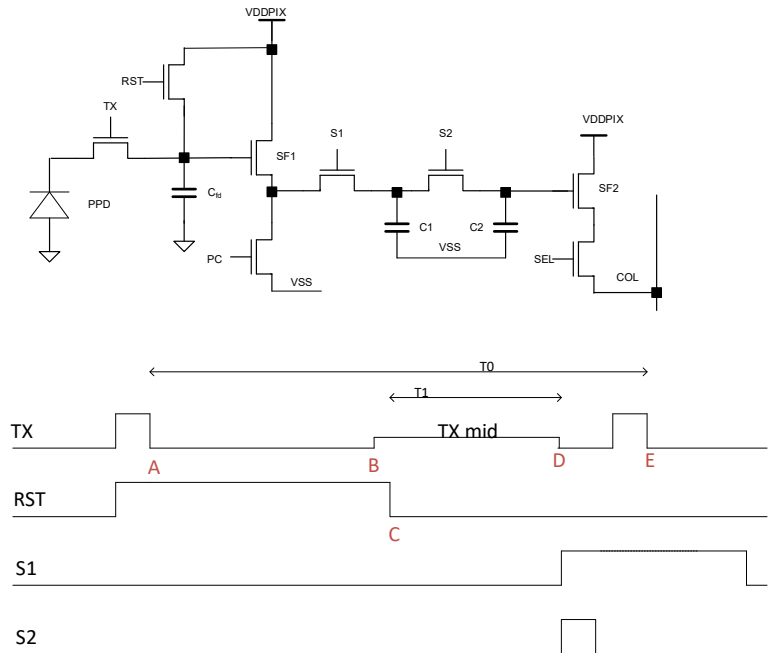
Fig. 2. Voltage domain GS pixel and HDR timing. C2 stores overflow signal (at high light) or reset signal (at low light). Reset level of overflow is read during row readout time. Only 2 capacitors are then needed to store reset low light, low light signal, overflow signal and reset of overflow. Overflow happens during short T1 at FD, avoiding need for large overflow cap and reducing DSNU and dark current.
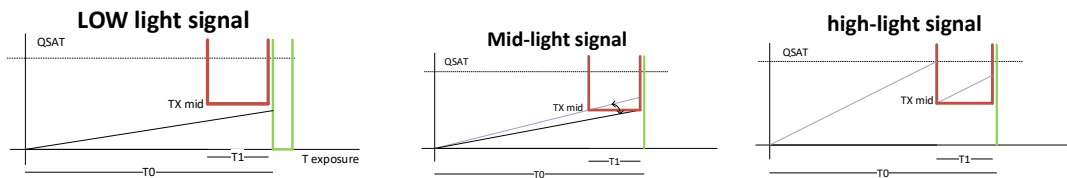

Fig. 3. During low light conditions (left plot) no signal overflows during T1. During mid light conditions (mid plot) overflow happens just after T1 starts. During high light conditions overflow happens before T1 starts.
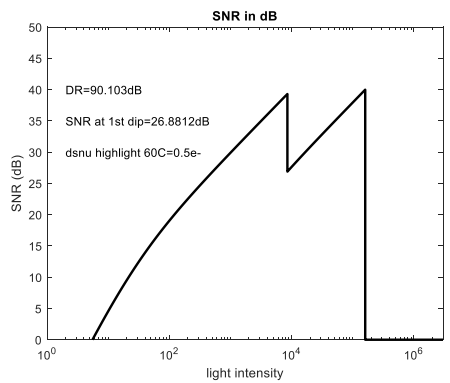

Fig. 4. ~90 dB DR is achieved with T0/T1=16x. Larger DR is possible by trading off with SNR dip. Part of data from the plot is simulated as full char report is not yet available at the moment of the writing of the text.

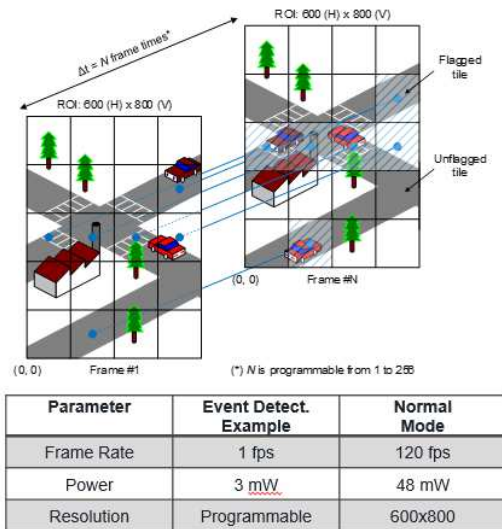| Parameter	Event Detect. Example	Normal Mode
Frame Rate	1 fps	120 fps
Power	3 mW	48 mW
Resolution	Programmable	600x800

Fig. 5. User programmable event detection mode operating at 1fps. When event is detected, the sensor switches to another user defined mode (e.g. 120fps, 10bit, 0.5MP)
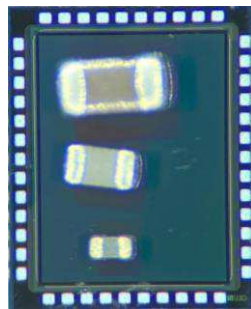

Fig. 6. 2.3mm x 2.8mm CSP version

Parameter	[1]	[2]	This work
technology	stacked 45nm-65nm	stacked 45nm-65nm	stacked 45nm + 40nm
Pixel pitch (um)	2.2	4	2.79
Resolution	640 x 480	1024 x 832	600 x 800
Shutter	Global VD	Global VD	Global VD
DR (dB)	61	90	90
Noise (e-)	2.3 (HCG mode)	4	5 (LCG mode)
Power (mW)	139	-	20mW @10b, 30fps, 60mW @120fps
Footprint (mm x mm)	2.6 x 2.95	8 x 8	2.3 x 2.8
QE 940nm (%)	38	40	36

Fig. 7. Comparison table

6   References

[1] Park et al. "A 2.2µm stacked back side illuminated voltage domain global shutter CMOS image sensor", IEDM19.

[2] Miyauchi et al. "4.0µm Stacked Voltage Mode Global Shutter Pixels with A BSI LOFIC and A PDAF Capability", IISW21

An InGaAs Multi-Functional Fast SWIR Imager with Event-based and Laser Multi-spot Sensing

C.G. Jakobson*, R. Fraenkel, N. Ben Ari, R. Dobromislin, N. Shiloah, T. Argov, W. Freiman, G. Zohar, L. Langof, O. Ofer, R. Elishkov,
E. Shunem, M. Labilov, M. Alcheck, M. Nitzani, Y. Karni, T. Markovitz

SCD Semiconductor Devices

P.O. Box 2250, Haifa, 31021, Israel - *Corresponding author: claudio@scd.co.il

Abstract - **A novel VGA, 10 micron pixel, multi-functional InGaAs imager is presented. This sensor combines standard imaging for daylight use and high-gain mode for low light level use with event-based imaging and laser multi-spot sensing. The fast imaging modes go up to 1.6 kHz for full-frame and are sensitive in the Visible to SWIR band (600–1700nm). The event-based mode and the laser multi-spot mode are able to operate simultaneously to the image with their output provided in a separate video channel. The event mode is responsive to image variations only, enabling to capture landscape events at a rate that is more than x10 faster than the standard imaging. The laser multi-spot mode responds to multiple laser illumination at a fast rate that enables to recognize the presence of laser spot as well as its pulse repetition frequency. This work elaborates on the multi-functional architecture, demonstrate its key features, and shows first measurements of the ROIC and sensor.**

Keywords - SWIR, ROIC, InGaAs, event-based vision, laser detection, laser multi-spot, multi-functional

I. INTRODUCTION

High Frame per Second (FPS) imaging in SWIR wavelengths is a highly desirable feature. For instance, autonomous navigation and collision awareness are emerging applications that require fast video imaging in addition to the known SWIR advantages at fog penetration and night vision [1-3]. In some scenarios fast integration and fast analog to digital column-parallel conversion are enough to meet the high-frame rate needs. However, these circuits present a bottleneck that practically limits video frame rate to hundreds of Hz [4]. On the other hand, and compared to their visible wavelength counterparts, SWIR pixels have a relatively large pitch that can be used to increase the pixel readout complexity [5]. For fast imaging applications, the pixel can combine multiple functions aimed at sensing fast events that may be present in the target scenario.

Event-based imaging is an emerging imaging paradigm focused only on variations in the pixel target that breaks the imaging speed bottleneck [6]. Typically event vision is intended for machine use, for example, event cameras can be used for movement detection, object recognition and tracking. Moreover, in defense and security applications its fast response allows to detect the presence of fast varying hostile threats. Navigation systems such as autonomous vehicles and drones can use event imaging for movement assistance and collision avoidance and at the same time assess motion-based depth information. Other processing options may enable its use to eliminate image artifacts caused by vibrations and turbulence.

On top of this, many lasers operate in the visible and SWIR wavelengths. Hence, laser pulses are a specific case of fast events particularly interesting in SWIR. These pulses are very fast, lasting a few hundreds of nanoseconds, and its reflection generates a small charge packet at the pixel level. Finding laser pulses in the image is a difficult task as this laser pulse reflections are slightly above the readout noise in standard imaging. Dedicated in-pixel circuitry can be used to enhance the response to fast pulses aiding to the laser detection process [5]. Moreover, there is an increased interest in decoding the laser pulse-repetition rates that are usually modulated up to 10 kHz. Laser decoding enables to identify between different laser sources and provides line-of-sight communication.

This works presents a novel imager, which is the first to introduce event-based imaging and laser-multispot detection in the SWIR wavelengths. This new imager provides a simultaneous conventional high FPS image synchronized to the event-based or to the laser multi-spot output. This imager follows a line of multi-functional InGaAs SWIR products. It is a fast multi-mode VGA imager implemented with a $10\mu m$ pitch sensor, while event-based and laser detection are shared between four adjacent pixels outputting a QVGA fast image. The imager is sensitive in the visible to SWIR bands, and provides full format integration of standard video at 1600 Hz with 11 bit output. The multi-spot Asynchronous Laser Pulse Detections (ALPD) works at a fast 50 kHz detection rate that enables the decoding of the laser Pulse Repetition Frequency (PRF) and distinguish between different lasers in the image scenario. Pixel techniques to combine imaging with event and ALPD are patented [7-8].

II. MULTI-FUNCTIONAL MODE OPERATION

Table 1 describes the modes of operation supported and their multi-functional combinations. All the imaging modes support global integration, the Standard Image Integration (SIM) mode integrates the image using a Direct Injection (DI) readout [4,9]; and may operate either as Integration Then Read (ITR) or as Integration While Read (IWR). In addition, a High Gain (HG) imaging mode uses a Charge Trans-Impedance Amplifier (CTIA) readout and Correlated Double Sampling (CDS) [5,9]. The HG imaging is ITR and is binned between four pixels. Moreover, the HG can be synchronized to an active illumination to provide active imaging over a time window of a few µsec.

	SIM	High Gain		No Imaging
	ITR & IWR	*ITR*	*Active*	
Event-based detection	YES			YESa
Laser Multi-spot detection	YES	NO		YES
No detection	YES	YES	YES	

a. Can work as image-disabled or event-only sub-modes

Table 1.   ROIC multi-functional modes

The event and laser multi-spot modes, referred also together as the "detection" modes, can operate simultaneously with the SIM imaging. The combination of the detection modes with the HG mode is disabled due to pixel hardware limitations.

The transition between the modes at Table 1 is very fast and can be done within one imaging frame by a simple control. As a particular case, the imaging and the detection modes can

operate separately. This provides, correspondingly, an imaging-only mode and a detection-only mode.

In most cases the functionality of each mode operating separately improves performance. This is the motivation to support detection-only and imaging–only modes. Particularly, at imaging-only mode, the SIM operation is significantly faster than its corresponding simultaneous imaging-detection operation. On top of this, the event-only mode not only improves event performance, but also allows to reconfigure the block and change its functionality. More details are provided in next section.

III. ROIC DESIGN

Fig. 1 shows the simplified schematic circuit for the imaging modes. At DI mode, the DI-bias transistor together with the V_{DETCOM} voltage determine the bias to the diode, while the diode current is injected to the integration capacitor and subsequently read and drive to the column through a Source Follower (SF). IWR mode is enabled by using multiple integration capacitors, which enable to read from one capacitor while the other is integrating the signal. The imaging output signal is driven from the SF to column-parallel ADCs that provide conversion in the 11-13 bit range. The maximum frame rate is achieved with 11 bit operation and can reach 1.6 kHz.

At HG mode, the CTIA is enabled and the charge is integrated at the feedback integration capacitor. Four diodes are binned to a single feedback capacitor that provides the high gain and proper impedance-matching for CDS operation, thus the pixel effective area for HG operation is 20 µm. The HG mode is ITR only, and provides the best dynamic range and low noise for low illumination scenarios. The CTIA shares the same SF than the DI, but in this case converting to 11bit is enough to meet the dynamic range requirements. The HG mode achieves 2 kHz.
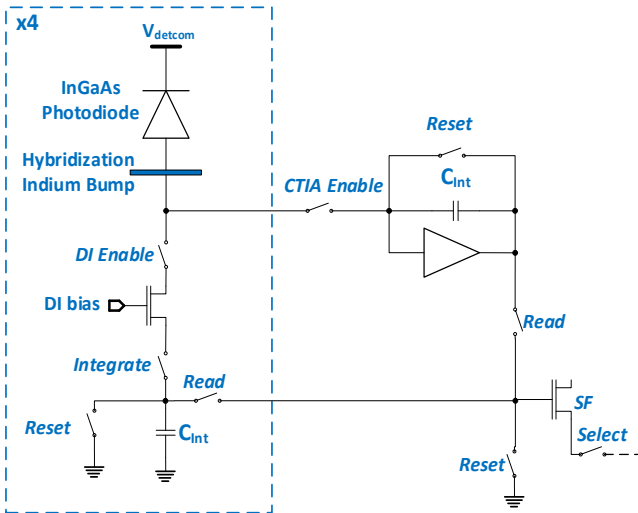

Fig. 1. Simplified imaging schematics

The main parameters for both imaging modes are summarized at Table 2.

Mode	Parameter		Units	Value
SIM	Format		N.A.	VGA
	Full well capacity	IWR	ke-	500
		ITR		900
	Readout Noise	IWR	e-	150
		ITR		260
	Max. Frame Rate (13/11 bit)	No Detection		800/1600
		With Detection	FPS	200/250
HG	Format		N.A.	QVGA
	Full well capacity	ITR	ke-	50
	Readout Noise		e-	50
	Max. Frame Rate (11 bit)		FPS	2000

Table 2. ROIC main imaging parameters

The detection modes are combined within the schematics shown at Fig. 1. The signal reconfiguration that enables the detection modes reuse the CTIA amplifier. Due to this constraint the detection modes can operate only simultaneously to the SIM mode.
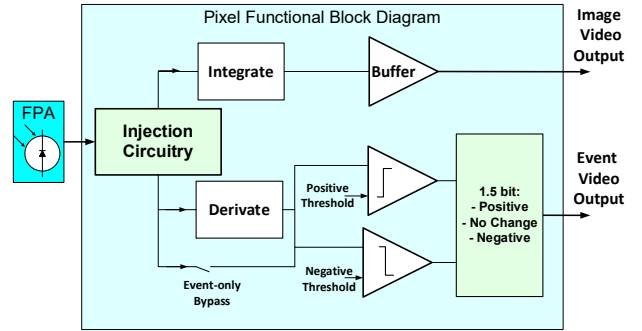
(a)
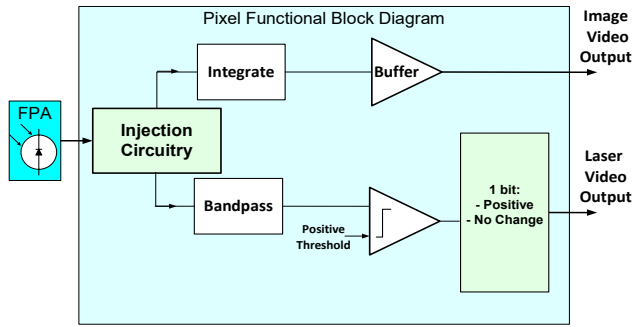
(b)

Fig. 2. Pixel functional block diagram: (a) event detection (b) laser multi-spot detection

Fig. 2 describes the detection modes by functional block diagrams. The upper figure (a) shows the block diagram for the event mode. The signal is injected into the DI circuit and simultaneously split in the frequency domain by two filters [8]. A low-pass filter enables the signal integration for the imaging, while a high-pass filter provides signal derivation for the event channel. Following, the variations in the derivative channels are compared to a positive and negative channel and an event is recognized as positive or negative if

it generates a signal above controlled positive and negative thresholds. From the information point of view, this channel provides 3 levels (1.5 bits) that indicate the presence of a negative or positive event, or no change, as usually done in dynamic vision systems [6].

The image can be disabled to provide an event-only mode. At first, this can be done to improve event sensitivity and eliminate possible parasitic coupling between the event sensing and the image read. However, disabling the image integration performed by the low pass filter also enables to bypass the high-pass filter as this last one is no further required for signal splitting. To this end a bypass filter switch has been implemented as shown in Fig. 2. (a). While usually the event mode is targeted high frequency fast variations, this variant of the event-only mode enables to extend it to respond to low frequency variations.

The laser multi-spot detection is done by reconfiguring the filters as shown in Fig. 2 (b). The main differences are in replacing the derivative channel by a band-pass filter [7], and in disabling the negative comparator as the laser signal is only positive. The end of this reconfiguration is to provide the best filter noise matching in order to detect low energy laser signals.

Table 2 summarizes the main parameters for the detection modes

Mode	Parameter	Units	Value
Event	*Format*		QVGA
	Event rate	kHz	0.5-25
	Sensitivity	e⁻	1500
Laser multi-spot	*ALPD/SLPD Sensitivity*	e⁻	2000
	ALPD Rate	kHz	0.5-50
	ALPD duty cycle	%	100

Table II. ROIC main detection parameters

Fig. 3 shows the ROIC floor plan, the VGA pixel channel is connected to column-parallel detection channels and column-parallel detection channels. The matrix read is done by rolling and after conversion it is driven by four video channel outputs.
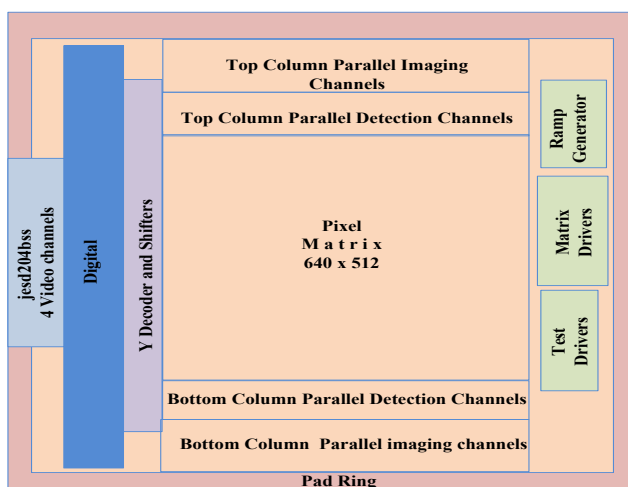

Fig. 3.   ROIC Floor plan

Fig. 4 shows the packaged prototype used for first measurements and characterizations described in the next section.
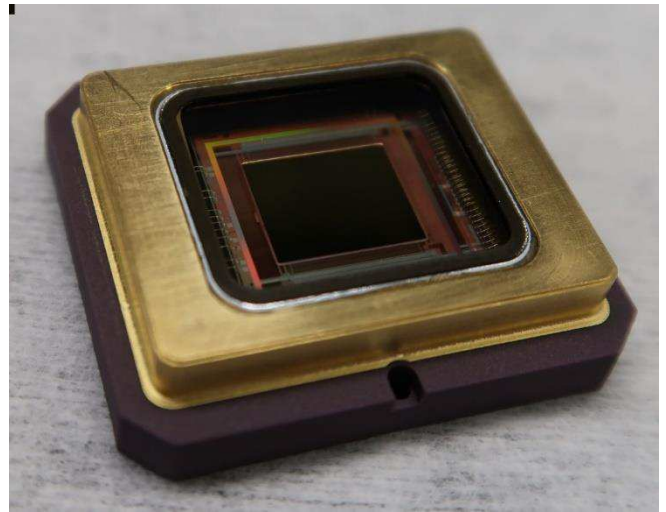

Fig. 4.   Packaged prototype photo

IV.   MEASUREMENTS

Fig. 5. shows an image captured by the new imager at SIM mode. Fig. 6. Shows the SIM noise floor histogram.
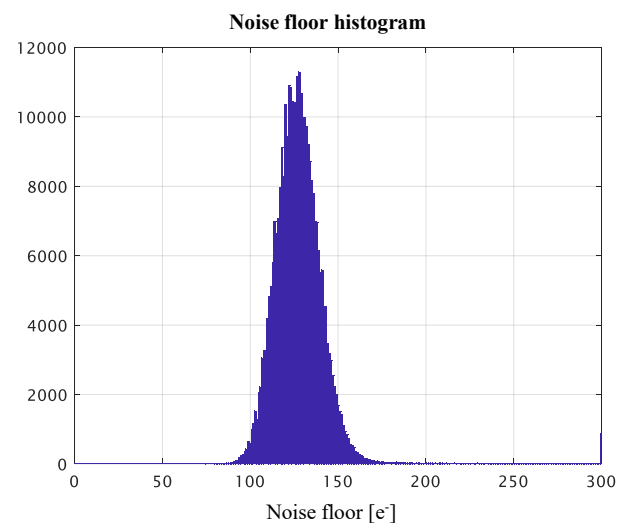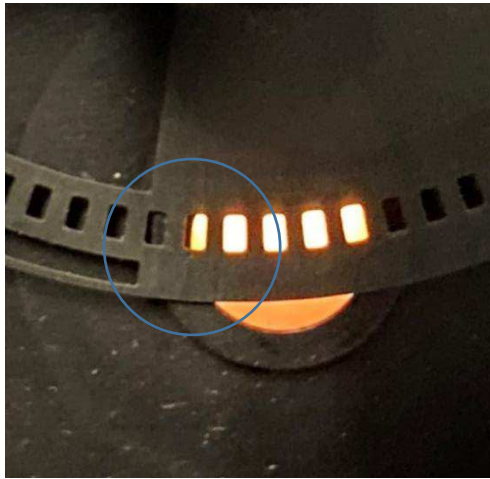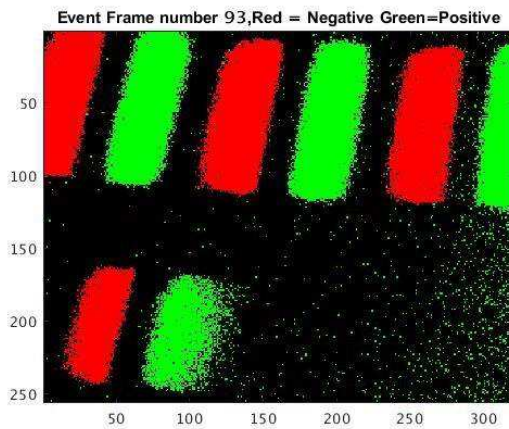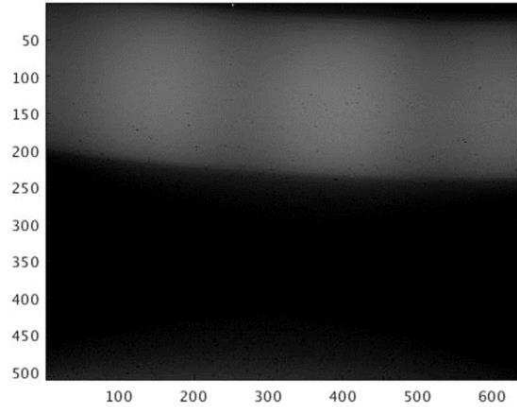

Fig. 5.   SWIR image in SIM mode

Fig. 6.   Dark noise floor histogram for the SIM image

recognized. Fig. 7 (c) shows the simultaneous SIM image captured by the imager, which, as expected, is blurred due to the fast rotation of the chopper.

Fig. 8. Shows a laser spot captured at 10kHz. The laser energy generates 4 ke$^-$ for the four shared pixels (1 ke$^-$ per diode). Despite the low laser energy, the signal is detected with a very low number of false alarms.
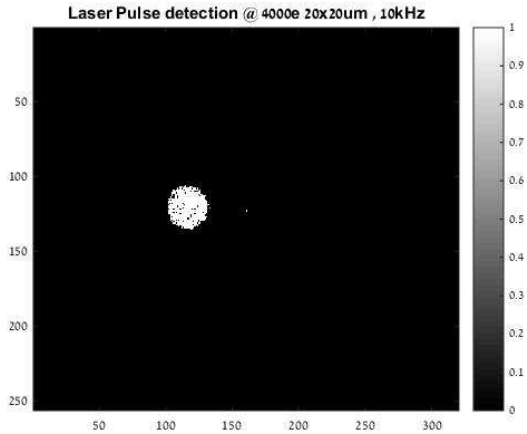

Fig. 8.   Laser spot captured at 10 kHz, the FAR observed is <0.1%

V.   SUMMARY

A novel VGA, 10 micron pixel, multi-functional InGaAs imager has been presented. The multi-functional modes enable simultaneous combination of SIM mode with detection modes. In addition, an HG mode can be used at low light level conditions. The imager prototype and first measurements have been presented, showing simultaneous SIM and detection, event-based imaging and fast laser detection.

REFERENCES

[1]   O.Ofer et al, "Performance of low noise InGaAs detector", Proc. SPIE 11741, Infrared Technology and Applications XLVII, on-line, 2021

[2]   C.G. Jakobson, "Development of Low Noise InGaAs Image Sensors for Short Wave Infrared (SWIR)", Invited Talk, Image Sensors Europe 2018, London, 2019

[3]   L. Shkedy et al. "Multifunction InGaAs detector with on-chip signal processing", Proc. of SPIE Vol. 8704 87042I-1, 2013.

[4]   C.G.Jakobson et al, "A 5Mpix 5 um 140 fps MWIR Focal Plane Array and Readout Integrated Circuit at 150K", IISW 2021, On-line, 2021

[5]   C.G. Jakobson et al, "A 10 μm pitch, SXGA Multifunctional IRFPA ROIC with In-Pixel Laser Event Detection and High Dynamic Range Imaging", IISW 2017, Hiroshima, 2017

[6]   J.A. Leniero-Bardallo et al, "Applications of event-based image sensors—Review and analysis", Int J Circ Theor Appl. 2018;46:1620–1630, 2018

[7]   S. Elkind, E. Ilan, R. Dobromislin, "Detector pixel signal readout circuit using an AC signal component in implementing an event detection mode per pixel", US Patent US9215386 B2, 2015.

[8]   C.G.Jakobson, N. Ben-Ari, I. Nevo, N. Shiloah, " Pixel readout circuit and a method for imaging", US patent US11533448 B2, 2019

[9]   A. Rogalski, "Infrared detectors: an overview", Infrared Physics & Technology 43 pp. 187–210, 2002

Fig. 7.   Event imaging: (a) measurement setup, (b) event channel output, (c) simultaneous SIM image

Fig. 7 (a) shows the event setup used to test the prototype and a single frame capture the event video. The setup consists of a rotating chopper for event generation that blocks a background light source. The rotating chopper has openings of different sizes to verify event response with different frequencies. Fig. 7 (b) shows the captured events, where green indicates a positive variation and red indicates a negative variation. The area captured by the event image is denoted by the blue circle in Fig. 7 (a). The inner shutter (the upper in the figure) generate events at a rate of 3500 events per second, and the outer shutter (the lower in the figure) does it at 70 events per second. Both event edges are well

Guided Flash Lidar: A Laser Power Efficient Approach for Long-Range Lidar

Filip Taneski
Institute for Integrated Micro and Nano Systems
University of Edinbugh
Edinburgh, United Kingdom
ORCID: 0000-0002-3073-5855

Istvan Gyongy
Institute for Integrated Micro and Nano Systems
University of Edinburgh
Edinburgh, United Kingdom
ORCID: 0000-0003-3931-7972

Tarek Al Abbas
Ouster Automotive
Ouster, Inc.
Edinburgh, United Kingdom
ORCID: 0000-0002-5486-8791

Robert K. Henderson
Institute for Integrated Micro and Nano Systems
University of Edinbugh
Edinburgh, United Kingdom
ORCID: 0000-0002-0398-7520

Abstract—Solid-state flash lidar holds the potential for low-cost, scalable depth-sensing in self-driving vehicles. However, the conventional approach of storing and processing all photon arrivals becomes impractical over long (+200 m) distances, and alternative partial histogram solutions offer poor laser power efficiency. We propose a new approach, *guided* flash lidar, allowing other on-board sensors to narrow down the depth search space for a power-efficient flash lidar solution. We use a SPAD sensor containing 64-by-32 macropixels fabricated in a standard 40 nm CMOS process. Each macropixel is capable of timing and storing photon arrivals into 8 bins within an independently programmable time window to enable guiding. A pair of vision cameras guide each macropixel to a depth window of interest by providing stereo depth estimates. The system is shown to operate outdoors over a distance of 75 m while running at 3 fps. This is a 40-times laser cycle reduction over using a sliding partial histogram approach with the same sensor, and a 25-times data reduction over using a conventional approach. The capability for guided flash lidar to mitigate multipath reflections is also demonstrated by ranging through a glass door.

Keywords—Lidar, time-of-flight, 3D vision, stereo depth.

I. INTRODUCTION

Combining lidar with other sensors such as cameras and radar, is widely considered to be the most safe and reliable design approach for self-driving vehicles [1, 2]. Automated driving safety frameworks such as those published by BMW and Volkswagen all follow a multi-sensory approach [3].

While traditional mechanical scanning lidar can achieve the necessary ranging performance, its high manufacturing cost, poor reliability, and frame rate limitations have made it less practical for use in commercial self-driving vehicles. On the other hand, solid-state *flash* lidar is a low-cost, scalable solution used in many indoor lidar applications such as smart phones and robotics. However, self-driving vehicles require lidar to perform over long distances (200 m) in outdoor environments. For flash lidar to meet these requirements, two critical challenges remain: high laser power consumption and large data volume. As illustrated in Fig. 1(a), to measure over longer distances requires the lidar sensor to count, store and process more photons over a wider temporal window. Novel
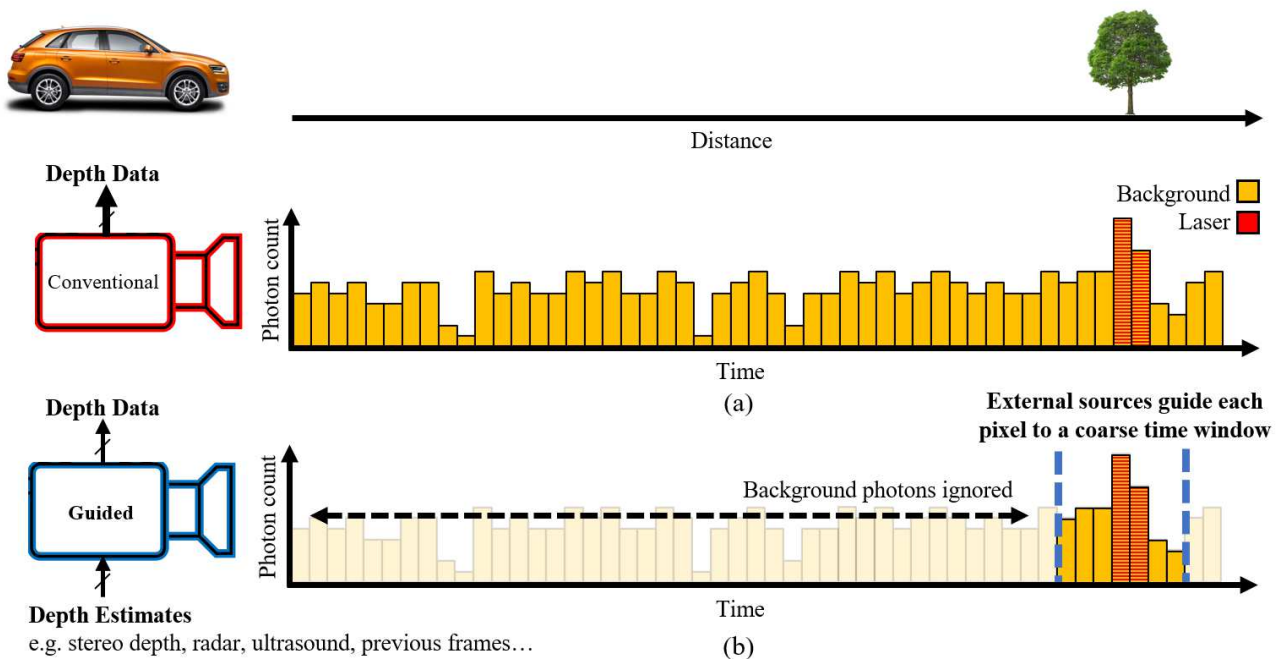

Fig. 1: Counting photon arrival events in (a) conventional flash lidar compared (b) to the proposed guided flash lidar approach.

"partial histogram" lidar sensors attempt to resolve this issue but do so at the cost of a severe laser power penalty [4].

This work showcases a new approach: *guided* flash lidar, which enables external sources to guide each pixel of the lidar to a coarse time window of interest (Fig. 1(b)). By utilizing data from other sensors on-board a self-driving vehicle to reduce the temporal search space, we propose that guided flash lidar can achieve the low power and data output required.

II. RELATED WORK

Various alternatives over the conventional histogram approach (Fig. 1(a)) have been proposed to reduce the amount photon arrival data stored and processed on the lidar sensor. These *partial histogram* approaches can be grouped into two categories: *zooming* and *sliding*. Both are illustrated in Fig. 2.
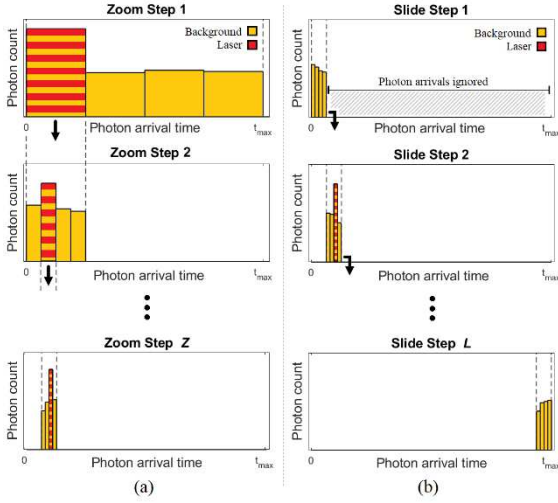

Fig. 2: Illustration of (a) zooming and (b) sliding histogram approaches.

In the zooming approach [5-8], the temporal histogram is initially spread across the full distance range. After multiple laser cycles, the peak (signal) bin is identified, and the histogram is zoomed in to a new, narrower time window. Multiple zoom steps can be performed until the required precision is achieved. In contrast, the sliding approach [9], spreads the histogram across only a subset of the sensing range and gradually slides the time window to cover the full range.

A major disadvantage of all partial histogram approaches is the severe laser cycle penalty they incur. In zooming, the wider time window used in earlier zooming steps leads to a high background photon count in the signal bin. As a result, many laser cycles are required to identify the peak bin. In the case of sliding, most laser pulses returning from the target fall outside of the window being observed at any one time. As a result, every L sliding steps used leads to an L-times increase in laser cycles and power. An in-depth review and analysis of partial histogram approaches is given in [4].

While partial histogram approaches may be sufficient for indoor applications, a more power-efficient solution is required for operating over long distances and under high ambient conditions. Although lidar is an effective solution for measuring depth to within centimeter precision, other sensors on-board a vehicle are also capable of providing range data. These include cameras, ultrasound, radar, and GPS/mapping data. A guided approach can use this data to reduce the depth search space, instead of the inefficient partial histogram approach of relying exclusively on the lidar to do so.

III. GUIDED LIDAR SYSTEM

A. Guided Lidar Sensor

The sensor used to demonstrate this approach (Fig. 3) was implemented in a standard 40 nm CMOS process and contains 64×32 macro pixels. Each macropixel is comprised of 4×4 SPADs alongside processing for timing and storing photon events into 8×12-bit time bins. The primary function of this sensor, originally presented in [10], allows each pixel to independently scan through time windows until a target (peak) is detected, at which point scanning stops and the peak is tracked as it moves into adjacent time windows. In this work, the sensor is reengineered to allow the time window of every pixel to be continuously and independently programmed.
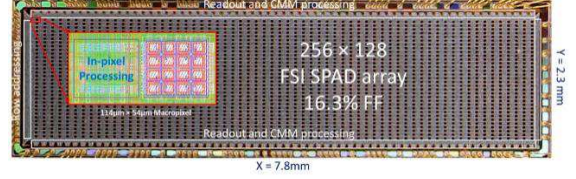

Fig. 3: Micrograph of the sensor used to demonstrate guided flash lidar

B. Depth Estimates Source: Stereo Depth

The presented implementation uses a pair of imaging cameras to provide stereo depth estimates for guiding the lidar sensor. This utilizes the resulting shift in pixel value (termed disparity) of matched points between each camera image to estimate distance. Distance z is given as a function of disparity d, camera baseline B and focal length f by:

$$z = \frac{fB}{d} \qquad (1)$$

Relying on pixel disparity results in discrete depth estimates which limits depth resolution. This resulting depth accuracy Δd is given by:

$$\frac{\Delta z}{\Delta d} = \frac{fB}{d^2} \implies |\Delta z| = \frac{z^2 \Delta d}{fB} \qquad (2)$$

Through interpolation, sub-pixel disparity Δd as low as 0.1 pixels is achievable. In practice, accuracy is limited by the performance of the adopted disparity matching algorithm. For the purpose of demonstrating guided flash lidar, the well-established semi-global matching (SGM) algorithm [11] is used as a convenient and time-efficient solution.
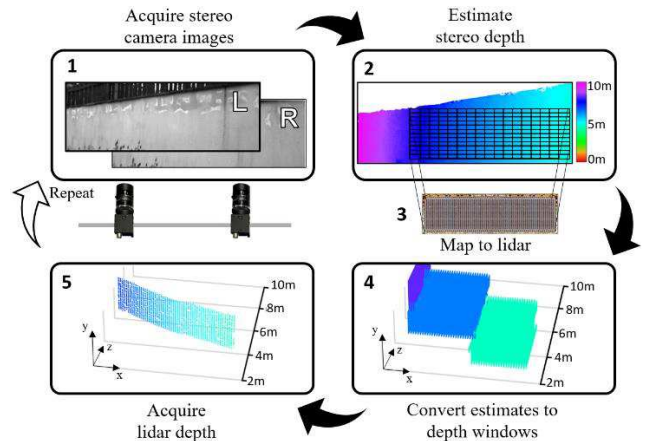

Fig. 4: Process flow of guided flash lidar using stereo depth estimates.

C. Process Flow

The processes involved in guiding the flash lidar sensor are shown in Fig. 4. The process begins by acquiring images from the stereo cameras, followed by solving the disparity between images to estimate depth. By prior calibration of the cameras and lidar using checkerboard images, the intrinsic parameters of all cameras and their pose with respect to each other is determined. This allows accurate mapping of the stereo depth image to each pixel on the lidar sensor. Finally, the depth estimate assigned to each pixel is converted to a time (depth) window and the guided sensor only counts returning photons within its assigned depth window.

D. Setup

The guided lidar setup, running off an Intel Core i7 8th generation laptop, is show in Fig. 5. It consists of:

- the lidar sensor mounted with 25 mm lens and 940 nm bandpass filter (10 nm full width half maximum)
- a 940nm laser module running at 80 kHz
- 2 × FLIR Blackfly BFS-U3-16S2M-CS cameras with 12 mm lenses, mounted on a meter-length rail
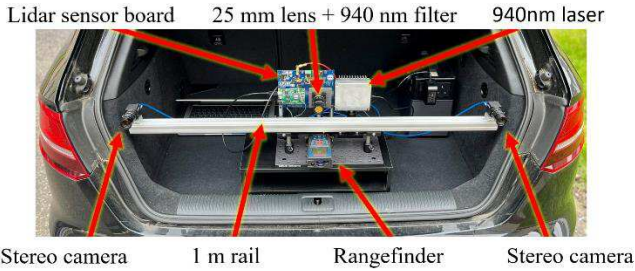- a Bosch GLM250VF rangefinder for ground truth

Fig. 5: Guided lidar setup

The sensor is configured to use 3 m wide depth windows (8×0.375 m bins) with a 1.125 m overlap. A maximum of 128 windows provides 240 m of unambiguous range. Wider depth windows (and bins) would give a greater safety margin for the guiding depth estimates at the cost of lidar accuracy.

IV. GUIDED LIDAR PERFORMANCE

A. Dynamic Outdoor Scene

Fig. shows a sample frame captured by the guided lidar system operating outdoors (15 klux) at 3 fps. It shows the stereo cameras creating depth estimates which are converted into coarse depth windows across objects in the scene such as the van. Every pixel then resolves any signal peak found in its allotted window to produce an accurate depth map. Further frames from the scene are given in Fig. 7 showing a sample pixel guided to follow the van as it drives away as far out as 75 m. A sliding partial histogram approach would otherwise require stepping through each of the 40+ depth windows, resulting in a 40 × laser cycle and power increase. Alternatively, a conventional histogram approach would require storing 200 bins per macropixel as opposed to only 8.
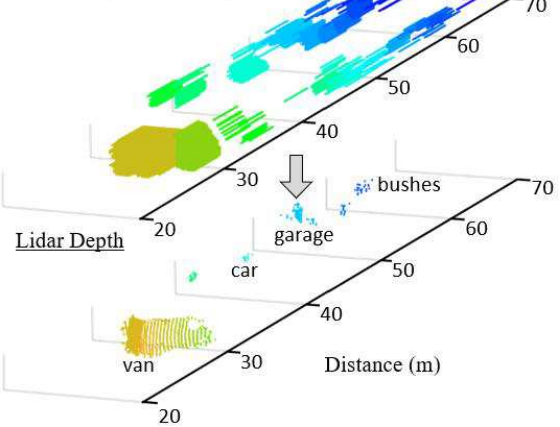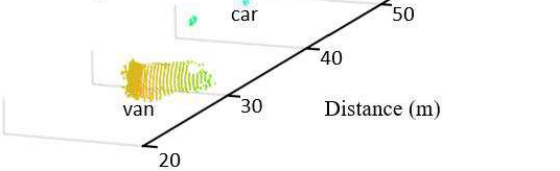

Fig. 6: One frame from the guided lidar system broken down into its component parts. Here the system is operating outdoors at 3 fps.

The processing time for each step of the system within a single frame is provided in Fig. 8, showing the lidar exposure period to be dominate most of the frame time. In short range indoor settings, runtimes exceeding 5 fps are achievable.
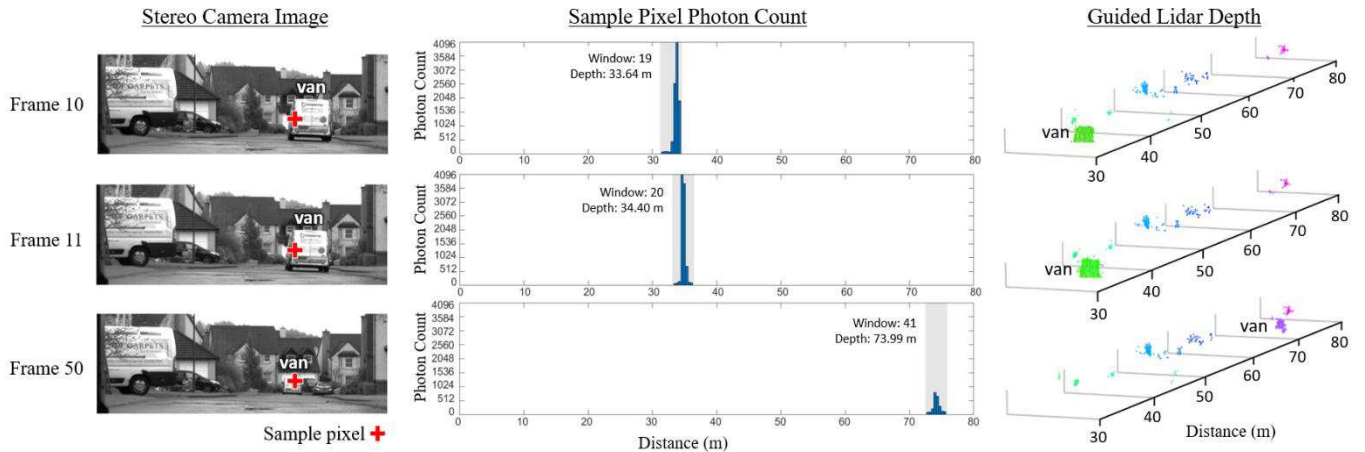

Fig. 7: Three further frames from the scene in Fig. 6 showing the time window of a sample pixel guided to track the moving van.
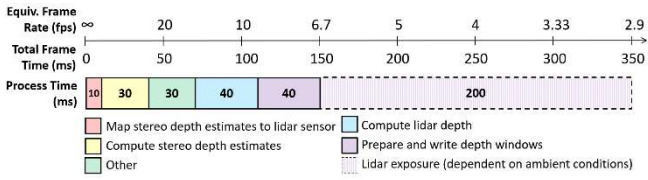
Fig. 8: Processing time of each step in one frame of the guided lidar system.

B. Outdoor Accuracy

The same setup running at 3 fps was used to perform a distance sweep up to 50 m under a higher ambient light condition of 72 klux. The combined precision and accuracy were evaluated by ranging a human target at 81 different points (3×3 pixels over 9 frames). The results are shown in Fig. 9, demonstrating that while the stereo depth accuracy deteriorates with distance, the guided lidar system maintains a root-mean-squared error of less than 20 cm.
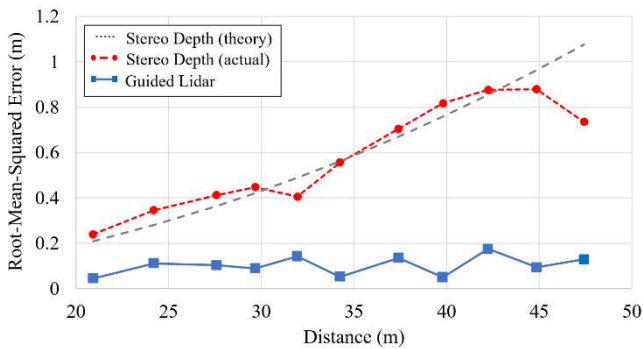

Fig. 9: Outdoor (72 klux) ranging performance operating at 3 fps. Theoretical stereo depth accuracy as given by (2) assumes sub-pixel disparity of 0.25.

C. Guiding Through Glass

Another benefit of guided flash lidar is mitigating multipath reflections. These can lead to ranging artefacts in the face of transparent surfaces or glare [12]. Fig. 10 shows the system is guided to the human figure beyond the glass door which is otherwise obscured if only the first peak is captured i.e. using zooming partial histogram approaches.
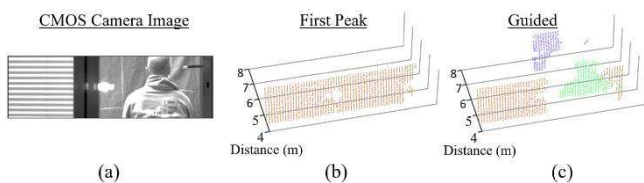

Fig. 10: (a) A human behind a glass screen is (b) obscured when using the first detected laser peak (c) revealed using guided lidar.

V. SUMMARY

The first ever guided flash lidar system has been demonstrated. The presented system is capable of operating at 3 fps under high ambient daylight conditions of 72 klux. Outdoor ranging up to 75 m is demonstrated, with each macropixel being guided to one of over 40 separate depth windows. This results in a 40× reduction in laser cycles (and laser power) over a sliding partial histogram approach. An equivalent performance using a conventional histogram approach would require the sensor to accommodate 25× more bins of storage per pixel. Guided flash lidar is also shown to mitigate against multipath reflections, successfully ranging though a glass door.

To improve performance, future design iterations could explore alternative guiding sources and parallelizing tasks within a frame. Combining a guided lidar approach with the increased sensitivity of state-of-the-art SPAD processes [13, 14] would greatly reduce the required exposure time to range over longer distance and/or using shorter exposure periods.

REFERENCES

[1] S. Rangwala, "Automotive LiDAR Has Arrived." [Online]. Available: https://www.forbes.com/sites/sabbirrangwala/2022/05/23/automotive-lidar-has-arrived/?sh=1f164cf013de

[2] Z. Wang, Y. Wu, and Q. Niu, "Multi-Sensor Fusion in Automated Driving: A Survey," *IEEE Access,* vol. 8, pp. 2847-2868, 2020, doi: 10.1109/ACCESS.2019.2962554.

[3] Aptiv *et al.*, "Safety First For Automated Driving [White Paper]," 2019. [Online]. Available: https://group.mercedes-benz.com/documents/innovation/other/safety-first-for-automated-driving.pdf

[4] F. Taneski, T. A. Abbas, and R. K. Henderson, "Laser Power Efficiency of Partial Histogram Direct Time-of-Flight LiDAR Sensors," *Journal of Lightwave Technology,* vol. 40, no. 17, pp. 5884-5893, 2022, doi: 10.1109/JLT.2022.3187293.

[5] C. Zhang, S. Lindner, I. M. Antolović, J. M. Pavia, M. Wolf, and E. Charbon, "A 30-frames/s, 252 x 144 SPAD Flash LiDAR With 1728 Dual-Clock 48.8-ps TDCs, and Pixel-Wise Integrated Histogramming," *IEEE Journal of Solid-State Circuits,* vol. 54, no. 4, pp. 1137-1151, 2019, doi: 10.1109/JSSC.2018.2883720.

[6] B. Kim *et al.*, "7.2 A 48×40 13.5mm Depth Resolution Flash LiDAR Sensor with In-Pixel Zoom Histogramming Time-to-Digital Converter," in *2021 IEEE International Solid- State Circuits Conference (ISSCC)*, 13-22 Feb. 2021 2021, vol. 64, pp. 108-110, doi: 10.1109/ISSCC42613.2021.9366022.

[7] C. Zhang *et al.*, "A 240 x 160 3D Stacked SPAD dToF Image Sensor with Rolling Shutter and In Pixel Histogram for Mobile Devices," *IEEE Open Journal of the Solid-State Circuits Society,* pp. 1-1, 2021, doi: 10.1109/OJSSCS.2021.3118332.

[8] S. Park *et al.*, "5.3 An 80×60 Flash LiDAR Sensor with In-Pixel Histogramming TDC Based on Quaternary Search and Time-Gated Δ-Intensity Phase Detection for 45m Detectable Range and Background Light Cancellation," in *2022 IEEE International Solid- State Circuits Conference (ISSCC)*, 2022.

[9] D. Stoppa *et al.*, "A Reconfigurable QVGA/Q3VGA Direct Time-of-Flight 3D Imaging System with On-chip Depth-map Computation in 45/40nm 3D-stacked BSI SPAD CMOS," in *International Image Sensor Workshop 2021*, 2021.

[10] I. Gyongy *et al.*, "A Direct Time-of-flight Image Sensor with in-pixel Surface Detection and Dynamic Vision," *IEEE Journal of Selected Topics in Quantum Electronics,* pp. 1-12, 2023, doi: 10.1109/JSTQE.2023.3238520.

[11] H. Hirschmuller, "Stereo Processing by Semiglobal Matching and Mutual Information," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 30, no. 2, pp. 328-341, 2008, doi: 10.1109/TPAMI.2007.1166.

[12] I. Gyongy, N. A. Dutton, and R. K. Henderson, "Direct Time-of-Flight Single-Photon Imaging," *IEEE Transactions on Electron Devices,* pp. 1-12, 2022, doi: 10.1109/TED.2021.3131430.

[13] O. Kumagai *et al.*, "7.3 A 189×600 Back-Illuminated Stacked SPAD Direct Time-of-Flight Depth Sensor for Automotive LiDAR Systems," in *2021 IEEE International Solid- State Circuits Conference (ISSCC)*, 13-22 Feb. 2021 2021, vol. 64, pp. 110-112, doi: 10.1109/ISSCC42613.2021.9365961.

[14] K. Morimoto *et al.*, "3.2 Megapixel 3D-Stacked Charge Focusing SPAD for Low-Light Imaging and Depth Sensing," in *2021 IEEE International Electron Devices Meeting (IEDM)*, 11-16 Dec. 2021 2021, pp. 20.2.1-20.2.4, doi: 10.1109/IEDM19574.2021.9720605.

FAD-SPADs: a New Paradigm for Designing Single-Photon Detecting Arrays

Mel White[1†], Tianyi Zhang[1†], Akshat Dave[1], Shahaboddin Ghajari[2], Alyosha Molnar[2], and Ashok Veeraraghavan[1]

[†]Indicates equal contribution
[1]Rice University, Houston, TX
[2]Cornell University, Ithaca, NY

Abstract—**We present a novel architecture for the design of single-photon detecting arrays that captures relative intensity or timing information from a scene, rather than absolute. The proposed method for capturing relative information between pixels or groups of pixels requires very little circuitry, and thus allows for a significantly higher pixel packing factor than is possible with per-pixel TDC approaches. The inherently compressive nature of the differential measurements also reduces data throughput and lends itself to physical implementations of compressed sensing, such as Haar wavelets. We demonstrate this technique for HDR imaging and LiDAR, and describe possible future applications.**

Index Terms—**SPAD, LiDAR, HDR, compressed sensing, computational imaging**
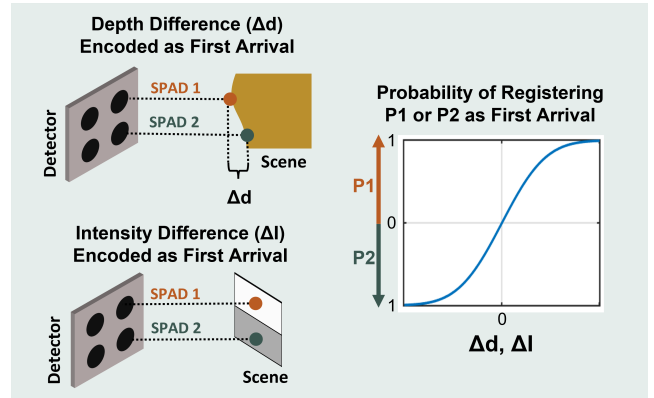
Fig. 1. FAD-SPAD operation principle. 1: Either depth intensity differences can be encoded with the first arrival of a photon within a time window. 2: The relationship between the relative flux and the probability of recording an up or down count with digital circuitry is nonlinear.

I. INTRODUCTION

Single-Photon Avalanche Diodes (SPADs) are highly sensitive photodiodes which can detect individual photons with extremely fast response and high precision. For this reason, they have become the gold standard in many photon-limited imaging applications [1], [2]. However, current SPAD array designs suffer from low spatial resolution due to complex circuitry and high data throughput needed for capturing absolute timestamps or photon counting, which limits their usage in many downstream applications. We present a novel, lightweight readout architecture which overcomes existing challenges in SPAD array designs: first arrival differential SPADs (FAD-SPADs). Our technique differs from previously proposed hardware solutions such as Time to Digital Converter (TDC) sharing, adaptive sensing [3], data sketching [4], and sensor fusion [5], [6] because it is not TDC-based, and instead relies on small circuits that perform data compression at the sensor.

FAD-SPADs record differential measurements between pixels, either in intensity or time of flight. Our key insight is that rich information is encoded in the relative timing of the first photon captured within a time window (See Fig. IV), and this information can be captured by small and simple digital circuitry. This method can also provide gains in certain imaging metrics, including significantly reduced circuit footprint and better pixel packing, orders of magnitude data size reduction, and improved dynamic range.
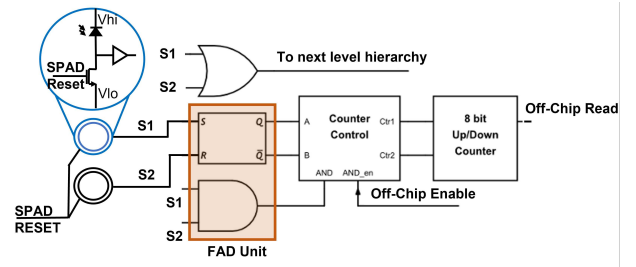

Fig. 2. FAD-SPAD readout circuitry. The SPADs are grouped, and each group is connected to the SR-Latch. Based on which group detects a photon first the counter will count up or down.

II. KEY CONCEPT: THE FAD UNIT

Let us first consider the case of only two SPAD pixels. We connect their readout to the inputs of an SR latch, as shown in Fig. 2. The first photon event on either of the two pixels triggers the latch, which holds its state regardless of subsequent photon events on either pixel. This information is passed to an up/down counter, which counts up if pixel 1 saw the first event, down if pixel 2 saw the first event, and holds its state if no events were detected. After a short time window (T), the latch and SPADs are reset to the "listening" state, and this process is repeated for N cycles. With the counter

Mel White is the corresponding author and can be reached at: email: mel.white@rice.edu
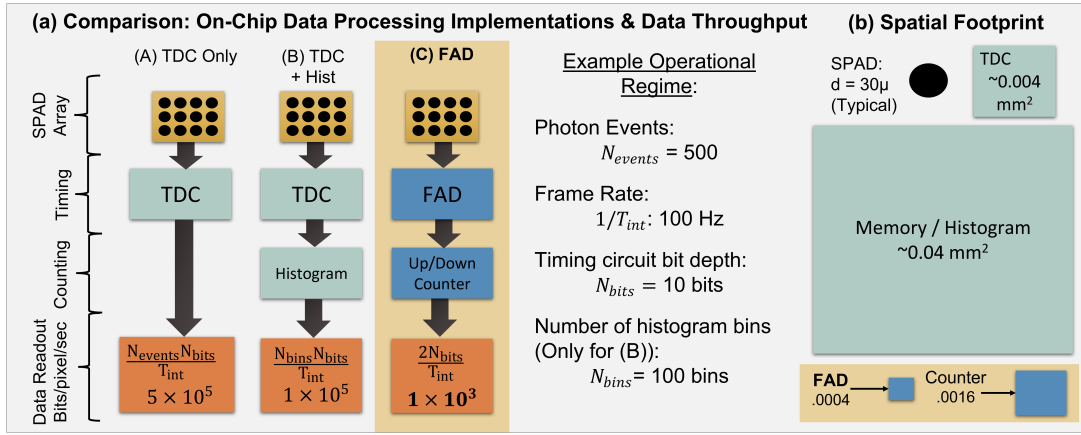
Fig. 3. A comparison of the estimated data throughput and circuit footprint for conventional approaches compared to ours (in highlighted gold boxes). At the operating conditions given in the figure, our data throughput is two orders of magnitude smaller while only using only 5% of the circuit footprint.

initialized to a midpoint, the resulting readout from the counter measures the relative flux between the two pixels. An OR gate connected to both SPADs allows us to pass this information along to other groups of pixels, or even a cumulative output. With the addition of an optionally-enabled AND gate, we can further distinguish between dual events (where both SPADs see a photon) and single events (only one SPAD sees a photon).

These circuits are extremely compact and throughput-efficient (see Fig. 3 for details) and thus scalable to large arrays. Critically, only the FAD unit need be placed inside the array and near the actual SPADs for accuracy; the counter and other supporting circuitry can be outside of the array, and do not impact the fill factor or spatial resolution. The FAD unit can be constructed of only 12 transistors: 8 for a NAND-type SR latch, and 4 for the AND gate. We contrast this to the use of per-pixel TDCs, which are often 3x-4x larger than the SPAD itself, and must be placed immediately adjacent to the pixel for accurate measurements.

A. Relationship between differential flux and counter readout

Uncertainty from photon noise provides a nonlinear relationship between the relative fluxes at each pixel and the recorded count. This relationship is expressed mathematically as the difference between the probability of an up or down count times the number of cycles. Taking the fluxes at the two SPADs to be Φ_1 and Φ_2 (photon flux per cycle), and assuming Poisson arrival processes at both SPADs, this gives:

$$\mathbb{E}(\mathbb{D}) = N_{cycles}[1 - e^{-(\Phi_1+\Phi_2)T}]\left(\frac{\Phi_1 - \Phi_2}{\Phi_1 + \Phi_2}\right) \quad (1)$$

Where T is the period of the detection cycle.

B. Connectivity schemes

To build upon the core concept described for two SPADs, pixels can be grouped together via OR gates at their output. Thus, if the first photon hits any pixel in group A before group B, the SR latch will count up, and vice versa for group B. Moreover, SPADs may be connected to many different groups simultaneously, allowing for simultaneous differential measurements.

Some possible configurations are shown in Fig. 4. Local groupings will suppress local background, while other kinds of clustering mimicking Haar or Hadamard transforms enable data compression at the sensor.
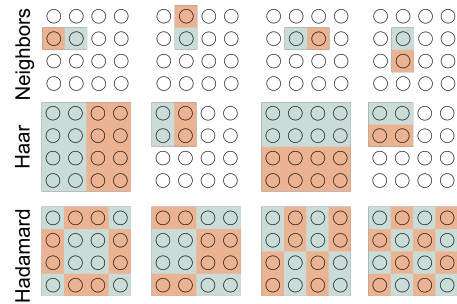

Fig. 4. Some examples of 'OR' connected groups in a 4×4 array. The choice of grouping will impact the data bandwidth and SNR of the reconstructed image.

III. ENCODING FIRST ARRIVAL AS INTENSITY DIFFERENCE

In a passive lighting arrangement, the differential information reflects the intensity of the scene. A SPAD saturates if it detects one photon, and so it may seem counter-intuitive to use them in high dynamic range imaging applications. Yet, several recent works [7], [8] have shown that single-photon sensitivity and non-linear behavior of SPADs can be exploited for high dynamic range imaging. The major drawback of existing SPAD array architectures is that they suffer from limited photon counter bit depth; under high photon flux, the individual photon counters saturate, and any local differences in intensity are lost. Our approach uses relative rather than absolute timing information, which will not saturate under small local gradients, as illustrated in Fig. 5. This can also be inferred from Equation 1. As fluxes Φ_1 and Φ_2 increase, the exponential term disappears, but the differential term in the numerator remains. As long as the difference does not
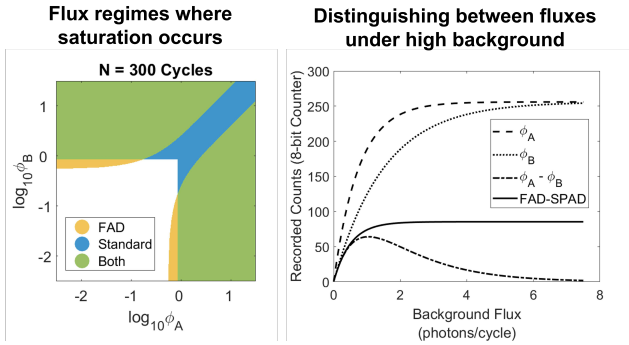
Fig. 5. Left: when the fluxes at pixels A and B are both high, regular counters will saturate, but the FAD does not. Right: under high background lighting conditions, the difference between SPADs with counters vanishes; FAD-SPADs preserve the difference.
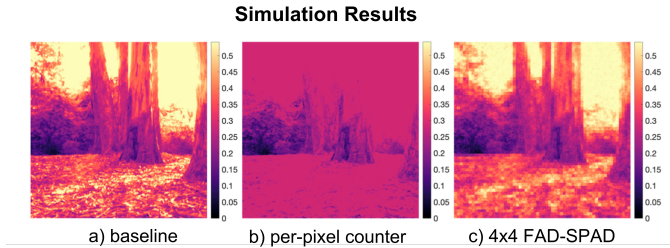

Fig. 6. A simulation showing how FAD-SPADs preserve high dynamic range of an image. The "blocky" artifacts in our result are due to the fact that we simulated a scanning 4x4 array (to mimic our prototype).

exceed the limits of the counter, the difference is preserved, theoretically, under infinite background flux. Fig.6 illustrates how small flux differences are preserved, even under high-flux background conditions. The dynamic range of FAD-SPADs is in practice determined by the largest and smallest detectable signal, limited by SNR. The interesting case is under high flux conditions (at low flux, it behaves exactly as two independent SPADs).

To show that this can be practically achieved, we have implemented a proof-of-concept prototype containing 16 pixels (4x4) using a conventional 180 nm CMOS process with a Haar grouping scheme, shown in Fig. 7 [9]. While this implementation includes only two layers of hierarchy of Haar wavelets, the small size of the local digital circuitry means this design is scalable to any $2^N \times 2^N$ array. Notably, the Haar pixel grouping scheme combined with the compact size of the support circuitry results in minimal impact on device's footprint; with no additional design effort, we achieve a 34% fill factor.

IV. ENCODING FIRST ARRIVAL AS DEPTH DIFFERENCE

Flash LiDAR systems utilize SPAD arrays to perform single-shot 3D imaging without the need for mechanical scanning [10]. However, SPAD arrays require per-pixel timing circuits (TDCs) with high spatial footprint and data throughput limiting the spatial and temporal resolution of such systems. In contrast to TDCs, the FAD units are more lightweight and are capable of depth difference between pairs of pixels.
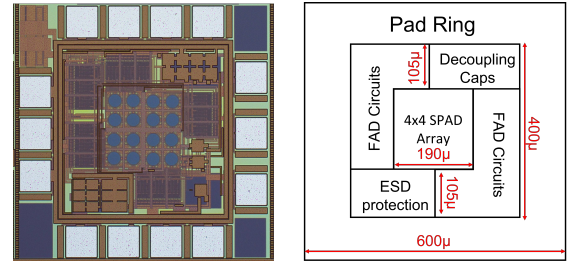

Fig. 7. A micrograph of the fabricated prototype in 180 nm CMOS. The 4x4 array (excluding the pad ring and supporting circuitry) measures $190\mu m \times 190\mu m$ and has a fill factor of 34%.

Using the FAD units, we design a flash LiDAR system that can perform high-resolution 3D imaging and scene inference [11]. The core idea is that FAD captures the relative order of photon arrivals at the two pixels. There exists a one-to-one mapping between this differential measurement and depth differences between the two pixels. FAD-LiDAR enables 3D inference tasks, including depth edge detection, depth-based segmentation, surface normal estimation, and depth imaging with only a few TDCs as shown in Fig. 8. It is worth noting that for cases where absolute depth is not required, our approach can be implemented without any TDCs.

Consider that the laser and detector are collocated. Assume SPAD pixel 1 points to a scene location that is closer to the detector, and SPAD pixel 2 to a farther location (in a setup shown in the bottom left section of Fig.). Then, within a time window, photons reaching SPAD pixel 1 are more likely to arrive earlier than photons from SPAD pixel 2. Over a large number of cycles, the relative frequency of first arrival photons between the pixels can capture information about depth difference Δd. This leads to a monotonic mapping between the FAD measurements, FAD, and the depth difference, $\Delta \tau$, as

$$ FAD \propto -N_{\text{cycles}} \alpha_1 \alpha_2 \left(\frac{\Delta \tau}{2\sigma} \right) \qquad (2) $$

where N_{cycles} is the total number of laser cycles and α_1, α_2 are the photon flux (per cycle) at the two pixel locations 1 and 2. We can acquire intensity estimates $\hat{\alpha}_1$, $\hat{\alpha}_2$ by using intensity measurements. To decouple illumination effects caused by single photon arrivals, we (1) enable the AND gate so that only when both pixels receive returning photons does FAD perform a comparison (2) measure intensity values at each pixel and factor them out. After these operations, we reach a normalized FAD measurement $nFAD$ that is only dependent on the relative depth between two pixels.

$$ nFAD = \frac{FAD}{N_{\text{cycles}} \hat{\alpha}_1 \hat{\alpha}_2} \qquad (3) $$

$$ nFAD \propto -\left(\frac{\Delta \tau}{2\sigma} \right) \qquad (4) $$

Using FAD units and nearest-neighbor connectivity, we directly perform 3D scene inference tasks as shown in (Fig. 8). Tasks such as depth edge detection, depth-based segmentation, and normal estimation are sufficient with FAD measurements
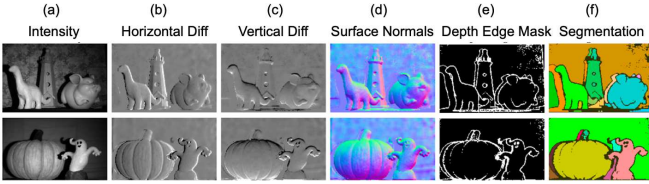
Fig. 8. 3D imaging applications of FAD LiDAR. Column (a): intensity view of the scenes. Columns (b) to (f) correspond to different 3D applications as labeled in the figure.
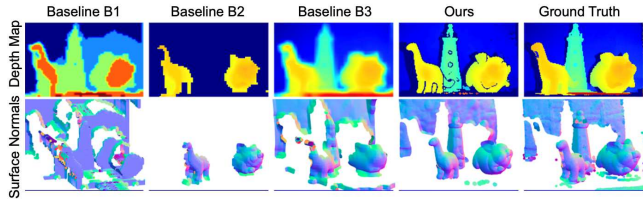

Fig. 9. High-resolution 3D imaging and surface normals by FAD LiDAR emulated using a single-pixel SPAD setup. Reconstruction from a scanning-based LiDAR design is denoted as Ground Truth. Conventional flash LiDAR designs B1, B2, and B3 (see [11] for details) suffer from performance tradeoffs resulting in poor depth resolution (B1), range (B2), or spatial resolution (B3). Our differential flash design offers significantly better reconstruction quality for the same data throughput as conventional baselines.

(relative depth information) and per-pixel intensity estimates. Depth edge detection and segmentation can be performed by appropriately thresholding nFAD measurements, while normal estimation requires a two-step procedure: first inverting relative depth difference from nFAD, then performing Poisson integration [12] to generate clean surface normal estimates.

To reconstruct absolute depth across the scene, we can capture a few absolute ToF measurements by sparsely distributing TDCs across the FAD array. We show that FAD-based LiDAR provides depth maps at higher resolution and range than existing TDC-based flash LiDAR, and demonstrate via emulation that FAD-LiDAR provides improved performance for the same data bandwidth (Fig. 9).

V. CONCLUSIONS AND DISCUSSION OF FUTURE DIRECTIONS

In this paper we present techniques and supporting analysis for a novel type of SPAD array design based on differential sensing. We also present two applications: HDR imaging and 3D imaging, featuring two architectures (Haar and nearest neighbor). However, this is only a small sample of the capabilities enabled by FAD architectures.

For example, the differential nature of FAD units is inherently amplifying of local differences and thus could enhance contrast in bioimaging applications, such as the loss in contrast due to scattering.

There is also fertile space for analysis of other differential connectivity schemes to enable compressed sensing of images. Binary compressed sensing matrices (eg, Hadamard transforms) can be implemented similarly to our Haar example for HDR by simply changing the connected groups. The

differential grouped measurements could also be used to do adaptive sensing in sparse image acquisition. For example, if a large differential signal is found in one region of an image, then a smart sensor could continue to collect finer-grained measurements in that region, and not collect redundant data in a region of the image that lacks contrast.

Finally, the gains in circuit footprint and scalability of the concept we show here could facilitate the development of larger and denser SPAD arrays with high photon detection probability.

REFERENCES

[1] C. Bruschini, H. Homulle, I. M. Antolovic, S. Burri, and E. Charbon, "Single-photon avalanche diode imagers in biophotonics: review and outlook." *Light: Science & Applications*, vol. 8, no. 87, 2019.

[2] F. Villa, F. Severini, F. Madonini, and F. Zappa, "SPADs and SiPMs arrays for long-range high-speed light detection and ranging (LiDAR)," *Sensors*, vol. 21, no. 11, 2021.

[3] C. Zhang, S. Lindner, I. M. Antolović, J. M. Pavia, M. Wolf, and E. Charbon, "A 30-frames/s, 252×144 SPAD flash LiDAR with 1728 dual-clock 48.8-ps TDCs, and pixel-wise integrated histogramming," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 4, pp. 1137–1151, 2018.

[4] M. P. Sheehan, J. Tachella, and M. E. Davies, "A sketching framework for reduced data transfer in photon counting LiDAR," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 989–1004, 2021.

[5] D. B. Lindell, M. O'Toole, and G. Wetzstein, "Single-photon 3D imaging with deep sensor fusion." *ACM Trans. Graph.*, vol. 37, no. 4, pp. 113–1, 2018.

[6] Y. Liu, F. Gutierrez-Barragan, A. Ingle, M. Gupta, and A. Velten, "Single-photon camera guided extreme dynamic range imaging," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 1575–1585.

[7] A. Ingle, A. Velten, and M. Gupta, "High flux passive imaging with single-photon sensors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6760–6769.

[8] M. Zarghami, L. Gasparini, M. Perenzoni, and L. Pancheri, "High dynamic range imaging with TDC-based CMOS SPAD arrays," *Instruments*, vol. 3, p. 38, 08 2019.

[9] M. White, S. Ghajari, T. Zhang, A. Dave, A. Veeraraghavan, and A. Molnar, "A differential SPAD array architecture in 0.18 μm CMOS for HDR imaging," in *2022 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2022, pp. 292–296.

[10] F. Villa, F. Severini, F. Madonini, and F. Zappa, "Spads and sipms arrays for long-range high-speed light detection and ranging (lidar)," *Sensors*, vol. 21, no. 11, p. 3839, 2021.

[11] T. Zhang, M. J. White, A. Dave, S. Ghajari, A. Raghuram, A. C. Molnar, and A. Veeraraghavan, "First arrival differential LiDAR," in *2022 IEEE International Conference on Computational Photography (ICCP)*, 2022, pp. 1–12.

[12] R. T. Frankot and R. Chellappa, "A method for enforcing integrability in shape from shading algorithms," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 10, no. 4, pp. 439–451, 1988.

Physical Modeling and Parameter Extraction for Event-based Vision Sensors

Andreas Suess, Menghan Guo, Rui Jiang, Xiaozheng Mou, Qiping Huang, Wenlei Yang, Shoushun Chen

OMNIVISION, 4275 Burton Drive, Santa Clara, CA 95054 USA

Abstract—Latency and noise are crucial aspects of Event-based Vision Sensors. Yet, in simulators used to create synthetic event data these effects are predominantly modeled phenomenologically and are rarely calibrated to actual measurements or circuit simulations. This work presents a physics-based latency and noise model achieving strong resemblance with circuit simulations and measurements. The model is computationally efficient enough to be suitable for camera simulation. This enables accurate training data synthesis for algorithm development and guides sensor design.

I. INTRODUCTION

Event-based Vision Sensors (EVS) experience an increasing commercial interest [1]–[6]. Simulators are commonly employed to synthesize events from videos connecting sensor design and algorithm development depending on application requirements [7]–[10]. Today's simulators, however, are mostly phenomenological and lack calibration against actual sensor designs using simulation or measurements [7]–[9]. [7], [8] introduced adaptive frame-upsampling before event generation and randomized the contrast threshold to allow for generalization of trained algorithms to sensor variability. [9] introduced modeling of pixel non-idealities. Pixel latency was modeled by a single pole IIR filter whose bandwidth scales proportional to the light level. A pixel-wise Gaussian-distributed contrast threshold models sensor non-uniformity. Periodic leak events and Poisson sampled noise events were introduced into the event stream, whereas the noise events were scaled with luma level to capture light dependence. [10] extended the attempt towards realistic event simulation by calibrating a more complex semi-empirical pixel latency model to circuit simulations. Furthermore, refractory period and latency from the peripheral circuitry were modeled. [10] presented reasonable matching of pixel latency at higher photocurrents, but matching at low light was limited. Conversely to [9], [10] did not inject noise events into a pre-computed noise-free event stream, but directly added random noise to the signal amplitude before the comparators. However, bandwidth was not considered. Modeling the noise directly in the voltage domain is crucial to match effects such as time-stamp uncertainty caused by noise, or fire probabilities in case the signal is in proximity of the contrast threshold. Leak events should also be modeled in the voltage domain to reflect the signal dependence of leak events.

In order to advance event simulators, this work derives physical models for latency and noise considering bias and temperature conditions (Sec. II and Sec. III). It will be shown that the models surpass the matching to circuit simulations achieved in [10] (Sec. II). Views on parameter extraction will be outlined and the model is validated against measurements (Sec. V).

II. LARGE-SIGNAL PIXEL LATENCY MODEL

Fig. 1 depicts a simplified schematic of the analog portion of an event-based vision pixel. The photodiode current, $I_{PD} = I_{photo} + I_{dark}$, is converted into a logarithmic voltage measured at node V_{FE}. M_{SF} decouples the front-end circuit from a difference-detecting switched-capacitor filter of gain G. Subsequent comparators determine if a temporal contrast beyond a predefined
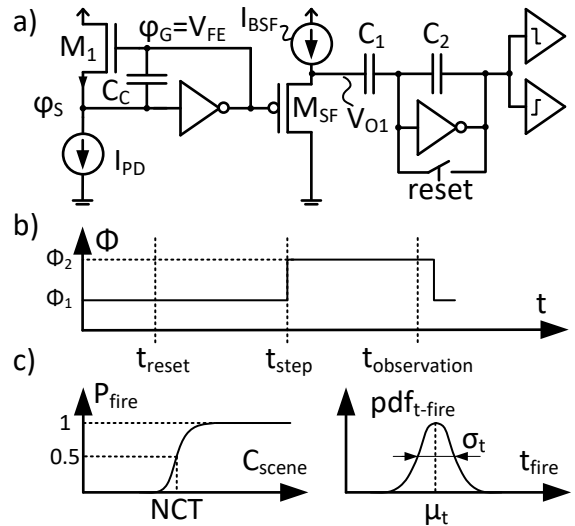

Figure 1. a) analog EVS pixel circuit schematic, b) temporal contrast measurement timing diagram, and c) fire probability "S-curve" and time-stamp distributions.

threshold is detected. Fig. 1b illustrates the temporal contrast step measurement technique [11], [12]. Fig. 1c depicts that at increasing contrast step, $C = [\Phi_2 - \Phi_1]/\Phi_1$, a corresponding increase in event firing probability is observed. The contrast at which the probability amounts to 50% is termed the Nominal Contrast Threshold (NCT) [3]. This measure is of relevance as it indicates the effective contrast threshold of a given pixel at a given measurement condition. Ideally, the transition from low firing probability to high firing probability is sharp around that NCT value. However, temporal noise causes randomness in trigger probability and results in the shape that coins this characteristic "S-curve." Clearly, the slope of the S-curve is important to characterize a pixel. The right part of Fig. 1c illustrates the event firing time probability distribution. The firing time has an offset latency μ_t determined by transients to charge/discharge nodes in the analog portion of the event pixel, as well as propagation and scan delays from the digital peripheral readout. Event pixels exhibit a timing uncertainty σ_t which is characterized by electronic noise and randomness in the peripheral readout. For S-curve characterization, it is preferred to select a small region of interest to separate the readout delay and uncertainty from pixel latency. It is clear that merits like NCT or firing time strongly depend on: radiance, the time interval of pixel reset and applied scene contrast step, as well as the overall observation time under which e.g. noise is permitted to elevate above a contrast threshold. This implies that such merits do not have full predictive value at operation under varying conditions. Therefore, a model-based approach is required.

All transistors are assumed to operate in weak inversion and

saturation and are modeled as:

$$I_D = I_0 \cdot \frac{W}{L} \cdot \exp\left(\frac{\varphi_G - \zeta \cdot \varphi_S}{\zeta \cdot V_T}\right) \quad \text{(NMOS)}$$

$$I_D = I_0 \cdot \frac{W}{L} \cdot \exp\left(\frac{-\varphi_G + \zeta \cdot \varphi_S}{\zeta \cdot V_T}\right) \quad \text{(PMOS)}, \quad (1)$$

with drain current I_D, specific current I_0, width W and length L, gate and source potentials φ_G and φ_S, slope-factor $\zeta \approx 1..1.4$ and thermal voltage $V_T = k_B \cdot \theta / q$ at temperature θ, with Boltzmann constant k_B and elementary charge q. The inverting amplifier in the feedback path of the logarithmic amplifier is modeled by an affine-linear relation: $\varphi_{G1} = A \cdot [V_x - \varphi_{S1}]$, with amplifier gain A and some offset term V_x. With $I_{D1} = I_{PD}$ the stationary solution is given by:

$$t \to \infty \Rightarrow \varphi_{G1} = \frac{A \cdot \zeta \cdot V_x}{\zeta + A} \cdot \left[1 + \ln\left(\frac{I_{PD}}{I_0 \cdot \frac{W_1}{L_1}}\right)\right]. \quad (2)$$

By rearranging $I_D(t)/I_D(t_0)$ one yields:

$$I_{D1}(t) = I_{D1}(t_0) \cdot \exp\left(\frac{\Delta\varphi_{G1}(t)}{\zeta \cdot V_T} - \zeta \cdot \frac{\Delta\varphi_{S1}(t)}{\zeta \cdot V_T}\right) \quad (3)$$

$$= I_{D1}(t_0) \cdot \exp\left(\frac{\zeta + A}{A \cdot \zeta \cdot V_T} \cdot \Delta\varphi_{G1}(t)\right). \quad (4)$$

Using Kirchoff's current law, one can now describe the deterministic large-signal circuit behavior through the following ordinary differential equation:

$$\frac{d\Delta\varphi_{FE}}{dt} + \frac{A \cdot I_D(t_0)}{[1 + A] \cdot C_C} \cdot \exp\left(\frac{\zeta + A}{A \cdot \zeta \cdot V_T} \cdot \Delta\varphi_{FE}(t)\right)$$
$$= \frac{A}{[1 + A] \cdot C_C} \cdot I_{PD}. \quad (5)$$

Similarly, the source follower buffer is modeled by:

$$\frac{d\Delta\varphi_{O1}}{dt} = \frac{1}{C_{LSF}} \cdot \left[I_{SF}\right.$$
$$\left. - I_{D2}(t_0) \cdot \exp\left(\frac{\zeta \cdot \Delta\varphi_{O1}(t)}{\zeta \cdot V_T} - \frac{\Delta\varphi_{FE}(t)}{\zeta \cdot V_T}\right)\right], \quad (6)$$

with bias I_{SF} and C_{LSF} being the load seen by the buffer. Note, that this derivation assumes a PMOS buffer, but the derivation for an NMOS buffer can be made analogously.

One critical scenario used for characterization is the response to a temporal contrast change (see Fig. 1b). Using separation of variables it can be shown that the ODE in Eq. 5 has an analytical solution. The input signal:

$$I_{photo}(t) = \begin{cases} I_{photo-0} & \text{for } t = t_0 \\ I_{photo-1} & \text{for } t > t_0 \end{cases} \quad (7)$$

corresponds to a linear scene-level temporal contrast step:

$$C_{scene} = \frac{I_{photo-1} - I_{photo-0}}{I_{photo-0}}. \quad (8)$$

Assuming the starting conditions: $I_{D1}(t_0) = I_{photo}(t_0)$ and $\Delta V_{FE}(t_0) = 0$, the solution to Eq. 5 is given by:

$$\Delta v = \frac{\Delta V_{FE}}{V_T \cdot \frac{A \cdot \zeta}{\zeta + A}}$$
$$= \ln\left(\frac{I_{photo-1}}{I_{photo-0}} \cdot \frac{1}{1 + C_{scene} \cdot \exp[-[t - t_0]/\tau]}\right), \quad (9)$$

with time-constant:

$$\tau = \frac{1 + A}{A} \cdot \frac{A \cdot \zeta}{\zeta + A} \cdot \frac{V_T \cdot C_C}{I_{photo-1}}. \quad (10)$$

Unfortunately, separation of variables of ODE in Eq. 6 leads to transcendental terms prohibiting an analytical solution for the response of V_{O1} to a temporal contrast step in I_{photo}. Furthermore, Eq. 5 also leads to transcendental terms if driven by other input signals than contrast steps. Numerical solutions are required. At low photocurrents, the logarithmic amplifier dominates the pixel latency and the SF buffer can be approximated by its DC gain of $1/\zeta$.

Fig. 2 shows that this model yields a significant matching improvement against circuit simulations compared to [10]. Forward-Euler integration is used to solve the ODEs.

III. SMALL-SIGNAL PIXEL NOISE MODEL

Noise is assumed to be a small-signal phenomenon. Simplified schematics of the logarithmic amplifier and source-follower and their small-signal equivalent representations are given in Fig. 3. Here, channel length modulation is neglected.

For NMOS i_D is given by:

$$i_D = g_{mG} \cdot \varphi_G + g_{mS} \cdot \varphi_S \quad (11)$$

$$g_{mG} = \frac{\partial I_D}{\partial \varphi_G} = \frac{I_D}{\zeta \cdot V_T} \quad (12)$$

$$g_{mS} = \frac{\partial I_D}{\partial \varphi_S} = -\frac{I_D}{V_T} = -g_{mG} \cdot \zeta. \quad (13)$$
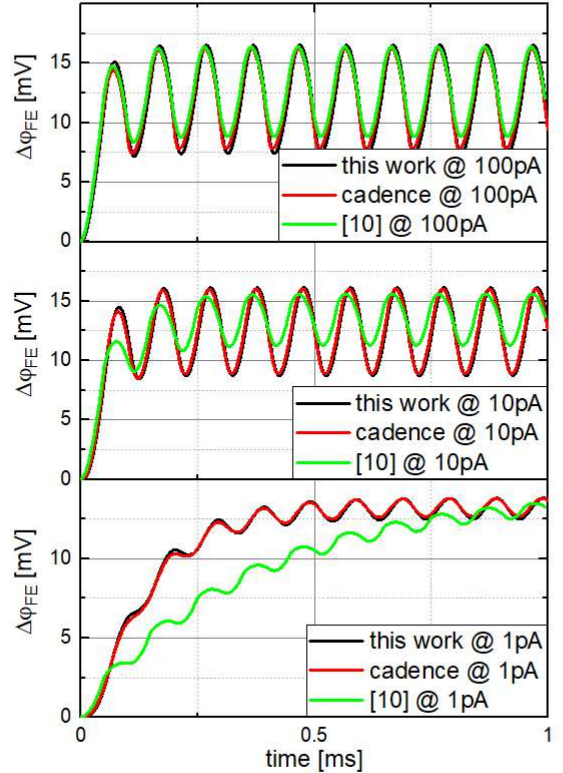
PMOS characteristics can be derived analogously.

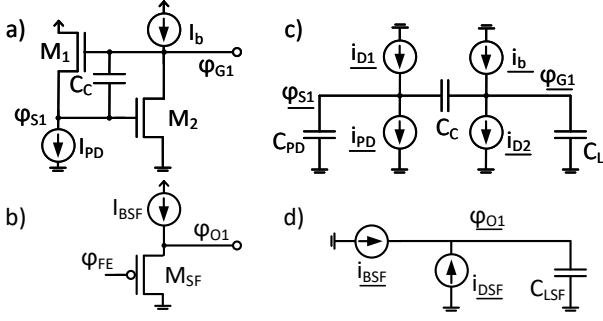Figure 2. Evaluation of the transient model.

Figure 3. Simplified log-amplifier schematic a) and SF schematic b) and their respective small signal circuits in c) and d), respectively.

It is assumed that $I_b \gg I_{photo}$ such that $g_{mG2} \gg g_{mG1}$. Using dominant pole approximation[1], φ_{G1} results in:

$$\underline{\varphi_{G1}} = \frac{i_{PD}}{g_{mG1}} \cdot \frac{1 - j\omega/z_1}{[1 + j\omega/p_{d1}] \cdot [1 + j\omega/p_{nd1}]}$$
$$+ \frac{i_b}{g_{mG2}} \cdot \frac{1 + j\omega/z_2}{[1 + j\omega/p_{d2}] \cdot [1 + j\omega/p_{nd2}]}, \quad (14)$$

with the zeros $z_1 = \frac{g_{mG2}}{C_C}$ and $z_2 = \frac{g_{mG1} \cdot \zeta}{C_C + C_{PD}}$, dominant poles $p_{d1} = p_{d2} = \frac{g_{mG1}}{C_C}$ and non-dominant poles $p_{nd1} = p_{nd2} = \frac{g_{mG2}}{C_{PD} \cdot \frac{C_C + C_{PD} + C_L}{C_C}}$. With $p_{d1} \ll p_{nd1} \ll z_1$, the first transfer function can be approximated simply by neglecting p_{nd1} and z_1 yielding a single-pole system. As $z_2 \ll p_{d2} \ll p_{nd2}$, the second transfer function first increases gain between z_2 and p_{d2} and then drops like a single pole system above p_{nd2}. Hence, we approximate $\frac{1 + j\omega/z_2}{1 + j\omega/p_{d2}} \approx \frac{p_{d2}}{z_2}$ yielding:

$$\underline{\varphi_G} \approx \frac{i_{PD}}{g_{mG1}} \cdot \frac{1}{1 + j\omega \underbrace{C_C/g_{mG1}}_{=\tau_1}}$$
$$+ \frac{i_b}{g_{mG2}} \cdot \frac{(C_C + C_{PD} + C_L)/C_C}{1 + j\omega C_{PD} \cdot \underbrace{\frac{C_C + C_L}{C_C}/g_{mG2}}_{=\tau_2}}. \quad (15)$$

Analogously, φ_{O1} can be derived to:

$$\underline{\varphi_{O1}} = \frac{1}{\zeta} \cdot \frac{\varphi_{FE}}{1 + j\omega C_{LSF}/[\zeta \cdot g_{mG}]}$$
$$+ \frac{1}{\zeta \cdot g_{mG}} \cdot \frac{i_{BSF}}{1 + j\omega \underbrace{C_{LSF}/[\zeta \cdot g_{mG}]}_{=\tau_3}}. \quad (16)$$

Shot noise is considered to be the primary noise source for a standard pixel under regular operating conditions. In contrast, flicker noise often is negligible. However, as the power spectral density of flicker noise scales inversely with transistor size and for small devices can vary by several orders of magnitude, it can contribute to noisy pixels. Especially I_b and M_2 have to be designed with care. The goal of this work is to derive a noise model for a "standard pixel." Thus flicker noise is neglected. For weak inversion operation in the saturation domain the single-sided power spectral density is $2qI$. M_1 exhibits shot noise

[1]Dominant pole approximation expresses a transfer function by dominant p_d and non-dominant p_{nd} poles: $\frac{1}{1 + j\omega\alpha + (j\omega)^2 \cdot \beta} \approx \frac{1}{[1 + j\omega/p_{pd}] \cdot [1 + j\omega/p_d]}$. With $\frac{1}{[1 + j\omega/p_{pd}] \cdot [1 + j\omega/p_d]} \approx \frac{1}{1 + j\omega/p_d + (j\omega)^2/(p_{pd} \cdot p_d)}$, $p_d = 1/\alpha$ and $p_{pd} = \alpha/\beta$ results.

of the same magnitude as the photodiode given that they carry the same current under static operation and both noise sources share the same transfer characteristic. However, they are assumed uncorrelated. Similarly, M_2 and the bias I_b share the transfer function and noise magnitude and, again, are uncorrelated. Thus, the autocorrelation functions of additive noise sources at V_{FE} and V_{O1} are:

$$R_{FE,FE}(\Delta t) = \zeta \cdot \frac{k_B \cdot \theta}{C_C} \cdot e^{-\frac{|\Delta t|}{\tau_1}}$$
$$+ \zeta \cdot \frac{k_B \cdot \theta}{\frac{C_{PD} \cdot C_C \cdot [C_C + C_L]}{[C_C + C_{PD}]^2}} \cdot e^{-\frac{|\Delta t|}{\tau_2}} \quad (17)$$

and

$$R_{O1,O1}(\Delta t) = \frac{k_B \cdot \theta}{C_{LSF}} \cdot e^{-\frac{|\Delta t|}{\tau_3}}. \quad (18)$$

IV. AUTOREGRESSIVE MODEL

Fig. 4 illustrates the signal flow used by the presented EVS model. The photodiode current $I_{PD} = I_{photo} + I_{dark}$ is propagated to V_{FE} using ODE model (Eq. 5) solved using Forward-Euler integration. Noise is added in a that matches the autocorrelation function Eq. 17. The superposition of signal and noise is then fed into the SF-buffer using Eq. 6. Now the noise contribution of the buffer is added in a way its autocorrelation function matches Eq. 18. The final result is then used to drive the difference detecting circuit which is also modeled by a simple differential equation in order to model signal dependent leak events and high-pass filter characteristic proposed in [5].
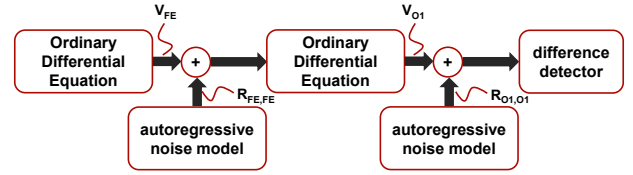

Figure 4. Pixel model signal flow diagram combining large-signal deterministic and small-signal noise behavior.

By construction, the noise models Eq. 17 and Eq. 18 describe single pole characteristics. Using Forward-Euler integration, the response X of such a system to noise e describes a first-order autoregressive model [13]:

$$X(n) = X(n-1) + [e(n) - X(n-1)] \cdot \frac{\Delta t}{\tau} \quad (19)$$
$$= \Phi_{1,1} \cdot X(n-1) + f(n), \quad (20)$$

where τ is the system time constant and Δt is the Forward-Euler time-step. $e(n) \sim \mathcal{N}(0, \sigma)$ is derived from a random number generator in order to yield a white Gaussian noise process $f(n) \sim \mathcal{N}\left(0, \sigma \cdot \frac{\Delta t}{\tau}\right)$. Choosing $E[X(0)] = 0$, it can be shown that $X(n)$ is wide-sense stationary

$$\Rightarrow E[X(n)] = 0 \quad (21)$$

$$\text{var}[X] = \Phi_{1,1}^2 \cdot E[X^2(n-1)] + \sigma^2 \cdot \frac{[\Delta t]^2}{\tau^2} \quad (22)$$

$$\Rightarrow \text{var}[X] = \sigma^2 \cdot \frac{\left[\frac{\Delta t}{\tau}\right]^2}{1 - \left[1 - \frac{\Delta t}{\tau}\right]^2}, \quad (23)$$

with $R_{X,X}(n) = \Phi_{1,1}^n \cdot \sigma_X^2$. Furthermore, we can show that the autocorrelation function of such an auto-regressive process $R_{X,X\text{-a.r.}}$ converges against the time-continuous $R_{X,X\text{-t.c.}}$ modeling white noise in first-order low-pass-filter systems:
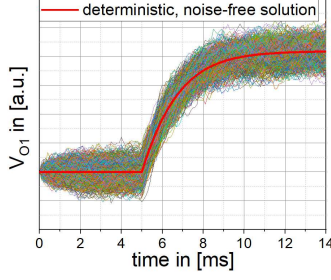
Figure 5. Sample paths of contrast step response using the deterministic ODE and the autoregressive noise model.
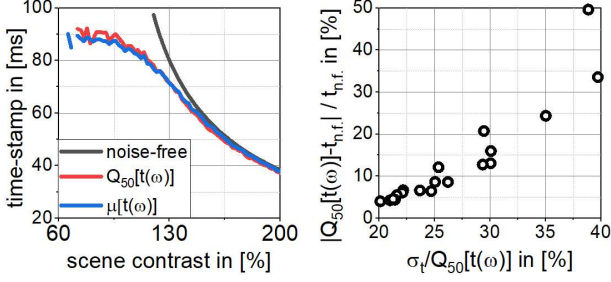

Figure 6. Comparison of noise-free time-stamp $t_{\text{n.f.}}$, median $Q_{50}[t(\omega)]$ and mean $\mu[t(\omega)]$ vs. contrast (left), and illustration of relative time-stamp variability $\sigma_t/Q_{50}[t(\omega)]$ as indicator of confidence (right).

$$R_{X,X-\text{t.c.}}(\Delta t \cdot m) = \sigma_X^2 \cdot \exp\left(-\frac{\Delta t \cdot m}{\tau}\right) \quad (24)$$

$$\approx \sigma_X^2 \cdot \left[1 - \frac{\Delta t}{\tau}\right]^m \quad (25)$$

$$= \sigma_X^2 \cdot \Phi_{1,1}^m = R_{X,X-\text{a.r.}}(m). \quad (26)$$

Rearranging Eq. 23 for σ as function of desired σ_X and τ (cf. Eq. 17, Eq. 18) and step-size Δt yields physical noise behavior of desired magnitude and bandwidth.

V. PARAMETER EXTRACTION AND VALIDATION

Fig. 5 shows Monte-Carlo sample paths as well as a noise-free solution of the ODE. Fig. 6 makes the crucial observation that the noise-free time-stamp can be approximated using the mean or median of the noise-affected time-stamp, provided the scene contrast is significant enough. This in turn allows the utilization of time-stamps from contrast step measurements for model parameter estimation using an optimization approach $\vec{p}^* = \arg\min_{\vec{p}} \left\| \vec{t_i} - t_{\text{model}}(I_{\text{photo-0}}, I_{\text{photo-1}}, \vec{p}) \right\|$. In the right part of Fig. 6, it is illustrated that the relative time-stamp variability $\sigma_t/\text{median}[t(\omega)]$ can be used as indicator of confidence in this approximation. Thus it can guide the selection of data points used for model parameter extraction. Three different paths are conceivable. Firstly, the analytical solution can be utilized for parameter optimization in a way Monte-Carlo trials can be omitted (Eq. 9). Rearranging Eq. 9 for the time-stamp at which events are triggered yields

$$t_{\text{event}} = \tau \cdot \ln\left(\frac{C_{\text{scene}} \cdot [1 + C_\infty]}{C_{\text{scene}} - C_\infty}\right), \quad (27)$$

with $C_\infty = \exp\left[\frac{V_{\text{threshold}}}{V_{\text{T}} \cdot (1+\zeta) \cdot G}\right] - 1$ and G being the difference detector gain and $V_{\text{threshold}}$ the trigger level of the comparators. t_{event} shows a dependency of τ and the characteristic threshold C_∞ that can be interpreted as contrast threshold if the pixel would exhibit
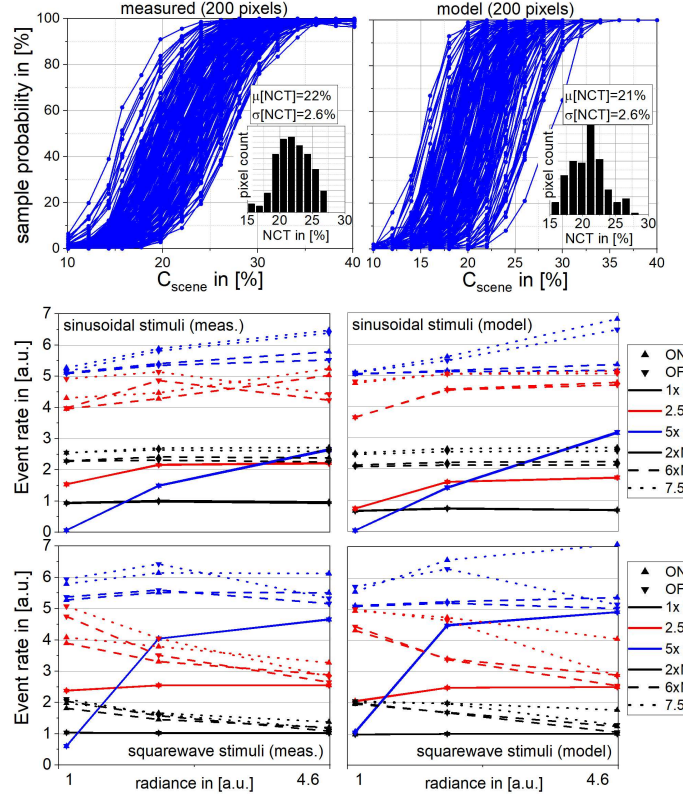

Figure 7. Comparison of measured pixel-level S-curves (top-left) vs. model (top-right), measured response to sinusoidal stimulus (middle-left) vs. model (middle-right) and measured response to squarewave stimulus (bottom-left) vs. model (bottom-right) across a 4.6x radiance, a 7.5x contrast and a 5x frequency range.

no noise. Combining a set of contrast steps at varying starting radiance yields independent measurements suitable to extract both parameters jointly. Secondly, a numerical solution can take into account the transient impact of M_{SF}. Lastly, to utilize low-contrast steps, the impact of noise needs to be considered using the autoregressive model. Note that alternatively to fitting time-stamps, one can also fit event firing probabilities. However, here, clearly noise can never be neglected. We measured devices of [5], extracted parameters, and compared measurements vs. simulations in Fig. 7 yielding good resemblance.

VI. CONCLUSION

We present a physical model for EVS pixels improving accuracy for latency and noise phenomena. We outline three paths for parameter extraction and demonstrated good model resemblance to circuit simulations as well as measurements.

REFERENCES

[1] P. Lichtensteiner et al., PhD thesis, ETH, 2006.
[2] G. Gallego et al., IEEE TPAMI, Vol. 44, No. 1, pp. 154-180, Jan. 2022.
[3] T. Finateu et al., IEEE ISSCC, pp. 112-114, Feb. 2020.
[4] Y. Suh et al., IEEE ISCAS, pp. 1-5, Oct. 2020.
[5] M. Guo et al., IEEE ISSCC, pp. 90-92, Feb. 2023.
[6] K. Kodama et al., IEEE ISSCC, pp. 92-94, Feb. 2023.
[7] H. Rebecq et al., CoRL, Oct. 2018.
[8] D. Gehrig et al., IEEE CVPR, pp. 3586-3595, June 2020.
[9] Y. Hu et al., IEEE CVPR, pp. 1312-1321, June 2021.
[10] X. Mou et al., Proc. IS&T EI, ISS-242, Jan. 2022.
[11] C. Posch et al., IEEE ISCAS, pp. 1572-1575, May 2011.
[12] B. McReynolds et al., Opt. Eng., 61(7) 074103, July 2022.
[13] K.S. Shanmugan and A.M. Breiphol, *Random Signals*, Wiley, July 1988.

Exploiting Alternating DVS Shot Noise Event Pair Statistics to Reduce Background Activity Rates

Brian McReynolds, Rui Graca, Tobi Delbruck

Sensors Group, Inst. of Neuroinformatics, UZH-ETH Zurich, Zurich, Switzerland

bmac,rpgraca,tobi@ini.uzh.ch, https://sensors.ini.uzh.ch

Abstract—Dynamic Vision Sensors (DVS) record "events" corresponding to pixel-level brightness changes, resulting in data-efficient representation of a dynamic visual scene. As DVS expand into increasingly diverse applications, non-ideal behaviors in their output under extreme sensing conditions are important to consider. Under low illumination (below $\approx$10 lux) their output begins to be dominated by shot noise events (SNEs) which increase the data output and obscure true signal. SNE rates can be controlled to some degree by tuning circuit parameters to reduce sensitivity or temporal response bandwidth at the cost of signal loss. Alternatively, an improved understanding of SNE statistics can be leveraged to develop novel techniques for minimizing uninformative sensor output. We first explain a fundamental observation about sequential pairing of opposite polarity SNEs based on pixel circuit logic and validate our theory using DVS recordings and simulations. Finally, we derive a practical result from this new understanding and demonstrate two novel biasing techniques shown to reduce SNEs by 50% and 80% respectively while still retaining sensitivity and/or temporal resolution.

Index Terms—dynamic vision sensor, event camera, DVS, noise statistics

I. INTRODUCTION

Dynamic Vision Sensors (DVS), or event cameras, efficiently encode dynamic visual information into a sparse stream of ON (increasing brightness) and OFF (decreasing) events with high temporal resolution. This sensing paradigm has several benefits including wide dynamic range, high temporal resolution, and low power consumption. DVS have already proven useful for many applications related to machine vision [1]. Despite these benefits, physical noise sources cause erroneous events even when there are no brightness changes in the scene, and elevated noise rates when illumination is low have thus far hindered widespread adoption in applications requiring high performance in dim lighting.

Under low illumination, Shot Noise Event (SNE)s dominate DVS noise [2]–[4], and denoising DVS output has been the focus of numerous efforts [5]–[7]. Although many custom denoising strategies have been developed, none explicitly consider noise event-pair statistics. Many aspects of DVS noise remain difficult to predict, but recent work has made significant progress toward understanding of the processes and trade-offs that influence SNEs [0], [8], [9].

We expand on these efforts and explain a simple yet previously unreported behavior inherent to the self-timed reset

necessary for DVS pixel operation. In Sec. II we describe the basic functionality of the DVS pixel with an emphasis on the circuit behavior that influences noise statistics. Sec. III describes the observation that SNEs tend to occur in opposite polarity (ON/OFF) pairs, and explains this behavior based on pixel reset logic. Sec. IV then demonstrates a practical result of this observation by demonstrating two sensor bias techniques that reduce SNE rates by directly manipulating noise statistics.

II. DVS PIXEL OPERATION

The first practical DVS pixel was introduced in [10], and modern event camera pixels are based on the same fundamental stages described in Fig. 1. These core components are a logarithmic transimpedance photoreceptor which generates an output voltage, V_{pr}, proportional to log photocurrent, a change amplifier that amplifies signal changes around a fixed reference point, two independent comparators for generating ON and OFF output events when the signal changes by a tunable threshold value, and a circuit to reset the change amplifier after each event to allow the pixel to respond to changes around a new reference level. In most cases, this new reference is approximately the signal level that generated the previous event. Readout circuits in the focal plane periphery record and timestamp the resulting sequence of ON and OFF events to encode pixel-level brightness changes.

Pixel behavior is refined by adjusting programmable biases (highlighted in red and depicted as current sources in Fig. 1), allowing the user to tune performance for a variety of sensing tasks. I_{pr} and I_{sf} adjust the temporal response of the photoreceptor, which is also limited by the background photocurrent. The effects of these two biases on SNE rates are extremely complex, and thoroughly described in [11]. The next set of biases define the independent ON and OFF thresholds, θ_{ON} and θ_{OFF}, which are proportional to $\log(\frac{I_{on}}{I_d})$ and $\log(\frac{I_d}{I_{off}})$ respectively. After each event, M_r shorts the input and output of the change amplifier to prevent subsequent events during a refractory period or "dead-time", and opens again as the reset node rises. I_{refr} controls the rate at which the reset node charges, and can be tuned to increase or decrease the maximum firing rate for individual pixels. Composite effects of these biases are further detailed in [9].
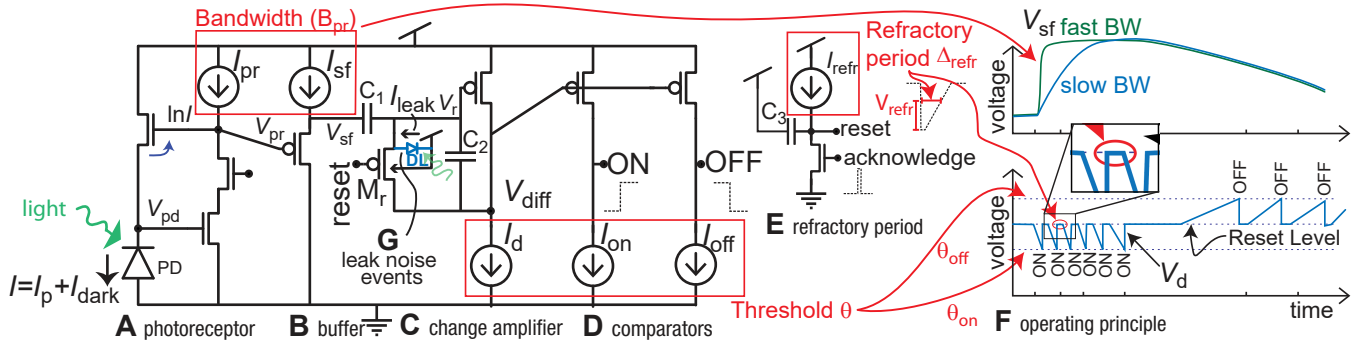
Fig. 1: Typical DVS pixel circuit schematic. The active logarithmic photoreceptor front-end (**A-B**) drives a cap-feedback change amplifier (**C**) with output V_{diff}. When V_{diff} deviates by either an ON or OFF threshold, comparators (**D**) report an event, and after a finite refractory period (**E**) the change amplifier is reset. **F:** After each reset, the pixel again responds to signal changes around a new reference level. This reset logic explains the key observation of this paper.
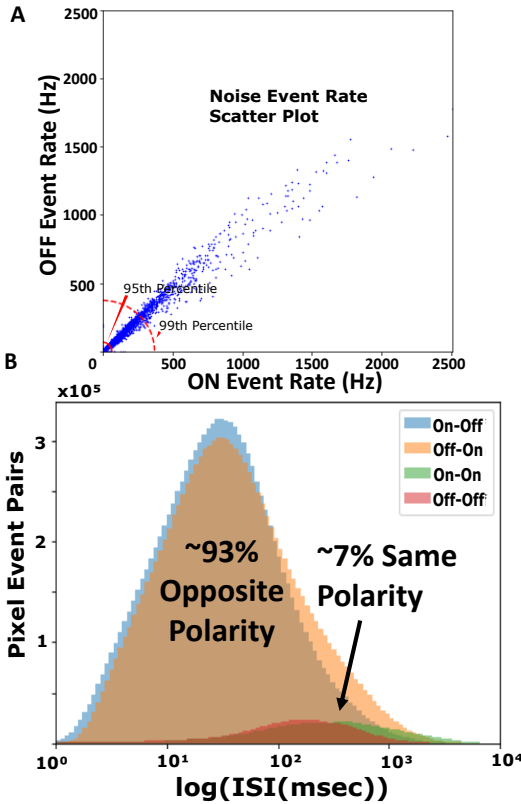

Fig. 2: Recorded DAVIS346 SNEs under 10 mlux illumination with high bandwidth biases. **A:** Per pixel ON and OFF SNE rates are nearly balanced, even for pixels with an abnormally high noise rates. **B:** Inter-Spike Interval (ISI) histograms reveal that over 90% of pixel SNE pairs are opposite polarity and occur at shorter time intervals than like polarity pairs.

III. SHOT NOISE EVENT PAIRS

To better understand the root causes of DVS SNEs, examining the scatter plot of ON and OFF noise events shown in Fig. 2 reveal ON and OFF events are nearly balanced in each pixel. At first glance, this result is counter-intuitive given the well-known mismatch in independent ON and OFF threshold levels [4], [10]. Specifically, noise rates are known to increase dramatically with sensitivity [0]. Because θ_{ON} and θ_{OFF} are independent, it is extremely unlikely that a pixel with a low θ_{ON} will also have an extremely low θ_{OFF}. In Fig. 2, the 99th percentile is depicted by the outer dashed red arc, and ON and OFF SNE rates of each type are still roughly balanced for pixels outside this region, indicating a dependency that is not explained by prior reasoning.

To further explore and illustrate this phenomenon, we calculated the ISI between consecutive event-pairs in each pixel and specifically examined the polarities of the pairs. Examining the ISI distribution in Fig. 2B reveals that over 90% of sequential noise event pairs are of opposite polarity **and** these pairs typically occur at shorter time intervals ($\approx 1/10$). Both of these observations about SNE pairs are in contrast with previous assumptions, which predict noise events should be independent of pixel history.

Fig. 3 explains how this behavior is a direct result of the pixel's self-timed reset. Events are generated when the signal deviates from a memorized reference level by more than an ON (θ_{ON}) or OFF (θ_{OFF}) threshold. Considering a filtered white gaussian noise pattern, each event resets the pixel's reference to a level offset from the mean noise value. Since gaussian noise tends to return to its mean value, this new reference increases the probability of an event of opposite polarity happening within relatively short time. This hypothesis is upheld in 4, which demonstrates how improving the v2e DVS simulator [2][1] by injecting white noise prior to the event generation block accurately models observed noise statistics.

IV. BIAS ADJUSTMENTS FOR SNE RATE REDUCTION

When operating in dim conditions, noise rates are typically managed by reducing sensitivity or photoreceptor bandwidth, but true signal is suppressed as changes too small or fast for the selected biases are missed completely. If an application

---

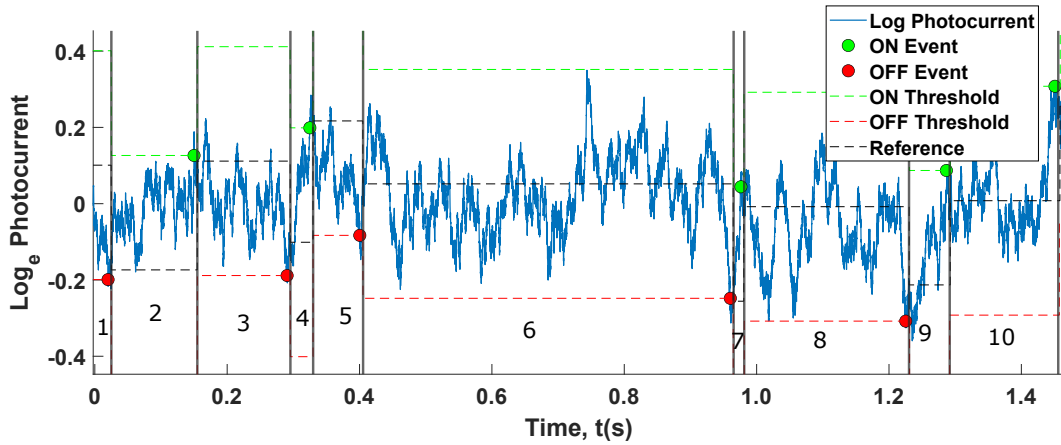[1]v2e on github - see –*photoreceptor_noise* option

Fig. 3: Explanation of noise event pairing. Each time window (labeled 1-10) terminates with an event when the noisy signal crosses either the **ON** or **OFF** threshold. Shortly after each event (dependent on refractory period), the reference level resets near the signal level that generated the previous event, increasing the probability that an opposite polarity noise event occurs. In the example, 8 of 10 event pairs are opposite polarity and occur on shorter time scales than like-polarity pairs.
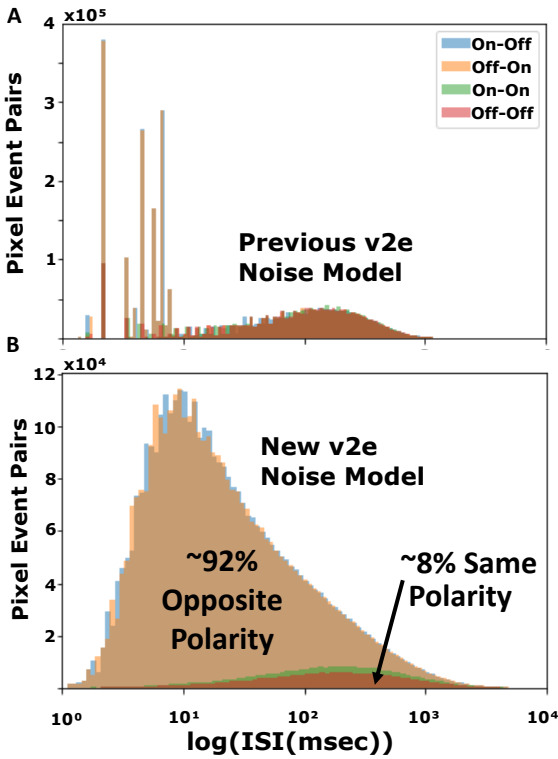

Fig. 4: Comparison of old and new v2e [2] noise models. **A:** The previous model did not accurately capture noise statistics. **B:** Adding Gaussian white noise to a DC signal and allowing the event generation model to generate noise events produces realistic DVS noise statistics.

requires detecting fast moving or dim objects/features, moderately elevated noise rates can be accepted and aggressive denoising applied after reading events off-chip at the cost of increased latency, power, computation, and data bandwidth. Alternatively, reasoning from Fig. 3 reveals two novel biasing

strategies to reduce background noise rates while still allowing pixels to be biased for high sensitivity and bandwidth.

The first strategy is to increase the refractory period. Fig. 5 demonstrates that this method decouples the reset level from the signal level that generated the previous noise event and reduces overall noise rates. Fig. 5A shows more than 50% reduction in noise rates and Fig. 5B demonstrates decoupling of ON/OFF pairs with a longer refractory period. Simulations suggest that in order for this decoupling to occur, the refractory period must be $\geq \frac{1}{2\pi f_{3dB}}$, where f_{3dB} is the low-pass corner frequency of the photoreceptor/source follower combination.

The second technique is deliberately applying a large imbalance in ON and OFF thresholds to force the reference level to settle near the extreme of the noise distribution corresponding to the more sensitive threshold. This large imbalance reduces the probability that a subsequent noise event will occur to reset the reference, thus breaking the event-pair cycle. In practice, Fig. 6 shows that this method works well when ON is much more sensitive than OFF, and Fig. 6B demonstrates up to an 80% reduction in noise event rates, even with an expected increase in sensitivity to ON changes.

V. CONCLUSION

SNE rate is an important consideration for expanding the utility of DVS into diverse applications in challenging lighting conditions. In this paper, we identified a key observation about how SNEs tend to occur in pairs of opposite polarity, and explained this phenomenon based on pixel architecture and logic. Leaning on this explanation, we propose and demonstrate two novel bias techniques for reducing SNE rates. Limiting noise rates in dim lighting conditions improves DVS Signal to Noise Ratio (SNR), and the techniques we describe facilitate direct manipulation of noise statistics. Further exploration of the benefits of these techniques should be explored in task specific scenarios. After achieving desired SNR performance, a deeper understanding of the resulting noise statistics can
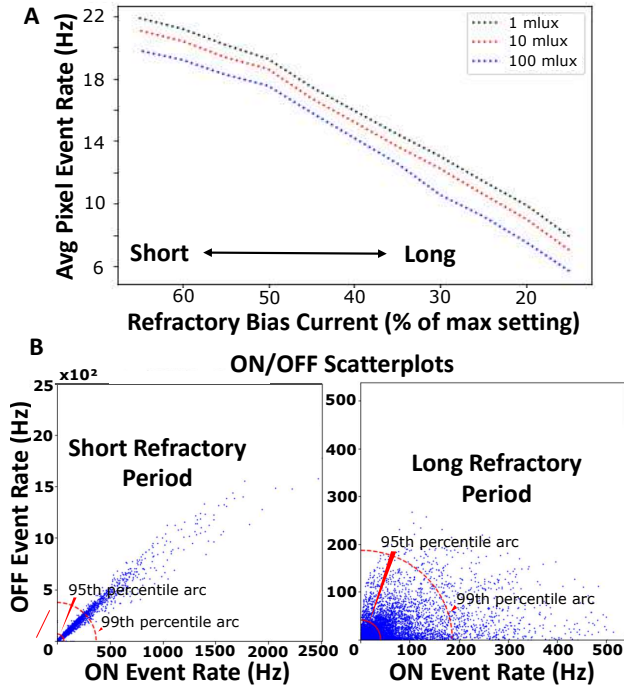
Fig. 5: **A:** Noise measurements taken from a FSI DAVIS346 [12] at three different illumination levels (1, 10 and 100 mlux). Horizontal axis is refractory period bias current ranging from 118pA (far left) to 8.7nA (far right). Increasing the refractory period reduces average noise rates by more than 50% for all illumination levels. **B:** Scatter plots of individual pixel ON and OFF noise events validate that the longer refractory period decouples ON and OFF event pairs.

aid in more efficient and effective denoising strategies, and inform improvements to already effective machine learning-based denoisers such as [6].
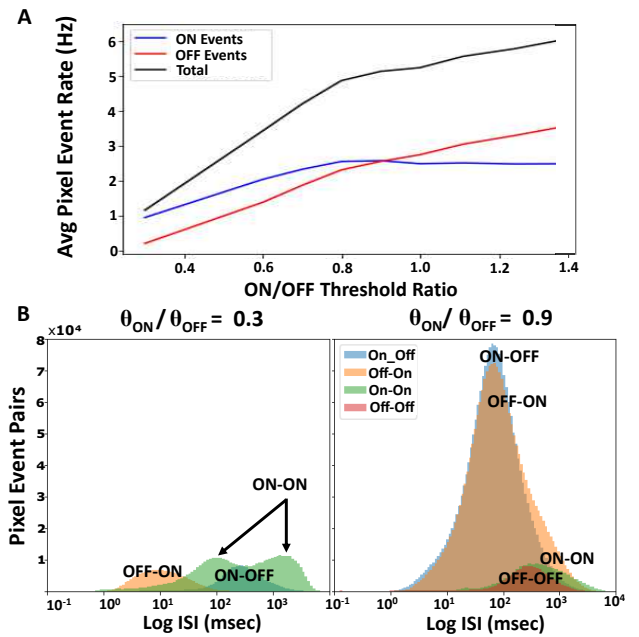

Fig. 6: **A:** Noise measurements from a FSI DAVIS346 [12] at 1 mlux for varied ratios of ON/OFF thresholds. Total SNE rates drop when the ON threshold (θ_{ON}) is lower than the OFF threshold (θ_{OFF}). **B:** ISI histograms show a shift in event-pair distributions, resulting in greater than 80% reduction in overall noise rates when $\theta_{ON}/\theta_{OFF} \approx 0.30$.

REFERENCES

[1] G. Gallego, T. Delbruck, G. Orchard, *et al.*, "Event-Based vision: A survey," en, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 1, pp. 154–180, Jul. 2020, ISSN: 0162-8828. DOI: 10.1109/TPAMI.2020.3008413.

[2] Y. Hu, S.-C. Liu, and T. Delbruck, "V2e: From video frames to realistic DVS events," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2021, pp. 1312–1321. DOI: 10.1109/CVPRW53098.2021.00144.

[3] Y. Suh, S. Choi, M. Ito, *et al.*, "A 1280×960 dynamic vision sensor with a 4.95-μm pixel pitch and motion artifact minimization," in *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, Sevilla: IEEE, Oct. 2020, pp. 1–5, ISBN: 9781728133201. DOI: 10.1109/iscas45731.2020.9180436.

[4] T. Finateu, A. Niwa, D. Matolin, *et al.*, "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," in *2020 IEEE International Solid- State Circuits Conference - (ISSCC)*, Feb. 2020, pp. 112–114. DOI: 10.1109/ISSCC19947.2020.9063149.

[5] A. Khodamoradi and R. Kastner, "O(N)-Space spatiotemporal filter for reducing noise in neuromorphic vision sensors," *IEEE Transactions on Emerging Topics in Computing*, vol. PP, no. 99, pp. 1–1, 2017, ISSN: 2168-6750. DOI: 10.1109/TETC.2017.2788865.

[6] S. Guo and T. Delbruck, "Low cost and latency event camera background activity denoising," en, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PP, Feb. 2022, ISSN: 0162-8828. DOI: 10.1109/TPAMI.2022.3152999.

[7] D. Czech and G. Orchard, "Evaluating noise filtering for event-based asynchronous change detection image sensors," in *2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, Jun. 2016, pp. 19–24. DOI: 10.1109/BIOROB.2016.7523452.

[0] R. Graca and T. Delbruck, *Unraveling the paradox of intensity-dependent DVS pixel noise*, 2021. DOI: 10.48550/ARXIV.2109.08640.

[8] T. Delbruck, R. Graca, and M. Paluch, "Feedback control of event cameras," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2021, pp. 1324–1332. DOI: 10.1109/CVPRW53098.2021.00146.

[9] R. Graca, B. McReynolds, and T. Delbruck, "Shining light on the DVS pixel: A tutorial and discussion about biasing and optimization," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2023. DOI: 10.48550/arXiv.2304.04706.

[10] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128×128 120 db 15 μs latency asynchronous temporal contrast vision sensor," *IEEE J. Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, Feb. 2008, ISSN: 0018-9200, 1558-173X. DOI: 10.1109/jssc.2007.914337.

[11] R. Graca, B. McReynolds, and T. Delbruck, "Optimal biasing and physical limits of DVS event noise," in *2023 International Image Sensor Workshop (IISW)*, May 2023. DOI: 10.48550/arXiv.2304.04019.

[12] G. Taverni, D. Paul Moeys, C. Li, *et al.*, "Front and back illuminated dynamic and active pixel vision sensors comparison," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 65, no. 5, pp. 677–681, 2018. DOI: 10.1109/TCSII.2018.2824899.