A 250 Gph/s digital SiPM for high-speed FLIM with 300 ps minimum gate width

Alessandro Tontini, Gianpietro Maddinelli, Alessandro Esposito, Conor A. Treacy, and Leonardo Gasparini

I. Abstract

We present a 100×100 CMOS-SPAD array for high-speed, single-shot fluorescence lifetime imaging microscopy (FLIM) in scanning systems. Fluorescence decay is estimated by summing photon detections from the array of SPADs using a time-gated approach with 4 gates, with a measured 49 ps FWHM skew and a 300 ps minimum gate width. Each SPAD cell includes a 4-bit memory, detecting up to one photon per laser shot per gate, and the global sum is performed on-chip via a fully parallel, pipelined architecture supporting up to 10,000 photons per laser shot. With a 25 MHz acquisition rate, the sensor reaches 25 Mpix/s, 250 Gph/s - the highest reported for scanning microscopy. A per-pixel 1-bit SRAM also enables noise reduction and acquisition of images.

II. Introduction

FLIM enables the analysis of biological samples by exploiting the fluorescence lifetime decay of specific markers, thereby improving contrast compared with conventional intensity-only techniques [1]. Two main architectures are used to measure fluorescence lifetimes: direct timestamping or time-gating [2]. Direct timestamping provides the highest temporal resolution but suffers from large pixel sizes, low fill factor, and excessive data rates that limit fast, high-resolution FLIM imaging. The time-gated approach achieves higher fill factor and array resolution with simpler pixels but often relies on only 1–2 gates [4-6], requiring repeated acquisitions and increasing the risk of photobleaching. In this work, we present a fully digital 100×100 SPAD-based array with 4 time-gates that enables high-speed, single-shot FLIM. The sensor is designed as a single-point digital SiPM for scanning microscopy systems and also supports imaging by enabling one SPAD cell at a time.

III. Sensor architecture

The sensor is composed of a 100×100 array of digital cells. Each cell contains one SPAD, a 1-bit SRAM to enable/disable the cell, 4 time-gated sampling circuits, a 4-bit memory, and two half-adder circuits. The complete schematic of the SPAD-based digital cell is shown in Figure 1.

A. Tontini, G. Maddinelli and L. Gasparini are with the Center for Sensors and Devices, Fondazione Bruno Kessler, Trento, Italy (e-mail: tontini@fbk.eu; gmaddinelli@fbk.eu; gasparini@fbk.eu).

C.A. Treacy and A. Esposito are with the Center for Genome Engineering and Maintenance, Brunel University of London, Uxbridge, UK (e-mail: alessandro.esposito@brunel.ac.uk; conor.treacy@brunel.ac.uk).

We thankfully acknowledge the financial support from the European Commission through the Grant Agreement No. 101135034 (project HILIGHT).

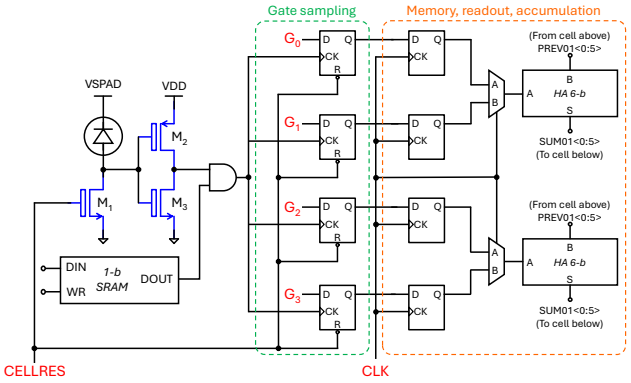

Fig. 1: Schematic of the SPAD-based digital cell. The SPAD is synchronously precharged at the beginning of each measurement, and thanks to the inverter M2-M3, no analogue voltage is required in the front-end circuit. The in-pixel 4-bit memory enables fully parallel acquisition and readout. The sum of photon counts is computed by each cell along the columns by means of 6-bit half adders.

The SPAD is charged synchronously at the beginning of each measurement by the *CELLRES* signal via a thick-oxide transistor M1. The avalanche pulse is then converted into a 1.2 V digital pulse by the inverter composed of thick-oxide transistors M2/M3. The output of the front-end circuit can be masked by an AND gate controlled by the 1-bit SRAM cell to remove the contribution of highly noisy SPADs or to build images by scanning the SPAD array enabling one cell at a time. Four D-type flip-flops (DFFs) sample the status of 4 globally distributed gating signals $G_{0..3}$ when a photon is detected by the SPAD. To avoid the complexity of distributing extremely short digital pulses, the cumulative gating approach in Figure 2 has been employed, in which sub-ns temporal windows are obtained by adjusting the delay between subsequent gating signals rather than controlling their temporal width. The individual window counts are then obtained by subtraction.

To withstand an operation rate of up to 25 MHz in a scanning microscope, the acquisition and readout phases are executed in a fully parallel, pipelined approach. Before the end of the measurement at clock cycle n, the 4-bit result is sampled by 4 additional DFFs, making the time-gated sampling logic available for the next measurement cycle n+1. The summation of the time-gated measurements to build the final 4-bin histogram of photon counts starts at cell level and propagates along the columns. Each cell updates the summation of photon

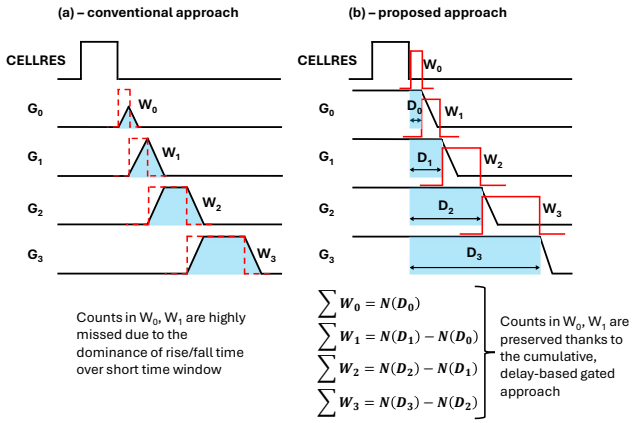

Fig. 2: Comparison between a conventional implementation (a) and the proposed cumulative gating approach (b) to achieve sub-ns gating windows without the need to propagate extremely short pulses. With this approach narrow time windows $W_{0..3}$ are defined by the delays $D_{0..3}$ between subsequent gating signals $G_{0..3}$. With this solution, only the falling edge is relevant for the purpose of the measurement.

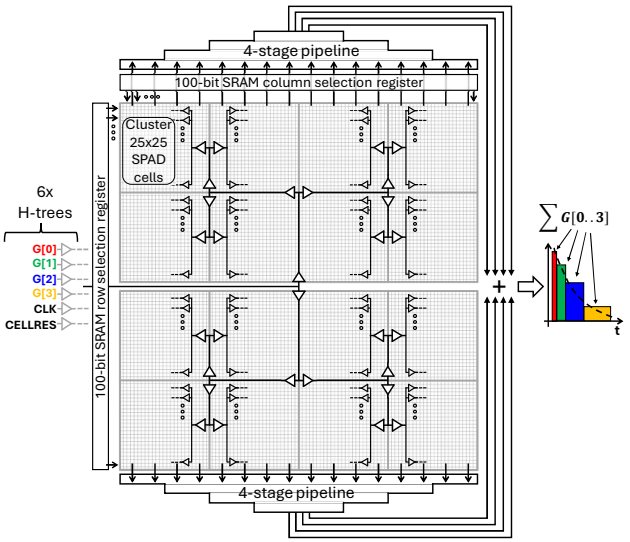

Fig. 3: Chip architecture. Timing-critical signals $G_{0..3}$, *CLK* and *CELLRES* are distributed by six independent, four-branch H-trees. Each terminal branch serves a 25×25 cell cluster with a 1×25 tree.

counts over each gating window and propagates the result to the next cell. To improve the fill factor, each cell implements only two adders that are 2-to-1 time-multiplexed. Moreover, considering that each cell can sum at most a single bit to the summation result, half-adders instead of full-adders have been employed. To reduce the propagation time required to compute the sum of photon counts at each column and simplify layout complexity, the array is logically split into two halves of 50×100 cells each, operating in parallel. After the partial summation result is ready at the column level, 3 additional clock cycles are required to compute the final, global 4-bin histogram for a total latency of 4 clock cycles. To minimise skew caused by the distribution of timing-critical signals, the array has been organised into 16 clusters of 25×25 cells each. Signals CELLRES, CLK, and $G_{0..3}$ are distributed by a 4-branch H-tree, shown in Figure 3. Each terminal branch of the H-tree serves a 25×25 cell cluster, where signals are distributed to each row by a 1×25 tree. The chip has been fabricated in a 110 nm, 4M CIS technology, taking an overall area of $\approx 4 \times 4.4$ mm^2 ($\approx 3.1 \times 3.2$ mm^2 for the array of SPADs). The micrograph is shown in Figure 4.

IV. Characterization Results

The median Dark Count Rate (DCR) and peak PDP of the SPADs at room temperature, with the chip operating at the maximum acquisition rate of 25 MHz (25 Mpix/s), is 437 cps/53% for an excess bias of 2 V and 696 cps/62% for an excess bias of 4 V. The DCR population for a chip together with PDP characterization is shown in Figure 5 for multiple excess bias voltages.

The timing skew of the distribution of the gating signals has been characterized by sweeping a 470 nm, 80 ps FWHM pulsed laser in steps of 10 ps. For each delay step, 10^5

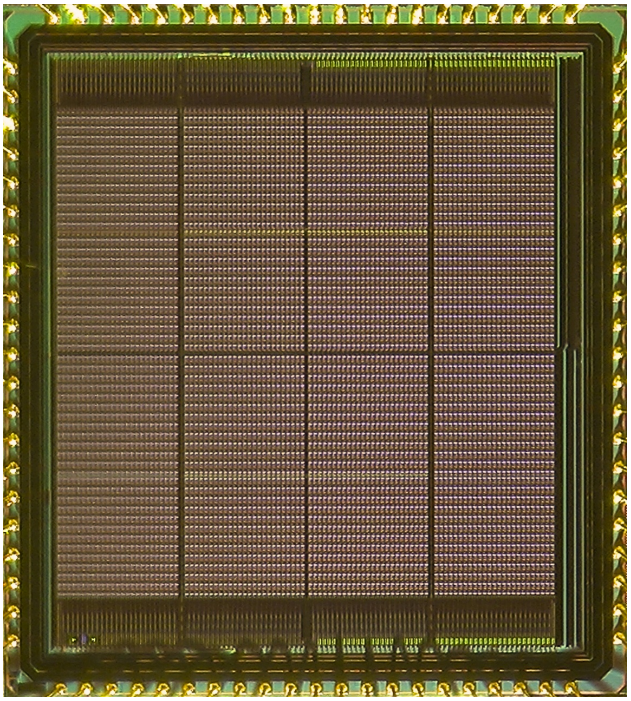

Fig. 4: Chip micrograph, showing the presence of 4×4 clusters of 25×25 cells each that allow the routing of the H-trees for signal distribution. At the top and bottom of the array, the pipeline architecture for computing the 4-bin histogram directly on chip is also visible.

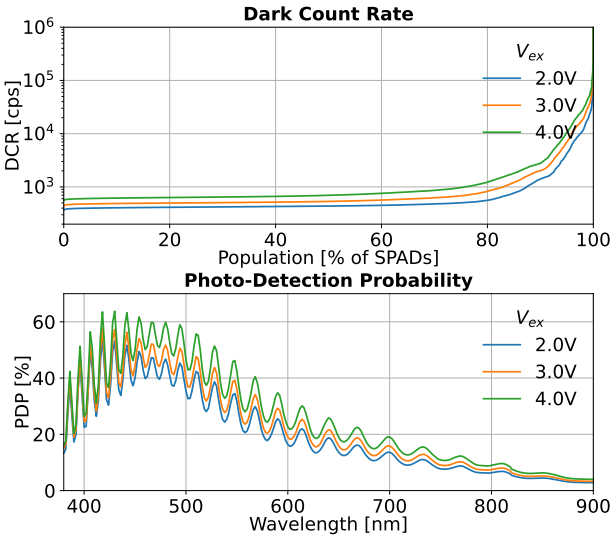

Fig. 5: DCR population and average PDP of a chip sample at multiple excess bias voltages. The median DCR is below 700 cps at up to 4 V, with almost 80% of the SPAD population exhibiting a DCR below 1000 cps for all excess bias voltages. The peak PDP at 4 V excess bias is ≈63% at 420 nm.

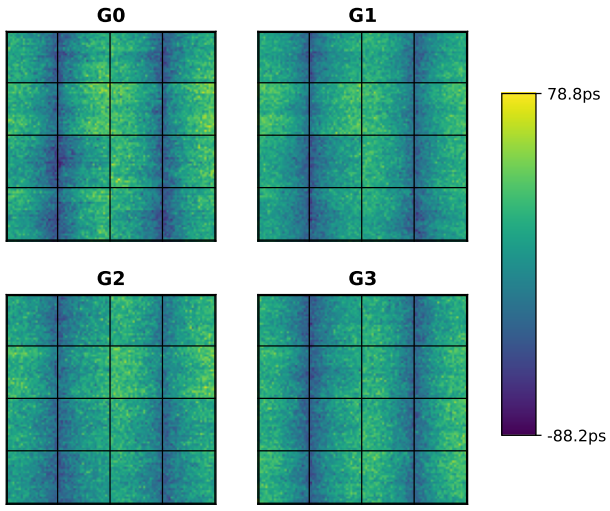

Fig. 6: Map of the falling edge skew of the gates $G_{0..3}$, normalized with respect to the mean. The spatial distribution of falling edge skew follows the signal distribution architecture, highlighting the presence of the H-tree up to each 25×25 cells cluster.

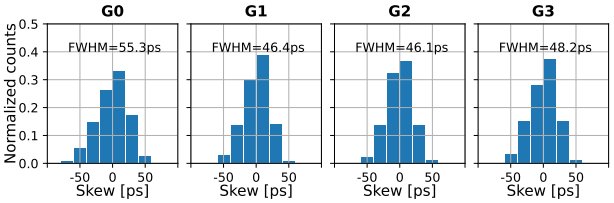

Fig. 7: Histograms of the distribution of the falling edge position measured for each SPAD cell.

acquisitions have been recorded, with the sensor operating below pile-up distortion. The rise and fall edge times are defined as the 50% crossing point of the resulting histogram of photon counts. Because the gating windows are defined between the falling edges of two consecutive gating signals, we focus on the timing performance of those edges only. Figure 6 reports the spatial distribution of the falling edge position of the gates, normalized with respect to the mean value. The measurement results, numerically summarized in Table I, match the signal distribution architecture, with the H-tree that propagates up to a terminal cluster of 25×25 cells, after which signals are distributed by means of a 1×25 tree to address layout complexity. As expected, the H-tree ensures that the skew is accumulated only along the rows of each 25×25 cluster. The histograms of the distribution of the falling edge position over the entire array are shown in Figure 7. The average skew of the gates is 49 ps FWHM, with a max-to-min variation below 170 ps for the worst-case gate (G_0).

TABLE I: Statistics of the falling edge skew of the gating signals $G_{0..3}$ over the entire array. The measured skew averaged over the four gates is 49 ps FWHM, with a min-to-max variation of less than 170 ps considering the worst gate G_0.

Gate	G_0	G_1	G_2	G_3
Falling edge skew [ps]				
FWHM	55.3	46.4	46.1	48.2
Min–Max	167.0	135.1	134.0	136.2

The minimum gate width has been characterized by sweeping the pulsed laser across gates G_0 and G_1 in steps of 10 ps, recording the differences between the counts in G_1 and the counts in G_0. Results are shown in Figure 6, with a minimum gate width of ≈300 ps, with average rise and fall times of 72 ps and 68 ps, respectively.

The chip has been tested in the field to acquire a sample of *Convallaria majalis* in a FLIM setup employing a 470 nm, 80 ps pulsed laser. Figure 9 shows both the intensity (photon-counting) and lifetime images obtained from the experiment. The results have been obtained with the integration of 2500 frames, each with $\approx 60 \cdot 10^4$ laser shots by using the sensor in imaging mode, therefore requiring a 1-to-1 SPAD scanning using the integrated SRAM. By properly combining the intensity and lifetime images it is possible to maximize contrast and visual information, as shown in Figure 9.

V. Conclusion

A 100×100 d-SiPM for time-gated FLIM applications has been realized in a 110 nm CIS process, achieving the shortest gate width, lowest skew, and highest sampling rate reported to date. The sensor supports both single-point acquisition for scanning systems and imaging operation, thanks to the possi-

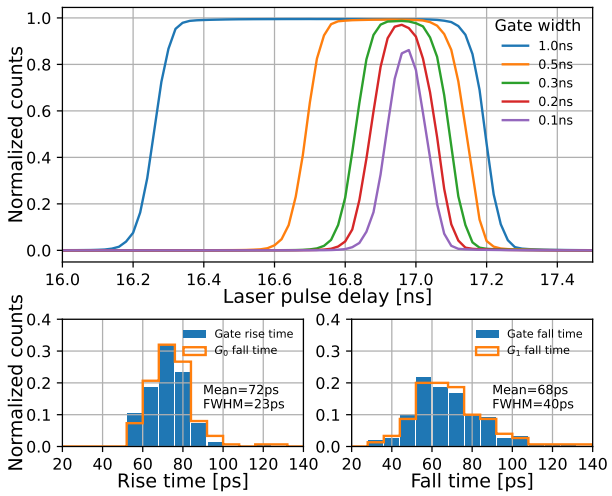

Fig. 8: Upper plot: characterization of the gate width (G_0 to G_1) over the entire array. The measured minimum width with negligible (<0.5%) plateau attenuation is as low as 300 ps. Lower plot: distribution of rise time (G_0 falling edge) and fall time (G_1 falling edge), with 23 ps and 40 ps FWHM, respectively.

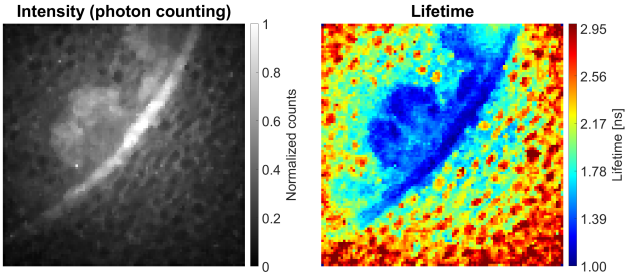

Fig. 9: Intensity (left) and lifetime (right) information of a sample of *Convallaria majalis* with 50× magnification in a custom FLIM setup. The images have been obtained using the sensor in imager mode, thereby enabling SPADs one-by-one using the integrated SRAM. The average lifetime value of the imaged sample is 2.2 ns with a standard deviation of 0.40 ns. The darker areas near the corners are due to non-uniform illumination of the sample by the pulsed laser.

bility of independently addressing individual cells. The device has undergone full electrical and timing characterization and has been successfully demonstrated in a real FLIM setup operating in imager mode. Table II summarizes the sensor performance and provides a comparison with the current state of the art.

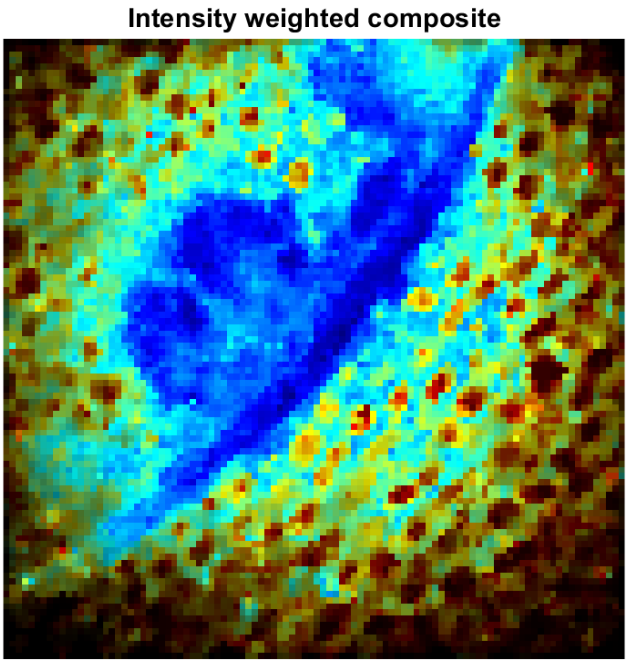

Fig. 10: Combined FLIM image (intensity + lifetime) of a sample of *Convallaria majalis*. The darker areas near the corners are due to non-uniform sample illumination.

TABLE II: Comparison with state-of-the-art.

	Dutton 2015 [3]	**Gyongy 2018 [4]**	**Ulku 2019 [5]**	**Wayne 2022 [6]**	**This work**
Process	130 nm	130 nm	180 nm	180 nm	**110 nm**
Array size	32×32	256×256	512×512	500×500	**100×100**
Pixel pitch	21 µm	16 µm	16.38 µm	16.38 µm	31.2 µm
Fill factor	43%	61%	10.5%	10.5%	**26%**
Peak PDP	–	40% @ 550 nm V_{ex} = 6.5 V	50% @ 520 nm V_{ex} = 6.5 V	50% @ 520 nm	**35% @ 550 nm V_{ex} = 4.0 V**
Median DCR/SPAD	–	3 cps V_{ex} = 3.3 V	7.5 cps V_{ex} = 6.5 V	10.2 cps	**696 cps V_{ex} = 4.0 V**
Power	14.1 mW	–	26.7 mW[a]	–	85 mW
Min. gate width	71 ps[b]	4.0 ns	6.0 ns	1.0 ns	**0.3 ns**
Gate skew FWHM	–	–	264 ps	153 ps	**49 ps**[c]
Sampling rate	1.7 Gph/s (8 bit)	6.5 Gph/s (1 bit)	25.6 Gph/s (1 bit)	12.5 Gph/s (2 bit)	**250 Gph/s (4 bit)**

[a]In dark mode (<0.4% count rate); [b]direct timestamping (8 bit TDC); [c]average of falling edge over $G_{0..3}$

References

[1] K. Suhling et al., "Fluorescence lifetime imaging (flim): Basic concepts and some recent developments," *Medical Photonics*, vol. 27, 2015.

[2] C. Bruschini et al., "Single-photon avalanche diode imagers in biophotonics: Review and outlook," *Light: Science Applications*, vol. 8, Sep. 2019. DOI: 10.1038/s41377-019-0191-5.

[3] N. A. W. Dutton et al., "11.5 a time-correlated single-photon-counting sensor with 14gs/s histogramming time-to-digital converter," in *2015 IEEE International Solid-State Circuits Conference - (ISSCC)*, 2015, pp. 1–3. DOI: 10.1109/ISSCC.2015.7062997.

[4] I. Gyongy et al., "A 256×256, 100-kfps, 61% fill-factor spad image sensor for time-resolved microscopy applications," *IEEE Transactions on Electron Devices*, vol. 65, pp. 547–554, 2018.

[5] A. C. Ulku et al., "A 512 × 512 spad image sensor with integrated gating for widefield flim," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 25, no. 1, pp. 1–12, 2019.

[6] M. Wayne et al., "A 500×500 dual-gate spad imager with 100% temporal aperture and 1 ns minimum gate length for flim and phasor imaging applications," *IEEE Transactions on Electron Devices*, vol. 69, no. 6, pp. 2865–2872, 2022.