Direct Time of Flight Single Photon Imaging

Direct Time of Flight Single Photon Imaging

Istvan Gyongy, Neale A.W. Dutton, *Member, IEEE* and Robert. K. Henderson, *Fellow, IEEE*

***Abstract*—This paper provides a tutorial introduction to the direct Time of Flight (dToF) signal chain and typical artifacts introduced due to detector and processing electronic limitations. We outline the memory requirements of embedded histograms related to desired precision and detectability which are often the limiting factor in the array resolution. A survey of integrated CMOS dToF arrays is provided highlighting future prospects to further scaling through process optimization or smart embedded processing.**

***Index Terms*—direct time-of-flight (dTOF), light detection and ranging (LiDAR), single photon avalanche diode (SPAD), silicon photo multiplier (SiPM), CMOS Image Sensor (CIS), SPAD array, 3D ranging**

I. Introduction

Solid-state Time-of-Flight (ToF) sensors provide compact, accurate and low-cost solutions for three-dimensional (3D) imaging applications in the consumer, automotive and industrial fields. Such systems extract distance by estimating the time that modulated or pulsed light takes to travel from an emitter to a target and back again to a time-resolved optical receiver. Compared to other contactless optical distance measurement techniques such as triangulation [1], pattern projection [2] and stereoscopic [3], ToF imaging has leveraged advances in CMOS technology scaling and custom fast photodetector arrays specifically designed for depth capture.

Fig. 1 shows various schemes by which ToF sensors integrate photons reflected from a target into a number of synchronous time bins C_i with time resolution T_{bin}. Indirect (iToF) systems emit 50% duty cycle square or sinusoidal light and employ homodyne photo-demodulator pixel structures to extract the phase offset which is used to calculate distance. Typically, only a few bins (2, 3 or 4) sample the optical waveform allowing small pixel pitches to be attained in analogue implementations. iToF sensors have been integrated in miniaturized modules and have capabilities to operate at fast frame rate, high resolution and wide field of view at modest power levels. These features have enabled mass consumer deployment for short range applications such as mobile Simultaneous Localization and Mapping (SLAM), virtual reality/augmented reality (VR/AR), computer games, gesture control and robotics [4]. iToF involves an inherent compromise between detection range and precision due to the direct and inverse proportionality respectively of those quantities to the illuminator modulation frequency. iToF sensors also have limited ability to distinguish multi-path

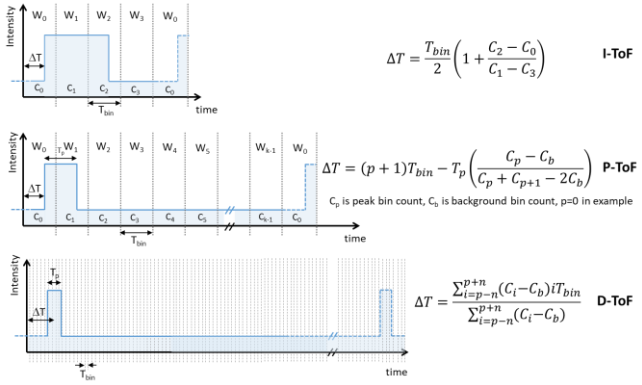


Fig. 1 The progression from indirect to direct time of flight with distance estimation formulae (a) indirect time of flight with 50% duty cycle square wave (b) short-pulsed time of flight with pulse length approaching integration window duration (c) direct time of flight with low duty cycle pulse of a few bin duration using center of mass time estimation.

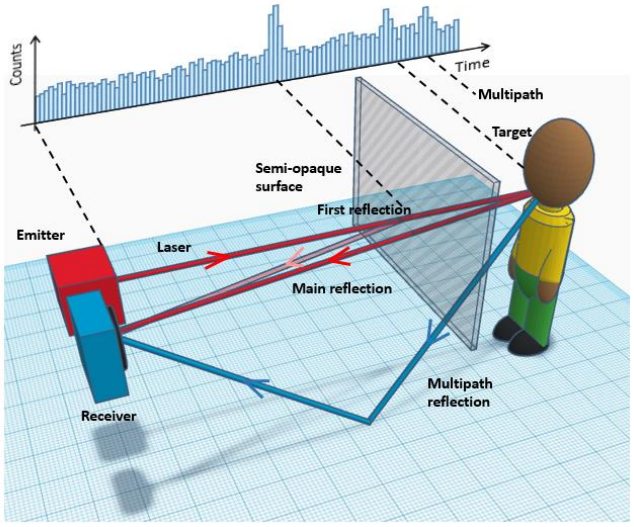


Fig. 2 Illustration of the key advantages of dToF; fine histogram bins allow to distinguish objects behind semi-opaque surfaces and to separate multipath reflections from the main target reflection.

I. Gyongy and R. K. Henderson are with the University of Edinburgh, Edinburgh EH9 3JL, U.K. (e-mail: istvan.gyongy@ed.ac.uk; robert.henderson@ed.ac.uk). I. Gyongy is funded by a Engineering and Physical Sciences Research Council (EPSRC) UKRI fellowship (EP/S001638/1)

N. A. W. Dutton is with the Imaging Division, STMicroelectronics, Edinburgh EH3 5DA, U.K. (e-mail: neale.dutton@st.com).

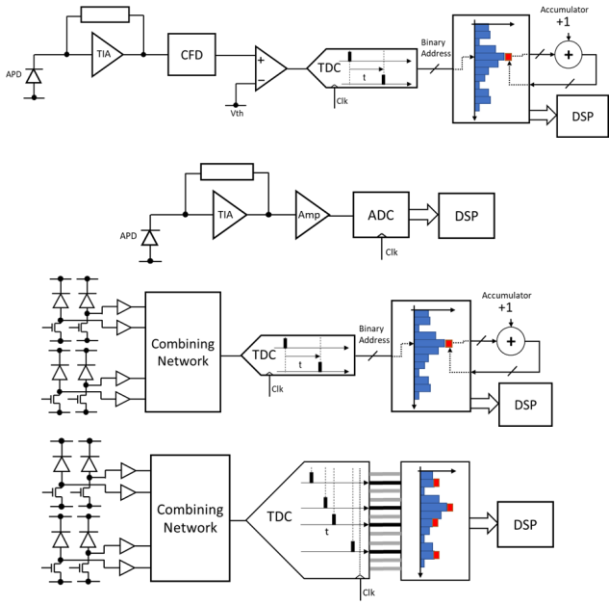


Fig. 3 APD and SPAD direct time of flight implementations (a) event based APD (b) continuous sampling APD (c) event based SPAD (d) continuous sampling SPAD.

reflections or two close objects within the same pixel field of view [5]. Multi-frequency or signal processing approaches have been applied to address these problems but increase system complexity or require serial acquisition of sub-frames introducing motion artifacts. iToF sensors are generally limited in range to tens of meters.

Pulsed (pToF) sensors have been proposed to address the range and range resolution tradeoff of iToF by operating multi-tap versions of photo-demodulator pixels with short-pulse modulation in combination with windowed sub-ranges to disambiguate the target distance [6]. Distance calculation is based on extrapolating the proportion of the total pulse energy falling within a time window. This affords the ability to avoid phase wrapping ambiguity and provides ambient background tolerance. However, the number of taps of pToF sensors is practically limited to around 8 by fill-factor constraints imposed by multiple photo-storage sites competing for photo-sensitive pixel area. High speed burst-mode CCDs and CMOS image sensors integrate hundreds of taps with commensurately low fill-factor but are unable to integrate on-chip over successive illumination cycles. [7].

Direct (dToF) exploits fast on-chip timing electronics in conjunction with avalanche detectors to measure the round-trip time of low duty cycle laser pulses. It enhances the distance precision by integration of the detection timestamps over multiple laser cycles in a histogram memory. The histogram usually requires a large number of bins set by the maximum round trip time divided by the time resolution. dToF provides a simple discrimination of multipath echoes by suitable interpretation of the multiple peaks within the ToF histogram [8-9] (Fig. 2).

dToF methods have their origins in laser rangefinders (LiDARs) employing linear detection methods based typically on an avalanche photo diode (APD) or PIN detectors [10]. More recently Single Photon Avalanche Diodes (SPADs) have been favored due to their high sensitivity, fast reaction time, low timing jitter and improving CMOS realizations. They are employed in association with statistical photon techniques such as Time Correlated Single Photon Counting (TCSPC). Initially these systems have employed board-level instrumentation from fundamental physics and life sciences. Examples applications are fluorescence lifetime imaging [11], earth mapping [12], time resolved Raman spectroscopy [13], and spacecraft navigation and landing [14]. The emergence of arrays of low cost single photon detectors allied to VCSEL laser arrays have allowed LIDAR systems to transition from scientific or military applications to mass market consumer imaging. SPAD-based dToF sensors are now embedded in mobile phones offering auto-focus assist function and applying multipath discrimination to discard early returns from the cover glass under which they must be embedded for cosmetic reasons. These devices have transitioned to low resolution few meter dToF imaging arrays in e.g. ST VL53L1X, a 16 × 16 array or the 24 × 24 pixel Apple LiDAR with recent announcements of higher resolution imaging arrays [15-17].

The longer measurement range of dToF (limited by optical power budget) and array implementation for faster frame rate has led widespread adoption in the burgeoning automotive LiDAR field for autonomous vehicles (AVs) and advanced driver assistance systems (ADAS). SPAD based dToF is now embedded in a variety of automotive LiDAR prototypes with numerous approaches to light projection and scanning [18-19].

Our aim in this article is to review the challenges to *array format* dToF imaging due to the large area and high power consumption of the time-to-digital converters and histogram memories as well as the associated high data rates. We hold that dToF is uniquely placed to exploit Moore's law scaling trends in digital CMOS processes as well as recent progress in 3D stacked CIS technologies. These advances provide prospects for smaller pitch, higher pixel count SPAD arrays, more compact histogram memories, faster photon timestamping and more complex processing electronics. The architecture of efficient dToF imaging systems poses one of today's most demanding but rewarding problems to semiconductor process engineering, microelectronic design, optical systems engineering and digital signal processing. Our paper is structured as follows; Section II gives background on dToF circuit architectures, histogram artifacts and precision and detectability, Section III looks at CMOS technology implementations of dToF receivers, Section IV surveys published literature and state of the art on dToF published research and accessible commercial literature.

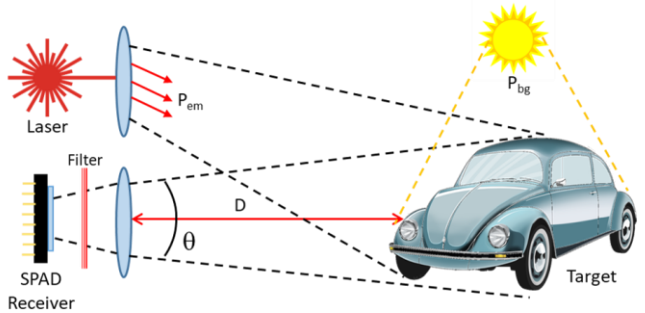
II. DToF Background

A. dToF Implementations

Fig. 3 shows the typical implementations of both linear-mode avalanche and Geiger-mode avalanche receivers. The front-ends of the detectors are quite different, linear mode detectors need fast and sensitive front-end amplifiers to amplify the APD current pulses. These are often accompanied by subsequent pulse shaping circuits such as Constant Fraction Discriminators (CFD) to avoid *walk error* due to widely changing return signal amplitude related to the inverse-square law [20]. Geiger-mode avalanche diodes are more commonly arranged as detector arrays digitally combined to provide increased sensitivity and more efficient utilization of the TDC.

Two general approaches can be applied to dToF histogram capture, applicable to either mode of avalanche detection: event-driven and continuous sampling. In the event-driven approach (Fig. 3a and Fig. 3c), a time to digital converter (TDC) is triggered on each avalanche event creating a time-stamp which is then used to index and increment a histogram location in a memory. TDCs readily provide fine timing resolution (10's picoseconds) at the cost of high power consumption and so must be activated sparingly or shared amongst detectors in array implementations. In higher photon fluxes the dead time of the conversion process leads to an effect called *pile-up* whereby only the first arriving photon can be captured and later photons are missed. The pile-up effect results in distorted histograms and failure to detect weak signals at longer ranges. Approaches such as time gating [21] or time offset time reference [22] can be taken to alleviate these effects at the cost of optical power.

The second architecture (Fig. 3b and Fig. 3d) uses continuous sampling to provide greater robustness to high background rates. For linear mode avalanche detectors this involves a multi-bit analog-to-digital converter (ADC) which must sample at 100's MHz or GHz rates to capture the nanosecond laser pulses. Geiger mode detectors require multi-hit TDC architectures based on parallel sampling of event sequences passed through delay lines or shift registers. The multi hit sequences must then be applied to increment histogram memories before subsequent laser cycles can proceed which places a high-speed requirement on the histogram generation memory unit. Aggregation over many laser cycles improves distance precision and extends range. Continuous sampling comes at a cost of high continuous receiver power consumption due to the necessity of high frequency global clock and data distribution and continuous memory operation. This places limitations on array sizes and temporal resolution compared to event-driven approaches.



$$\phi_{pixel} = \frac{\lambda A_{pixel} FF \rho \tau_{opt}}{4hcF_{\#}^2}\left(\frac{P_{em}}{\left(2D\tan\frac{\theta}{2}\right)^2} + P_{bg}\right) \quad (1)$$

Fig. 4 Photon rate at a SPAD pixel from the LIDAR equation

Fig. 4 shows the LiDAR equation Eqn.(1) which allows a simple calculation of the pixel photon rate ϕ_{pixel} due to laser peak power P_{em} and reflected ambient power P_{bg} assuming Lambertian scattering from the target. The target distance D, angular field of view θ, pixel area A_{pixel}, pixel fill-factor FF, target reflectivity ρ, optical efficiency t_{opt}, lens f-number $F_{\#}$, laser wavelength λ, Planck's constant h, speed of light c [23].

Single photon dToF systems conventionally operate with pulsed narrow linewidth lasers and are shielded behind bandpass optical filters (typically a few tens of nanometers). These optical filters are critical to prevent saturation of the SPAD detectors whose dead time limits the maximum photon flux ϕ_{pixel} to 10's of mega counts per second before paralysis occurs. dToF system parameters must be chosen such that the maximum solar flux (typically scaled to 100kLux) passing to the pixel as ϕ_{pixel} does not exceed the paralysis rate. The background rate can be estimated from the LIDAR equation referring to ASTM G-173 solar irradiance charts [24]. Typical wavelengths for silicon detectors are selected to fall in solar notches (850nm, 905nm or 940nm) related to atmospheric water absorption bands or the 1300-1550nm range for InGaAs or Ge on Si SPADs.

B. Histogram Memory and Peak Identification

As the dToF histogram occupies such a significant proportion of the pixel silicon area it is useful to estimate the size of the memory A_{hist} required for a given ranging scenario. Fig. 5a shows a histogram showing a simple model of top-hat laser peak return. We assume high ambient illumination to model a Gaussian distribution of photon counts in both signal and background bins.

$$A_{hist} = \frac{A_{bit}\log_2(MN_{LRR})2D_{max}c}{a}, \quad (1)$$

N_{LRR} is the number of laser repetitions per pixel, M is the number of combined SPADs per pixel, A_{bit} is the memory area per bit, D_{max} is the maximum range, a is the TDC resolution. The maximum value of N_{LRR} depends on the exposure time available per pixel T_{pixel} (Eqn. 2), which may equal to the frame time T_{frame} in a flash system or T_{frame}/N_{pos} in the case of a scanning system where N_{pos} is the number of distinct scan positions within a frame.

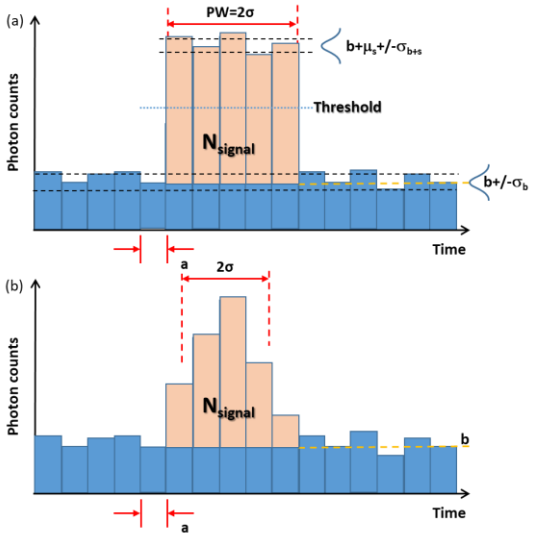


Fig. 5 Histogram peak definitions for (a) top-hat and (b) Gaussian pulse shapes

$$N_{LRR} = \frac{T_{pixel}c}{2D_{max}}, \tag{2}$$

N_{LRR} may be reduced over the frame time bounds if more laser peak power P_{em} is available such that a certain probability of detection P for minimum reflective objects at maximum range D_{max} is met. The minimum average number of signal photons per bin N_{signal} in the histogram peak can be found from Eqn (3) by applying theory from photon shot noise limited bit error rate (BER) in optical communications [25]. Assuming Gaussian distributions of the photons per bin due to background b (Fig. 5) then

$$N_{signal} = 2Q\sqrt{b} + Q^2, \tag{3}$$

where Q can be determined from Eqn.(4)

$$Q = \sqrt{2}\,\mathrm{erfc}^{-1}(2 - 2P), \tag{4}$$

The final parameter to determine histogram area is the minimum TDC resolution a. Assume that the signal peak is spread over multiple histogram bins so that sub-bin precision can be obtained in the estimate for the temporal position of the peak (Fig. 5b). Under the assumption of the signal peak (or instrument response function, IRF) having a Gaussian profile, the uncertainty, or standard deviation δ, in the estimate can be approximated by Eqn.(5), adopted from single molecule localization microscopy [9,26]:

$$\delta = \sqrt{\frac{\sigma^2 + a^2/12}{N_{signal}} + \frac{4\sqrt{\pi}\sigma^3 b}{aN_{signal}^2}}, \tag{5}$$

where σ is the standard deviation of the IRF. Eqn.(5) can be re-written as:

$$\delta = \frac{\sigma}{\sqrt{N_{signal}}}\sqrt{1 + \frac{1}{12}\left(\frac{a}{\sigma}\right)^2 + 4\sqrt{\pi}\left(\frac{\sigma}{a}\right)\frac{b}{N_{signal}}}, \tag{6}$$

The second and third terms in the above expression represent the excess noise in the peak estimate, arising due the discretization in the histogram, and background photons, respectively. For bin widths $a < \sigma$, the contribution from histogram discretisation (or TDC resolution) rapidly diminishes, which implies that there no benefit in improving the TDC resolution beyond a certain value. It is therefore proposed in literature that the bin width should fall in the range of $\sigma < a < 2\sigma$ [27].

It must be noted that the detector may at times be subject to high signal returns from close or retro-reflective targets. This can in turn distort and narrow the IRF (as explained in Section C below), thereby requiring higher temporal resolution for the signal peak to be adequately captured (i.e. with sub-bin precision) and the range walk error resulting from the distortion to be compensated for [28]. Techniques for peak extraction include iterative curve fitting [29], as well as filtering the LIDAR waveform (histogram) using a finite impulse response filter (FIR) matching the temporal profile of the anticipated signal peak [30]. It has been shown that even the computationally modest approach of local centroiding of

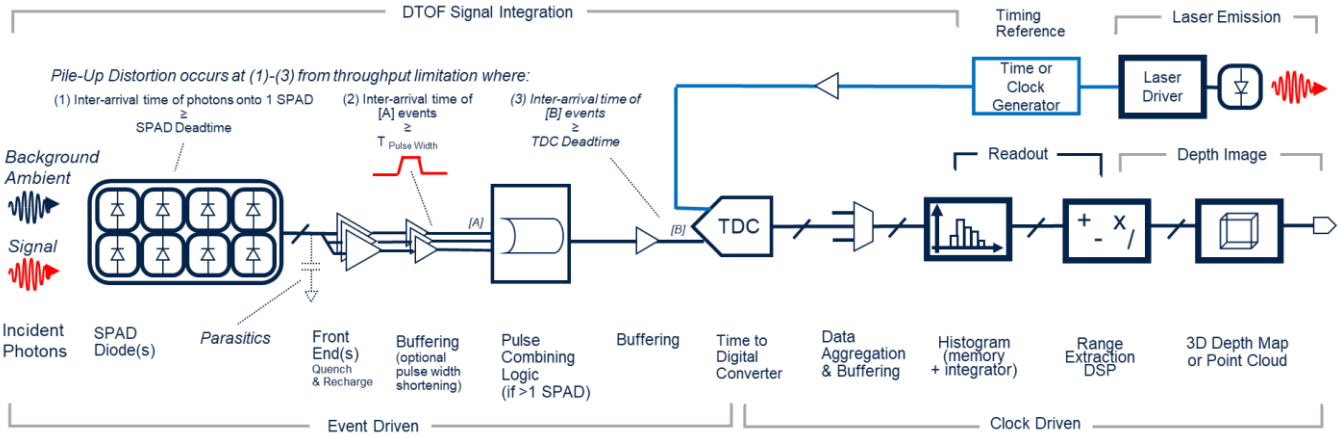


Fig. 6 Sources of pile-up distortion in a typical CMOS SPAD dToF signal chain due to timing throughput limitations.

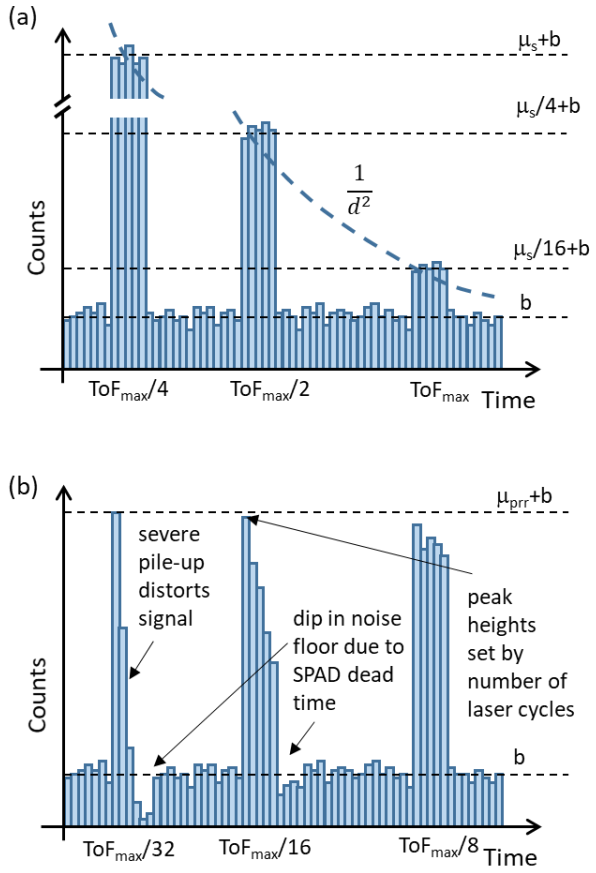


Fig. 7 (a) 1/d^2 signal amplitude at long range or fractional signal photon/bin/cycle regime (b) pile-up artifacts at short range or many signal photon/bin/cycle regime.

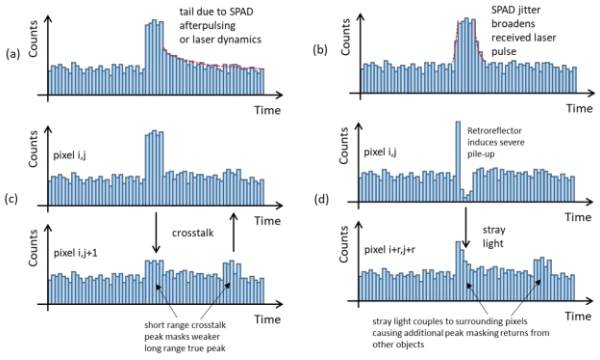


Fig. 8 Artifacts induced in dToF histograms due to SPAD or optics non-idealities (a) afterpulsing (b) jitter (c) crosstalk (d) dead time and stray light

the histogram (following background compensation) can result in a performance approaching the Cramér-Rao bounds that define the lowest possible variance for an unbiased estimator (see, e.g. [31]).

C. dToF Histogram Artifacts

Fig. 6 shows the signal chain of a typical SPAD dToF receiver highlighting areas where throughput limitations in processing photon events give rise to pile-up distortions. A number of common distortions in dToF histograms are illustrated qualitatively in Figs. 7 and 8 assuming an ideal top-hat laser pulse. When operating at the limit of detectability and minimum emitter peak power the average signal peak at ToF_{max} may operate at a signal to background ratio (*SBR*) close to unity. *SBR* is defined here as:

$$SBR = \frac{N_{signal}a}{\sigma b} \tag{7}$$

This definition is drawn from the precision Eqn (6) as the contribution of background photons to the excess noise in the peak estimate depends on the parameter $\sigma b/aN_{signal}$. Note that $\sigma b/a$ relates to the number of background photons in the histogram peak (σ/a being the normalised width of the IRF).

Fig. 7 shows the signal peak growing in height as the target approaches the LiDAR, initially following the inverse square law and thereafter showing saturation of the SPAD and eventually pile-up. In the pile-up condition induced by close or retro-reflective targets the received pulse height clips at $M \times N_{LRR}$ losing information on target reflectivity. Received histogram profiles exhibit a trailing edge that is distorted with an exponential decay. The pulse centroid deviates by up to half the pulse width representing an inaccuracy of many 10's of centimeters for typical few nanosecond laser pulses. In these cases, the leading edge of the pulse still conveys high precision distance information with a walk error related to the SPAD avalanche onset time resulting in an accuracy deviation of a few centimeters [32]. Moreover there is a trough in the probability of background photon detections subsequent to the peak due to all SPADs being simultaneously within their dead time which can mask secondary targets.

Fig. 8 shows artifacts introduced into dToF histograms by SPAD or optics non-idealities. Fig. 8a shows a tail introduced to a peak due to SPAD afterpulsing nanoseconds or due to slow diffusion-dominated carrier transport which occurs in both non-fully depleted SPADs and laser tailing dynamics which may extend over 100's of nanoseconds [33-34]. Fig. 8b shows peak broadening due to jitter with a SPAD diffusion tail on the falling edge, such effects extending the received pulse by only a few hundred picoseconds. Potentially more serious distortions are shown in Fig. 8c and Fig. 8d. Fig. 8c shows SPAD optical crosstalk causing peaks from one pixel histogram spreading into a neighboring pixel resulting in spurious detections. Fig. 8d is an example of time-domain veiling glare induced by stray light from an intense return signal from a retroreflector (e.g. street signs). The retroreflector signal exhibits strong pile-up artifacts and the stray light spreads a proportion of that return to a wide surrounding region of neighboring pixels causing peaks at similar range offsets. This veiling glare phenomena is interpreted as a halo like disk artifact in the point cloud information [35].

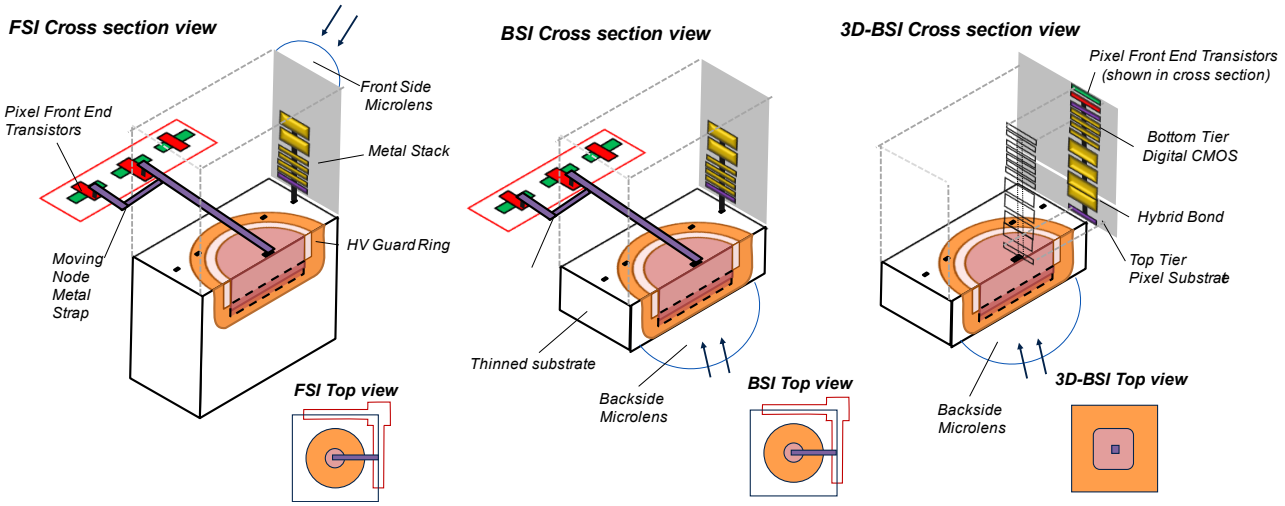


Fig. 9. dToF receiver technologies in cross section from left to right: front-side illumination (FSI), back-side illumination (BSI), 3D stacked BSI top tier with digital CMOS bottom tier.

III. dToF Receiver Trends and Analysis

A. Technology Trends

Following similar technology trends as CIS, CMOS dToF sensors in front-side illuminated (FSI) technologies exhibit few percent PDE at NIR wavelengths [36] due to thick BEOL stacks, and large pixel pitch as pixel transistors must be placed near or beside the pixel photodetector as shown in Fig. 9. The location of pixel front-end (FE) transistors and back-end (BE) metal routing is key to SPAD pixel performance as parasitic capacitance on the SPAD moving node affects multiple system parameters (afterpulsing, charge per avalanche, jitter, etc.) In addition, the matching of pixel BE routing is challenging in FSI processes. Back-side illuminated (BSI) technology with optimized optical stack improved the PDE from 600nm to 1000nm [37] but did not address the location of pixel circuits remaining physically isolated from the high voltage SPAD guard rings. Aull et al. trialed bump-bonded 3D-stacked BSI SPADs [38]. Yet, the most recent advance in CMOS technology for dToF is 3D-stacking with face to face bonding of a top-tier BSI SPAD wafer with an advanced digital CMOS wafer bottom-tier (3D-BSI) [39]. The BSI SPAD is placed directly above the pixel circuit, a recent example showing a 90nm 1ML / 45nm 11ML stack is given in [40]. There are further benefits to stacked technology for dToF: cost reduction or performance increase and independent technology development and optimization of digital CMOS and SPAD diode processes.

Fig. 10 shows the chronological trend of CMOS dToF pixel shrink indicating the larger pixel pitch of FSI versus the shrink offered by 3D-stacking technology. The pitch reduction is dictated by both SPAD diode and pixel circuit, and the overall sensor digital logic area for TDC and histogram. Future projections (made initially in 2016 [41]) and the linear trends shown in this figure indicate that denser digital nodes combined with innovations in SPAD diode shrink will fuel the dToF pixel race of commercial dToF sensors below 10 µm pitch in years to come [21,30,42-44].

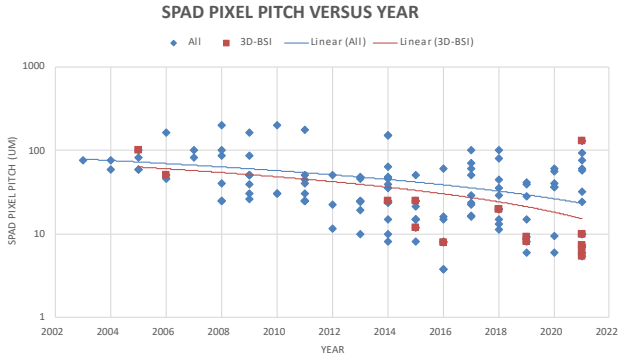


Fig. 10. Pixel pitch (µm) versus year of publication indicating the trend of pixel shrink with two linear trend lines drawn for all pixels and for only 3D-BSI pixels.

B. Pile Up Distortion in the CMOS dToF Signal Chain

The typical CMOS dToF signal chain is shown previously in Fig. 4. Distortion in the signal chain can occur throughout the event-driven section: three points where time-domain pile-up distortion occurs are indicated as (1) –(3). Signal losses in the dToF signal chain occur when the throughput rate of one component is less than or equal to its input rate. The ideal dToF signal chain has an increasing throughput rate for each component. The losses occur as missed dToF measurements where not every photon is processed; manifesting as pile-up distortions and causing errors in the computed distance measurement. The TDC conversion rate (point 3) or combining logic maximum rate (point 2) must be higher than the SPAD maximum count rate for pile-up not to occur at high event rates. In systems with low TDC conversion rate or limited data readout rate, the only option left to the user is to optically reduce the input photon rate: the rule of thumb is for the maximal photon rate to be $1/20^{th}$ to $1/10^{th}$ of the system conversion rate [11]. The SPAD maximum count rate is

Combining Logic	**Pixel Input Pulse Generator**	**Ref of First Use**	**Throughput Model of Combining Logic (Events/sec)**		**Example reference of technique in use**	**Photons / TDC Sample in example ref.**	**Calculated Dynamic Range in 1 laser pulse of 100ns in example ref.**	**Dominant Pile Up Location(s) in example ref.**	**Pile Up Condition in example ref.**
No combining logic: 1 SPAD to 1 TDC	N/A	From TCSPC methods [11]	1 / (e . SPAD dead time) * or 1 / (TDC dead time) if lower		[53]	1	0dB for single event TDC	TDC and off-chip data transfer	Photon rate > TDC sample rate
Shared Bus	NMOS pull down, and Monostable	[86]	1 / (e . monostable pulse width)		[79]	≤1	0dB for 1 SPAD <0dB for photon rate > Histogram memory rate	Shared Bus combining logic, and Shared Histogram Memories	Photon rate > Histogram memory rate
Co-indicidence (where k = number of coincident pulses)	Monostable	[47]	1 / (k. e . monostable pulse width)		[67]	2 - 7	No activity for input events < k 6-17 dB for input events ≥ k	Co-incidence detection combining logic	Incident photons < k
OR tree	Monostable	[48]	1 / (e . monostable pulse width)		[73]	1	20dB	OR Tree combining logic & TDC	Total SPAD rate > OR bandwidth or TDC rate
XOR Tree	Toggle flip flop	[49]	1 / (XOR gate delay)		[58]	1	30.4dB	XOR Tree combining logic	Total SPAD rate > XOR bandwidth
Synchronous summation technique (SST)	Clock-driven Flop	[83]	N SPADs x Clock freq.		[31]	81	71.1dB	SPADs only	Photon rate > SPAD max count rate
					[84]	100	74dB		

Table 1. Left hand side: throughput of pulse combining techniques with analytical modelling equations of each where e is Euler's constant. Right hand side: Examples of each with calculated dynamic range in 100ns laser pulse period and location of pile-up (*) Assumes passive recharge.

determined by the front end (FE) circuit. Many FE circuits are described in the literature namely passive quench/recharge and active approaches with higher count rates [45-46]. Gated front-ends serve to confine Geiger-mode operation within a time-window correlated to laser emission [21,42] which can reduce event rates outside of a temporal region of interest and so, in effect, reduce pile up distortion.

Two papers [49] and [50] describes the range of combining techniques of multiple SPADs to one TDC and here is included in Table 1 where each combinational logic method is shown and a throughput analytical model is given from those works. In addition examples of each technique are given with the photons sampled for each TDC sample. Furthermore, to allow benchmarking of the effectiveness of pile up reduction of these four techniques, a dynamic range measure of the photons sampled in a laser pulse repetition is calculated as:

$$DR_{100ns\,period} = 20.\log\left(N_{photons\_per\_pulse}\right) \quad (8)$$

Where 100ns is chosen as a fixed value to allow fair comparison and the photons per pulse is calculated for each reference based on the lowest throughput section of the signal chain.

There are numerous architectures of TDCs for dToF described in the literature [51]. They can be categorized into two main categories from the perspective of pile-up distortion, and conversely dynamic range, by the number of photons processed per laser emission cycle (laser shot): first-photon or single event TDCs can process 1 photon per laser shot [52] (an equivalent of 0dB dynamic range per laser repetition), and multiple event TDCs that can process 2 or more photons per laser shot with higher dynamic range, reduced pile-up distortion but at the cost of higher downstream data rate. The TDC-in pixel sensor architecture directly connects the buffered output of one SPAD FE to one TDC at the cost of severely constrained photon rate by data readout limitation and so sensitive to distortion above the published TDC conversion rate [53-55] so suitable only for photon-starved applications. TDCs that are shared between multiple active pixels may result in lower than single event conversion rate [56] and lowest dynamic range. In addition to this in [79], the histogram memories are shared between multiple TDCs further adding pile up distortion for SPAD event rates > histogram memory bandwidth. Whereas, multiple event TDCs have been demonstrated from 3 to 33 events per laser shot with 10GS/s conversion rate [57-58] designed to mitigate pile up distortion with 30dB dynamic range per laser pulse. Finally, the SST-TDC technique, proposed in [83], is a combination of oversampling TDC and combination logic that has shown to have the highest photon throughput with 100 photons simultaneously digitized per bin [84] and 81/bin [30] providing the best system for pile up distortion with none in the signal chain (except for the SPAD diode itself) with the highest dynamic range per shot and a distance range extension over those architectures with lower throughput [50]

C. Noise in the CMOS dTOF Signal Chain

Fig. 11 illustrates the noise sources in the CMOS dToF signal chain (corresponding to Fig. 4). On the left of the dotted line, the physical and optical noise sources are indicated which

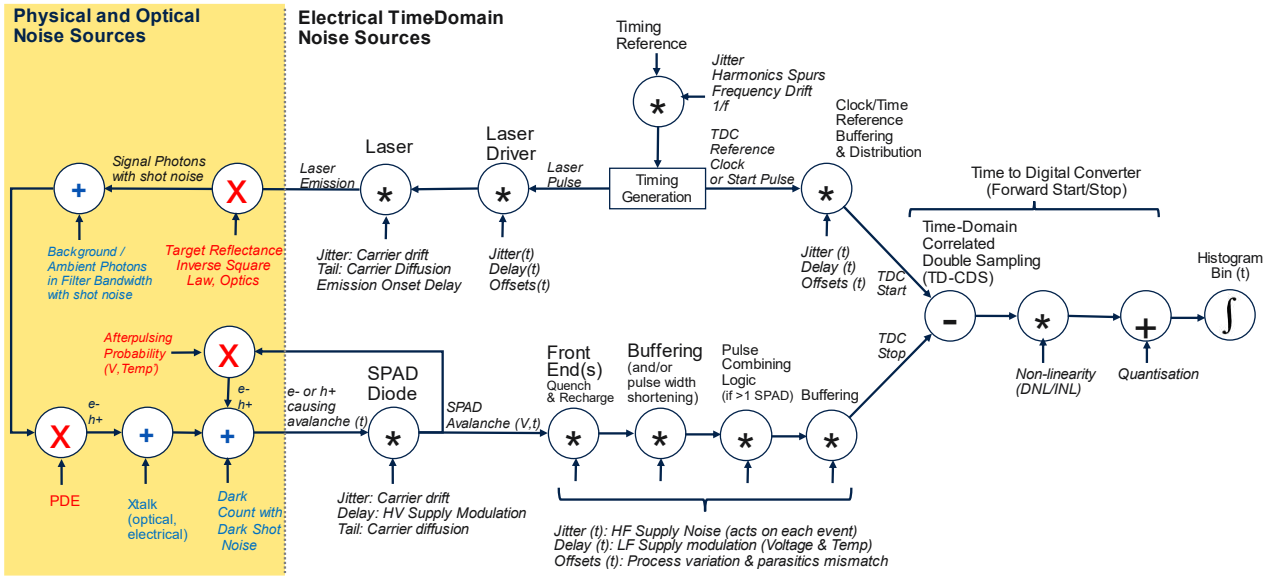


Fig. 11. Noise sources in the DTOF Signal Chain: left are physical and optical noise sources, and right are electrical time-domain noise sources.

add additional photons, cross talk or dark events not correlated to the time-correlated laser emission ToF photons. Afterpulsing is shown as a feedback loop where an afterpulsing probability of less than 1 creates a tail from the damped time-domain response. Background ambient signal may be reduced by a decrease in the optical filter bandwidth. Reduction of cross-talk, after-pulsing and DCR and increase in PDE is an on-going technology improvement activity [36][40]. On the right of the dotted line, the electrical time-domain noise sources are shown with two paths. The upper path is from the timing generation for the laser pulse and the TDC reference clock. The lower path depicts the time-domain noise sources from the event-driven section of the signal chain such as jitter, delay (proportional to temperature and voltage) and time offsets. In both upper and lower paths, the delay and offsets may be minimized by reduction of the total path length and absolute delay (this will also have a power reduction benefit). The jitter may be minimized by increasing the slew rate at all points in the signal chain to reduce noise injection at the zero-crossing point of each logic gate in the path.

The TDC performs a time-domain correlated double sampling operation (TD-CDS) and the observant reader will notice both that the paths drawn are not matched for noise subtraction and that in contrast to a CMOS image sensor, the signal integration is after the TD-CDS and data converter quantization. To alleviate the issue of non-matched noise sources and to perform a true TD-CDS measurement, the designer of the dToF system must take care to perform calibration (one-off or continual either foreground or background) and/or replica path design where a second optical feedback path is created in the sensor module and packaging with its own dToF signal chain and histogram to provide a baseline zero distance that is equally affected by voltage and temperature time-domain noise effects [85].

IV. dToF Sensor Architectures

dToF SPADs may be categorized by the level of processing carried out on the sensor. In the simplest case, the outputs of individual SPADs are read out directly, and processed externally using TDCs implemented in ASIC or FPGA [58-59]. The number of SPADs that can be used concurrently is then limited by the number of output lines available. For the case of a 2D array, there may be multiplexing logic enabling the selection of different groups of SPADs within the array. While the availability of raw SPAD data is useful for the evaluation of different forms of photon processing [50], there has been a trend to integrate an increasing level of processing into SPAD sensors, in order to develop single-chip receiver solutions, in large array format, capable of operating in a flash (scan-less) modality. Driving factors behind these developments include a desire for solid-state dToF systems with fast acquisition over a large field of view, reduced system power consumption, and increased robustness to ambient light. We can consider the following hierarchy of different levels of processing.

A. dToF SPADs with integrated photon timing

These SPADs include photon timing circuits, serving individual pixels or groups of pixels. Timing may be achieved using TDCs or, less commonly, using TACs. In the former case, the reference signal for timing is typically provided by an internal gated ring oscillator [61], a delay line [62], or a global high frequency clock [63]. The output is a digital time code representing the time of arrival of (typically) the first detected photon. In a TAC-based architecture, pixels sample a voltage ramp when they detect a photon [64-65]. The timing information is thus stored as an analogue voltage value, which

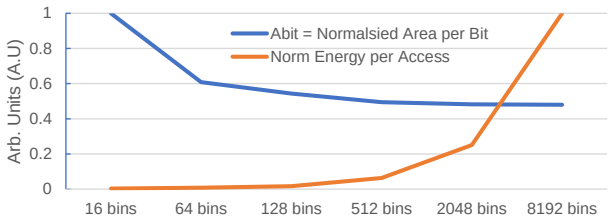


Fig. 12. Modelling study indicating the normalized area/ normalized energy tradeoff met in the design of histogram generation circuits based on serially-access single memory SRAM instances of 7b/word.

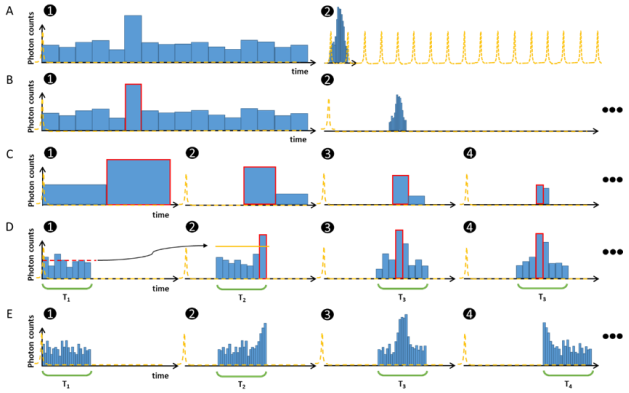


Fig. 13 On-chip histogramming approaches with reduced number of bins: (a) two-step approach a large bin size for range disambiguation is followed by a small bin size for precise peak extraction [74] (b) multi-step approach, in each step, the peak bin is identified, and the histogramming logic zooms in on the corresponding time range, through appropriate filtering of the time stamps from the TDC [79] (c) similar approach to (b), but using only 2 bins [80] (d) the histogram is shifted in time to track peaks and peak detection is based on an estimate of the background level that is updated after every time shift [80] (e) histogram is swept through the full time range [82]. The dashed, yellow lines indicate the timing of the laser pulses with respect to the histogram time range.

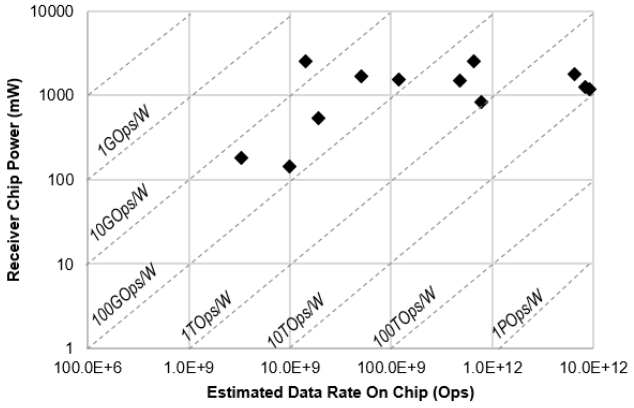


Fig. 14. Power versus data rate and indicative power efficiency of sensors included in Table 2.

~~is usually then digitized using column parallel ADCs during~~ readout.

A sub-category of SPADs with integrated TDCs feature coincidence detection [14,66-68], which requires multiple SPAD firings from a pixel or group of pixels, within a certain time window, for an event to be recorded. The purpose of this functionality is to filter out background photons and hence minimise TDC pile-up effects which could mask signal photons. The photon threshold and time window should ideally be set according to the signal and background photon rates, so the scheme requires an adaptive mechanism for optimal operation [69]. Related to coincidence detection, the scheme of [69] uses analogue summation of SPAD currents to measure activity levels across the array and implement row and frame skipping to accelerate read out.

The storage, processing and/or integration of digital time-stamp data is an active area of research in scientific and PET sensors. In burst-mode applications (e.g. PET imaging) time-stamp data can be locally stored but necessitates high memory read/write bandwidth: 100MS/s [48] to ~400MS/s [71-72].

B. dToF SPADs with on-chip histogramming

In SoC physical implementation, close physical placement of interconnected components and minimized wire-lengths is the primary pathway to minimized energy per operation and maximal operating frequency. The same principle applies in the physical implementation of the dToF signal chain with interconnection from pixel to TDC to histogram, and the histogram generation circuit architecture. Therefore, there has been a trend in dToF SPADs not just to time photons on-chip, but also to generate photon timing histograms physically close to the SPAD array, which results in considerable data compression, and hence alleviates readout bottleneck issues, allowing higher photon throughputs to be achieved, and thus faster data acquisition [57,73]. Moreover, combining on-histogramming with a multi-event TDC [58,74-76], capable of registering multi-events per laser cycle, has been demonstrated to reduce TDC pile-up distortion under high ambient conditions [31].

The generation of histograms can be carried out in-pixel, or outside the array, potentially as column parallel logic. The advantage of in-pixel processing is that it avoids any bottlenecks in transferring data out of the array. However, there is then limited space for histogram storage, which impacts the number of histogram bins that can be accommodated (and hence the timing range). The sensor in [74], for example, features 16 bins, which, assuming a 1ns bin size, equates to just 2.4 meters. Capturing over long distances therefore requires multiple exposures for range disambiguation.

On the other hand, if the processing is carried out outside the array, then the pixel array can be made dense and compact, but in the case of a large array, there could be bottleneck issues under high ambient conditions, for example when multiple SPADs in a column are sending events to the same shared

Ref.	[78]	[84]	[30]	[58]	[73]	[74]	[79]	[75]	[80]	[81]	[82]	[87]
Arch.	Full Histogram On-Chip						Partial Histogram On-Chip					
Author	Niclass	Van Blerk-om	Kumag-ai	Al Abbas	Erdogan	Hutchi-ngs	Zhang	Seo	Kim	Gyongy	Stoppa	Zhang
Techno.	180nm FSI	40nm FSI	90nm/40nm 3D-BSI	130nm FSI	130nm FSI	40nm 3D BSI	180nm FSI	110nm FSI	110nm FSI	40nm FSI	90/40 nm 3D-BSI	65/65 nm 3D-BSI
Histogram Channels	16	128	384	1	512	4096	36288	36	1920	2048	4800	2400
On-Chip Data Storage Mem' (kb)	352	1536	9108	4.125	176	896	5670	N/A	33.75	192	1800	2508
Histo Memory	SRAM	SRAM	SRAM	Ripple Counter	Ripple Counter	Ripple Counter	SRAM	Ana' Counter	Ripple Counter	Ripple Counter	SRAM (est.)	SRAM
Data Rate On-Chip	19.2G	6.4T	9.2T (est)	10G	51.2G	8.2T	14.4G	3.3G	768G	655G	480G	1.9T
Total Histo Area (est.) (μm^2)	15.2 $x10^6$	19.7 $x10^6$	23.0 $x10^6$	297 $x10^3$	18.1 $x10^6$	3.72 $x10^6$	60 $x10^6$	8.2 $x10^6$	7.6 $x10^6$	1.12 $x10^6$	12 $x10^6$	6.6 $x10^6$
Equivalent Area per Bit (μm^2)	42.2	12.5	2.47	70.4	101	4.0	10.3	4000	222	5.7	6.51	10.7
Power inc SPADs (mW)	530	1792	1311 (max est. *)	144	1680	1258 (max *)	2538	180	840	2566 (max *)	1500 (max)	1530 (max est. *)
Power per dToF Channel (mW)	33.1	14.0	3.1	144	3.28	0.3	0.07	5	0.43	1.3	0.3	0.6
Power Efficiency (est.) (Ops/W)	36.2G	3.6T	7.0T	69.4G	30.5G	6.5T	5.7G	18.4G	914G	4T	320G	78G

Table 2. Selected references comparison table of five full histogram on chip sensors compared to seven partial histogram on chip sensors. (*) power scaled by exposure time to obtain continuous operation power.

TDC [77]. Although the histogram memory is no longer constrained by pixel size, the use of a large number of histogram bins may lead to a requirement to transfer large amounts of data in and out of memory, which could have frame rate and power consumption implications. The histogram generation unit area can be several times the dimension of a single SPAD pixel depending on Eqn. (1). For example in SRAM-based histograms, there is tradeoff between area per histogram bit (A_{bit}) and the energy per access as shown in Fig. 12 indicating that for small histogram units the memory periphery logic dominates the area whereas for large memories the energy per access becomes a dominant source of power consumption.

C. dToF SPADs with on-chip histogramming and histogram processing

Further data compression and area reduction may be obtained by combining embedded histogram generation with the processing of these histograms (Fig. 13). An early example of such an architecture can be found in [30], where an FIR filter is used to detect segments in the histogram containing peaks. Only the identified segments are subsequently read out. In [79], the bin width of 16 bin histograms is progressively reduced, for increased depth precision, by "zooming in" on the peak bin in each step. A similar approach is taken in [80] but using only 2 (in-pixel) bins, the final successive approximation step being followed by depth computation (interpolation). [81] also features a partial (8 bin) histogram in each pixel but rather than adjusting the bin width, the histogram is shifted automatically in time to locate and track peaks. Sub-bin resolution peak extraction is provided by column parallel logic, which accounts for the ambient level. In [82] a 32 bin partial histogram is swept through the full time range. The chip features 80×60 macropixels of 4×4 SPADs each, with QVGA image resolution being obtained by 16:1 multiplexing of SPADs. On-chip peak extraction is implemented outside the focal plane array [30] generating full sized histograms, which are then processed using an FIR filter to detect peaks and extract the peak bin.

D. Discussion

Table 2 provides a comparison of selected references of full histogram on chip versus partial histogram on chip. The SST-TDC techniques [30][84] offer high power efficiency at the cost of modest resolution and high silicon area. For higher resolutions, partial histogramming can reduce the histogram

area requirements in the chip by at least an order of magnitude. Fig. 14 plots the power and data rate on chip to visually analyze the best power efficiencies reported in the literature of the selected references of Table 2. Power savings, quantified in terms of power per histogram, can also be significant. However, it should be noted that such a measure does not take into account of the potential increase in the overall acquisition time due to data being captured in multiple exposures (nor the wasted illumination power whenever the return signal falls outside the current timing range). The longer the overall time range (or distance range) to be covered, the more steps it takes to scan or search through it, and the shorter, in relative terms, the temporal aperture during which signal photons are collected. In this respect, a "vertical search" (Figs. 13a-c), which zooms into the peak bin in every step can be more efficient than a "horizontal search" that sweeps across the time range with a fixed bin size (Fig. 13d). However, searching vertically may be challenging under high ambient levels, when there is a large build-up of background counts for wide bin widths.

We argue that to realize the potential power savings offered by partial histogramming, a chip requires "smart" peak scanning and illumination strategies, which could include:

(1) Increasing the temporal aperture using pixels that lock onto peaks and track them rather than continually scan the whole time range [81].

(2) Adapting the exposure time/illumination power for pixels that are peak searching (assuming an illuminator with addressable elements), and only reading out pixels where a peak has been detected.

Whether a chip features partial or full histogramming, further power savings may be attained by only acquiring/reading out histogram data when a change in the scene is detected. Change detection can potentially be implemented via a much lower power, passive, intensity imaging modality [88].

V. Conclusion

.SPAD dToF sensors are rapidly advancing in levels of integration and performance driven by smart pixel architectures, advanced CIS process technology and digital processing. They can be expected to achieve the practical array sizes and photon event processing rates required to achieve unambiguous depth imaging at high frame rates for volume products in the consumer, industrial and automotive sectors. With power efficiencies breaking through the 1Tops/W barrier, denser digital nodes will improve this further and allow higher sensor resolutions for the same power budget. "It is worth reflecting on the indirect CO_2 emission from power consuming dToF receivers in mass market products globally as the volume of LIDAR systems rapidly accelerates: it is desirable that design teams of dToF receivers take steps to lower the on-chip power and improve power efficiencies to lower indirect CO_2 emissions. Moreover, with the best reported power per dToF channel still in 0.1mW's range there are orders of magnitude power reduction still required if dToF sensor resolutions are to scale up further in the future.

Acknowledgment

The authors thank Tarek Al Abbas for use of his data on SPAD pixel pitch.

References

[1] R.I. Hartley, P. Sturm, "Triangulation", Comput. Vis. Image Underst. 1997, vol. 68, pp. 146–157.

[2] S. Zhang, "High-speed 3D shape measurement with structured light methods: a review," Opt. Laser Eng. 106, 119–131 (2018)

[3] M. Bertozzi, A. Broggi, A. Fascioli and S. Nichele, "Stereo vision-based vehicle detection," Proceedings of the IEEE Intelligent Vehicles Symposium 2000 (Cat. No.00TH8511), 2000, pp. 39-44, doi: 10.1109/IVS.2000.898315.

[4] C. Debeunne and D. Vivet, "A Review of Visual-LiDAR Fusion based Simultaneous Localization and Mapping," Sensors, vol. 20, no. 7, p. 2068, Apr. 2020, doi: 10.3390/s20072068.

[5] C. S. Bamji et al., "A 0.13 µm CMOS System-on-Chip for a 512 × 424 Time-of-Flight Image Sensor With Multi-Frequency Photo-Demodulation up to 130 MHz and 2 GS/s ADC," in IEEE Journal of Solid-State Circuits, vol. 50, no. 1, pp. 303-319, Jan. 2015, doi: 10.1109/JSSC.2014.2364270.

[6] Y. Shirakawa, K. Yasutomi, K. Kagawa, S. Aoyama, and S. Kawahito, "An 8-Tap CMOS Lock-In Pixel Image Sensor for Short-Pulse Time-of-Flight Measurements," Sensors, vol. 20, no. 4, p. 1040, Feb. 2020, doi: 10.3390/s20041040.

[7] L. Wu, D. San Segundo Bello, P. Coppejans, J. Craninckx, A. Süss, M. Rosmeulen, P. Wambacq, and J. Borremans, "Analysis and Design of a CMOS Ultra-High-Speed Burst Mode Imager with In-Situ Storage Topology Featuring In-Pixel CDS Amplification," Sensors, vol. 18, no. 11, p. 3683, Oct. 2018 [Online]. Available: http://dx.doi.org/10.3390/s18113683

[8] G. Chen, C. Wiede and R. Kokozinski, "Data Processing Approaches on SPAD-Based d-TOF LiDAR Systems: A Review," in IEEE Sensors Journal, vol. 21, no. 5, pp. 5656-5667, 1 March1, 2021, doi: 10.1109/JSEN.2020.3038487.

[9] L. J. Koerner, "Models of Direct Time-of-Flight Sensor Precision That Enable Optimal Design and Dynamic Configuration," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-9, 2021, Art no. 8502609, doi: 10.1109/TIM.2021.3073684.

[10] J. Kostamovaara et al., "On Laser Ranging Based on High-Speed/Energy Laser Diode Pulses and Single-Photon Detection Techniques," in IEEE Photonics Journal, vol. 7, no. 2, pp. 1-15, April 2015, Art no. 7800215, doi: 10.1109/JPHOT.2015.2402129.

[11] W. Becker, "Advanced Time-Correlated Single Photon Counting Techniques", Springer Series in Chemical Physics (2005). 81. I-387. 10.1007/3-540-28882-1_9.

[12] C. L. Glennie,W. E. Carter, R. L. Shrestha and W. E. Dietrich, "Geodetic imaging with airborne LiDAR: the Earth's surface revealed", Rep. Prog. Phys. 76 086801 (2013).

[13] F. Madonini and F. Villa, "Single Photon Avalanche Diode Arrays for Time-Resolved Raman Spectroscopy," Sensors, vol. 21, no. 13, p. 4287, Jun. 2021 [Online]. Available: http://dx.doi.org/10.3390/s21134287

[14] M. Perenzoni, D. Perenzoni, and D. Stoppa, "A 64×64-pixels digital silicon photomultiplier direct TOF sensor with 100-MPhotons/s/pixel background rejection and imaging/altimeter mode with 0.14% precision up to 6 km for spacecraft navigation and landing," IEEE J. Solid-State Circuits, vol. 52, no. 1, pp. 151–160, Jan. 2017.

[15] VL53L1X Datasheet, 3rd ed., ST, Geneva, Switzerland, Nov. 2018.

[16] M. Vogt, A. Rips, and C. Emmelmann, "Comparison of iPad Pro®'s LiDAR and TrueDepth Capabilities with an Industrial 3D Scanning Solution," Technologies, vol. 9, no. 2, p. 25, Apr. 2021 [Online]. Available: http://dx.doi.org/10.3390/technologies9020025

[17] D. Stoppa at al., “A Reconfigurable QVGA/Q3VGA Direct Time-of-Flight 3D Imaging System with Onchip Depth-map Computation in 45/40nm 3D-stacked BSI SPAD CMOS”, in Proc. International Image Sensor Workshop, 2021.
[18] R. Roriz, J. Cabral and T. Gomes, "Automotive LiDAR Technology: A Survey," in IEEE Transactions on Intelligent Transportation Systems, doi: 10.1109/TITS.2021.3086804.
[19] J. Lambert et al., "Performance Analysis of 10 Models of 3D LiDARs for Automated Driving," in IEEE Access, vol. 8, pp. 131699-131722, 2020, doi: 10.1109/ACCESS.2020.3009680.
[20] M. Hintikka, L. Hallman, and J. Kostamovaara , "Comparison of the leading-edge timing walk in pulsed TOF laser range finding with avalanche bipolar junction transistor (BJT) and metal-oxide-semiconductor (MOS) switch based laser diode drivers", Review of Scientific Instruments 88, 123109 (2017) https://doi.org/10.1063/1.4999253
[21] K. Morimoto, A. Ardelean, Ming-Lo Wu, A. C. Ulku, I. M. Antolovic, C. Bruschini, and E. Charbon, "Megapixel time-gated SPAD image sensor for 2D and 3D imaging applications," Optica 7, 346-354 (2020)
[22] A. Gupta, A. Ingle and M. Gupta,"Asynchronous Single-Photon 3D Imaging," in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019 pp. 7908-7917. doi: 10.1109/ICCV.2019.00800.
[23] A. Tontini, L. Gasparini, and M. Perenzoni, “Numerical Model of SPAD-Based Direct Time-of-Flight Flash LIDAR CMOS Image Sensors,” Sensors, vol. 20, no. 18, p. 5203, Sep. 2020 [Online]. Available: http://dx.doi.org/10.3390/s20185203
[24] American Society for Testing and Materials, 2012. ASTM G173 - 03 “Standard Tables for Reference Solar Spectral Irradiances: Direct Normal and Hemispherical on 37° Tilted Surface”
[25] “Noises in Optical Communications and Photonic Systems” au. Le Nguyen Binh (Boca Raton: CRC Press, 07 Nov 2016.
[26] R. E. Thompson, D. R. Larson and W.W. Webb, “Precise nanometer localization analysis for individual fluorescent probes”, Biophysical journal, 82(5), pp.2775-2783 (2002).
[27] N. Hagen, M. Kupinski, and E.L. Dereniak, ”Gaussian profile estimation in one dimension”, Applied Optics, 46(22), pp.5374-5383 (2007).
[28] J. Rapp, Y. Ma, R.M. Dawson and V. K. Goyal, High-flux single-photon lidar”, Optica, 8(1), pp.30-39 (2021).
[29] G. Tolt, C. Grönwall, and M. Henriksson, “Peak detection approaches for time-correlated single-photon counting three-dimensional lidar systems” Optical Engineering, 57(3), p.031306 (2018).
[30] O. Kumagai, J. Ohmachi, M. Matsumura, S. Yagi, K. Tayu, K. Amagawa, T. Matsukawa, O. Ozawa, D. Hirono, Y. Shinozuka, and R. Homma, “A 189x600 Back-Illuminated Stacked SPAD Direct Time-of-Flight Depth Sensor for Automotive LiDAR Systems”, In 2021 IEEE International Solid-State Circuits Conference (ISSCC) (Vol. 64, pp. 110-112).
[31] I. Gyongy, S.W. Hutchings, A. Halimi, M. Tyler, S. Chan, F. Zhu, S. McLaughlin, R.K. Henderson and J. Leach, “High-speed 3D sensing via hybrid-mode imaging and guided upsampling”, Optica, 7(10), pp.1253-1260 (2020).
[32] J. Huikari, S. Jahromi, J.-P. Jansson, J. Kostamovaara, "Compact laser radar based on a subnanosecond laser diode transmitter and a two-dimensional CMOS single-photon receiver," Opt. Eng. 57(2) 024104 (19 February 2018) https://doi.org/10.1117/1.OE.57.2.024104
[33] R. Michalzik, “VCSELs: Fundamentals, Technology and Applications of Vertical-Cavity Surface-Emitting Lasers,” Springer-Verlag, Berlin & Heidelberg (2013). https://doi.org/10.1007/978-3-642-24986-0.
[34] F. Ceccarelli, G. Acconcia, A. Gulinatti, M. Ghioni, I. Rech and R. Osellame, “Recent Advances and Future Perspectives of Single-Photon Avalanche Diodes for Quantum Photonics Applications”, Adv. Quantum Technol., 4: 2000102 (2021). https://doi.org/10.1002/qute.202000102
[35] Eric C. Fest, “Stray Light Analysis and Control” SPIE Press PM229, (2013).
[36] S. Pellegrini and B. Rae "Fully industrialised single photon avalanche diodes", Proc. SPIE 10212, Advanced Photon Counting Techniques XI, 102120D (1 May 2017); https://doi.org/10.1117/12.2264364
[37] E. A. G. Webster, J. A. Richardson, L. A. Grant, D. Renshaw and R. K. Henderson, "A Single-Photon Avalanche Diode in 90-nm CMOS Imaging Technology With 44% Photon Detection Efficiency at 690 nm," in IEEE Electron Device Letters, vol. 33, no. 5, pp. 694-696, May 2012, doi: 10.1109/LED.2012.2187420
[38] B. Aull, “Geiger-Mode Avalanche Photodiode Arrays Integrated to All-Digital CMOS Circuits,” *Sensors*, vol. 16, April 2016.
[39] M.J. Lee and E. Charbon, “Progress in single-photon avalanche diode image sensors in standard CMOS: From two-dimensional monolithic to three-dimensional-stacked technology”, Jpn. J. Appl. Phys. 57, 1002A3, 2018.
[40] K. Ito et al., "A Back Illuminated 10μm SPAD Pixel Array Comprising Full Trench Isolation and Cu-Cu Bonding with Over 14% PDE at 940nm," 2020 IEEE International Electron Devices Meeting (IEDM), 2020, pp. 16.6.1-16.6.4, doi: 10.1109/IEDM13553.2020.9371944.
[41] N. Dutton, T. Al Abbas, I. Gyongy, F. Mattioli Della Rocca, and R. Henderson, “High Dynamic Range Imaging at the Quantum Limit with Single Photon Avalanche Diode-Based Image Sensors,” Sensors, vol. 18, no. 4, p. 1166, Apr. 2018 [Online]. Available: http://dx.doi.org/10.3390/s18041166
[42] N.A.W. Dutton, L. Parmesan, A.J. Holmes, L.A. Grant, and R.K. Henderson, “320×240 Oversampled Digital Single Photon Counting Image Sensor,” *In Proceedings of VLSI Symposium*, 2015.
[43] T. Al Abbas, N. A. W. Dutton, O. Almer, S. Pellegrini, Y. Henrion and R. K. Henderson, "Backside illuminated SPAD image sensor with 7.83μm pitch in 3D-stacked CMOS technology," 2016 IEEE International Electron Devices Meeting (IEDM), 2016, pp. 8.1.1-8.1.4, doi: 10.1109/IEDM.2016.7838372
[44] J. Ogi et al., "7.5 A 250fps 124dB Dynamic-Range SPAD Image Sensor Stacked with Pixel-Parallel Photon Counter Employing Sub-Frame Extrapolating Architecture for Motion Artifact Suppression," 2021 IEEE International Solid- State Circuits Conference (ISSCC), 2021, pp. 113-115, doi: 10.1109/ISSCC42613.2021.9365977
[45] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, "Avalanche photodiodes and quenching circuits for single-photon detection," Appl. Opt. 35, 1956-1976 (1996)
[46] A. Eisele, et al. “185 MHz Count Rate, 139 dB Dynamic Range Single-Photon Avalanche Diode with Active Quenching Circuit in 130 nm CMOS Technology” in proc. International Image Sensor Workshop 2011.
[47] P. Padmanabhan, C. Zhang, and E. Charbon, “Modeling and Analysis of a Direct Time-of-Flight Sensor Architecture for LiDAR Applications,” Sensors, vol. 19, no. 24, p. 5464, Dec. 2019 [Online]. Available: http://dx.doi.org/10.3390/s19245464
[48] L. H. C. Braga et al., “A Fully Digital 8x16 SiPM Array for PET Applications With Per-Pixel TDCs and Real-Time Energy Output,” JSSC, vol. 49, no. 1, pp. 301–314, 2014.
[49] S. Gnecchi et al., "Digital Silicon Photomultipliers With OR/XOR Pulse Combining Techniques," in IEEE Transactions on Electron Devices, vol. 63, no. 3, pp. 1105-1110, March 2016, doi: 10.1109/TED.2016.2518301.
[50] S. Patanwala, I. Gyongy H. Mai, A. Aβmann, N.A.W. Dutton, B.R. Rae, R.K. Henderson, “A High-Throughput Photon Processing Technique for Range Extension of SPAD-based LiDAR Receivers”, Accepted for publication, IEEE OJSSC, Sep. 2021.
[51] D. P. Palubiak and M. J. Deen, "CMOS SPADs: Design Issues and Research Challenges for Detectors, Circuits, and Arrays," in IEEE Journal of Selected Topics in Quantum Electronics, vol. 20, no. 6, pp. 409-426, Nov.-Dec. 2014, doi: 10.1109/JSTQE.2014.2344034.
[52] S. Jahromi, J. Jansson, P. Keränen and J. Kostamovaara, "A 32 × 128 SPAD-257 TDC Receiver IC for Pulsed TOF Solid-State 3-D Imaging," in IEEE Journal of Solid-State Circuits, vol. 55, no. 7, pp. 1960-1970, July 2020, doi: 10.1109/JSSC.2020.2970704.
[53] F. Villa, B. Markovic, S. Bellisai, D. Bronzi, A. Tosi, F. Zappa, S. Tisa, D. Durini, S. Weyers, U. Paschen, and W. Brockherde, “SPAD Smart Pixel for Time-of-Flight and Time-Correlated Single-Photon Counting Measurements,” IEEE Photonics J., vol. 4, no. 3, pp. 795–804, Jun. 2012.
[54] M. Gersbach et al., "A Time-Resolved, Low-Noise Single-Photon Image Sensor Fabricated in Deep-Submicron CMOS Technology," in IEEE Journal of Solid-State Circuits, vol. 47, no. 6, pp. 1394-1407, June 2012, doi: 10.1109/JSSC.2012.2188466.
[55] M. Perenzoni, N. Massari, D. Perenzoni, L. Gasparini, and D. Stoppa, “A 160×120 Pixel Analog-Counting Single-Photon Imager With Time-Gating and Self-Referenced Column-Parallel A/D Conversion for

Fluorescence Lifetime Imaging," *IEEE J. Solid-State Circuits*, vol 51., Jan. 2016.

[56] A. Carimatto et al., "11.4 A 67,392-SPAD PVTB-compensated multi-channel digital SiPM with 432 column-parallel 48ps 17b TDCs for endoscopic time-of-flight PET," 2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers, 2015, pp. 1-3, doi: 10.1109/ISSCC.2015.7062996.

[57] C. Niclass, M. Soga, H. Matsubara, M. Ogawa and M. Kagami, "A 0.18um CMOS SoC for a 100-m-Range 10-Frame/s 200 x 96-Pixel Time-of-Flight Depth Sensor," in IEEE Journal of Solid-State Circuits, vol. 49, no. 1, pp. 315-330, Jan. 2014, doi: 10.1109/JSSC.2013.2284352.

[58] T. Al Abbas, N. A. W. Dutton, O. Almer, N. Finlayson, F. M. D. Rocca and R. Henderson, "A CMOS SPAD Sensor With a Multi-Event Folded Flash Time-to-Digital Converter for Ultra-Fast Optical Transient Capture," in IEEE Sensors Journal, vol. 18, no. 8, pp. 3163-3173, 15 April15, 2018, doi: 10.1109/JSEN.2018.2803087.

[59] C. Niclass, A. Rochas, P. -. Besse and E. Charbon, "Design and characterization of a CMOS 3-D image sensor based on single photon avalanche diodes," in IEEE Journal of Solid-State Circuits, vol. 40, no. 9, pp. 1847-1854, Sept. 2005, doi: 10.1109/JSSC.2005.848173.

[60] F. Borghetti, D. Mosconi, L. Pancheri, and D. Stoppa, "A CMOS single-photon avalanche diode sensor for fluorescence lifetime imaging" In *IEEE International Image Sensors Workshop* (pp. 7-10), 2007.

[61] J. Richardson et al., "A 32×32 50ps resolution 10 bit time to digital converter array in 130nm CMOS for time correlated imaging," 2009 IEEE Custom Integrated Circuits Conference, 2009, pp. 77-80, doi: 10.1109/CICC.2009.5280890.

[62] I. Nissinen, J. Nissinen, P. Keränen, A. Länsman, J. Holma and J. Kostamovaara, "A 2\times (4)\times 128$ Multitime-Gated SPAD Line Detector for Pulsed Raman Spectroscopy," in IEEE Sensors Journal, vol. 15, no. 3, pp. 1358-1365, March 2015, doi: 10.1109/JSEN.2014.2361610.

[63] B. Aull et al., "Laser Radar Imager Based on 3D Integration of Geiger-Mode Avalanche Photodiodes with Two SOI Timing Circuit Layers," 2006 IEEE International Solid State Circuits Conference - Digest of Technical Papers, 2006, pp. 1179-1188, doi: 10.1109/ISSCC.2006.1696163.

[64] D. Stoppa et al., "A 32x32-pixel array with in-pixel photon counting and arrival time measurement in the analog domain," 2009 Proceedings of ESSCIRC, 2009, pp. 204-207, doi: 10.1109/ESSCIRC.2009.5325970.

[65] L. Parmesan, N.A. Dutton, N. Calder, N. Krstajic, A. J. Holmes, LA.. Grant, and R.K. Henderson "A 256 x 256 SPAD array with in-pixel Time to Amplitude Conversion for Fluorescence Lifetime Imaging Microscopy". In International Image Sensor Workshop 2015.

[66] M. Beer, C. Thattil, J. F. Haase, W. Brockherde and R. Kokozinski, "2×192 Pixel CMOS SPAD-Based Flash LiDAR Sensor with Adjustable Background Rejection," 2018 25th IEEE International Conference on Electronics, Circuits and Systems (ICECS), 2018, pp. 17-20, doi: 10.1109/ICECS.2018.8617905.

[67] P. Padmanabhan et al., "7.4 A 256×128 3D-Stacked (45nm) SPAD Flash LiDAR with 7-Level Coincidence Detection and Progressive Gating for 100m Range and 10klux Background Light," 2021 IEEE International Solid- State Circuits Conference (ISSCC), 2021, pp. 111-113, doi: 10.1109/ISSCC42613.2021.9366010.

[68] J.F. Haase, S. Grollius, S. Grosse, A. Buchner and M. Ligges, "A 32x24 pixel SPAD detector system for LiDAR and quantum imaging" In Photonic Instrumentation Engineering VIII (Vol. 11693, p. 116930M). International Society for Optics and Photonics, March 2021.

[69] M. Beer, J. Haase, J. Ruskowski, and R. Kokozinski, "Background Light Rejection in SPAD-Based LiDAR Sensors by Adaptive Photon Coincidence Detection," Sensors, vol. 18, no. 12, p. 4338, Dec. 2018, doi:10.3390/s18124338

[70] M. Zarghami et al., "A 32 × 32-Pixel CMOS Imager for Quantum Optics With Per-SPAD TDC, 19.48% Fill-Factor in a 44.64-µm Pitch Reaching 1-MHz Observation Rate," in IEEE Journal of Solid-State Circuits, vol. 55, no. 10, pp. 2819-2830, Oct. 2020, doi: 10.1109/JSSC.2020.3005756.

[71] A. Carimatto et al., "Multipurpose, Fully Integrated 128 $\times$ 128 Event-Driven MD-SiPM With 512 16-Bit TDCs With 45-ps LSB and 20-ns Gating in 40-nm CMOS Technology," in IEEE Solid-State Circuits Letters, vol. 1, no. 12, pp. 241-244, Dec. 2018, doi: 10.1109/LSSC.2019.2911043.

[72] A. R. Ximenes, P. Padmanabhan, M. Lee, Y. Yamashita, D. N. Yaung and E. Charbon, "A 256×256 45/65nm 3D-stacked SPAD-based direct TOF image sensor for LiDAR applications with optical polar modulation for up to 18.6dB interference suppression," 2018 IEEE International Solid - State Circuits Conference - (ISSCC), 2018, pp. 96-98, doi: 10.1109/ISSCC.2018.8310201.

[73] A. T. Erdogan et al., "A CMOS SPAD Line Sensor With Per-Pixel Histogramming TDC for Time-Resolved Multispectral Imaging," in IEEE Journal of Solid-State Circuits, vol. 54, no. 6, pp. 1705-1719, June 2019, doi: 10.1109/JSSC.2019.2894355.

[74] S. W. Hutchings et al., "A Reconfigurable 3-D-Stacked SPAD Imager With In-Pixel Histogramming for Flash LIDAR or High-Speed Time-of-Flight Imaging," in IEEE Journal of Solid-State Circuits, vol. 54, no. 11, pp. 2947-2956, Nov. 2019, doi: 10.1109/JSSC.2019.29390

[75] H. Seo et al., "Direct TOF Scanning LiDAR Sensor With Two-Step Multievent Histogramming TDC and Embedded Interference Filter," in IEEE Journal of Solid-State Circuits, vol. 56, no. 4, pp. 1022-1035, April 2021, doi: 10.1109/JSSC.2020.3048074.

[76] F. Severini, V. Sesta, F. Madonini, A. Incoronato, F. Villa, and F. Zappa, "SPAD array for LiDAR with region-of-interest selection and smart TDC routing", In Quantum Optics and Photon Counting 2021 (Vol. 11771, p. 117710F). International Society for Optics and Photonics.

[77] A. Carimatto et al., "11.4 A 67,392-SPAD PVTB-compensated multi-channel digital SiPM with 432 column-parallel 48ps 17b TDCs for endoscopic time-of-flight PET," 2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers, 2015, pp. 1-3, doi: 10.1109/ISSCC.2015.7062996.

[78] C. Niclass, et al., "A 0.18µm CMOS SoC for a 100m Range 10-Frames/s 200 x 96- Pixel Time-of-Flight Depth Sensor," in IEEE Journal of Solid-State Circuits, vol. 49, no. 1, Jan. 2014, doi: 10.1109/JSSC.2013.2284352.

[79] C. Zhang, S. Lindner, I. M. Antolović, J. Mata 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," in IEEE Journal of Solid-State Circuits, vol. 54, no. 4, pp. 1137-1151, April 2019, doi: 10.1109/JSSC.2018.2883720.

[80] B. Kim, S. Park, J. -H. Chun, J. Choi and S. -J. Kim, "7.2 A 48×40 13.5mm Depth Resolution Flash LiDAR Sensor with In-Pixel Zoom Histogramming Time-to-Digital Converter," 2021 IEEE International Solid- State Circuits Conference (ISSCC), 2021, pp. 108-110, doi: 10.1109/ISSCC42613.2021.9366022.

[81] I. Gyongy, A. Erdogan, N.A.W. Dutton, H. Mai, F.M. Della Rocca, R.K. Henderson, "A 200kFPS, 256×128 SPAD dToF sensor with peak tracking and smart readout", In International Image Sensor Workshop 2021.

[82] D. Stoppa, S. Abovyan, D. Furrer, R. Gancarz, T. Jessenig, R. Kappel, M. Lueger, C. Mautner, I. Mills, D. Perenzoni, G. Roehrer, P. Taloud, "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.

[83] S. Patanwala, I. Gyongy, N.A.W. Dutton, B.R. Rae, R.K. Henderson, "A Reconfigurable 40nm CMOS SPAD Array for LiDAR Receiver Validation" In International Image Sensor Workshop, 2019.

[84] D. van Blerkom "Modelling TDC Circuit Performance for SPAD Sensor Arrays" In International SPAD Sensor Workshop, 2020.

[85] F. Martin et al. "An all-in-one 64-zone SPAD-based Direct-Time-of-Flight Ranging Sensor with Embedded Illumination" in IEEE Sensors 2021.

[86] C. Niclass et al. "A 128 128 Single-Photon Image Sensor With Column-Level 10-Bit Time-to-Digital Converter Array" in IEEE Journal of Solid-State Circuits, vol. 43, no.12, pp 2977-2989, Dec. 2008.

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

[88] F. Mattioli Della Rocca et al., "A 128 × 128 SPAD Motion-Triggered Time-of-Flight Image Sensor With In-Pixel Histogram and Column-Parallel Vision Processor," in IEEE Journal of Solid-State Circuits, vol. 55, no. 7, pp. 1762-1775, July 2020, doi: 10.1109/JSSC.2020.2993722.