A 128 x 128 SPAD Dynamic Vision-Triggered Time of Flight Imager

A 128 x 128 SPAD Dynamic Vision-Triggered Time of Flight Imager

Francesco Mattioli Della Rocca[1,2], Hanning Mai[1], Sam W. Hutchings[1], Tarek Al Abbas[1,*], Andreas Tsiamis[1], Peter Lomax[1], Istvan Gyongy[1], Neale A. W. Dutton[2] and Robert K. Henderson[1]

1 – School of Engineering, Institute for Integrated Micro and Nano Systems, University of Edinburgh, King's Buildings, Scottish Microelectronics Centre, Alexander Crum Brown Road, Edinburgh, EH9 3FF
2 – STMicroelectronics, 1 Tanfield, Edinburgh, EH3 5DA
*now with Sense Photonics, Edinburgh, UK
Email: francesco.mattiolidellarocca@ed.ac.uk

***Abstract*— A 128 x 128 SPAD motion detection-triggered time of flight (ToF) sensor is implemented in 40nm CMOS. The sensor combines vision and ToF ranging functions to only acquire depth frames when inter-frame intensity changes are detected. The 40µm x 20µm pixel integrates two 16-bit time-gated counters to acquire ToF histograms and repurposes them to compare two vision frames without requirement for additional out-of-pixel frame memory resources. An embedded ToF and vision processor performs on-chip vision frame comparison and binary frame output compression as well as controlling the time-resolved histogram sampling. The sensor achieves a maximum 20kfps in vision modality and 500fps in motion detection-triggered ToF over a measured 2.55m range with 1.6cm accuracy. The vision function reduces the sensor power consumption by 70% over continuous ToF operation and allows the sensor to gate the ToF laser emitter to reduce the system power when no motion activity is observed.**

Keywords—Time of flight, vision, CMOS, SPAD

I. Introduction

State of the art vision cameras have accomplished solutions to high-frame rate motion detection imaging at low power consumptions [1, 2] by selectively reading out and processing frames or pixels when intensity changes are detected. Multi-modal motion detection vision cameras have been proposed to rapidly switch from a low-power reduced data readout regime for idle activity to a high resolution intensity image [3, 4]. Despite the advances of vision techniques, time of flight (ToF) cameras, both indirect (IToF) [5, 6] and direct (DToF) [7, 8], operate continuously regardless of motion activity in the scene. For applications such as IoT, security and AR/VR, combining vision motion detection and ToF promises a new paradigm in always-on depth sensing. In this paper, we present a motion detection-triggered ToF camera based on single photon avalanche diode (SPAD) technology to address this application.

Time-resolved SPAD imagers have data loads exceeding hundreds of kilobytes per frame [7, 8] to deliver single photon counting resolution at high frame rate and spatial resolution. ToF cameras have been proposed that histogram time-correlated single photon counting data at each pixel to pre-process photon time stamps [7] and in doing so reduce the output data volume for each frame. Other SPAD ToF cameras can output direct depth maps [6] or embed histogram processing to output the position of the peak [8]. Despite these compression techniques, ToF systems still read out high spatial and temporal resolution data for every frame irrespective of scene activity. This continuous operation in ToF systems results in high power consumption due to the uninterrupted triggering of the laser emitter [5], generation of high frequency time-gates in-pixel and the continuous despatch of ranging data to an external processor, the highest power contributors in a ToF system [9].

We propose a scheme where the ToF camera is triggered upon motion detection allowing a reduction in the system

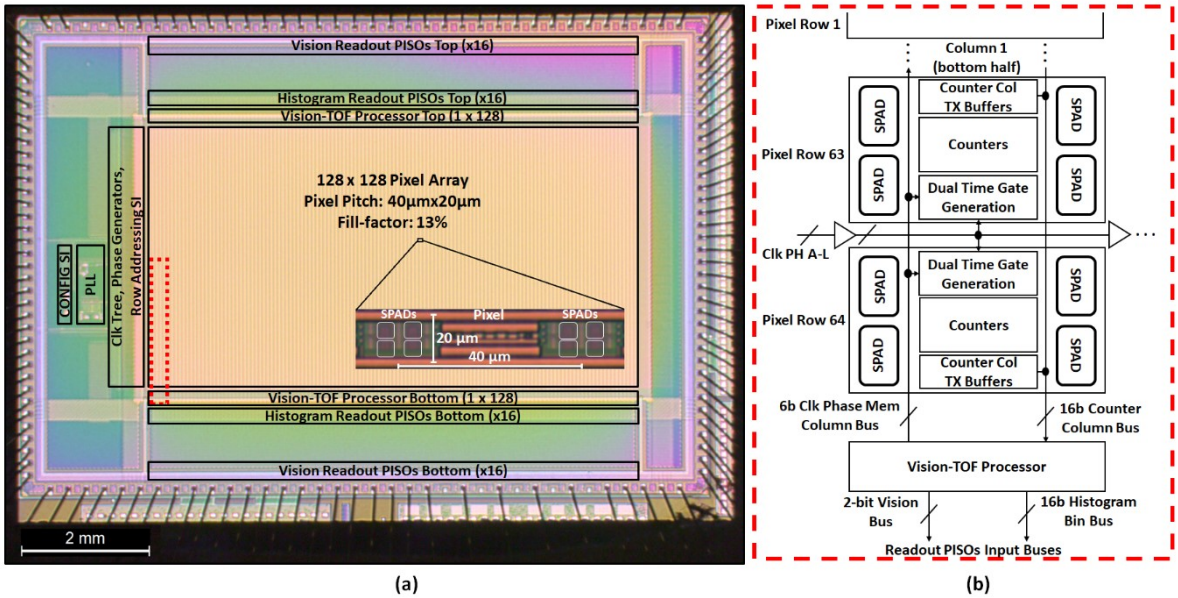

Fig 1. (a) Block diagram of the sensor overlaid on chip micrograph with pixel inset. (b) Communication between column-parallel Vision-ToF Processor and pixel half-column depicted in red dotted box on micrograph.

power by gating the laser emitter and by avoiding readout and processing of high resolution ToF frames with no motion activity. An embedded column-parallel processor performs vision frame comparison on-chip reading out binary frames signalling presence or absence of inter-frame activity. Once a vision event is registered, the camera is seamlessly switched to ToF operation and captures a time-resolved histogram from in-pixel photon sampling time gates. In this way, only ToF frames containing motion information are ever read out for processing, thus reducing the data transferred during idle operation and eliminating the power consumption contribution of the ToF laser emitter and external processor in the absence of motion activity.

II. Sensor Architecture

A block diagram of the sensor overlaid on the chip micrograph is shown in Fig. 1a. The imager is fabricated in 40nm CMOS technology optimized for SPADs [10] and comprises a 128 x 128 array of pixels. Each half-column of the imager array is mapped to a corresponding Vision-ToF processor as shown in Fig. 1b. The array can be operated in rolling or global shutter exposure. The column-parallel processors sample pixel frames on a rolling row-by-row scheme. A PLL generates a global high-frequency clock with range 500MHz-1GHz. This clock or a divided version of it (÷2/8/16) is distributed to the edge of the array via a clock tree. The high-frequency clock is then further divided into 12 edge-shifted clock phases, as shown in Fig. 2, distributed to the array horizontally, two rows of pixels sharing the same 12 clock phase lines. The 12 clock phases are used by each pixel to generate dual time-gates for time-resolved in-pixel histogram sampling of SPAD events. A configuration shift register controls the operating mode of the sensor. A total of 32 parallel input to serial output (PISO) shift registers read out either the vision or time of flight data, 8 imager columns sharing one readout serial output data line.

III. Pixel Architecture

A diagram of the pixel is shown in Fig. 2. The pixel is 40µm by 20µm in pitch with 13% fill factor. Each pixel comprises 4 passively quenched SPADs split either side of the pixel electronics. Pairs of SPADs from neighboring pixels are arranged in a column-wise well-sharing layout. The pitch of the pixel electronics matches the SPAD pitch for compatibility to stacked processes. The digital output pulses from the SPADs are shortened and combined by an OR tree into a single stream of event pulses. From the 12 clock phases arriving to the pixel, 3 clocks are selected to generate two time-gates, TG1 and TG2, each spanning two clock phase shifts. The 2 gates can be scanned across the temporal range in steps of 1 phase shift by writing to a 6-bit in-pixel phase selection memory storing the selection of the clock phase centred between the two time gates. SPAD events are quantized into either time-gate and two 15-bit counters count the events occurring within the respective time-gate. An additional bit for each counter locks the SPAD sampling to avoid counter rollover. The outputs of the SPAD circuits can be selectively masked to avoid sampling events from high dark count SPADs. While most frame-comparison vision cameras require an out-of-pixel frame memory to store the previous frame, in this sensor the time-gated counters for ToF operation are repurposed in the vision modality to store both intensity frames for direct processing by the embedded column-parallel processors.

IV. Vision-Driven ToF

The Vision-ToF processor is a digitally-synthesized and automatically place-and-routed logic block occupying an area of 40µm by 80µm matching the horizontal pitch of the pixel columns. The logic integrated in the processor is shown in Fig. 3 and the flow diagram in Fig. 4 explains the vision-triggered ToF operation.

In vision mode the pixel toggles sampling of frame exposure in either one of the two counters. The two counter values are read into the processor sequentially by row at the end of every exposure into two 16-bit registers. The vision processor on-chip calculates the threshold at each pixel for a valid vision event to be a fraction $1/2^n$ of the previous frame pixel photon count where n is a programmable global integer coefficient. The vision threshold at each pixel therefore changes dynamically and is proportional to the photon activity at each pixel [1]. The difference between the two counters Δ is compared against the threshold and the processor outputs one of three possible logic states: *0* to indicate that Δ has not

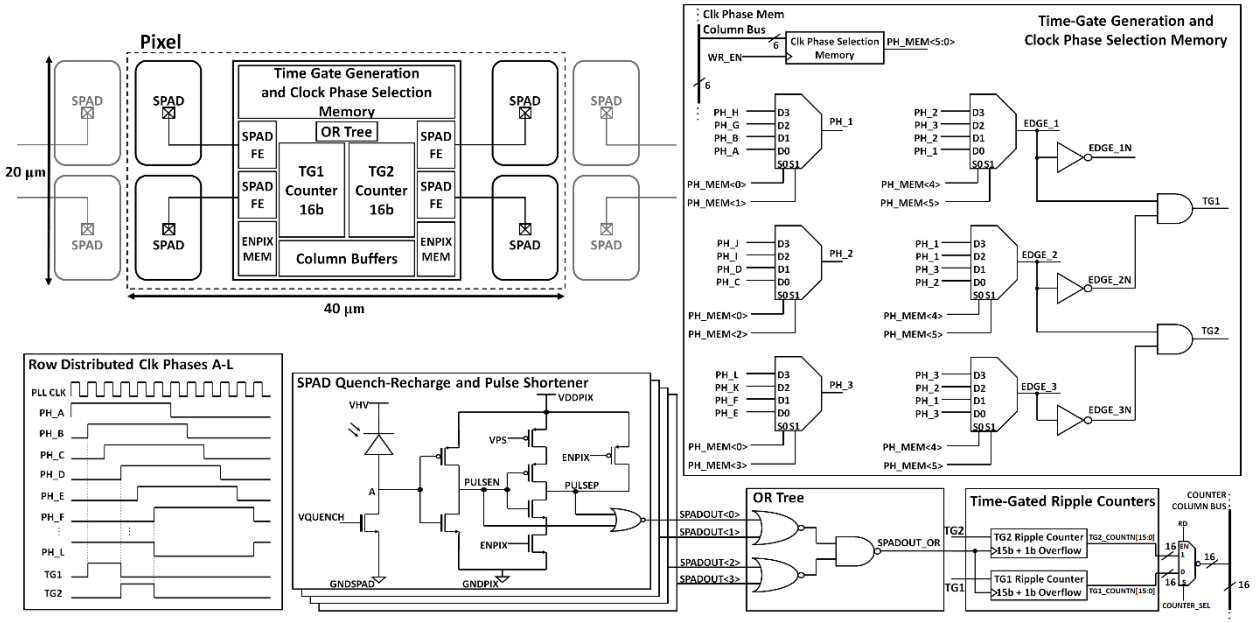

Fig. 2. Pixel block diagram.

reached the threshold, *1* for the case that Δ > *threshold,* corresponding to the current frame detecting an increased intensity from the previous frame, and *2* if Δ < *-threshold,* corresponding to a decreased intensity. The vision output of each pixel is encoded in a 2-bit number which is read out through the 32 output serial lines at 50 MHz clock rate corresponding to a maximum frame rate of 20kfps.

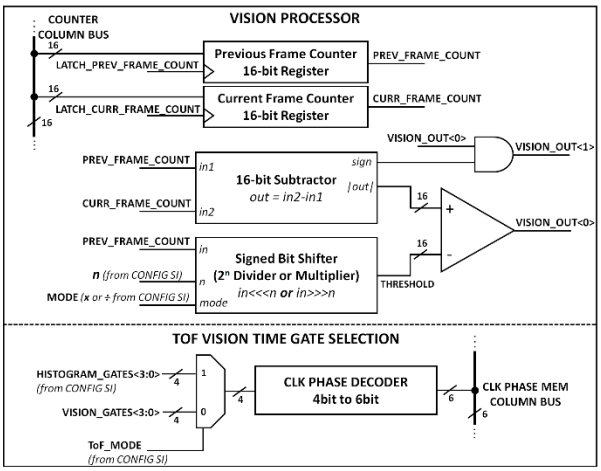

Fig. 3. Vision-ToF processor diagram of circuits integrated in column-parallel block.

Vision frames are read out to an FPGA programmed to switch the sensor to ToF mode upon detecting a programmable number or spatial pattern of vision events in each frame. Upon such condition, a single trigger configures the chip in ToF mode and the sensor performs a sequential scan of the pixel bins reading out 3 frames for a total of 6 histogram bin photon counts. For each frame, pixels are configured by the Vision-ToF processor to globally shift TG1 and TG2 by two clock phase edges thus covering the full temporal range of the histogram in 3 frame readouts. Alternatively, the sensor can be configured for dual-bin IToF ranging by generating TG1 and TG2 to cover the entire temporal range and reading out the 2 ToF bins in a single frame. After the ToF frame has completed the sensor is triggered to revert to scanning the scene in vision mode until the next vision event.

The sensor controls the laser trigger for ToF operation. The laser trigger is masked during vision operation thus saving power on the pulsed laser emitter. In vision mode the PLL clock is switched to its lowest frequency and divider setting, outputting a 30MHz clock to reduce the power consumed in the distribution of the high frequency phase clocks for in-pixel gate generation, only necessary for ToF operation.

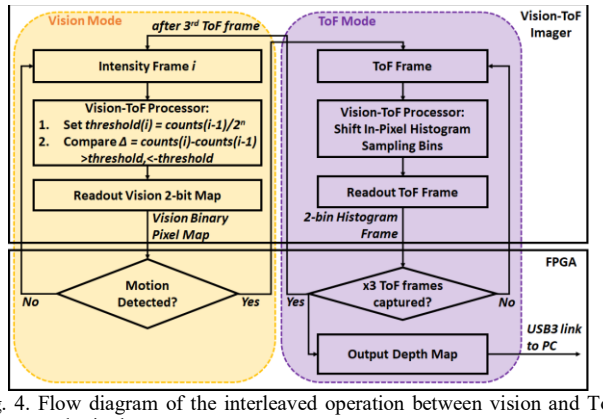

Fig. 4. Flow diagram of the interleaved operation between vision and ToF sensor modes in the sensor.

V. Experimental Results

The vision and ToF performances of the camera are presented in the following section. The profiles of the ToF time-gates of the sensor were characterized by delaying a Hamamatsu picosecond laser (PLP-10, 483nm, 70ps FWHM) across the timing range. The timing profile of the time-gates is shown in Fig. 5. The six gates were generated with 4ns FWHM with a worst case DNL averaged across the pixel array of 80ps equivalent to 2% of the time gate width.

The ranging capability of the sensor was evaluated by measuring the distance of a physical target using a Picoquant pulsed laser (840nm, 6ns FWHM). The measured distance and error are shown in Fig. 6. The sensor achieves a 1.6cm rms error over a 2.55m range showing an accuracy of 0.6% of the range. A depth map triggered by motion of a rotating fan in the scene is shown in Fig. 7.

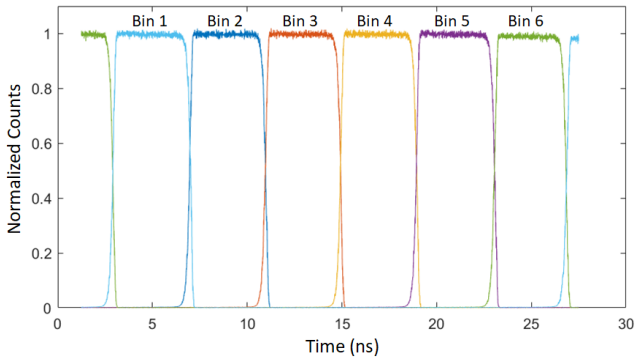

Fig. 5. Timing profile of the in-pixel generated time gates for ToF histogram bin sampling. Counts normalized to average photon count rate.

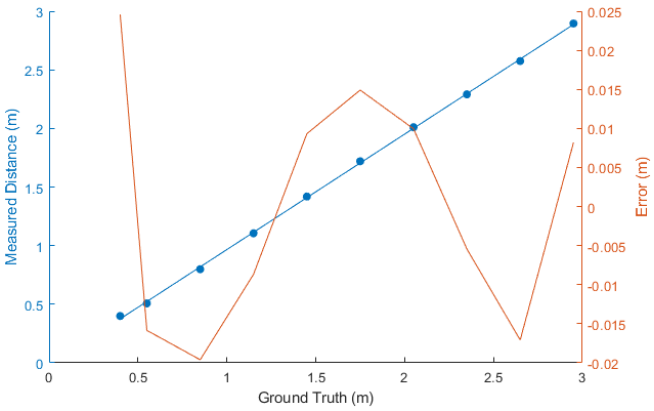

Fig. 6. Target distance measurement and error over range.

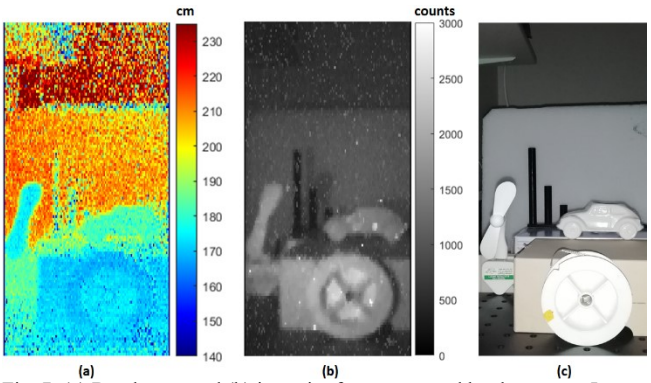

Fig. 7. (a) Depth map and (b) intensity frame captured by the sensor. Images scaled to aspect ratio of sensor. (c) Reference image captured from a standard camera.

The vision function of the sensor was tested by imaging a spinning disk made of gray-scale sections with incremental percentage darkness interspaced by white sections as shown in Fig. 8. The experiment allows verification of vision sensitivity over different intensity changes. Vision binary frames show the sensor detecting changes at each section transition for different values of threshold sensitivity to pixel intensity *n*. The spatial smearing of motion detection events across pixels observed at the section transitions is due to the difference between the rotating disk speed of 1200rpm and the sensor frame rate used in the experiment of 1kfps. With higher sensitivities smaller deviations in frame intensity including photon shot noise are captured as vision events.

Table I shows a comparison of the imager with state-of-the-art vision and ToF sensors. The sensor has a power consumption of 185mW when continuously ranging at 100fps with 1klux illumination, while it consumes 55mW when operating in vision-triggered ranging due to the reduced frequency of the distributed phase clocks and the lower output data volume in idle state. This represents a 70% sensor power saving for motion-activated ranging excluding the system power reduction from emitter gating and data processing.

TABLE I. COMPARISON OF STATE-OF-THE-ART TOF AND VISION IMAGERS

Parameter	**ToF**				
	This Work	**[5]**	**[6]**	**[7]**	**[8]**
Techn. (nm)	40	65 BSI	130	40/90 BSI	180
Detector	SPAD	QEM-PD	SPAD	SPAD	SPAD
Resolution	128x128	1024x 1024	128x96	64x64	252x144
Pixel Pitch (µm)	40 x 20	3.5	44.65	38.4	28.5
Fill Factor (%)	13	~100	3.17	51	28
Frame Rate (fps)	500	30	20	760	30
Motion Detection	Yes	No	No	No	No
Power Cons. (mW)*	185	650	40	77.6	2540
Range (m)	0.15-1.65 (min) 4.8-52.8 (max)	0.4-4.8	0-45	0-50	2-50
ToF Accuracy (% range)	0.6	0.2	6	0.34	0.17
Parameter	**Vision**				
	This Work	**[4]**	**[3]**	**[2]**	**[1]**
Techn. (nm)	40	65	90/40 BSI	350	350
Detector	SPAD	CIS	CIS	CIS	CIS
Resolution	128x128	32x20	160x154	64x64	128x128
Pixel Pitch (µm)	40 x 20	1.5	1.5	26	40
Fill Factor (%)	13	-	-	25	8.1
Frame Rate (fps)	20000	170	10	8000	10000
Dynamic Range (dB)	90	64.3	96	100	120
Ranging	Yes	No	No	No	No
Power Cons. (mW)*	55	4.5	1.1	0.03	30

* Power consumption measured at 1klux illumination and 100fps.

VI. CONCLUSION

Embedded on-chip processing integrated in advanced nanometer technologies allows the integration of multiple imaging modes on the same sensor. By interleaving vision and ToF acquisitions the presented camera achieves motion-triggered depth imaging thus opening opportunities for applications of ToF image sensors in low-power IoT systems.

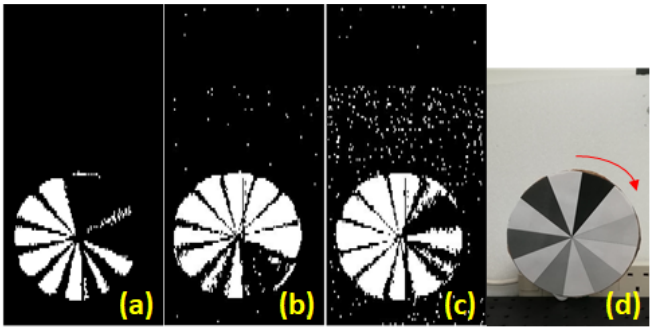

Fig. 8. Vision binary frames of rotating disk at different threshold sensitivities: (a) low sensitivity (*n=1*), (b) medium sensitivity (*n=2*), (c) high sensitivity (*n=3*). (d) Reference image captured from a standard camera.

ACKNOWLEDGMENTS

The authors thank STMicroelectronics for the fabrication of the integrated circuit and the POLIS project and BBSRC project (BB/R004226/1) for funding.

REFERENCES

[1] P. Lichtsteiner, C. Posch and T. Delbruck, "A 128 x 128 120db 30mW asynchronous vision sensor that responds to relative intensity change," *2006 IEEE International Solid State Circuits Conference - Digest of Technical Papers*, San Francisco, CA, 2006, pp. 2060-2069.

[2] N. Massari, M. De Nicola and M. Gottardi, "A 30µW 100dB contrast vision sensor with sync-async readout and data compression," *2010 Proceedings of ESSCIRC*, Seville, 2010, pp. 138-141.

[3] O. Kumagai *et al.*, "A 1/4-inch 3.9Mpixel low-power event-driven back-illuminated stacked CMOS image sensor," *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, San Francisco, CA, 2018, pp. 86-88.

[4] K. D. Choo *et al.*, "5.2 Energy-efficient low-noise CMOS image sensor with capacitor array-assisted charge-injection SAR ADC for motion-triggered low-power IoT applications," *2019 IEEE International Solid-State Circuits Conference - (ISSCC)*, San Francisco, CA, USA, 2019, pp. 96-98.

[5] C. S. Bamji *et al.*, "1Mpixel 65nm BSI 320MHz demodulated TOF Image sensor with 3µm global shutter pixels and analog binning," *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, San Francisco, CA, 2018, pp. 94-96.

[6] R. J. Walker, J. A. Richardson and R. K. Henderson, "A 128×96 pixel event-driven phase-domain ΔΣ-based fully digital 3D camera in 0.13µm CMOS imaging technology," *2011 IEEE International Solid-State Circuits Conference*, San Francisco, CA, 2011, pp. 410-412.

[7] R. K. Henderson *et al.*, "5.7 A 256×256 40nm/90nm CMOS 3D-stacked 120dB dynamic-range reconfigurable time-resolved SPAD imager," *2019 IEEE International Solid- State Circuits Conference - (ISSCC)*, San Francisco, CA, USA, 2019, pp. 106-108.

[8] 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," in *IEEE Journal of Solid-State Circuits*.

[9] J. Noraky and V. Sze, "Low power depth estimation for time-of-flight imaging," *2017 IEEE International Conference on Image Processing (ICIP)*, Beijing, 2017, pp. 2114-2118.

[10] S. Pellegrini *et al.*, "Industrialised SPAD in 40 nm technology," *2017 IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, 2017, pp. 16.5.1-16.5.4.