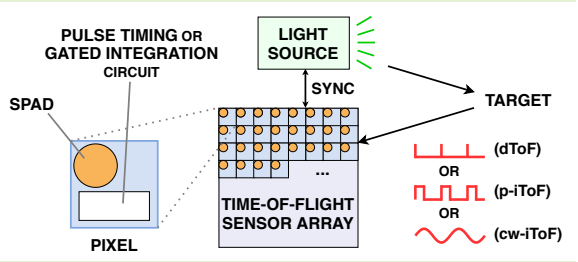
A Review of Single-Photon Avalanche Diode Time-of-Flight Imaging Sensor Arrays

François Piron, Daniel Morrison, *Student Member, IEEE*, Mehmet R. Yuce, *Senior Member, IEEE*, and Jean-Michel Redouté, *Senior Member, IEEE*

***Abstract*— Time-of-flight (ToF) sensors using light pulses or continuous waves allow accurate distance measurements. Three-dimensional imagers can be based on an array of timing or time-gated integration pixels. Single-photon avalanche diodes (SPADs) have been increasingly chosen as the pixel's photodetector device to develop fast, long-range ToF sensors. Solid-state ToF cameras are replacing other alternatives, showing attractive characteristics and bringing up new potential applications. This paper presents the technical evolution of SPAD ToF 3D imaging sensors, and provides insight into their development over the last few decades. Starting with the first SPAD imagers reported in the early 2000's, various direct and indirect arrays up to present day state-of-the-art prototypes are referenced. The existing methods, options and possible implementations are described, addressing their advantages and drawbacks, and showing the progress yet to be made. The performance of the different presented approaches are given and compared.**



***Index Terms*— 3D Imaging, 3DIC, CMOS imagers, Direct Time-of-Flight (dToF), Depth Image Sensors, Indirect Time-of-Flight (iToF), Light Detection and Ranging (LiDAR), Time-of-flight (ToF), Single-Photon Avalanche Diode (SPAD)**

I. Introduction

TIME-OF-FLIGHT (ToF) sensors are able to capture distances by measuring the travel time of an emitted signal. An optical ToF sensor emits a light wave or pulse, and measures the duration between the time this signal is emitted, and the moment it hits the target and is reflected back to the sensor. As the speed of light is constant in air, the distance of the observed object is directly proportional to this duration. An array of pixels able to independently measure ToF form the core of a solid-state three-dimensional (3D) imager.

Other optical techniques exist to measure distances without contact [1]. Triangulation-based methods require a geometrical separation between the laser emitter and imager, and fast mechanical laser scanning over the field of view. Pattern projection methods suffer from a lower accuracy at longer distances. Stereoscopic systems need two distant imagers and usually require computationally intensive post-processing techniques to extract the depth data. Many different approaches have been proposed and combined, but ToF has become one of the preferred choices for depth sensing over the last decades.

The work of F. Piron was supported by a FRIA Grant funded by the FNRS, Belgium.

F. Piron and J.-M. Redouté, are with the Department of Electrical Engineering and Computer Science, University of Liège, 4000, Belgium (francois.piron@uliege.be).

D. Morrison and M. R. Yuce, are with the Department of Electrical and Computer Systems Engineering, Monash University, Melbourne, VIC 3800, Australia (daniel.morrison@monash.edu).

F. Piron and D. Morrison have contributed equally to this work.

ToF provides a compact and accurate solution to 3D ranging. As imager ICs became low-cost and widespread, solid-state indirect ToF sensors started to be developed [2], either using conventional CMOS and CCD imagers or specific pixel designs with embedded photo-demodulation [3]–[5]. Their analog separation, modulation and accumulation of photocharges led to limited bandwidth and sensitivity, noise and other error sources.

A single-photon avalanche diode (SPAD) is a photodetector device, consisting of a p-n junction reverse-biased above its breakdown voltage. In this unstable state, called Geiger-mode, a single photon hitting the junction can trigger a self-sustaining avalanche, generating a detectable current flow. These perceived single photon events, outputted as digital pulses, can then be timed or counted by accompanying electronics.

The main characteristics of SPADs include photon detection probability (PDP), which is the probability of breakdown for each impinging photon, usually from a few percents up to 50%, depending on the wavelength. After a detection, the bias voltage must be reduced to quench the avalanche and reset the SPAD. The duration of this operation, during which the SPAD is unable to detect photons, is called dead-time, and is typically a few nanoseconds long. The SPAD can also be spuriously triggered by thermally generated carriers instead of photons. This effect is measured by the dark count rate (DCR). Carriers trapped during an avalanche can also be spontaneously released after the SPAD has been reset, causing afterpulsing, i.e. a new undesired avalanche [6].

High sensitivity, fast reaction time, and low timing jitter make SPADs excellent photodetector candidates for time-dependent applications, such as 3D imaging. However, SPADs require a guard ring to contain the high electric field, prevent premature edge breakdown and reduce the probability of crosstalk. SPAD pixels also require a quenching circuit and can include timing, counting, and buffering circuitry within the pixel. This reduces the fill-factor (FF) of the sensor, i.e. the ratio of active area to the total pixel area, causing SPADs to be relatively large compared with other photodetector structures.

3D cameras offer many possible applications, in the automotive, security, industrial, gaming and medical fields among others [7]. SPADs were first proposed and developed for pulse shape analysis [8], fluorescence decay measurements [9], positron emission tomography (PET) and single photon counting [10]. CMOS SPAD arrays for ToF started to appear in [11]–[13]. Nowadays, their application is not limited to 3D imaging, but also fluorescence lifetime imaging microscopy (FLIM), PET, Raman spectroscopy, fluorescence correlation spectroscopy (FCS), near-infrared optical tomography (NIROT), etc. The properties and challenges of SPADs, their integration, associated front-end circuit, and imaging arrays intended for such applications, are reviewed in depth in [14], [15]. Sensors designed for one application can often be repurposed for another, e.g. [16] and [17].

ToF sensors are divided into two main categories. Direct time-of-flight (dToF) sensors directly measure the ToF of a light pulse and are discussed in depth in Section II. On the other hand, indirect time-of-flight (iToF) sensors, detailed in Section III, measure light intensity in predefined time windows, computing depth as a post-processing step.

II. Direct Time of Flight Depth Sensors

The underlying concept of a dToF sensor involves a pulsed emission of light where the duration between emission and detection is directly measured using a timing circuit, usually a time-to-digital converter (TDC) or a time-to-amplitude converter (TAC). In the context of 3D imaging applications the light pulse is emitted toward the scene of interest, scatters off the target, and is detected by the sensor. Since the light pulse travels to the target and back, the distance (d) between the sensor and the target is given by [18]

$$d = \frac{ct_m}{2}, \tag{1}$$

where t_m is the time measured by the sensor and c is the speed of light. If the timing information is recorded with a precision of Δt, then the single-shot distance resolution (Δd) achieved is given by $\Delta d = c\Delta t/2$. A centimeter level distance resolution requires timing information on the order of nanoseconds, while a millimetric resolution requires a picosecond level timestamp.

The single-shot distance resolution can be improved upon by capturing the frame multiple times and recording a histogram of the time measurements. This improves the distance resolution according to $\Delta d_N = \Delta d/\sqrt{N_{frames}}$ [19]. While multiple measurements allow an arbitrary distance resolution to be achieved if a large number of frames are measured, the scattered light pulse is an extremely weak signal especially when capturing over long distances. To resolve the distance measurement to a millimetric precision, within a time period short enough for video capture, an extremely sensitive optical sensor with low timing jitter as well as a high resolution timing circuit is required. SPADs have been shown to be an ideal sensor device for dToF sensors, with excellent PDPs and timing jitter. However, the relatively low FF of SPADs impact the effective sensitivity, and background noise photons or DCR causing dead-time can prevent the detection of signal photons. The time at which the SPAD events occur need to be captured with a high resolution timing circuit. The ability to integrate both SPADs and CMOS circuitry [20] on the same silicon die has opened up the ability to implement this timing circuitry in many different ways, all that playing into the same trade-off between timing performance and FF.

Since 2016, a range of single point SPAD dToF range finders began to become commercially available with maximum ranging capabilities up to 4 m [21], [22]. SPAD dToF sensors in the 3D image format also became commercially available recently [23]–[25].

Table I provides a comparison of the performance metrics for the SPAD dToF sensors that have been presented over the last two decades. The following sections review these designs, demonstrating the progress in addressing the challenges of implementing large scale SPAD-based dToF depth sensors.

A. Sensor Architecture

The first SPAD dToF sensors presented in literature used arrays of SPADs multiplexed to external TDC ICs. Reference [18] implemented an 8×4 dToF sensor and [26] presented a 32×32 dToF sensor. These sensors started the TDCs synchronized to the laser emission and stopped the TDCs on the first SPAD avalanche event to occur for the selected SPAD. Full depth images were built up by sequentially, measuring each pixel 10,000 times to achieve a millimetric precision. These sensors demonstrated excellent timing performance with 120 ps resolution and 300 ps jitter, owing to the dedicated TDC IC. Only seven transistors were present in the per-pixel circuitry allowing a compact 58 µm×58 µm pixel with an FF of 1.1%. However, the multiplexing nature of these designs requires many measurements for every pixel, so this structure was not scalable for a practical high resolution sensor.

The next generation dToF sensors moved the timing circuits onto the same silicon die as the SPAD sensors, starting the trend of time-stamping SPAD events in parallel allowing a much higher degree of scalability and a lower acquisition time. Reference [27] implemented a 128×128 array with 32 10-bit 100 ps resolution TDCs shared between groups of four columns. Only one row was activated at a time, so 128 pixels could be read out simultaneously. Scanning was used to capture other rows in subsequent frames with a complete frame acquired in 50 ms. This design demonstrated the trade-off between low pixel complexity to achieve high FFs and a high level of parallelism to achieve low acquisition times.

Capturing timing information for every pixel in the array simultaneously requires a smart pixel architecture where timing circuits are included within the pixel layout. These

architectures have the highest degree of parallelism at the cost of a reduced FF and a high power consumption. The highest performing smart pixel designs implement the timing circuitry co-located in the SPAD array, shown in Fig. 1a, preventing loss in time precision due to routing parasitics and process variation. However, this can introduce dead-zones within the pixels further reducing the FF. Flash dToF sensors have strict requirements on the size of the timing circuitry so that the design has an acceptable FF.

Many examples of smart pixel dToF sensors demonstrate the reduced FF. Reference [28] presented a 32×32 pixel sensor utilizing an analog integration timing technique and achieved a timing resolution of 160 ps with a 50 µm pitch and approximately 1.05% FF pixels. Reference [29] integrated 160×128 pixels with 55 ps 10-bit TDCs in each 50 µm pixel, achieving an FF of approximately 1%. References [30] and [31] both presented prototype dToF pixels, with [30] presenting a 0.5% FF, 10 ps resolution pixel and [31] presenting a 3.14% FF pixel with a timing resolution of 312 ps. The design in [32] achieved an FF of 2.76% in a 64 µm pixel with a 145 ps timing resolution. All of these designs have FFs less than 4%.

There are three ways the FF of a smart pixel dToF sensor have been improved: reducing the per-pixel complexity, 3D-stacking, and microlensing. Reduction of the per-pixel complexity has been achieved by sharing timing circuit resources between multiple SPADs, and by reducing the area of the timing circuit. The trade-offs of these aspects are discussed in Section II-B and Section II-C respectively. 3D-stacking can be used to integrate the SPAD devices on the top silicon layer, while the bottom layer implements the functionality of the smart pixel, depicted in Fig. 1b. Reference [33] presented the first 3D-stacked SPAD architecture with a 4×200 array of backside illuminated (BSI) SPADs on a 47 µm pitch and an FF of 23.3%. Each group of 2×4 SPADs on the top layer shared a 49.7 ps resolution TDC on the bottom layer. The quenching circuits for the SPAD were implemented on the top layer, but did not impact the FF of the linear array. [34] presented a 3D-stacked frontside illuminated (FSI) SPAD dToF pixel on a 50 µm pitch. The top layer was dedicated to the SPADs and was bonded to the bottom layer containing the quenching circuit and timing circuitry with through-silicon-vias (TSVs). Reference [35] implemented a 19.8 µm, 256×256 SPAD array with an FF of 31.3%. Each block of 2×8×8 SPADs was connected to a shared, continuously running, TDC through a low skew arbiter tree. Reference [36] presented a 256×256 BSI SPAD array on a 9.18 µm pitch with a 50% FF. Each block of 4×4 SPADs was connected to an advanced functionality pixel circuit allowing photon counting at the SPAD level and dToF at the block level, using TDCs with a timing precision of 38 ps. 3D-stacking has also allowed the SPAD devices and the CMOS electronics to be implemented in different process nodes, so that the SPAD layer can be implemented in an custom optimized SPAD process while the electronics can be integrated in an advanced technology node to allow more complex per-pixel circuitry. Reference [36] utilized a 40 nm electronics layer and a 90 nm SPAD layer and implemented five different operating modes. Reference [37] corrected uniformity and skew using in-pixel digital signal processing. Implementing features such as these without 3D-stacking would heavily impact the sensor FF.

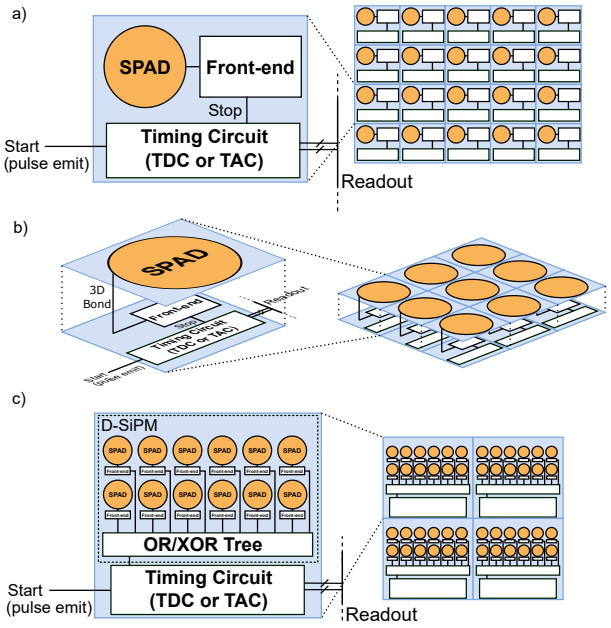


Fig. 1: Array and pixel architectures for smart pixel dToF sensors. a) Planar b) 3D stacked c) D-SiPM Array

Microlensing is another technique that has commonly been used to improve the effective FF of SPAD arrays by redirecting photons away from dead-zones within the pixel and toward the active area of the SPAD. References [35], [38], [39] all implemented SPAD arrays with microlenses fabricated on top of the SPAD smart pixel dToF sensors. Reference [38] used diffractive and refractive microlens arrays on the 3.14% SPAD pixel reported in [31] and found a maximum improvement of 16 times the original FF. Reference [39] used cylindrical microlenses to improve the FF from 13% to 42%. Reference [35] implemented a microlens array on a 3D-stacked SPAD array, improving the FF from 31.3% to 50.6%.

B. D-SiPM Arrays and Macro Pixels

Smart pixel designs can also be structured using macro pixels comprised of digital silicon photomultipliers (D-SiPMs) as shown in Fig. 1c. This technique utilizes multiple SPADs as microcells and combines the output of each with an OR or XOR tree allowing all microcells to act as a single pixel. It allows a larger total active area and also helps to reduce the limitation of the SPAD dead-time. The concept of D-SiPMs was first proposed in [40] with a sensor containing four D-SiPMs arrays with 64×32 SPADs each, and an FF of 50%. Reference [41] presented an updated design using D-SiPMs with 64×100 SPADs. A dToF imaging sensor can be created by associating D-SiPMs with a timing circuit, rather than the individual microcells, and forming a 2D array out of these macro pixels. D-SiPM arrays can achieve a relatively high FF due to the high number of SPADs within each pixel. However, each individual pixel is much larger than other smart pixel SPAD arrays, so the spatial resolution of the dToF sensor is much lower.

Reference [42] presented a 14×10 pixel array, where each pixel consisted of 32 SPADs, was 145 µm×215 µm, and achieved an FF of 29%. Groups of four D-SiPMs were connected to a TDC with a timing resolution of 325 ps. This design was extended in [43] with an implementation of a 8×16 pixel array with a pixel area of 570 µm×610 µm. Each pixel consisted of four 180-SPAD D-SiPMs with an FF of 35.7% and was connected to two 65 ps TDCs as well as two 7-b counters for photon counting. Reference [44] implemented a 4×4 pixel array with 800 µm×780 µm pixels. Instead of using an OR tree as in the previous designs, every fourth microcell in each column of the D-SiPM shared a 44 ps resolution TDC. This allowed the design to be more tolerant to DCR compared with using a single TDC. The FF achieved was 57%. Reference [45] was a further development of this design with a 9×18 pixel array using the same technique. Reference [46] presented a larger array compared with other D-SiPM arrays, with a 64×64 pixel sensor. Each pixel had an FF of 26.5% and consisted of eight SPADs which were connected with an OR-tree to a 250 ps resolution TDC. Reference [47] implemented a 16×8 D-SiPM array with 30 SPAD pixels, with a 32.1% FF and contained an 80 ps TDC and 5-b counter for photon counting. A 2×2 D-SiPM array with 432 SPADs per D-SiPM and a 37% FF was presented in [48]. This design utilized fast gated SPADs, where pairs of SPADs shared a differential frontend block. The outputs of each SPAD frontend were connected through a balanced OR tree to a single TDC to timestamp the first photon event. In [49] a 9×9 SPAD array was constructed out of four sub-arrays. A ten-channel TDCs was used to timestamp all elements of a selected subarray. The tenth channel was used to time the pulse time of the on-chip laser driver. This design achieved an FF of 51.4%.

Macro pixels can also provide additional functionality not available when using single SPAD pixels. Reference [50] introduced a correlation technique that required two or more microcells to fire simultaneously to record the timestamp. Since there is a strong correlation between multiple SPADs firing and the laser emission, each pixel can reject random noise events. This design implemented a sensor with 32 macro pixels each with 12 SPADs and was able to operate with a resolution better than 10 cm over 100 m despite 80 klux of background noise. A coincidence detection circuit presented in [51] also prevented timing unless multiple SPADs fire simultaneously but within a 2x2 macro pixel.

C. Timing Circuit Architecture

The timing circuit is the most critical component of dToF sensors. Numerous challenges exist in implementing the timing circuit, particularly for smart pixel sensors, where it must be compact to minimize the loss in FF, low power so that the design is scalable, and resistant to PVT variation across the array; all while achieving a high linearity and timing precision. Many different timing architectures for smart pixel dToF sensors have been proposed in literature, the simplest being a digital counter to count clock cycles of a high speed globally distributed clock. The digital counter is synchronized to start with the emission of the light pulse and stops on the first

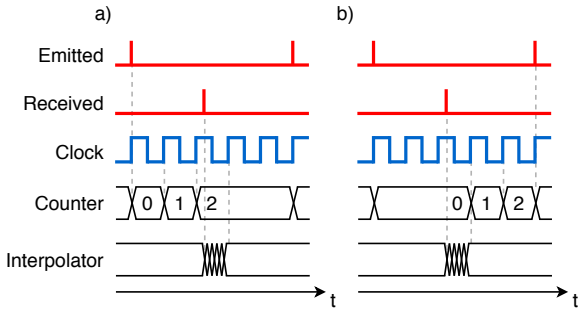


Fig. 2: Timing diagram of a dToF timing scheme with interpolation. a) Start-stop b) Stop-start

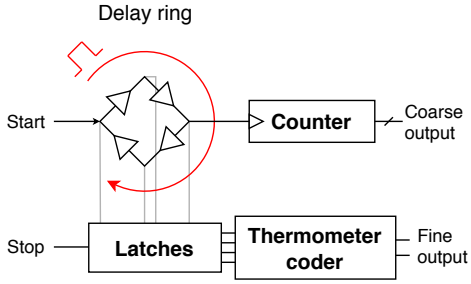


Fig. 3: Conceptual schematic of a ring oscillator (RO) TDC.

SPAD event providing a timestamp to the resolution of one clock cycle. Distributing a high speed clock to a large array of SPAD pixels with low skew is challenging and high clock rates lead to a high power consumption. Despite this nearly all dToF sensor designs include a digital counter to determine a coarse timestamp, typically a ripple counter or, where a synchronous counter is required, a linear feedback shift register (LFSR) where the counter order is converted to binary as a post processing step [52]. Each of these designs include additional circuitry to interpolate between clock cycles, reducing the clock frequency required for a high resolution. Fig. 2 shows the operation of a dToF timing scheme with interpolation to improve the timing resolution.

1) Ring Oscillators: RO-based TDC designs use a per-pixel RO to generate the clock signal as depicted in Fig. 3. At the start of the timing interval, the RO is initialized and a digital counter increments every time the RO loops. The RO is stopped at the end of the timing interval and additional information is interpreted based on the phases of the RO. When the RO is stopped, the phases are a thermometer code representation of the progression of the RO through the cycle which can be encoded into additional bits of timing information. This scheme was first conceived for early laser range finding systems [53] that had less stringent area restrictions. This TDC design used a 16-stage differential RO and a 10-b loop counter. It was expanded to a full range finding system including delay lines for improved timing precision of 250 ps in [54]. The first SPAD dToF sensor utilizing an RO [55] addressed the large integration area by simplifying the scheme, using a 4-stage RO and 7-b ripple counter to achieve a resolution of 50 ps, a differential non-linearity (DNL) of 0.5 LSB and integral non-linearity (INL) of 2.4 LSB. The

RO was started at the start of the frame and stopped upon the SPAD event. The same TDC structure was then used in [42] and [29] but starting the RO on the SPAD event and stopping at the end of the frame. This method of timing, known as stop-start timing, is able to reduce power consumption in the case of low light levels where the probability of detecting a photon is low. However, a stop-start timing scheme is relative to a future laser pulse rather than the laser pulse that illuminates the scene, thus causing the scheme to be susceptible to laser jitter and repetition rate instability. References [32], [43], [47] all used the same TDC structure and stop-start timing with coarse counter sizes of 7-b, 8-b and 9-b respectively.

Reference [33] extended this timing scheme by introducing a dual-speed RO to further reduce the power consumption of the TDC. Upon the SPAD event, the RO was started at a low frequency (246 MHz), capturing 5-b of timing information with a digital counter. The stop signal triggered the RO to be boosted to a high frequency (2.52 GHz). A high speed counter counted the frequency-boosted RO from the final falling edge of RO in the low frequency mode, to a delayed version of the stop signal, capturing an additional 4-b of timing data with a further 3-b interpreted from the phase of the RO. This technique allowed the RO to run at a high frequency for only a fraction of the measurement period, greatly reducing the TDC power consumption. The design achieved a timing resolution of 49.7 ps with a DNL of ±0.88 LSB and INL of ±0.47 LSB. This technique was also used in [56], except an additional stop signal was used at an integer multiple of the laser stop signal frequency, further reducing the amount of time the RO was active and hence further reducing the power consumption. This design achieved a resolution of 48.8 ps and a DNL of 0.48/−0.48 LSB and an INL of 0.89/−1.67 LSB.

Reference [46] opted for a simpler TDC, capturing only a single bit of phase information from a three stage RO. While the previously mentioned designs all used either differential or pseudo-differential buffers in the RO, [46] used a single-ended inverter generating a much higher frequency clock of 4 GHz for a 7-b or 15-b coarse counter depending on the operating mode. This achieved a 250 ps timing resolution. Reference [57] also used this design, albeit adding a RO phase encoder to obtain the additional 2-b of timing resolution. A timing resolution of 210 ps and peak-to-peak DNL of 1.28 LSB and INL of 1.92 LSB were achieved.

Reference [39] implemented a four stage pseudo-differential RO with a timing resolution of 33 ps to 120 ps and a peak-to-peak DNL of 0.9 LSB and INL of 5.64 LSB. This design used a separate supply network to allow the timing resolution to be tuned, as well as to prevent digital noise coupling into the operation of the RO. Calibration pixels were also included at the end of the rows that measured the number of RO loops for a global clock cycle, allowing the effects of PVT variation across the array to be calibrated out of the final time measurement.

References [37] and [34] used a vernier RO loop technique where each TDC contained a slow RO and a fast RO. The slow RO was started at the start of the timing interval and the fast RO was started at the end of the timing interval. The number of loops the fast RO required to catch up to the slow RO provided the time measurement with a resolution equal to the difference between the period of the slow RO and the fast RO allowing an extremely low timing resolution. These designs achieved timing resolutions of 10 ps and 15 ps respectively.

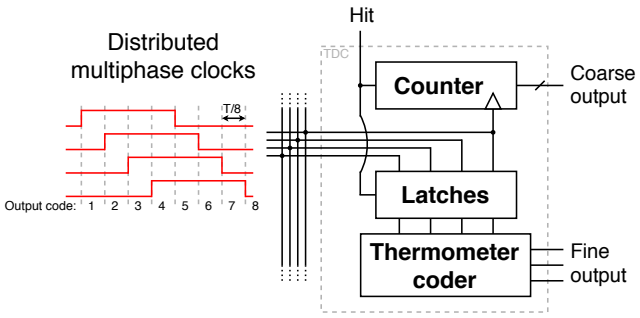


Fig. 4: Operation of a multiphase clock TDC.

The above designs demonstrate the trade-off between circuit complexity and timing resolution. The RO frequency is limited to hundreds of picoseconds by the intrinsic gate delay, the power consumption of the TDC, and the maximum frequency of the counting circuit. The more complicated phase detector or vernier loop structures have achieved sub 10 ps timing resolutions at the cost of area and power consumption. Improving the timing resolution limits the dynamic range of the TDC unless more bits are added to the counter. This also increases the area and power of the TDC. The trade-offs of the different RO TDC approaches are reviewed in depth in [58]. RO TDCs are susceptible to PVT variation. As the SPAD array size is scaled and smaller technology nodes are utilized, more emphasis will be needed on techniques such as those implemented in [39] for per-pixel calibration.

2) Multiphase Clocks: Another timing technique that has been used within SPAD dToF sensors is to distribute a multiple phase clock (MPC) to every TDC. As shown in Fig. 4, a phase locked loop (PLL) or delay locked loop (DLL) is used to generate the MPC and on the SPAD event, the state of each clock phase is recorded and can be interpreted to gain additional bits of timing information, similar to RO-based TDCs. A coarse counter is used to count the number of clock cycles from the start pulse until the stop pulse. This technique is able to achieve a better uniformity than ROs as only a single clock source is used so the effect of PVT variation is minimized. However additional routing resources are used in the distribution of the multiple clock phases.

The MPC technique is common in dToF sensors using a column-wise TDC. Reference [27] used a DLL to generate a 16-phase clock that was distributed to all column TDCs. A 2-b coarse counter counted the number of clock cycles of a 40 MHz clock and also latched the state of 16 phases of this clock, providing 4-b of timing data. A TDC level delay line was also used for an additional 4-b of timing information. Overall this implementation obtained a timing resolution of 97.66 ps with a DNL of 0.08 LSB and an INL of 1.89 LSB. Reference [59] presented a similar design, although in the form of a smart pixel. This design used a 4-b coarse counter with a 16-phase 100 MHz clock generated from a DLL allowing 4-b of phase timing information. The design also used a delay line for the fine interpolation, but with a vernier delay loop rather

TABLE I: Direct Time-of-Flight Sensor Arrays

	[26]	[27]	[28]	[42]	[29]	[44]	[50]	[43]	[60]	[19]	[32]	[45]	[61]
Year	2005	2008	2009	2011	2011	2013	2013	2014	2014	2014	2015	2015	2015
Tech. Node	0.8 µm	0.35 µm	130 nm	0.35 µm	130 nm	0.35 µm	180 nm	130 nm	130 nm	0.35 µm	180 nm	0.35 µm	130 nm
Illum. Dir.	FSI	FSI	FSI	FSI	FSI	FSI	FSI	FSI	FSI	FSI	FSI	FSI	FSI
Vert. Int.	Planar	Planar	Planar	Planar	Planar	Planar	Planar	Planar	Planar	Planar	Planar	Planar	Planar
Tim. Arch.	Ext. TDC	MPC	TAC	RO	RO	MPC	MPC	RO	TAC	MPC	RO	MPC	TAC
Array Size	32×32	128×128	32×32	14×10	160×128	4×4	32×1	8×16	3×3	32×32	64×64	9×18	256×256
# SPAD / Pixel	1	1	1	32	1	416	12	720	1	1	1	416	1
Tim. Circ. / Pixel	1/1024	1/512	1	1	1	48	1	1	1	1	1	48	1
Pixel Pitch (µm)	58×58	25×25	50×50	145×215	50×50	800×800	50	570×610	9.8×9.8	150×150	64×64	800×800	8×8
Fill Factor (%)	1.1	6	1.05	29	1	57	70[b]	42.6	3.27	3.14	2.76	57	19.63
Tim. Depth	8-b	10-b	7-b	10-b	10-b	17-b	12-b	12-b	–	10-b	11-b	17-b	–
Tim. Res. (ps)	120	97.7[c]	160	325	55	44	208	64	93	312	145	48.5	6.66
Tim. Range (ns)	20	100[c]	20[c]	330	55	5700	853	262	80[c]	3.12	297	6350	50[c]
INL (LSB)	–	1.89	1.9	–	2	3.5	0.73	2.3/-3.9	–	0.22	–	4/-2.1	–
DNL (LSB)	–	0.08	0.7	–	0.3	2	-0.52	0.28/-0.24	3.5	0.06	–	0.75/-0.75	–
Tim. Jitter (ps)	300	–	600	–	140	179	–	266	5100	600	290	327	368
Power (mW)	6	183	307	–	550	330	–	300	–	2800	–	990	–

	[33]	[46]	[35]	[39]	[36]	[47]	[56]	[51]	[37]	[57]	[62]	[63]	[48]
Year	2015	2017	2018	2019	2019	2019	2019	2019	2020	2020	2020	2020	2020
Tech. Node	130 nm	150 nm	45 nm / 65 nm	40 nm	40 nm / 90 nm	150 nm	180 nm	180 nm	65 nm	150 nm	0.35 µm	130 nm	350 nm
Illum. Dir.	BSI	FSI	BSI	FSI	BSI	FSI	FSI	FSI	FSI	FSI	FSI	FSI	FSI
Vert. Int.	2 Layer Cu-to-Cu	Planar	2 Layer Cu-to-Cu	Planar	2 Layer Cu-to-Cu	Planar	Planar	Planar	2 Layer TSV	Planar	Planar	Planar	Planar
Tim. Arch.	RO	RO	RO	RO	RO	RO	RO	TAC	RO	RO	MPC	TAC	DL
Array Size	1×400	64×64	256×256	192×128	64×64	16×8	252×144	113×129	16×16	32×32	32×128	64×64	2×2
# SPAD / Pixel	8	8	1	1	16	30	1	1	1	1	1	1	432
Tim. Circ. / Pixel	1	1	1/128	1	1	1	6/128	1	1	1	1/16	1	1/1728
Pixel Pitch (µm)	47	60×60	19.8×19.8	18.4×9.2	36.72 × 36.72	125×250	28.5×28.5	30×30	65×65	44.64 × 44.64	40×40	60×51	62.3× 202.4
Fill Factor (%)	23.3[b]	26.5	31.3 / 50.6[a]	13 / 42[b]	51	32.1	28	45	–	19.48	35[b]	0.66	37
Tim. Depth	12-b	16-b / 15-b	14-b	12-b	14-b / 4-b	10-b	12-b	16-b[d]	–	8-b	13-b	24-b[d]	7-b
Tim. Res. (ps)	49.7	250-20000	60	33-120	35/560	80	48.8	100	10	210	78	13	78
Tim. Range (ns)	200	6400 / 327700	1000	135-491	573	639	199.9	633.3	4	53	639	213333	9.98
INL (LSB)	0.47/-0.47	–	3.4/-0.8	5.64	0.1/-0.08	0.35/-2.4	0.89/-1.67	–	–	1.28	1.24	14.4	-0.27/+0.58
DNL (LSB)	0.44/-0.44	0.5/-0.4	0.8/-0.7	0.9	0.05/-0.05	0.2/-0.19	0.48/-0.48	–	–	1.92	0.833	0.47	0.039
Tim. Jitter (ps)	260	780	–	208	277	196	–	–	–	240	250	233	380
Power (mW)	7	93.5	–	1	77.6	57	2540	–	–	12.7	180	733	–

[a]With microlenses [b]Only includes light sensitive portion of linear array. [c]Adjustable [d]Conversion off-chip

than a conventional delay line. The implementation achieved a timing resolution of 10 ps and a peak DNL of 0.04 LSB.

A MPC timing scheme with a much faster base clock rate was used in [50] and [19] allowing the delay line to be omitted. Reference [50] used a 600 MHz clock with 8 phases and [19] used a 16-phase clock at 200 MHz. The resolution achieved by [50] was 208 ps, with a DNL of −0.52 LSB and an INL of 0.73 LSB, whereas [19] achieved a timing resolution of 312 ps and a DNL and INL of 0.06 LSB and 0.22 LSB respectively. This structure was also used in [62] but was extended to using 64 phases at 200 MHz. This design achieved a resolution of 78 ps with a DNL of 0.833 LSB and an INL of 1.24 LSB.

The effective number of phases in MPC schemes can be extended using comparators between adjacent phases. In [44] an 8-phase clock recorded a 5-b timestamp with a resolution of 44 ps. A follow up design in [45] used the same MPC scheme in a larger array with a resolution of 48.5 ps and a DNL of ±0.75 LSB and an INL of 4/−2.1 LSB.

3) Time to Amplitude Converters: SPAD dToF sensors have also been implemented using analog timing techniques known as TACs. TACs are able to achieve superior linearity, resolution and consume less power compared to TDCs, but are more susceptible to PVT variation, noise and transistor mismatch.

Integration-based techniques charge a capacitor so that the output voltage is proportional to the measured time. However, achieving uniformity with the PVT variation of the array and reducing susceptibility to noise typically requires unfeasibly large capacitors and current sources to achieve high linearity and resolution. A fully differential integration-based TAC was presented in [64] that simultaneously charged and discharged two capacitors. This achieved a resolution better than 27 ps and a DNL of 0.002 LSB. However, this TAC required an area of 440 µm×650 µm preventing integration in a high density SPAD array. References [28], [63], and [51] all presented TACs that improved the area by using reference integrations as comparisons to correct PVT variations. In [28] the time measurement was performed and then subsequently converted the voltage to a digital value utilizing a single-slope analog-to-digital converter (ADC) within the pixel and a current source matched to the time integration current source. In [63] a technique was proposed that also uses a per-pixel integrator to perform the time measurement, but then performed two reference integrations of one and two clock cycles respectively. All three voltages were measured using ADCs allowing the effect of PVT variation to be reduced. Reference [28] achieved a resolution of 160 ps and DNL and INL of 0.7 LSB and 1.9 LSB respectively. Reference [63] achieved a resolution of 13 ps, a DNL of 0.47 LSB, and an INL of 14.4 LSB. The TAC design in [51] introduced a time-to-charge technique where charge was transferred from a capacitor to two other capacitors. After a SPAD event the charge direction was swapped from one capacitor to the other and the ratio of voltages on the two capacitors represented the timing information. This technique relied on a constant and stable shutter duration and laser pulse delay but achieved a resolution of 100 ps.

References [60] and [61] took a different approach, by distributing a timing ramp waveform and sampling the waveform in each pixel at the SPAD event. This allowed a minimal per-pixel area footprint while still implementing a smart pixel design. The resolution achieved by [61] was 6.66 ps.

4) Multi-Event TDCs: Sensors that only capture a single photon per-pixel per-frame are susceptible to noise photons or dark count events that may be recorded before the arrival of photons from the light source. This effect, known as photon pileup, causes a low probability of detecting the signal from targets at late detection times, particularly in high ambient light conditions. Some designs have implemented multi-event TDCs allowing multiple timestamps to be generated per frame. Reference [65] implemented a multi-event folded flash TDC for a D-SiPM with an XOR tree readout. By encoding the time events with a one-hot encoding scheme and by using a pipelined processing chain, the design was able to continuously record timestamps at 10 GS/s. A similar design was used in a D-SiPM array in [36]. Reference [35] used a continuously running RO-based TDC shared between multiple SPADs that continuously recorded the timestamp and address of the firing SPAD.

D. On-Chip Histogramming

Histogramming the output data frames of a dToF sensor is typically one of the first data processing steps in the signal processing pipeline. Moving this step on-chip has been an aspect of several works. In [66] a line sensor implementing per-pixel histogramming was presented. This was implemented using a 10-b ripple counter for all 32 histogram bins. A further development was presented in [65] with a histogramming technique for a single D-SiPM and a multievent folded flash TDC. This design directly created a histogram for each phase of a 33-phase RO directly from the one-hot output of the folded flash TDC rather than capturing a timestamp for each individual event. Reference [36] also utilized this design when in histogramming mode. This design implemented an array of D-SiPM, so a full histogram was generated for each pixel within a single laser pulse. By only reading out the histogram of multiple laser pulses, the required data rate for readout is significantly reduced. Reference [56] further reduced the output data bandwidth by implementing a peak detection algorithm for a per-pixel histogramming technique for a 252×144 SPAD array. Only time bins surrounding a detected peak were offloaded from the chip. Reference [48] built a histogram by first converting the TDC output to a one-hot encoding and then incrementing the histogram counter corresponding to the one-hot signal.

The commercial sensor offerings in [21], [22], [25] all make use of on-chip histogramming in order to utilize an I^2C interface to output data from the sensor. Further data compression is achieved using target tracking algorithms on-chip.

III. Indirect Time of Flight Depth Sensors

Unlike dToF, iToF sensors do not directly measure time between emitted and received optical pulses. Rather, they collect time-dependent intensity information to deduce the delay between emitted and received signals. The received signal is integrated during specific time windows, given by reference

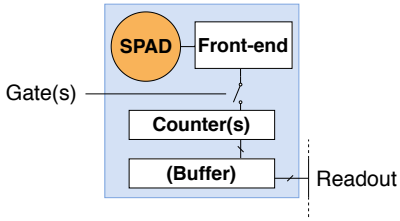


Fig. 5: Indirect pixel architecture

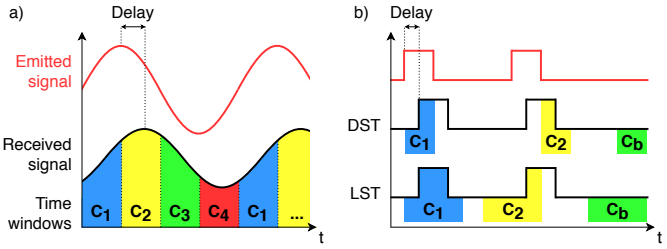


Fig. 6: Continuous wave (a) and pulse (b) iToF principle

clock(s) synchronized with the emitted signal. In a SPAD-based pixel, like in Fig. 5, this integration is performed by digitally counting incoming photons. The hit pulses are routed to one or several of the embedded counters, depending on the time window during which it happened. This process is often referred to as time-gated single photon counting (TGSPC). If the pixel design contains multiple counters, this operation can be performed simultaneously with multiple reference phase-shifted clocks. Otherwise, the necessary samples must be acquired successively.

The integration windows, processing steps, and performance depend on the optical signal shape. The two common signals used for iToF are the continuous wave (cw-iToF) and the square or pulse (p-iToF) [67]. At the sensor architecture level, trade-offs exist between accuracy and range (depending on signal frequency), frame rate (depending on integration time) and resolution (i.e. number of pixels, depending on active area and number and size of counters). Other limiting factors include the optical signal amplitude, PDP, background light, window lengths, and readout throughput. The unambiguous range usually does not exceed a few meters, but it can also be extended using multiple modulation frequencies [68]. Even when the entire observed scene is closer than the unambiguous range, interference in the received signal can also be created by multi-path reflections.

For cw-iToF, the sensor follows the principle of a lock-in amplifier measuring the phase shift between the emitted and received optical signals. Typically, the period is divided into four integration time windows [69], e.g. as shown in Fig. 6a. The phase shift can then be computed as such:

$$\phi = \text{atan}\left(\frac{C_4 - C_2}{C_1 - C_3}\right) \tag{2}$$

For best performance, the received optical sine wave should have the largest amplitude possible and be as undistorted as possible. Many ToF systems rely on on/off modulation to drive the light source, creating harmonics that impact the accuracy [69].

With square pulses, the multiple double short time integration (MDSI) method must be used [70]. A pulse burst is emitted and for each one, the signal is gated and integrated in two time windows. Thereby, the collected values are sufficient to get a depth measurement independent from the signal amplitude. Note that a third integration is required to remove background dependency, but it can take place outside the pulse burst. Several techniques are possible, as shown in Fig. 6b. The delay Δt for a pulse width T_p can then be computed in a straightforward way:

$$\text{DST:} \quad \Delta t = \frac{C_2 - C_b}{C_1 + C_2 - 2C_b} \times T_p \tag{3}$$

$$\text{LST:} \quad \Delta t = \frac{C_1 - C_2}{C_1 - C_b} \times T_p \tag{4}$$

Double sampling (DST) and long shutter (LST) both ideally provide the same result. However, the timing uncertainty of LST is analytically and experimentally inferior to DST [71].

An alternative method with pulses is the sliding window technique. An increasing delay is applied to the single fixed-length integrating window. The pixel value is read and reset at each delay increment. After all samples are gathered, the received signal over time can be reconstructed. As delay steps can be as low as a few picoseconds, the resulting depth resolution can be considerably better than other indirect methods, at the cost of an especially long frame duration. In an effort to increase the pixel FF, its memory can even be reduced to a single bit, making the acquisition time even longer. Yet, this kind of sensor is perfectly adequate in applications that do not require a high frame rate.

In low background conditions and with low signal distortion, p-iToF can give excellent accuracy. The integration windows can also be shortened to improve resolution. However, cw-iToF is more robust, and its accuracy is usually better and independent of measured distance [67]. The method choice is also affected by other practical considerations, such as the heavier required post-processing and the optical signal generation challenges for continuous wave.

An overview of the properties and performance of previously reported iToF pixels is available in Table II.

A. Indirect SPAD Pixels

Using SPADs to measure iToF was first reported in [80]. The presented 1×64 linear array was able to operate either in dToF or gated photon counting mode. The timing counters normally fed by clock signals were then instead directly incremented by photon hits. However, dedicated iToF sensors have smaller pixel circuits, allowing a larger FF and better performance.

The first SPAD array iToF sensor was introduced in [69]. The 60×48, 85 µm pitch pixels in 0.35 µm CMOS contained a 7 µm diameter SPAD and two 8-b counters. The 30 MHz infrared signal was generated by LEDs emitting an average optical power of 800 mW. After a 45 ms integration, the system achieved a 3.8 cm repeatability error at 2.4 m.

Reference [72] presented a 32×32 gated counter-based array in 0.35 µm high voltage CMOS. In every pixel, the single 8-b counter was accompanied by a latch register, allowing parallel measurement and readout, and thus global shutter operation.

TABLE II: Indirect Time-of-Flight Sensor Arrays

	[69]	[71], [72]	[73]	[74]	[75], [76]	[77]	[78]	[79]
Year	2009	2010	2011	2014	2014	2019	2020	2020
Method	cw-iToF	p-iToF	PD$\Delta\Sigma$[c]	cw-iToF	1-bit p-iToF	p-iToF	1-bit scan. win.	Analog p-iToF
Tech. node	0.35 µm	HV 0.35 µm	130 nm	HV 0.35 µm	130 nm	40 nm	0.18 µm	110 µm
Array Size	60×48	32×32	128×96	64×32	320×240	128×128	1024×1000	64×64
Pixel Pitch (µm)	85	100	44.7	150	8	40×20	9.4	32
Fill Factor (%)	0.5	3.14	3.17	3.14	26.8	13	7	17.3
Pixel counters	2×8 bits	1×8 bits	1×6 bits u/d[b] + 1×6 bits	2×9 bits u/d[b] + 1×9 bits	1 bit	2×16 bits	1 bit	1× analog
Illum. frequency (MHz)	30	6.67	3.33	7.5	–	–	40	25
Optical power (mW)	800	750	50	800	20	–	2	–
Optical wavelength (nm)	850	808	850	850	840	840	637	850
Max range (m)	5	22.5	45	20	–	48	–	40
Accuracy[a] μ_{error} (cm)	11	<28	0.5	80	<2	3.6	<1	90
Precision[a] σ_{error} (cm)	3.8	27	16	85	3.8	6.4	<0.8	20.4
Range of measurement (m)	2.4	14	2.4	20	0.6	3	1.6	40
Integration time (ms)	45	100	50	10	25 seconds	10	10's seconds	0.25
Frame rate (fps)	22	10	20	33k	130k	500	24k	90
Power consump. (mW)	35	20	40	50	69	185	284	20

[a]Measured in the *Range of measurement* [b]Up/down counters [c]Phase-Domain Delta-Sigma

The 100 µm pitch pixel contained a larger 20 µm diameter SPAD. This sensor was used in [81] with 4 samples per frame to recover the phase shift of a 20 MHz sine generated by red LEDs. After a total frame duration of 10 ms, it was able to get an error of 11 cm at 2.5 m. In [71], p-iToF was analyzed using a triple integration technique, both in LST and DST. A near-infrared laser generated 150 ns pulses at 1.67 MHz (25% duty cycle) with a 750 mW peak optical power. Each 100 ms long frame had a linearity error of around 6% over the 21 m range, both in DST and LST. Standard deviation was lower in LST, with a maximum around 25 cm at 10 m.

A 64×32 array of 150 µm pitch indirect pixels was presented in [74], using a 0.35 µm CMOS high voltage process. Along its very large 30 µm SPAD, it contained two 9-b up/down counters, enough to gather all information required to compute (2), and their associated buffer for global shutter operation. Up/down counters also increase the number of received photons before overflow. For a 10 ms integration time, the worst mean error was 80 cm and the precision 85 cm at 20 m.

Reference [73] showed a digital phase domain delta-sigma converter implemented in each pixel. Compared to conventional gated counting pixels, this approach allowed the output bandwidth to be greatly reduced and a pitch down to 45 µm. The 128×96 array, built in a 0.13 µm imaging CMOS process, included 8 µm diameter SPADs, for a 3.17% FF. A 50 mW 3.33 MHz square signal was generated by infrared LEDs. The sensor's repeatability error was 16 cm at 2.4 m after 50 ms, but achieved a excellent INL of 5 mm.

The 160×120 array presented in [82] had a 21% FF with 15 µm pitch pixels containing only 8 transistors in high voltage 0.35 µm CMOS. They performed gated analog counting between each column-parallel A/D conversion.

A similar charge transfer analog counting scheme had already been used in [83], with larger 25 µm pitch pixels, and more recently in [79], with 64×64 32 µm pitch pixels reaching a 2.25% INL and a 0.51% precision over the 40 m range.

Reference [76] achieved a 26.8% FF with 320×240, 8 µm pitch, 9 transistors, analog counting pixels, capable of switched current source and charge transfer amplifier counting methods.

A 3D-stacked BSI process was used in [84] to fabricate a 128×120 array with 45% FF. The imaging 65 nm top layer contained the SPADs and the 40 nm bottom one held the pixels electronics, reaching a 7.83 µm pitch, even though they included two 6-b counters. The sensor was further improved in [85], with a similar design based on a single 14-b counter and on-chip accumulation in SRAM.

In [77], two counters were each gated by a combination of 12 time-delayed external clocks. Time windows were configurable per-pixel. Thanks to the use of a 40 nm technology, the pixels achieved a 40 µm×20 µm size while containing two 16-b counters and with a 13% FF. The sensor achieved a 1.8 cm error over a 3 m range for 2 ms frames.

B. Single-bit SPAD Pixels

The thrive for low-power, small pitch and large FF pixels, essential for large-scale ranging arrays, led to simple single bit memory gated SPAD pixels. These can be classified as a kind of quanta image sensor (QIS), which was initially proposed in [86] and then developed with conventional imaging detector devices [87]. Time-gated SPAD-based QISs were also designed and they can be used for iToF. Their pixels typically increment external counters, e.g. in on-chip memory or an FPGA. The amount of gating windows can be arbitrarily large, depending on the method and desired range resolution. However, as readout needs to take place after each SPAD hit, the total frame time is inevitably much longer.

The memory of analog counting SPAD iToF pixels such as [76], [82], can also often operate digitally as single-bit pixels.

Reference [88] presented such a single bit memory SPAD pixel, in a 512×128 array. Built in a 0.35 µm CMOS technology, the 24 µm pitch pixels only reached a 5% FF.

In [89], a 512×512 array was reported in a 0.18 µm CMOS process, with a 16.38 µm pixel pitch and a 10.5% FF. The array could be operated at up to 97.7 kfps and the proposed p-i-n SPAD structure obtained a high 50% PDP and very low 7.5 dark counts per second.

Reference [16] also showed a 256×256 array of single bit pixels. Built in 0.13 µm CMOS, they reached a 16 µm pitch and a very high 61% FF, which even outperformed some CCD pixels. A sliding window method was used to measure depth. A 10 ns gating pulse was iteratively delayed by 10 ps steps. The 2.25 m range was thus divided into 1501 delay steps, for each of which up to 2000 binary frames were captured. This resulted in a millimeter scale depth resolution at the expense of a total acquisition time of up to several minutes, even with a binary frame rate of 100 kfps.

In [78], a 1024×1000 array in 0.18 µm CMOS was proposed. Its 9.4 µm pitch pixel design shared readout transistors, achieving a 13.4% FF and 24k binary frames per second. The 3.8 ns gate window was shifted by 36 ps delay steps.

Vertical avalanche photodiodes (VAPDs) in Geiger mode with a fast capacitive quenching method were reported in [90]. The 400×400 array of 6 µm pitch pixels, built in a 65 nm CMOS process, achieved a 70% FF. VAPDs were also used in the larger 1200×900 array of [91], able to measure both subrange (dToF) and iToF.

IV. Discussion

Since the first SPAD arrays, iToF has always been considered the best choice for short distances and small pixels in large arrays, with a better accuracy and lower power consumption. Counters triggered by photon hits are usually smaller and less power consuming than active TDCs, which require complex or sensitive timing circuits to reach a comparable accuracy. However, dToF does not suffer from the severe range limitation of iToF due to ambiguity and optical power constraints, nor from iToF multi-path reflections issues.

Throughout years of development, dToF sensors have evolved from relatively simple and small SPAD array structures with all complexity external to the sensor, to highly complex pixel structures with advanced functionality and large dense sensor arrays. In contrast, the simpler architecture of iToF SPAD pixels has not evolved significantly. Recently, improvements in dToF pixel designs, the emergence of SPAD arrays in finer technology nodes, and 3D ICs have overcome many dToF weaknesses, to the extent that it is now even sometimes preferred over iToF for low range applications.

The evolution of SPAD arrays is illustrated in Fig. 7 through their pixels' number and pitch.

However, this does not take into account the array performance. As many parameters come into play when comparing 3D imagers, their comparison is not straightforward. Some more comprehensive figures of merit were elaborated and evaluated in [92].

V. Conclusion

ToF sensor arrays first appeared with CMOS and CCD imagers able to integrate the optical signal within given time windows. These cameras, based on the same principles as conventional 2D imagers, provide great accuracy in a small pixel pitch but at a limited range due to finite optical power. Since their integration in CMOS ICs, SPADs have proved to provide good results in ToF arrays, with excellent timing jitter, low noise, single-photon sensitivity and digital operation. They enable fast and long-range dToF measurement, with pixels embedding time converters ranging from simple clock counters to oscillator-based systems with tens of picoseconds resolution. Their photon counting ability also allow them to replace the analog time-gated integration arrays in iToF sensors. However, large-scale integration is still a major challenge. Solutions like multi-layer ICs and microlenses were proposed to alleviate some common SPAD issues, but these come with an unavoidable increase in cost. Progress in manufacturing processes, TDC optimization, light emitting and processing systems, may lead the way to SPAD 3D imaging sensors becoming the standard for depth cameras. While these have already been used for other applications for a few years, SPAD ToF ranging cameras were only introduced in consumer electronics products by major manufacturers recently, both in low range dToF and iToF variants, for augmented reality, enhanced picture quality, or 3D modelling. With machines collecting a growing level of data about their environment, depth is valuable information not yet commonly available. Many fields would benefit from a solid-state, reliable and accurate source of 3D images, including autonomous vehicles, human-machine interfaces, robotic control, and many others.

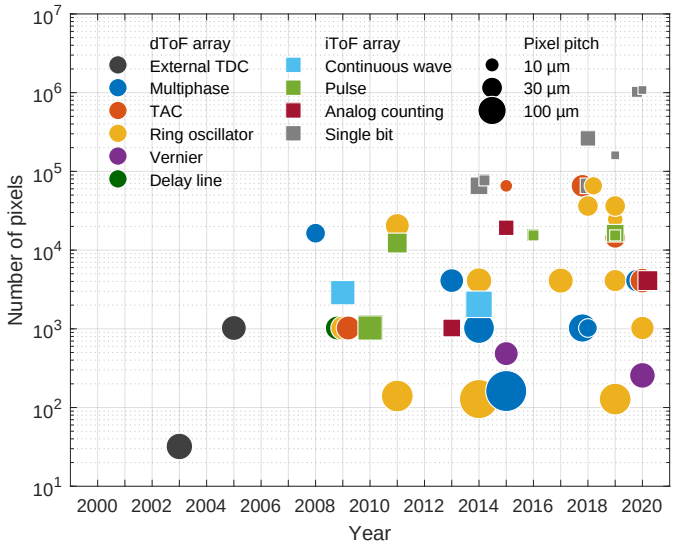


Fig. 7: Evolution of array size and pixel pitch in published SPAD arrays with different ToF methods

References

[1] F. Blais, "Review of 20 years of range sensor development," *J. Electron. Imaging*, vol. 13, no. 1, pp. 231 – 243, 2004.

[2] T. Spirig, P. Seitz, O. Vietze, and F. Heitger, "The lock-in CCD-two-dimensional synchronous detection of light," *IEEE J. Quantum Electron.*, vol. 31, no. 9, pp. 1705–1708, 1995.

[3] S. Foix, G. Alenya, and C. Torras, "Lock-in time-of-flight (ToF) cameras: A survey," *IEEE Sensors J.*, vol. 11, no. 9, pp. 1917–1926, 2011.

[4] M. Hansard, S. Lee, O. Choi, and R. P. Horaud, *Time-of-flight cameras: principles, methods and applications*. Springer Science & Business Media, 2012.

[5] R. Horaud, M. Hansard, G. Evangelidis, and C. Ménier, "An overview of depth cameras and range scanners based on time-of-flight technologies," *Machine Vision and Applications*, vol. 27, no. 7, pp. 1005–1020, Oct 2016.

[6] S. Cova, A. Lacaita, and G. Ripamonti, "Trapping phenomena in avalanche photodiodes on nanosecond scale," *IEEE Electron Device Lett.*, vol. 12, no. 12, pp. 685–687, 1991.

[7] F. Remondino and D. Stoppa, *TOF range-imaging cameras*. Springer-Verlag Berlin Heidelberg, Jul 2013.

[8] S. Cova, A. Longoni, and A. Andreoni, "Towards picosecond resolution with single-photon avalanche diodes," *Rev. Sci. Instrum.*, vol. 52, no. 3, pp. 408–412, 1981.

[9] S. Cova, A. Longoni, A. Andreoni, and R. Cubeddu, "A semiconductor detector for measuring ultraweak fluorescence decays with 70 ps FWHM resolution," *IEEE J. Quantum Electron.*, vol. 19, no. 4, pp. 630–634, 1983.

[10] R. McIntyre, "Recent developments in silicon avalanche photodiodes," *Measurement*, vol. 3, no. 4, pp. 146 – 152, 1985.

[11] B. F. Aull, A. H. Loomis, J. A. Gregory, and D. J. Young, "Geiger-mode avalanche photodiode arrays integrated with CMOS timing circuits," in *Proc. 56th IEEE Annual Device Research Conf. Digest*, Jun 1998, pp. 58–59.

[12] B. F. Aull *et al.*, "Geiger-mode avalanche photodiodes for three-dimensional imaging," *Lincoln laboratory J.*, vol. 13, no. 2, pp. 335–349, 2002.

[13] A. Rochas *et al.*, "Single photon detector fabricated in a complementary metal–oxide–semiconductor high-voltage technology," *Rev. Sci. Instrum.*, vol. 74, no. 7, pp. 3263–3270, 2003.

[14] D. P. Palubiak and M. J. Deen, "CMOS SPADs: Design issues and research challenges for detectors, circuits, and arrays," *IEEE J. Sel. Topics Quantum Electron.*, vol. 20, no. 6, pp. 409–426, 2014.

[15] C. Bruschini, H. Homulle, and E. Charbon, "Ten years of biophotonics single-photon SPAD imager applications: retrospective and outlook," in *Proc. SPIE Conf. Multiphoton Microscopy in the Biomedical Sciences XVII*, vol. 10069, 2017, pp. 213 – 233.

[16] X. Ren *et al.*, "High-resolution depth profiling using a range-gated CMOS SPAD quanta image sensor," *Opt. Express*, vol. 26, no. 5, pp. 5541–5557, Mar 2018.

[17] I. Gyongy *et al.*, "A 256 × 256, 100-kfps, 61image sensor for time-resolved microscopy applications," *IEEE Trans. on Electron Devices*, vol. 65, no. 2, pp. 547–554, Feb 2018.

[18] C. Niclass, A. Rochas, P.-A. Besse, and E. Charbon, "Toward a 3-D camera based on single photon avalanche diodes," *IEEE J. Sel. Topics Quantum Electron*, vol. 10, no. 4, pp. 796–802, Jul. 2004.

[19] F. Villa *et al.*, "CMOS imager with 1024 SPADs and TDCs for single-photon timing and 3-D time-of-flight," *IEEE J. Sel. Topics Quantum Electron*, vol. 20, no. 6, pp. 364–373, Nov. 2014.

[20] A. Rochas *et al.*, "First fully integrated 2-D array of single-photon detectors in standard CMOS technology," *IEEE Photon. Technol. Lett.*, vol. 15, no. 7, pp. 963–965, Jul. 2003.

[21] ST Microelectronics. Vl53l1x datasheet. [Online]. Available: https://www.st.com/resource/en/datasheet/vl53l1x.pdf

[22] AMS. Tmf8801 datasheet. [Online]. Available: https://ams.com/documents/20143/36005/TMF8801_DS00064_5-01.pdf/8a77012e-4b0f-2796-9682-4d61d73cc469

[23] Photon Force. Pf32 time-correlated single-photon counting camera. [Online]. Available: https://www.photon-force.com/pf32-time-resolved-single-photon-counting-camera/

[24] Horiba. Flimera brochure. [Online]. Available: https://static.horiba.com/fileadmin/Horiba/Products/Scientific/Molecular_and_Microanalysis/FLIMera/FLIMera_brochure.pdf

[25] ST Microelectronics. Vl53l5 data brief. [Online]. Available: https://www.st.com/resource/en/data_brief/vl53l5.pdf

[26] C. Niclass, A. Rochas, P.-A. Besse, and E. Charbon, "Design and characterization of a CMOS 3-D image sensor based on single photon avalanche diodes," *IEEE J. Solid-State Circuits*, vol. 40, no. 9, pp. 1847–1854, Sep. 2005.

[27] C. Niclass, C. Favi, T. Kluter, M. Gersbach, and E. Charbon, "A 128×128 single-photon image sensor with column-level 10-bit time-to-digital converter array," *IEEE J. Solid-State Circuits*, vol. 43, no. 12, pp. 2977–2989, Dec. 2008.

[28] D. Stoppa *et al.*, "A 32×32-pixel array with in-pixel photon counting and arrival time measurement in the analog domain," in *Proc. 35th IEEE Eur. Solid-State Circuits Conf.*, Sep 2009.

[29] C. Veerappan *et al.*, "A 160×128 single-photon image sensor with on-pixel 55ps 10b time-to-digital converter," in *Proc. IEEE Int. Solid-State Circuits Conf.*, Feb 2011.

[30] B. Markovic, S. Tisa, A. Tosi, and F. Zappa, "Smart-pixel for 3D ranging imagers based on single-photon avalanche diode and time-to-digital converter," in *Proc. SPIE Adv. Photon Counting Techn. V*, vol. 8033, May 2011.

[31] F. Villa *et al.*, "SPAD smart pixel for time-of-flight and time-correlated single-photon counting measurements," *IEEE Photon. J*, vol. 4, no. 3, pp. 795–804, Jun. 2012.

[32] I. Vornicu, R. Carmona-Galán, and Á. Rodríguez-Vázquez, "A SPAD-based 3D imager with in-pixel TDC for 145ps-accuracy ToF measurement," in *Proc. SPIE Image Sens. and Imaging Syst.*, vol. 9403, Mar 2015.

[33] J. M. Pavia, M. Scandini, S. Lindner, M. Wolf, and E. Charbon, "A 1×400 backside-illuminated SPAD sensor with 49.7 ps resolution, 30 pJ/sample TDCs fabricated in 3D CMOS technology for near-infrared optical tomography," *IEEE J. Solid-State Circuits*, vol. 50, no. 10, pp. 2406–2418, Oct. 2015.

[34] N. Roy, F. Nolet, F. Dubois, M.-O. Mercier, R. Fontaine, and J.-F. Pratte, "Low power and small area, 6.9 ps RMS time-to-digital converter for 3-D digital SiPM," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 1, no. 6, pp. 486–494, Nov. 2017.

[35] A. R. Ximenes, P. Padmanabhan, M.-J. 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 18.6 dB interference suppression," in *Proc. IEEE Int. Solid-State Circuits Conf.*, Feb 2018.

[36] 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," *IEEE J. Solid-State Circuits*, vol. 54, no. 11, pp. 2947–2956, Nov. 2019.

[37] F. Nolet *et al.*, "A 256 pixelated SPAD readout ASIC with in-pixel TDC and embedded digital signal processing for uniformity and skew correction," *Nucl. Instrum. Methods Phys. Res. A*, vol. 949, p. 162891, Jan. 2020.

[38] G. Intermite *et al.*, "Enhancing the fill-factor of CMOS SPAD arrays using microlens integration," in *Proc. SPIE Photon Counting Appl.*, vol. 9504, May 2015.

[39] R. K. Henderson *et al.*, "A 192×128 time correlated SPAD image sensor in 40-nm CMOS technology," *IEEE J. Solid-State Circuits*, vol. 54, no. 7, pp. 1907–1916, Jul. 2019.

[40] T. Frach, G. Prescher, C. Degenhardt, R. de Gruyter, A. Schmitz, and R. Ballizany, "The digital silicon photomultiplier - principle of operation and intrinsic detector performance," in *Proc. IEEE Nucl. Sci. Symp. Conf. Rec.*, 2009, pp. 1959–1965.

[41] T. Frach, G. Prescher, C. Degenhardt, and B. Zwaans, "The digital silicon photomultiplier - system architecture and performance evaluation," in *Proc. IEEE Nucl. Sci. Symp. Conf. Rec.*, 2010, pp. 1722–1727.

[42] L. H. C. Braga *et al.*, "A CMOS mini-SiPM detector with in-pixel data compression for PET applications," in *Proc. IEEE Nucl. Sci. Symp. Conf. Rec.*, Oct 2011.

[43] L. H. C. Braga *et al.* , "A fully digital 8×16 SiPM array for PET applications with per-pixel TDCs and real-time energy output," *IEEE J. Solid-State Circuits*, vol. 49, no. 1, pp. 301–314, Jan. 2014.

[44] S. Mandai and E. Charbon, "A 4×4×416 digital SiPM array with 192 TDCs for multiple high-resolution timestamp acquisition," *J. Instrum.*, vol. 8, no. 05, p. P05024, May 2013.

[45] A. Carimatto *et al.*, "A 67,392-SPAD PVTB-compensated multi-channel digital SiPM with 432 column-parallel 48ps 17b TDCs for endoscopic time-of-flight PET," in *Proc. IEEE Int. Solid-State Circuits Conf.*, Feb 2015.

[46] 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.

[47] E. Manuzzato *et al.*, "A 16×8 digital-SiPM array with distributed trigger generator for low SNR particle tracking," in *Proc. 45th IEEE Eur. Solid-State Circuits Conf.*, Sep 2019.

[48] E. Conca *et al.*, "Large-area, fast-gated digital SiPM with integrated TDC for portable and wearable time-domain NIRS," *IEEE J. of Solid-State Circuits*, vol. 55, no. 11, pp. 3097–3111, 2020.

[49] J. Jansson, S. Jahromi, K. Haapalainen, H. Juttula, A. Mäkynen, and J. Nissinen, "An integrated 9x9 SPAD array with a 10-channel TDC and a CMOS laser diode driver for a wearable time-domain diffuse optics

optode," in *Proc. IEEE Int. Instrum. and Meas. Tech. Conf.*, 2020, pp. 1–5.

[50] C. Niclass, M. Soga, H. Matsubara, S. Kato, and M. Kagami, "A 100-m range 10-frame/s 340×96-pixel time-of-flight depth sensor in 0.18-µm CMOS," *IEEE J. Solid-State Circuits*, vol. 48, no. 2, pp. 559–572, Feb. 2013.

[51] Y. M. Wang, L. Shi, C. Wang, K. O. Kim, I. Ovsiannikov, and S. Hwang, "A direct TOF sensor with in-pixel differential time-to-charge converters for automotive flash lidar and other 3D applications," in *Proc. Int. Image Sens. Workshop*, Jun. 2019.

[52] D. Morrison, D. Delic, M. R. Yuce, and J.-M. Redoute, "Multistage linear feedback shift register counters with reduced decoding logic in 130-nm CMOS for large-scale array applications," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 27, no. 1, pp. 103–115, Jan 2019.

[53] I. Nissinen, A. Mantyniemi, and J. Kostamovaara, "CMOS time-to-digital converter based on a ring oscillator for a laser radar," in *Proc. 29th IEEE Eur. Solid-State Circuits Conf.*, 2003, pp. 469–472.

[54] J. Nissinen, I. Nissinen, and J. Kostamovaara, "Integrated receiver including both receiver channel and TDC for a pulsed time-of-flight laser rangefinder with cm-level accuracy," *IEEE J. Solid-State Circuits*, vol. 44, no. 5, pp. 1486–1497, 2009.

[55] J. Richardson *et al.*, "A 32×32 50ps resolution 10 bit time to digital converter array in 130nm CMOS for time correlated imaging," in *Proc. IEEE Custom Integr. Circuits Conf.*, Sep 2009.

[56] C. Zhang, S. Lindner, I. M. Antolovic, 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 J. Solid-State Circuits*, vol. 54, no. 4, pp. 1137–1151, Apr 2019.

[57] 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," *IEEE J. Solid-State Circuits*, 2020.

[58] Z. Cheng, X. Zheng, M. J. Deen, and H. Peng, "Recent developments and design challenges of high-performance ring oscillator CMOS time-to-digital converters," *IEEE Trans. Electron Devices*, vol. 63, no. 1, pp. 235–251, 2016.

[59] B. Markovic, S. Tisa, F. A. Villa, A. Tosi, and F. Zappa, "A high-linearity, 17 ps precision time-to-digital converter based on a single-stage vernier delay loop fine interpolation," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 60, no. 3, pp. 557–569, Mar 2013.

[60] L. Parmesan, N. A. W. Dutton, N. J. Calder, A. J. Holmes, L. A. Grant, and R. K. Henderson, "A 9.8µm sample and hold time to amplitude converter CMOS SPAD pixel," in *Proc. 44th IEEE Eur. Solid-State Device Research Conf.*, Sep 2014.

[61] L. Parmesan *et al.*, "A 256×256 SPAD array with in-pixel time to amplitude conversion for fluorescence lifetime amplitude conversion for fluorescence lifetime imaging microscopy," in *Proc. Int. Image Sens. Workshop*, Jun. 2015.

[62] S. Jahromi, J.-P. Jansson, P. Keranen, and J. Kostamovaara, "A 32 × 128 SPAD-257 TDC receiver IC for pulsed TOF solid-state 3-D imaging," *IEEE J. Solid-State Circuits*, vol. 55, no. 7, pp. 1960–1970, Jul 2020.

[63] D. Morrison, S. Kennedy, D. Delic, M. R. Yuce, and J. M. Redouté, "A 64×64 SPAD flash lidar sensor using a triple integration timing technique with 1.95 mm depth resolution," *IEEE Sensors J.*, pp. 1–1, 2020.

[64] G. Acconcia, M. Crotti, S. Antonioli, I. Rech, and M. Ghioni, "High performance time-to-amplitude converter array," in *Proc. IEEE NoMe TDC*, 2013, pp. 1–5.

[65] 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," *IEEE Sensors J.*, vol. 18, no. 8, pp. 3163–3173, 2018.

[66] A. T. Erdogan, R. Walker, N. Finlayson, N. Krstajic, G. O. S. Williams, and R. K. Henderson, "A 16.5 giga events/s 1024×8 SPAD line sensor with per-pixel zoomable 50ps-6.4ns/bin histogramming TDC," in *Proc. Symp. VLSI Circuits*, 2017, pp. C292–C293.

[67] S. Bellisai, F. Villa, S. Tisa, D. Bronzi, and F. Zappa, "Indirect time-of-flight 3D ranging based on SPADs," in *Proc. SPIE Quant. Sens. and Nanophotonic Devices IX*, vol. 8268, Jan 2012, pp. 282 – 289.

[68] A. D. Payne, A. P. Jongenelen, A. A. Dorrington, M. J. Cree, and D. A. Carnegie, "Multiple frequency range imaging to remove measurement ambiguity," in *Proc. 9th Conf. Opt. 3-D Meas. Techn.*, Jul 2009.

[69] C. Niclass, C. Favi, T. Kluter, F. Monnier, and E. Charbon, "Single-photon synchronous detection," *IEEE J. Solid-State Circuits*, vol. 44, no. 7, pp. 1977–1989, 2009.

[70] M. L. Hafiane, W. Wagner, Z. Dibi, and O. Manck, "Analysis and estimation of NEP and DR in CMOS TOF-3D image sensor based on MDSI," *Sens. Actuators A, Phys.*, vol. 169, no. 1, pp. 66 – 73, 2011.

[71] S. Bellisai, D. Bronzi, F. A. Villa, S. Tisa, A. Tosi, and F. Zappa, "Single-photon pulsed-light indirect time-of-flight 3D ranging," *Opt. Express*, vol. 21, no. 4, pp. 5086–5098, Feb 2013.

[72] F. Guerrieri, S. Tisa, A. Tosi, and F. Zappa, "Two-dimensional SPAD imaging camera for photon counting," *IEEE Photon. J.*, vol. 2, no. 5, pp. 759–774, 2010.

[73] R. J. Walker, J. A. Richardson, and R. K. Henderson, "A 128×96 pixel event-driven phase-domain ΣΔ-based fully digital 3D camera in 0.13µm CMOS imaging technology," in *Proc. IEEE Int. Solid-State Circuits Conf.*, Feb 2011, pp. 410–412.

[74] D. Bronzi *et al.*, "100 000 frames/s 64 × 32 single-photon detector array for 2-D imaging and 3-D ranging," *IEEE J. Sel. Topics Quantum Electron.*, vol. 20, no. 6, pp. 354–363, 2014.

[75] N. A. W. Dutton *et al.* , "Oversampled ITOF imaging techniques using SPAD-based quanta image sensors," in *Proc. Int. Image Sens. Workshop*, Jun 2015.

[76] N. A. W. Dutton *et al.*, "A SPAD-based QVGA image sensor for single-photon counting and quanta imaging," *IEEE Trans. on Electron Devices*, vol. 63, no. 1, pp. 189–196, 2016.

[77] 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," *IEEE J. Solid-State Circuits*, vol. 55, no. 7, pp. 1762–1775, 2020.

[78] K. Morimoto *et al.*, "Megapixel time-gated SPAD image sensor for 2D and 3D imaging applications," *Optica*, vol. 7, no. 4, pp. 346–354, Apr 2020.

[79] B. Park, I. Park, W. Choi, Y. Na, and Y. Chae, "A 40-m range 90-frames/s CMOS time-of-flight sensor using SPAD and in-pixel time-gated pulse counter," *IEEE Solid-State Circuits Lett.*, vol. 3, pp. 422–425, 2020.

[80] D. Stoppa, L. Pancheri, M. Scandiuzzo, L. Gonzo, G. Dalla Betta, and A. Simoni, "A CMOS 3-D imager based on single photon avalanche diode," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 54, no. 1, pp. 4–12, 2007.

[81] S. Bellisai, F. Guerrieri, S. Tisa, and F. Zappa, "3D ranging with a single-photon imaging array," in *Proc. SPIE IS&T Electron. Imaging*, vol. 7875, Feb 2011, pp. 152 – 157.

[82] 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, no. 1, pp. 155–167, 2016.

[83] L. Pancheri, N. Massari, and D. Stoppa, "SPAD image sensor with analog counting pixel for time-resolved fluorescence detection," *IEEE Trans. on Electron Devices*, vol. 60, no. 10, pp. 3442–3449, 2013.

[84] 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," in *Proc. IEEE Int. Electron Devices Meeting*, Dec 2016, pp. 8.1.1–8.1.4.

[85] T. Al Abbas *et al.*, "A 128×120 5-wire 1.96mm2 40nm/90nm 3D stacked SPAD time resolved image sensor SoC for microendoscopy," in *Proc. IEEE Symp. VLSI Circuits*, Jun 2019, pp. C260–C261.

[86] E. R. Fossum, "What to do with sub-diffraction-limit (SDL) pixels? – a proposal for a gigapixel digital film sensor (DFS)," in *Proc. IEEE Workshop CCDs and Adv. Image Sensors*, June 2005.

[87] J. Ma, D. Starkey, A. Rao, K. Odame, and E. R. Fossum, "Characterization of quanta image sensor pump-gate jots with deep sub-electron read noise," *IEEE J. Electron Devices Soc.*, vol. 3, no. 6, pp. 472–480, 2015.

[88] S. Burri, Y. Maruyama, X. Michalet, F. Regazzoni, C. Bruschini, and E. Charbon, "Architecture and applications of a high resolution gated SPAD image sensor," *Opt. Express*, vol. 22, no. 14, pp. 17 573–17 589, Jul 2014.

[89] A. C. Ulku *et al.*, "A 512 × 512 SPAD image sensor with integrated gating for widefield FLIM," *IEEE J. Sel. Topics Quantum Electron.*, vol. 25, no. 1, pp. 1–12, 2019.

[90] Y. Hirose *et al.*, "A 400×400-pixel 6µm-pitch vertical avalanche photodiodes CMOS image sensor based on 150ps-fast capacitive relaxation quenching in geiger mode for synthesis of arbitrary gain images," in *Proc. IEEE Int. Solid-State Circuits Conf.*, Feb 2019, pp. 104–106.

[91] T. Okino *et al.*, "A 1200×900 6µm 450fps geiger-mode vertical avalanche photodiodes CMOS image sensor for a 250m time-of-flight ranging system using direct-indirect-mixed frame synthesis with configurable-depth-resolution down to 10cm," in *IEEE Int. Solid-State Circuits Conf.*, Feb 2020, pp. 96–98.

[92] D. Bronzi, F. Villa, S. Tisa, A. Tosi, and F. Zappa, "SPAD figures of merit for photon-counting, photon-timing, and imaging applications: A review," *IEEE Sensors J.*, vol. 16, no. 1, pp. 3–12, 2016.