A CMOS SPAD Sensor with a Multi-Event Folded Flash Time-to-Digital Converter for Ultra-fast Optical Transient Capture

A CMOS SPAD Sensor with a Multi-Event Folded Flash Time-to-Digital Converter for Ultra-fast Optical Transient Capture

Tarek Al Abbas*, *Student Member, IEEE*, Neale A.W. Dutton*, *Member, IEEE*, Oscar Almer, Neil Finlayson, *Member, IEEE*, Francescopaolo Mattioli Della Rocca, *Student Member, IEEE*, Robert Henderson, *Senior Member, IEEE*

***Abstract*—A digital silicon photomultiplier (dSiPM) in 130nm CMOS imaging technology implements time-correlated single photon counting (TCSPC) at an order of magnitude beyond the conventional pile-up limit. The sensor comprises a 32x32 43% fill-factor SPAD array with a multi-event folded-flash time to digital converter (TDC) architecture operating at 10GS/s. 264 bins x 16bit histograms are generated and read out from the chip at a maximal 188 kHz enabling fast time resolved scanning or ultrafast low light event capture. Full optical and electrical characterization results are presented.**

***Index Terms*—time to digital converter, single photon avalanche diode, optical sensor, CMOS, silicon photomultiplier.**

I. INTRODUCTION

Optical sensing systems designed to perform the technique of time correlated single photon counting (TCSPC) record ultra-fast optical phenomena with picosecond resolution providing the capability of capturing an optical signal with ultimate sensitivity in both spatial and temporal domains. Applications vary from Positron Emission Tomography (PET) and Fluorescence Lifetime Imaging Microscopy (FLIM) to Time of Flight (TOF) based Laser Detect and Range (LIDAR) [1]. This technique comprises three parts: single photon detection, measurement of photon arrival time, and arrival time data collection and processing. Most TCSPC experimental setups consist of many discrete components: pulsed laser, detector and processing electronics. Discrete high gain single photon sensitive photo-detectors are used such as an avalanche photo diode (APD), a single photon avalanche diode (SPAD) or a photomultiplier tube (PMT). The processing function is accomplished by a front-end circuit (amplifier, comparator etc.), followed by event driven time conversion circuit based on either Time to Analogue (TAC) or Time to Digital conversion (TDC), and finally a PC for data collection and post-processing, most commonly involving histogram generation. Systems based on discrete components are physically large and bulky with high overall cost, limiting their take up to military and scientific applications. Recent efforts leverage CMOS integration to place part, or all, of the TCSPC system on chip [2][3][4] bringing the size and cost down by orders of magnitude and into consumer TOF applications [5].

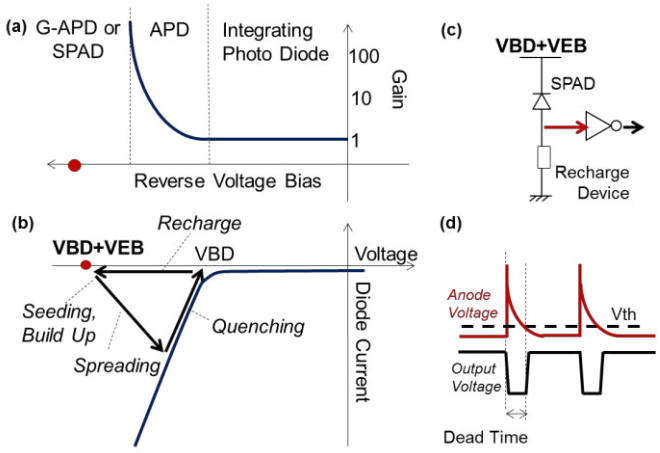

Fig. 1. SPAD Operation at high reverse voltage bias illustrated against (a) photo-detector gain and (b) photo-detector current with the five stages of a SPAD avalanche. (c) SPAD passive recharge circuit with CMOS output inverter and (d) timing waveform of anode and inverter output.

In this paper we examine each of the CMOS TCSPC system components in order to significantly increase the photon conversion rate and system throughput to capture fast optical phenomena. This is achieved by maximizing the operating rate in each of the four primary functions: single photon detection across multiple detectors in an optical sensing array, signal routing and combination logic in the array, time conversion circuitry and finally data processing. Both parallelisation and increased conversion rate are employed with the primary goal to mitigate time-domain pile-up distortion – the key distortion

Paper received 01 January 2017 and published 01 January 2017. This work was supported by STMicroelectronics and has received funding from EPRSC Grant (EP/K03197X/1).

* contributed equally to the work described in this paper.

N. A. W. Dutton is with the Imaging Division, STMicroelectronics, Edinburgh EH12 7BF, U.K. (e-mail: neale.dutton@st.com)

T. Al-Abbas, O. Almer, N. Finlayson, Francescopaolo Mattioli Della Rocca and R. Henderson are with the School of Engineering, University of Edinburgh, Edinburgh, Scotland EH9 3JL, U.K..

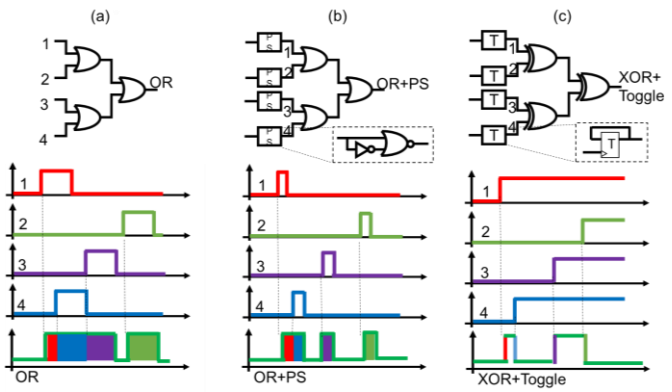

Fig. 2. (a) OR Tree (b) OR tree with monostable pulse shaper input. (c) This work: XOR Tree with toggle flip-flop input.

mechanism in TCSPC systems [6]. In this paper, we present a digital silicon photomultiplier based on a single channel TDC operating at maximum 10GS/s, an increase in TDC conversion rate by an order of magnitude over published works [7]. We report here a revised version of the original sensor [8] with a slightly lower conversion rate, improved PLL linearity and revised pixel routing for better timing performance. Extensive circuit description and characterization results are provided of this new device. We demonstrate high throughput TCSPC operation at beyond ten times the laser repetition rate or over a hundred times the conventional pile-up limit. In addition, we show streak camera like operation, whereby the capture of single-shot few photon transient events occur in nanosecond time scales. The IC has fully parallelised histogram generation logic on-chip to capture the maximal 10Gb/s data rate and substantially reduce the data transmission off-chip to remove any bottlenecks due to data handling and output which would impose a system-level readout pile-up distortion.

II. TCSPC Sensor Design

In this section, the design of each primary system component in a TCSPC optical receiver is discussed.

A. SPAD Array

SPADs have three significant advantages over other optical detectors: single photon detection, picosecond time resolution and ease of integration in modern deep sub-micron (DSM) CMOS allowing arrays of detectors and fast digital signal processing. The SPAD device is a class of APD operating in 'Geiger' mode (G-APD). The three regions of photo-diode operation are illustrated in the gain to reverse bias plot in Fig.1(a): integration, avalanche and 'Geiger' single photon avalanche mode. The SPAD is a reverse biased PN junction biased and operated above its breakdown voltage (V_{BD}) by an excess bias (V_{EB}) shown by the red dot on the I-V plot in Fig. 1(b) and in Fig. 1(c) showing a typical SPAD circuit diagram. Electron-hole pairs generated by absorption of photons may trigger a current avalanche in the active region of the device. The five stages of Geiger mode operation are seeding, current build-up, spreading, avalanche halting by quenching and finally recharge. The duration of the avalanche and recharge is known as SPAD dead time (indicated in Fig.1(d)), as the detector has a reduced sensitivity to incoming photons. It is controllable by the recharge resistance and is on the order of nanoseconds.

While conventional analogue SiPMs sum the currents of individual SPADs in an array, digital silicon photomultipliers (dSiPMs) [9] combine the individual digital pulses from each SPAD into a single pulse train. Both photon count and timing information are then processed by following electronics. The sensitivity of each SPAD pixel in the array is proportional to the fill-factor (the pixel area dedicated to the active photo-region). A major design challenge in dSiPMs is in combining these pulses together and routing them to the time converter whilst maintaining high fill factor and minimum degradation of timing precision.

B. Input Routing and Combination Logic

Fig. 2 illustrates three methods of combining pulses through a logic tree from multiple trigger inputs from an array of SPADs to a single output channel. Fig. 2 (a) illustrates the simplest technique, consisting of an OR tree network. A problem with this method is that simultaneous pulses from different SPADs coalesce during any overlaps in their dead-time. This limits the maximum rate of output events to a rate proportional to the reciprocal of the dead time. By adding a pulse-shortening monostable circuit [10] photon arrivals are represented by the rising edges of shortened output pulses (Fig. 2(b)). This reduces the chance of pulse coalescence and increases the maximum rate of output events by the ratio of SPAD dead time to monostable time. In this work, we propose an asynchronous dual-data-rate (DDR) encoding scheme which uses both rising and falling edges of the pixel output to represent a trigger output. This dual edge approach requires the use of a positive-edge triggered toggle flip-flop attached to each pixel output as shown in Fig. 2(c). Several inputs are combined by an XOR-based combination tree. It achieves at least twice the maximum output rate of the OR-tree architectures; the interested reader is directed to our previous work on comparison between XOR and OR trees [11]. Yet if two events close in time (within a gate delay) try to propagate through the same XOR gate the events will cancel out resulting in loss of data.

C. Multiple Event Time to Digital Converter

All data converters, including TDCs, naturally have a conversion dead time following the input sampling phase. The majority of TAC or TDC circuits for TCSPC have converter dead time limiting the system to one photonic event per laser excitation and time conversion cycle. This converter dead time, after a first input trigger event causes secondary events to be missed. The effect is shown in Fig.3 on an amplitude modulated received (RX) optical signal converted in a TCSPC system suffering from TDC pile-up distortion. TDC dead time is illustrated in Fig. 3(c) showing the missed secondary trigger events.

The toggling XOR tree output, which encodes photon

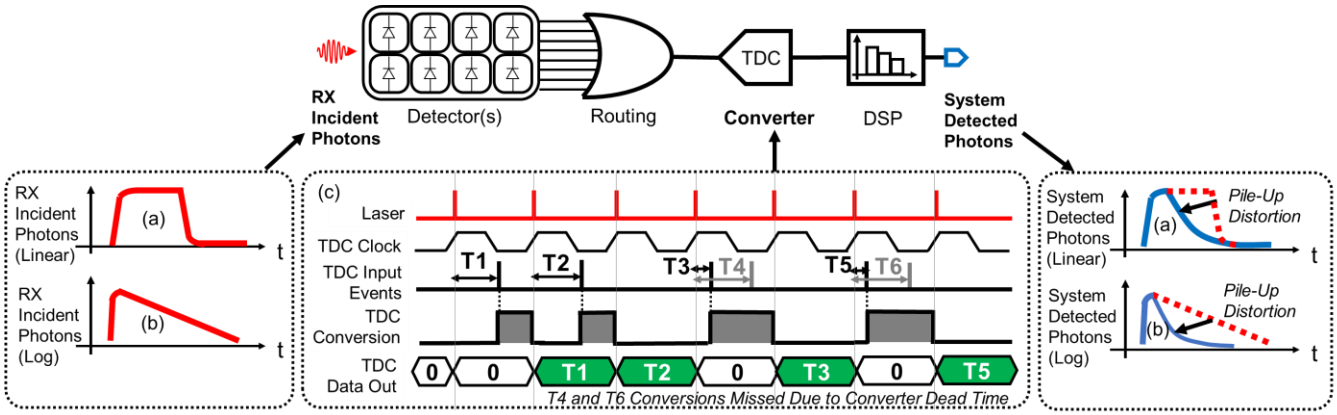

Fig. 3. Time to digital converter pile-up distortion example for incident photon and TCSPC system output for (a) a square pulse or (b) an exponential decay input to (c) a typical time converter circuit with converter dead-time after an input trigger event.

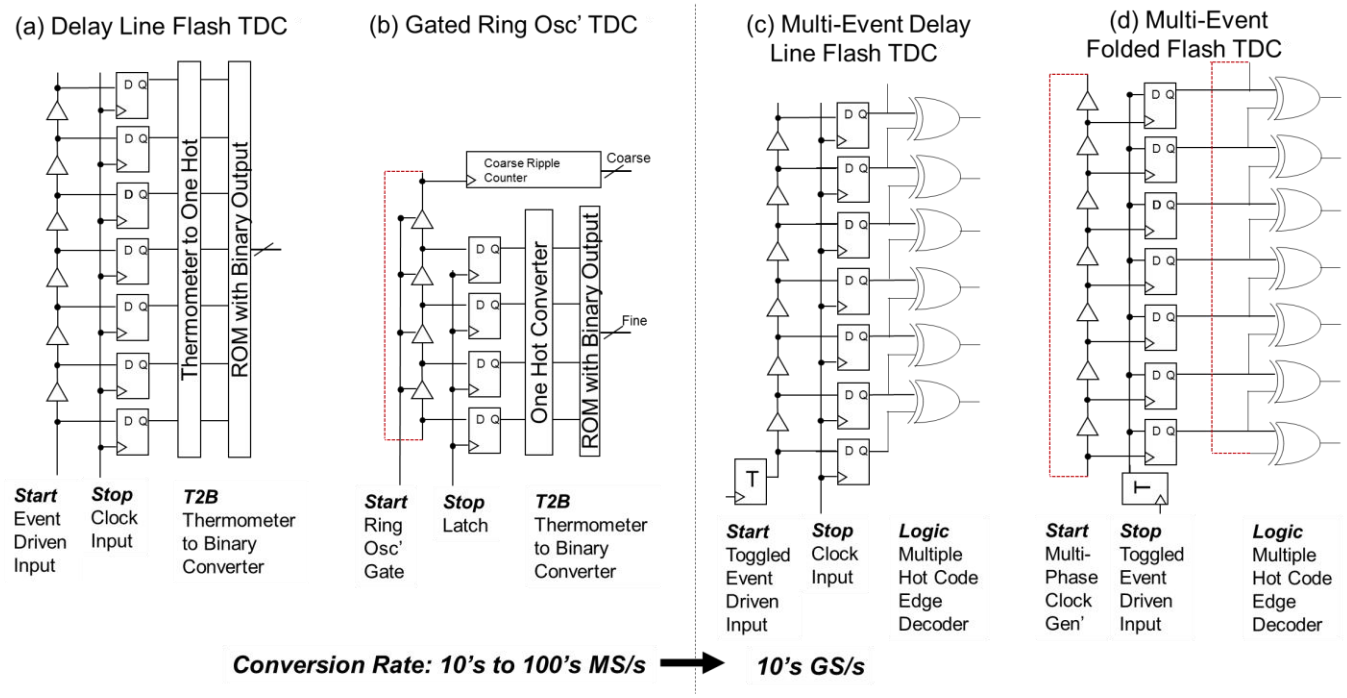

Fig. 4. TDC Architectures. (a) Flash delay-line TDC (b) Ring Oscillator TDC (c) Our previous work: the multiple event flash delay line TDC [12] (d) This work: the multiple event folded flash TDC. All delay chains shown can be embedded within a DLL or PLL for improved jitter performance.

arrivals on both positive and negative edges, motivates a new design of TDC capable of converting multiple events per laser cycle. The following section details the architecture of a so-called *multi-event folded-flash TDC* which eliminates converter dead time and attains conversion rates of the reciprocal of a gate delay.

Flash and ring-oscillator (RO) are two fundamentally similar architectures of TDC using the propagation delay time of a logic gate (inverter or buffer) to provide a timing window forming the Least Significant Bit (LSB) of the time to data conversion [13]. As illustrated in Figs. 4(a) and (b), an open-loop delay line or closed-loop ring of delay gates is employed, where each delay element is tapped to a sampling flip-flop.

In both flash and RO TDC architectures, an appropriate thermometer to binary (T2B) or one-hot thermometer to binary converter is employed to output a binary time-stamp value. This output code represents an integer number of timing windows between the input signal and the reference clock (or to use the common stopwatch analogy the 'start' and 'stop' inputs). One 'start' signal may be converted per 'stop' or conversion period. In the delay-line flash TDC, the input 'start' triggers a sequence of logical high transitions down the delay chain until the 'stop' signal captures the position of the rising edge. This binary converter then decodes the input from the TDC sampling flip-flops. By inspection the maximum rate of this logic converter is one conversion per clock period. As a result, the T2B logic converter imposes a significant restriction on the rest of the system: to avoid pile-up distortion the input rate must be less than the converter rate. Recent TDC examples go some way to overcome this by simply increasing

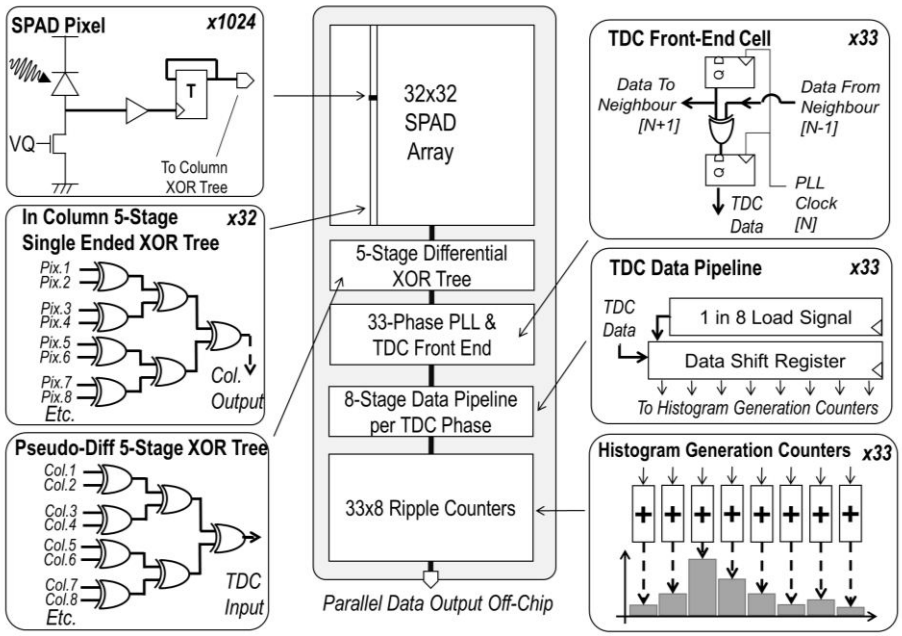

Fig. 5. Block diagram of the TCSPC receiver IC comprising: 1024 SPAD Pixels in a 32x32 Array with embedded column parallel 5-stage single-ended XOR tree. A second five-stage XOR tree connects to the input of the 33-phase folded flash TDC with 8b shift-register based pipeline connecting to integrating ripple counters forming the 264-bin histogram on-chip.(change figure to blue SPAD)

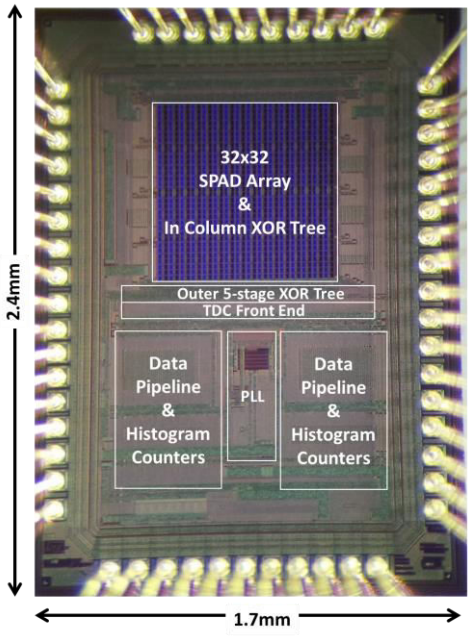

Fig. 6. Photomicrograph of the fabricated TCSPC receiver IC in STMicroelectronics imaging 130nm process.

the clock rate from 300MS/s [14] to 500MS/s [15]. However, a rate improvement step can be taken by redesigning the logic converter.

Considering first the delay-line flash TDC in Fig. 4(c), multiple input events per conversion cycle may be encoded using the toggled input encoding scheme. A first 'start' signal raises the input to the delay line causing a rising edge to propagate down the delay line. Then a second 'start' signal drops the input low forming a propagation of a falling edge, followed by a third input 'start' signal triggering a rising edge and so on. This may be encoded by a toggle flip-flop (as illustrated in the diagram) for a single input or by a toggle and XOR tree for more than one input. Multiple events are allowed to propagate through the delay line during a conversion cycle as the sampling flip-flops will capture the state of the multiple triggers across the delay line at the assertion of the clock pulse. Once the flip-flops have sampled the state of the delay line, each input trigger event is represented as a logical transition: '0→1' for a positive edge, '1→0' for a negative edge. An XOR gate per sampling flip-flop is employed to decode the transition. The bank of logic decoder gates in effect become a fully parallelised multiple-hot decoder, where the *positions* of each of the logical high outputs in the decoder represent the times of arrival of each photonic event. For comparison, the conversion rate in the delay line flash TDC with T2B logic converter in Fig. 4(a) is the sampling clock rate. Yet for the same delay line and sampling flip-flops with multiple edge decoding, shown in Fig. 4(c), the conversion rate is the reciprocal of a gate delay. As a comparison, our recent FPGA-based multiple event delay line flash TDC, with multiple hot output decoder, achieved 6.2GS/s [12] measured conversion rate over a conventional version at 300MS/s [14] using a similar reference clock (316MHz to 300MHz respectively) on the same Xilinx Virtex 5 FPGA architecture.

The RO TDC is also the basis for the new TDC design, detailed in this work, to encode multiple trigger inputs per reference clock period shown in Fig. 4(d). A multi-phase RO embedded in a PLL is connected to the TDC front end flip-flops. The same critical rate-increasing step is applied at the output of each flip-flop by replacing the T2B converter with a multiple hot edge detector. As the ring-oscillator is closed-loop, so also is the edge detector logic which folds the last sampling flip-flop output back to the input of the first XOR logic decoder, thus forming a *folded flash* TDC. By inspection, the LSB temporal resolution of this folded flash TDC is the time-difference between clock phases. The conversion rate of the folded flash TDC with dual-edge decoding, is the reciprocal of the time-difference between clock phases (t_{LSB}). For example, a 100ps temporal difference between clock phases creates a 10GS/s conversion rate. On the other hand, the dynamic range (the total temporal span that the front end flip flops can sample) is one clock period of the PLL. If there are 'N' clock phases then the front end dynamic range is N x t_{LSB}. Increasing the PLL clock rate to lower the time-resolution impacts the dynamic range, so to overcome this trade-off, a shift register is attached to each edge detector output forming a parallel bank of shift registers. Each shift-register clock is connected to the same clock phase as its companion TDC sampling flip-flop. The shift register length is "M" bits per clock phase, hence increasing the total dynamic range to M x N x t_{LSB}. This is illustrated in Fig. 7. Capturing and processing the Gb/s output data across multiple clock

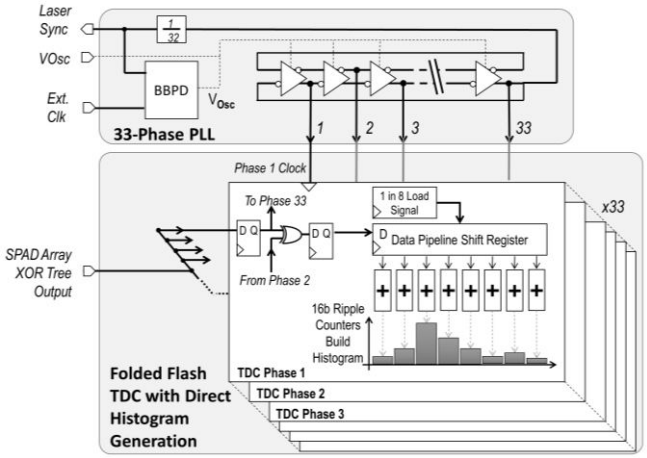

Fig. 7. Thirty-three multi-phase PLL connected to 33-phase folded flash TDC (N=33). Each TDC phase contains front-end sampling flip-flop, dual-edge detection logic connecting to the neighboring TDC phase, 8b shift register (M= 8) and 8 ripple counters directly generating a 264 bin histogram (M x N = 264) from the parallelised TDC output.

phases becomes a design challenge detailed in the next section.

D. Direct Histogram Generation Logic

In TCSPC systems, a common first step of data handling is to collect the TDC binary "time-stamp" data output codes into a histogram. An existing technique is to employ a RAM with an address per histogram bin [19]. The time-stamp data is used to address one RAM location with the respective memory location data value containing the integrated count for that bin. The three step histogram generation process is: first, RAM address look-up, second, increment the value by one, and third, write the new value. This process can be optimized by using a dual port RAM and address pipelining, so that the data read and write are pipelined to function in a single clock cycle [16]. Nonetheless this serial process is a rate-limiting factor in system throughput, especially if the system is partitioned with separate IC's for the TDC and histogram generation and the connection between these two IC's is I/O rate limited.

An alternative approach to histogram generation is proposed in this work, involving the addition of a counter to each of the parallelised shift-register outputs. This results in direct creation of the M x N bin histogram on-chip. Instead of a multi-bit binary value encoding the TDC data for a single event, the multiple-hot edge detector and shift register outputs are single bit, whereby the positions of the output and the respective connected counter denotes the TDC value.

The TDC is operated for an exposure time to build up the histogram, and following this the integrated values of the bank of counters are read sequentially off-chip. This on-chip integration of the TDC Gb/s output into a parallel counter bank significantly reduces the I/O data rate and associated power consumption.

III. Silicon Implementation

Fig. 5 illustrates a block diagram of the TCPSC sensor with 32x32 SPAD array, 10-stage XOR combination tree, 33-phase PLL (with VCO operating at typical 303MHz), 10GS/s folded flash TDC, shift-register based data pipeline and matching 10Gb/s, 264-bin histogram generation block formed by parallelised ripple counters. Fig. 6 shows the photomicrograph of the fabricated IC manufactured in STMicroelectronics imaging 130nm technology.

The optical detector array is formed by 32x32 SPAD pixels, each consisting of a front-end buffer connected to a toggle flip-flop. The pixel pitch is 21µm with 43% fill factor. In each array column, a five stage single-ended XOR tree logically combines all 32 SPAD pixel toggles into a centrally tapped output. This output is routed to the top of the array for connection to test functions (an on-chip counter and an output pad not shown in the diagram) and to the bottom of the array to the TDC. Each of the 32 column outputs connects to the pseudo-differential five-stage XOR tree. The final XOR stage connects to the data input of the front-end TDC flip-flops. Effectively, in Fig 4.(c) and (d) the toggle flip flop on the input is replaced by the Toggle and XOR Tree. The TDC is formed by 33 parallel flip-flops with common data input from the SPAD array, and each has a separate individual clock phase input generated from the 33 phase PLL operating at a range of 256 to 303MHz per phase. The data output from the flip-flops is passed through the multiple hot code edge detector logic converter. As displayed in Fig. 7, the PLL is a conventional multi-phase structure. The VCO is designed with a ring of inverters each with a minimum of 70ps gate delay in simulation (maximum VCO control voltage) providing a 432MHz PLL phase frequency, and typically 105ps (measured mid-range VCO control voltage) with 288MHz PLL frequency and 1/32 divider ratio. The output of the multiple event logic converter is re-sampled by a pipeline flip-flop and connected to an 8b shift register both on the same clock phase as the front-end flip flop. A 1 in 8 load signal is generated by a similar linear shift register. Each parallel output from the serial to parallel register is attached to a ripple counter representing one histogram bin. The dynamic range of the folded front end is extended 8 times by the shift register where each of the 33 front end flip-flops connects through its own shift register to 8 ripple counters with 16b depth, forming a total of 264 (33 × 8) histogram bins. Due to the folded nature of the front-end, the histogram is also folded such that the final histogram bin is contiguous with the first histogram bin.

The PLL reference clock input derives from an off-chip source typically an FPGA or laser trigger. A selectable PLL feedback divider provides a range of frequencies for the oscillator or laser trigger. A TDC gating signal is generated by a controlling FPGA which permits an integration of multiple laser repetitions to build up a TCSPC histogram. Exposure times can be as short as a single histogram cycle (27.78ns) to capture rapid single-shot transient optical events. In TCSPC mode the sensor is typically operated for many thousand cycles to build a histogram from sparse photons stimulated by

a synchronous laser source. At the end of an exposure cycle, the off-chip data readout is via a 16b parallel bus and one ripple counter is read per clock cycle. The off-chip readout operates at a maximum of 50MHz or 800Mb/s transferring a histogram (528 bytes) in a minimum of 5.3 µs to the FPGA for subsequent data transfer to PC.

IV. Measurement Results

Electrical and optical characterization of the sensor IC is performed across a range of measurements and the results are presented in this section. Unless stated otherwise, for all measurements herein the PLL was locked to a 9MHz clock (VCO-midrange) resulting in a temporal resolution of 105ps, temporal dynamic range of 27.78ns and sampling rate of 10GS/s.

A. TDC Linearity

The optical statistical code density test is performed to estimate the TDC linearity [17]. Continuous wave light from an LED biased with a DC source provides uncorrelated SPAD trigger events or white temporal noise to the TDC. The Differential and Integral Non-Linearity (DNL/INL) across the 33-phase TDC front end is characterized and shown in Figs. 8(a) and 8(b) respectively. The worst-case measured DNL is +0.53/-0.47 LSB and INL is +0.5/-0.19 LSB. The two spikes, at bins 9 and 26, are systematic and are attributed to the VCO layout comprising two interleaved banks of inverters and specifically the two routes at both ends connecting the two banks. Using the sensor for TCSPC applications, the non-linearity may be compensated off-chip by scaling the individual bin values by the inverse of the DNL.

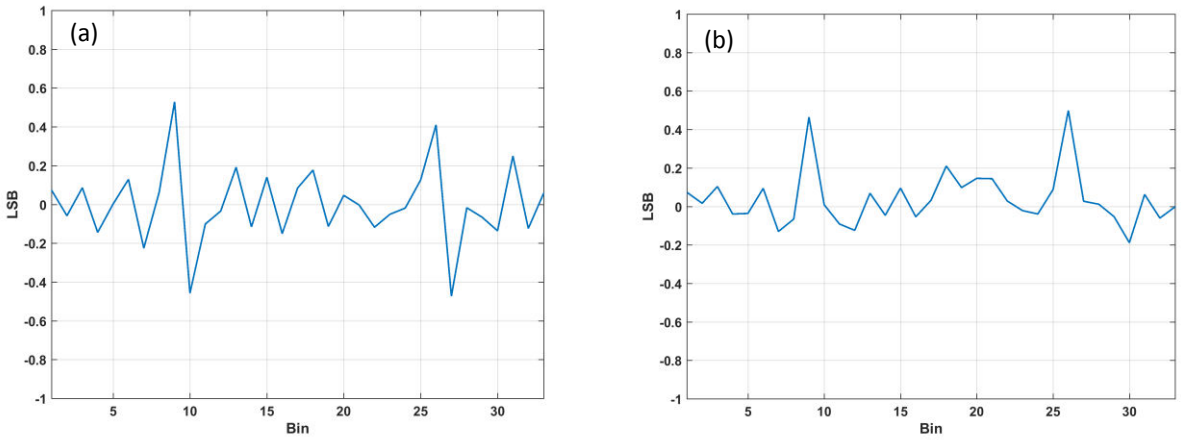

Fig. 8. Sensor overall non-linearity from code density test (a) DNL Max/Min: +0.53LSB/-0.47LSB (b) INL Max/Min: +0.50LSB/-0.17LSB.

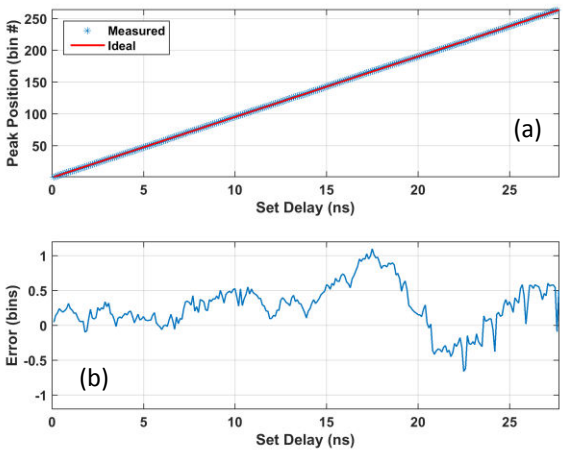

Fig. 9. (a) Pulsed laser IRF sweep (b) the calculated center histogram peak position against delay generator setting.

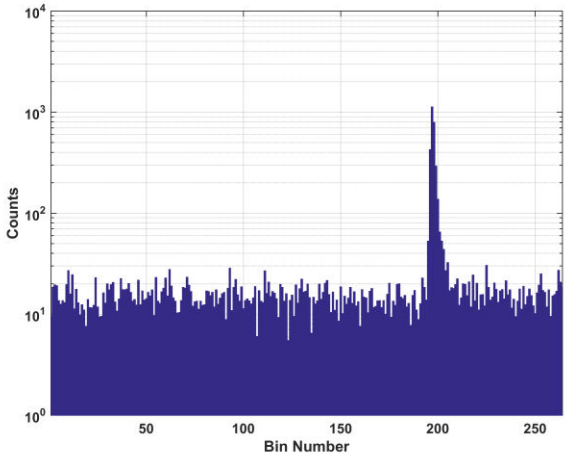

Fig. 10. Typical optical TCSPC histogram (log scale) from the sensor with 231ps FWHM IRF with a single SPAD enabled.

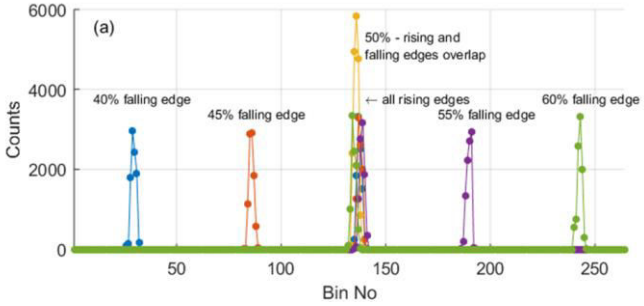

Fig. 11. Synchronous TDC electrical input with duty cycle variation from 40% to 60% duty cycle in 5% increments.

To characterize the TCSPC accuracy and precision of the sensor and to confirm the dynamic range, the PLL is configured as in the previous experiment and the TDC input is multiplexed back to the SPAD array. The laser synchronization trigger connects to a Stanford DG645 delay generator with minimum 5 ps time-step resolution to perform a time-sweep. This in turn connects to a Hamamatsu PLP10 driver with 443nm output head with 70ps FWHM quoted electrical integrated jitter. Both IC and laser face a fixed-distance white diffuser with no optical lensing.

An incremental delay sweep is performed capturing one histogram per 100ps delay step with 20ms exposure time. The centroid of the histogram peak is calculated by the Centre of Mass Method (CMM). The calculated average bin position is plotted against absolute delay in Fig. 9(a) and the error of calculated peak position to absolute delay in Fig. 9(b) indicating a TCSPC precision (σ) of 34ps and accuracy of +116ps, -70ps across the 27.78ns full-scale dynamic range. No dead zone is evident in the dynamic range confirming that the folded flash TDC architecture has its last histogram bin contiguous with the first bin.

A typical histogram from one 10ms exposure is shown in Fig. 10, and a Gaussian-fitted curve indicates a FWHM of 231ps with a single SPAD enabled. Subtracting the external sources of jitter and the SPAD jitter of ~200ps [23] in quadrature yields a similar result to the electrical test of 103ps. Table 1 shows how this optical IRF broadens as the number of enabled SPADs is increased related to path mismatches inside the XOR tree. There is room for design improvements by balanced routing and matched gate rising and falling edge delays to tighten this distribution as shown in [24].

Number of SPADs	Mean FWHM (bins)	Mean FWHM (ps)
1x1	2.2	231
2x2	2.6	273
4x4	2.8	294
8x8	3.4	357
16x16	4.9	514.5

Table 1. Relationship between the number of SPADs enabled in contiguous blocks and the mean FWHM IRF over all such blocks stepped across the whole 32x32 array. The IRF broadens due to the delay mismatch between pixels.

B. Electrical IRF and Linearity Measurements

To confirm the bin and integrated jitter resolution, the TDC and histogram generation logic are first tested employing a dual-channel LeCroy Wavestation 3082 signal generator. The first generator channel is used to generate the PLL reference clock and the second channel generates a synchronous electrical test signal with variable duty cycle.

For finer temporal resolution, in this experiment the reference clock is set at 9.475MHz and 1/32 feedback divider sets the PLL VCO frequency at 303.2MHz. Corresponding to a VCO period of 3.3ns and TDC dynamic range of 26.38ns where each histogram bin has a width of approximately 100ps. The TDC's input is now connected to an electrical test signal with a variable duty cycle from 40% to 60% in 5% increments at 9.475MHz such that 4 TDC conversions are performed per cycle of the TDC. The signal generator positive edge is fixed and synchronously aligned with the PLL reference clock, whereas the falling edge is varied. Rising edges are captured on the first TDC cycle and falling edges on the third TDC cycle. Fig. 11 shows histograms corresponding to rising and falling edge positions for the 5 different duty cycle settings. The rising edge for all settings remains static at the middle bin while the falling edge appears offset by the duty cycle delay deviation from 50%. Peak counts at exactly 50% duty cycle condition are a factor of two greater than for other duty cycle values as rising and falling edges coincide. A Gaussian fit on each of the histogram peaks yields an average 287ps FWHM, subtracting in quadrature the oscilloscope-measured signal generator jitter leaves a 103ps FWHM mean integrated electrical TDC jitter which is in the order of a single histogram bin.

C. Saturation Measurements

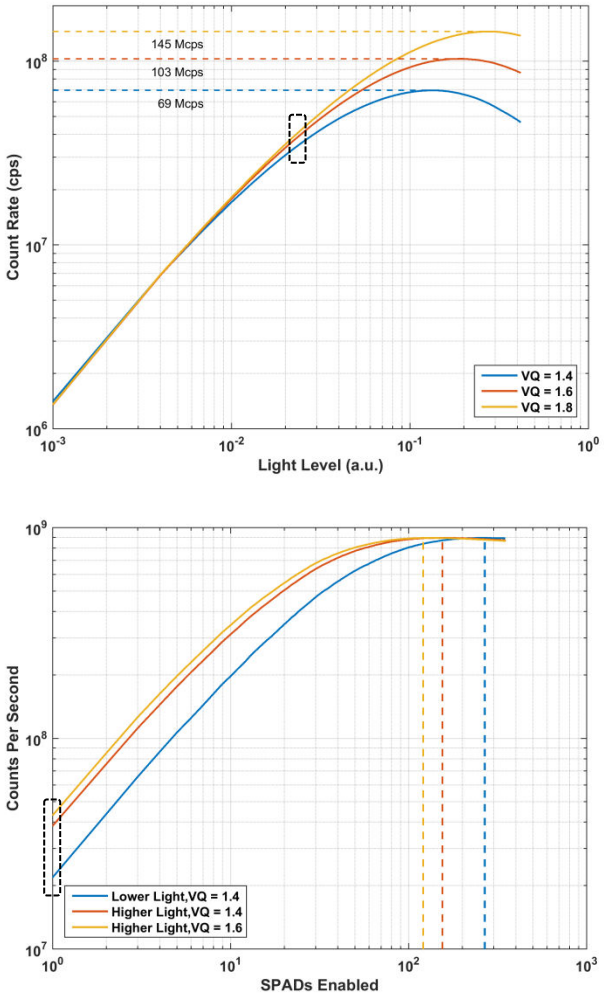

Fig 12. (a) Saturation limit of a single SPAD versus quench voltage (b) saturation limit of XOR tree versus number of SPADs enabled for different quench and light settings. Dotted lines indicate the number of SPADs needed to saturate the XOR channel. At higher ambient conditions, this number is influenced by the quench voltage as the SPADs are operating close to their saturation limit and experience different count rates as indicated by the dotted box on both graphs.

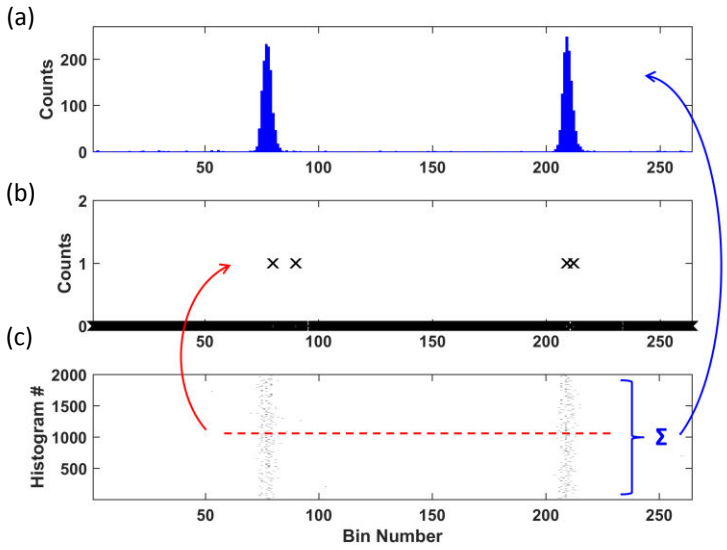

Fig. 13 (a) a histogram formed by summing 2,000 single-shot 30ns exposures (b) one example single-shot 30ns exposure showing 4 multiple photons captured where no TDC dead time is evident as two photons are captured in neighboring bins (c) 2,000 successive single-shot histograms.

Fig. 12(a) shows saturation limits for a single passively quenched SPAD, as VQ increases the dead time decreases resulting in a higher maximum count rate (saturation level). For a quench voltage VQ=1.6V the maximum count rate is around 100MHz, this is equivalent to a dead time around 3.5ns. Fig. 12(b) shows the saturation limit of the XOR tree combining several SPADs as 900MCounts/s related to the bandwidth limited by the parasitics of the metal lines [11].

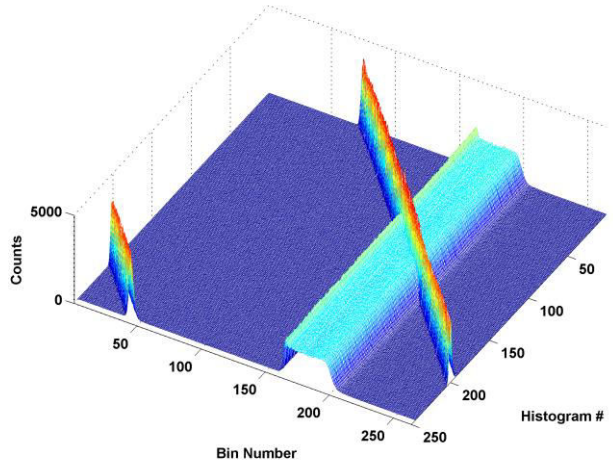

Fig. 14. Incremental delay sweep of one laser against a fixed delay of a second laser. The 3D plot has histogram bin on the x-axis, histogram frequency or count on the z-axis, and successive histograms on the y-axis.

D.Multiple Photons per Laser Excitation Cycle

To demonstrate the capability of the TCSPC sensor to capture multiple photons per laser excitation cycle, a laser repetition rate of 72MHz is selected, corresponding to pulse intervals of 13.89ns, or 2 laser pulses per one histogram cycle (27.78ns). This is achieved by locking the TDC PLL to a 9MHz clock from channel 1 of a Keysight 33250A function generator while using channel 2 to trigger the laser. Fig. 13(a) shows 2000 successive single shot histograms (arrayed vertically) captured with a single 30ns exposure (limited by firmware) where each dot represents a single time correlated photon. Fig. 13(b) displays one example histogram of one of the exposures demonstrating 4 photons captured. No TDC dead time is evident as two photons are captured in neighboring bins. The upper graph in Fig. 13(c) is a summation of the 2,000 single exposures, in effect a 60µs exposure histogram, showing the dual peaks from the lasers. The broader IRF observed for the two peaks is due to the frequency deviation between the channels of the function generator causing the TDC operation and the laser trigger to be slightly out of sync.

A second experiment is performed with a second laser added to the experimental setup, adding an incremental delay to this laser with respect to the first. Fig. 14 shows the successively captured histograms as a 3D plot, with histogram bin on the X-axis, histogram frequency on the Z-axis, and successive histograms from the incremental delay on the Y-axis. As the lasers cross each other in the histograms, there is no pile-up distortion evident which would be revealed as a reduction in the peak intensity of the two pulses when closely spaced in time.

E. Fast Temporal Dynamics and Sensor Bandwidth

Like image sensors, TCSPC sensors capturing fast temporal dynamics of an imaged subject require fast frame and readout rates. The Nyquist sampling rate of image sensors is simply the reciprocal of two frame periods. However, with the event-driven nature of single photon sensing, the trade-off is the exposure time needed for the target application to achieve the requisite number of captured photons per frame (or histogram

in this case) against the desired frame-rate for a high sampling rate.

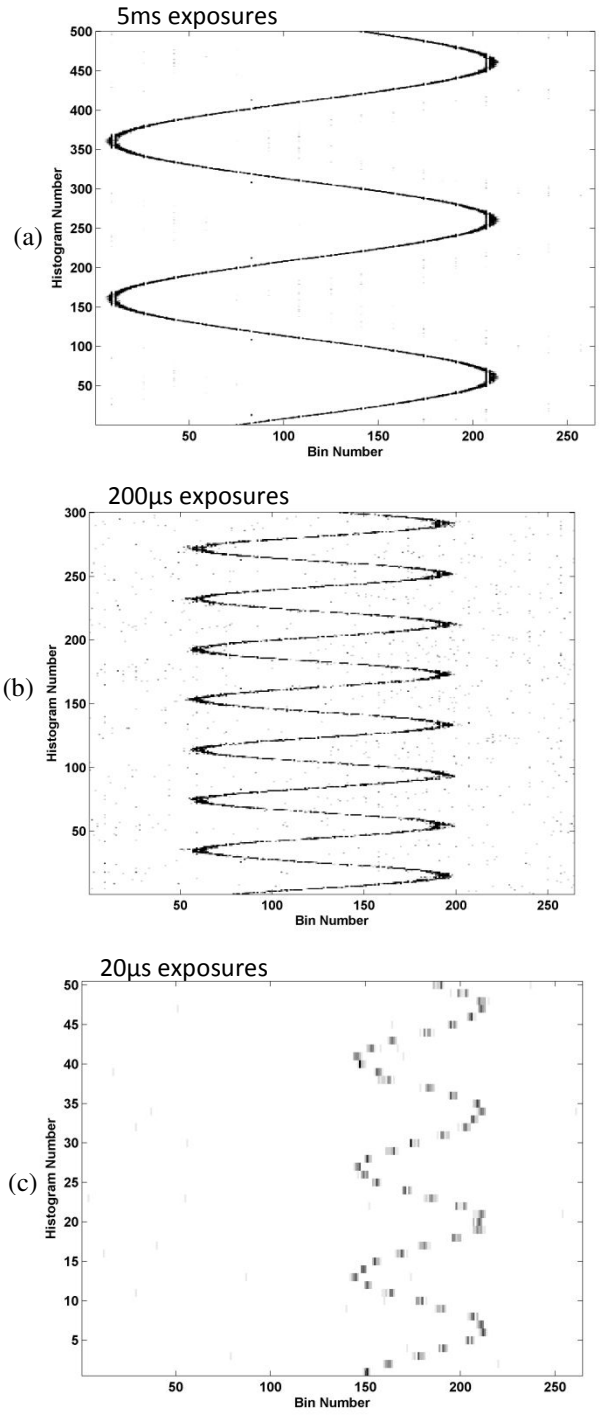

Fig 15. Bandwidth experiments with a sinusoidally modulated pulsed laser (a) 1 Hz with 5ms exposure time (b) 1KHz with 200µs exposure time (c) 10KHz with 20µs exposure time.

A sensor bandwidth experiment is performed to test these limits, with three examples shown in Fig. 15 displayed as a sequential continuous time-series of captured histograms. A Keysight 33250A arbitrary waveform generator is used to generate a 9MHz lock signal for the sensor PLL whilst a second output is frequency modulated around 9MHz and used to trigger the Hamamatsu PLP10 443nm laser. Fig. 15(a) shows a 1 Hz, frequency deviation 0.6Hz and exposure time 5ms, and clearly resolved laser peak tracking the sinusoidal waveform. Fig. 15(b) shows a 100 Hz input with 9 Hz deviation and 200 µs exposure waveform with lower peak height and more DCR in evidence. The final plot in Fig. 15(c) is of a 1 kHz input with 400Hz deviation and exposure time of 20 µs. At this point the readout dead time of 5.3 µs imposes gaps between samples and the waveform is more discretized. The laser can now be seen within a histogram to be streaking across bins at rates determined by the sine wave rate of change. Capture of this type of signal in a time of flight context represents high rates of velocity and acceleration [18]. The maximum sampling rate of the sensor is measured at 188kHz limited by the 5.3 µs readout time.

F. Power Measurement

Table 2 details the power consumption of each constituent part of the sensor giving a total of 170.6mW consumption at 10GS/s with conversion rates of 1GPhoton/s recorded and limited by the output rate of the XOR tree. This equates to a Figure of Merit (FOM) of 0.48pJ per TDC sample (S) or TCSPC time-stamp considering only the TDC front-end and PLL, and 16.1pJ/S for the whole optical sensor not including I/O and 178.8 pJ per photon with power measured at 899.1M photons per second.

System Block	Measured Power (mW)
TDC Pipeline and Histogram Counters at 10GS/s	144
I/O Pads at 50MHz	10
Second stage XOR tree (Stages 6-10)	10
TDC Front End at 10GS/s (including multiple edge detector)	2
PLL	2.4
Column-wise first stage XOR tree (Stages 1-5)	2.4
Total	**170.6**

Table 2. Power Consumption Breakdown

V. Discussion

The sensor is evaluated in a side by side comparison in Table 3. This work achieves the highest single channel TDC conversion rate of any CMOS sensor ASIC implementation for TCPSC. The single shot time resolution of 105ps and full scale range of 27.78ns are moderate in comparison to other works but adequate for many applications such as fluorescence lifetime or distance measurement. The temporal dynamic range (DR) of this architecture is constrained by the available histogram memory area as we require Y (number of bins) × Z (bin bit depth) × A (unit area per bit) of silicon real estate. In this work we have used ripple counters for the histogram memory while future implementations can take advantage of high density SRAM elements [19] which

	This Work	[4]	[20]	[15]	[21]	[22]
TC Architecture	Folded Flash TDC	Col // Flash TDC	Interleaved GRO-TDC	Interleaved Vernier TDC	Oversampling SRO-TDC	Column Parallel GRO-TDC
Application	TCSPC	TCSPC	TCSPC	TCSPC	ADPLL	TCSPC
Interleaved TDCs per Channel	1	4*	16	2	1	1
Parallel Channels	1	96	1	1	1	512
Data Processing on Chip	Histogram	Histogram	Fluorescence Lifetime	-	Mean Time Difference	Histogram
Tech.	130nm	180nm	130nm	130nm	90nm	130nm
Supply	1.2V	1.8V	1.2V	1.3V	1V	1.2V
Single Shot Resolution	100ps	208ps	52ps	31ps	156ps	50ps
Dyn. Range	26.8ns	853ns	3.6µs	2ns	2 to 840ns	3.2µs
External Cal. Needed	No	No	Yes	Yes	No	Yes
System Conv. Rate	10GS/s	6GS/s **	100MS/s	500MS/s	750MS/s	16.5GS/s (in Histogramming mode) 194MS/s (in TCSPC mode)
TDC Conv. Rate		62.5MS/s	12.5MS/s	250MS/s		32.2MS/s
Sensor Bandwidth	30kHz	60Hz **	-	-	-	-
Off-Chip Sensor Rate	188k Histogram / s	30 FPS	60k	500 M Ph/s	-	6.06M Histogram /s (194MS/s in TCSPC mode)
TDC Power	TDC 2 mW + PLL 2.4mW	-	1.8mW	1mW	2mW	1.8mW
TDC FOM1*	0.48 pJ / S	-	18.0nJ / S	2.00nJ / S	2.67nJ / S	36.0 nJ / S

Table 3. CMOS TDC Comparison Table * FOM1 = Power / Conversion Rate = J / Time Stamp.

significantly reduce the area overhead. Alternatively, the DR can be extended by reusing the same memory resources either by trading off temporal resolution for DR, or by employing time offsetting techniques (zooming) [22] to reallocate the memory to the temporal region of interest of the full DR. The extension of TCSPC to high sample rates is expected to improve the linearity of distance measurement by reducing pile-up distortion with high signal levels or tracking of high velocity objects. The sensor is expected to have significant advantages in scanning confocal microscopy by improving the image dynamic range in intensity and time-resolved modes. Used in a single shot mode the chip can operate as a solid-state streak sensor digitizing few-photon fast transient events.

VI. Conclusion

A TCPSC optical sensor is described with a novel multi-event folded-flash TDC capable of both ultra-fast conversion rate and direct histogram generation on-chip. The sensor breaks the traditional limitation of conventional TDCs allowing multiple time events to be digitized per master clock period. Coupled to a SPAD array, this TDC enables ultrafast low-light transient phenomena capture as well 3D ToF scanning systems with extended distance dynamic range.

VII. Acknowledgements

Salvatore Gnecchi contributed to the design of the sensor. Technical discussions with Bruce Rae, Sara Pellegrini and Kevin Moore have been influential in this research. We are grateful to STMicroelectronics for silicon fabrication and PhD student support for Francescopaolo Mattioli Della Rocca. Tarek Al Abbas acknowledges funding from The University of Edinburgh and PROTEUS project (http://proteus.ac.uk EPSRC grant number EP/K03197X/1)

VIII. References

[1] E. Charbon, “Single-photon imaging in complementary metal oxide semiconductor processes,” Phil. Trans. R. Soc. A, vol. 372, Feb. 2014.

[2] J. Richardson, R. Walker, L. Grant, D. Stoppa, F. Borghetti, E. Charbon, M. Gersbach, and R. K. Henderson, “A 32x32 50ps resolution 10 bit time to digital converter array in 130nm CMOS for time correlated

imaging," Cust. Integr. Circuits Conf. 2009. CICC '09. IEEE, pp. 77–80, 2009.

[3] D. Tyndall, B. R. Rae, D. D.-U. Li, J. Arlt, A. Johnston, J. A. Richardson and R. K. Henderson, "A High-Throughput Time-Resolved Mini-Silicon Photomultiplier with Embedded Fluorescence Lifetime Estimation in 0.13µm CMOS," IEEE Transactions on Biomedical Circuits and Systems, 6, 562–570 (2012).

[4] C. Niclass, M. Soga, H. Matsubara, S. Kato, and M. Kagami, "A 100-m Range 10-Frame/s 340x96-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.

[5] STMicroelectronics, VL6180 Datasheet (http://www.st.com/content/st_com/en/products/imaging-andphotonics-solutions/proximity-sensors/vl6180x.html).

[6] W. Becker, "Advanced Time-Correlated Single-Photon Counting Techniques", Springer, Berlin/Heidelberg/New York, 2005.

[7] Z. Cheng, X. Zheng, D. Palubiak, M. J. Deen and H. Peng, "A Comprehensive and Accurate Analytical SPAD Model for Circuit Simulation," in IEEE Transactions on Electron Devices, vol. 63, no. 5, pp. 1940-1948, May 2016.

[8] N. A. W. Dutton et al., "A time-correlated single-photon-counting sensor with 14 GS/S histogramming time-to-digital converter," in IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers, Feb. 2015, pp. 1–3.

[9] T. Frach et al., "The Digital Silicon Photomultiplier - Principle of Operation and Intrinsic Detector Performance," 2009 IEEE Nucl. Sci. Symp. Conf. Rec., pp. 1959–1965, Oct 2009.

[10] L. H. C. Braga et al., "A Fully Digital 8x16 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, 2014.

[11] S. Gnecchi et al., "Digital Silicon Photomultipliers with OR/XOR pulse Combining Techniques," IEEE Trans. Electron Devices, vol. 63, no. 3, pp. 1105–1110, Mar. 2016.

[12] N. Dutton et al., "Multiple-event direct to histogram TDC in 65nm FPGA technology", Proc. IEEE PRIME, pp. 1-5, Jun. 2014.

[13] G. Roberts, M. Ali-Bakhshian, "A brief introduction to time-to-digital and digital-to-time converters", IEEE Trans. Circuits Syst. II Exp. Briefs, vol. 57, no. 3, pp. 153-157, Mar. 2010.

[14] A.S. Yousif, et al., "A Fine Resolution TDC Architecture for Next Generation PET Imaging" IEEE Trans. Nuclear Science, vol. 54, no. 5, pp. 1574-1582, Oct. 2007.

[15] A. Elshazly, et al., "A 13b 315fsrms 2mW 500MS/s 1MHz Bandwidth Highly Digital Time-to-Digital Converter Using Switched Ring Oscillators," ISSCC Dig. Tech. Papers, pp. 464-465, Feb. 2012.

[16] T. P. Haraszti, CMOS Memory Circuits, MA, Norwell:Kluwer, 2000.

[17] J. Doernberg, H.-S. Lee, and D. A. Hodges, "Full-speed testing of A/D converters," IEEE J. Solid-State Circuits, vol. SSC-19, no. 6, pp. 820–827, Dec. 1984.

[18] N. Finlayson, T. Al Abbas, F. Mattioli Della Rocca, O. Almer, S. Gnecchi, N. A. W. Dutton, R. K. Henderson, "Hypervelocity time-of-flight characterisation of a 14GS/s histogramming CMOS SPAD sensor," Proc. SPIE 10111, Quantum Sensing and Nano Electronics and Photonics XIV, 101112Z (27 January 2017).

[19] C. Niclass, M. Soga, H. Matsubara, M. Ogawa and M. Kagami, "A 0.18-µm CMOS SoC for a 100-m-Range 10-Frame/s 200x96-Pixel Time-of-Flight Depth Sensor," in *IEEE Journal of Solid-State Circuits*, vol. 49, no. 1, pp. 315-330, Jan. 2014.

[20] D. Tyndall, B. R. Rae, D. D.-U. Li, J. Arlt, A. Johnston, J. A. Richardson, and R. K. Henderson, "A high-throughput time-resolved mini-silicon photomultiplier with embedded fluorescence lifetime estimation in 0.13 µm CMOS.," IEEE Trans. Biomed. Circuits Syst., vol. 6, no. 6, pp. 562–70, Dec. 2012.

[21] A.Elshazly, S.Rao, B.Young, P.K. Hanumolu, "A Noise-Shaping Time-to-Digital Converter Using Switched-Ring Oscillators—Analysis, Design,and Measurement Techniques" IEEE J. Solid-State Circuits, Vol. 49, No. 5, May 2014.

[22] A.T. Erdogan, R. Walker, N.Finlayson, N.Krstajić, G.O.S. Williams, R.K. Henderson1 "A 16.5 Giga Events/s 1024 × 8 SPAD Line Sensor with per-pixel Zoomable 50ps-6.4ns/bin Histogramming TDC" Proc. VLSI Symposia, June 2017.

[23] J. A. Richardson, L. A. Grant and R. K. Henderson, "Low Dark Count Single-Photon Avalanche Diode Structure Compatible With Standard Nanometer Scale CMOS Technology," in *IEEE Photonics Technology Letters*, vol. 21, no. 14, pp. 1020-1022, July15, 2009.

[24] M. A. Tétrault, É. D. Lamy, A. Boisvert, J. F. Pratte and R. Fontaine, "Real-time discreet SPAD array readout architecture for time of flight PET," *2014 19th IEEE-NPSS Real Time Conference*, Nara, 2014, pp. 1-3.

Tarek Al Abbas is a PhD candidate at The University of Edinburgh, UK, working on single photon avalanche diode (SPAD) sensors.

Neale A. W. Dutton is a Senior Analogue Design Engineer with STMicroelectronics and the Imaging Sub-Group. He completed his MEng and PhD degrees both at the University of Edinburgh. His PhD research into SPAD image sensors was with the CMOS Sensors and Systems Group, Institute for Integrated Micro and Nano Systems at the University of Edinburgh. He has authored or co-authored 30 papers and has 18 patents either granted or filed. His research interests are in SPAD and Time of Flight sensors. Neale is a member of the technical programme committees for VLSI Symposium and International Image Sensors Workshop.

Oscar Almer obtained his undergraduate degree in Computer Science with Electronics from the University of Edinburgh. He then earned his PhD from the same institution in 2012 and has since worked in the field of digital design for opto-electronics. He is currently with Semtech Corp.

Dr. Neil Finlayson is a Research Associate working in the CMOS Sensors and Systems Group, Institute for Integrated Micro and Nano Systems at the University of Edinburgh. Neil works on the Proteus project which is focused on molecular imaging of lungs tissue. His primary responsibilities are software and firmware development, optical characterisation and applications of next-generation time-resolved fluorescence/Raman spectroscopic sensors. In the last year Neil has developed time-resolving spectrometer sensor software and contributed to the development and optical characterisation of an ultrafast time-resolving spectrometer and scanning confocal imaging system. Over a thirty year engineering and research career in industry and academia, Neil has led teams and worked on projects in optoelectronic systems, energy engineering, Internet services and software development. Neil has authored or co-authored 54 peer-reviewed journal and conference papers.

Francescopaolo Mattioli Della Rocca received the MEng degree in Electronics and Electrical Engineering at the University of Edinburgh in 2015. He is currently working toward the PhD degree at the University of Edinburgh funded and jointly supervised by STMicroelectronics. His interests include single photon avalanche diode (SPAD) image sensors for time of flight (TOF) applications. He has been the recipient of the IEEE ISSCC 2017 Student Research Preview award.

Robert Henderson is a Professor in the School of Engineering at the University of Edinburgh. He obtained his PhD in 1990 from the University of Glasgow. From 1991, he was a research engineer at the Swiss Centre for Microelectronics, Neuchatel, Switzerland. In 1996, he was appointed senior VLSI engineer at VLSI Vision Ltd, Edinburgh, UK where he worked on the world's first single chip video camera. From 2000, as principal VLSI engineer in STMicroelectronics Imaging Division he developed image sensors for mobile phone applications. He joined Edinburgh University in 2005, designing the first SPAD image sensors in nanometer CMOS technologies in MegaFrame and SPADnet EU projects. In 2014, he was awarded a prestigious ERC advanced fellowship.