The final published version is available online at:
https://ieeexplore.ieee.org/document/7299269

Rights / License:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

When citing, please refer to the published version.

A 160×120-Pixel Analog-Counting Single-Photon Imager with Time-Gating and Self-Referenced Column-Parallel A/D Conversion for Fluorescence Lifetime Imaging

[1]Matteo Perenzoni, [1]Nicola Massari, [1]Daniele Perenzoni, [1]Leonardo Gasparini, [1]David Stoppa

[1]Fondazione Bruno Kessler (FBK), Trento, Italy

Abstract

A single-photon, time-gated, 160×120-pixel imager is presented for its application in fluorescence lifetime imaging microscopy. Exploiting Single-Photon Avalanche Diodes and an extremely compact pixel circuitry – only 7 MOSFETs and 1 MOSCAP – the imager is capable of gathering information about photon position, number and time distribution, enabling cost-effective devices for scientific imaging applications. This is achieved thanks to the photon counting and time-gating capabilities implemented in the analog domain, which in turn enable a 15-µm pixel with a 21% fill-factor. A reconfigurable column circuitry supports both the analog conventional readout and a self-referenced analog-to-digital conversion, able to cancel-out the pixel to pixel non-uniformities, and speeding up the framerate to 486 fps. The imager, featuring also a delay locked loop in order to stabilize the internal waveform generation for reliable timing performance, has been implemented in a standard high-voltage 0.35-µm CMOS technology. Measurements in a fluorescent lifetime setup have been performed, comparing the results with single-point acquisitions made with commercial time-correlated equipment.

Index Terms: analog counter; CMOS; FLIM; image sensor; lifetime; SPAD; time-gating.

Contact author

Matteo Perenzoni

Via Sommarive, 18 - 38123 Trento, Italy

Phone: +39 0461 314533; Fax: +39 0461 314591

E-mail: perenzoni@fbk.eu

I. INTRODUCTION

Scientific imaging applications often require time resolving capability, either as a necessary parameter or as an additional source of information about the sample or process under observation. Some examples include fluorescence lifetime imaging microscopy (FLIM), Raman spectroscopy, time-resolved near-infrared spectroscopy, but also consumer applications like 3D imaging based on Time-of-Flight. In particular, fluorescence microscopy is an established technique for the analysis of biological processes and relies on the measurement of the fluorescence intensity upon an excitation [1] at different wavelengths. Fluorescent light emission from biological samples is a phenomenon occurring in a variety of temporal scales, from nanoseconds and upwards; the fluorescence lifetime measurement adds valuable additional information to the analysis of the process, but requires specific hardware and setup to be implemented. Therefore, the availability of an integrated solution, preferably in a low-cost CMOS technology, would enable the development of reliable diagnostic tools at an affordable price, to be used for example by the family physician in order to give results in few minutes. At the same time, a custom FLIM sensor could be employed into scientific instrumentation where its flexibility would enable a whole new set of experiments, for example acquiring at the same time spectral- and polarization-resolved images [2] with optical elements exploiting one array dimension. In order to achieve this, a robust and reliable sensor, with low sensitivity to process, voltage and temperature variations, having a high throughput and a relatively small pixel pitch for increased spatial resolution, is needed.

The reconstruction of the fluorescent emission can be done by means of the Time-Correlated Single Photon Counting technique (TCSPC) [3], which allows recording the time of arrival of each photon after the sample excitation with a very short light pulse. With TCSPC it is possible

to retrieve any type of decay function, from single exponential to multiple exponentials. For this technique, the solid-state device of choice is usually a Single-Photon Avalanche Diode (SPAD), a special photodiode operated in the Geiger region, able to convert any detected photon into a fast digital pulse [4]. However, in order to perform imaging, TCSPC requires either an event-driven approach [5] with the dynamic allocation of several time-to-digital converters (TDCs) to the triggered pixels, or a per-pixel TDC implementation inside the focal plane [6]-[8] or in the close periphery [9]. These approaches are very effective in the precision of the measurement but have a number of drawbacks, i.e. large pixel and reduced fill-factor, poor scalability to larger arrays, high power consumption and the need for advanced technology nodes. Alternative solutions have been proposed to reduce the pixel size and improve the sensitive area by employing time-to-analog converters (TACs) [10],[11], but this solution still has not been presented in moderate to high resolution imagers, even though it has a high potential for pixel pitch reduction and fill-factor increase. Eventually, the TCSPC technique capability of reconstructing the whole time waveform overestimates the need for data, as for the fluorescence phenomena, the interesting parameter is often the time constant of a single or double exponential.

Therefore, measurements exploiting the control of the observation time in the nanosecond scale can be employed, such as photo-demodulating and time-gated devices. Photo-demodulating pixels, originally developed for 3D time-of-flight measurements, in the FLIM context base their operation on the low-pass effect of the fluorophore response, and therefore the phase shift between the modulated excitation light and the fluorescent emission is measured [12]-[15]. With the knowledge of the excitation frequency the extraction of a single exponential time constant is straightforward, while multi-exponentials require additional measurements at different frequencies. On the one hand, they have the advantage of being devices with a good fill-factor

and compact pixel pitch, but on the other hand imagers based on photo-demodulators suffer from sensitivity issues as the modulated light source has smaller energy per unit time with respect to a pulsed source, and both the source and the device need to be modulated at challenging high speed, several tens of MHz, to recover time constants in the ns range.

Time-gated imaging is based on the repeated integration in determined time windows of nanosecond-range duration. It can be realized in several ways: by increasing the complexity and cost of the optical setup, it is possible to use gated optical intensifiers, which is an additional optical element to be added in front of a conventional camera, e.g. a charge-coupled device (CCD) camera [16]. Alternatively, the complexity can be moved into the sensor itself, therefore exploiting the advantages of integrated circuit production to lower the cost of the FLIM system. Due to the short gating time required, the approach must be strongly device-oriented: in particular, fast photogates and SPAD devices are the current choice. Photogates can be engineered in such a way that they can perform fast charge transfer only during the gating interval [17], and so performing multiple small integrations. While pitch and fill-factor are good, anyway their sensitivity is limited by the readout noise and may require a customized process in order to obtain the desired modulation speed characteristics. On the other hand, SPAD devices have been demonstrated in a large number of conventional CMOS technologies, although when developed in custom processes they can achieve excellent photon detection efficiencies and noise performance [18]. However, in a time-gated imager the gating window is so short that the typical dark count rate (DCR, average rate of noisy avalanches not related to photons) of standard CMOS SPADs can be tolerated. The main advantages of SPAD are their single-photon sensitivity, extremely useful to detect the weak fluorescent emission, and the digital nature of their output. On the other hand, the digital circuitry needed to determine whether a photon is or

not within a time window requires a very large area often not practical. A possible workaround relies on moving this processing on the periphery of the pixel array [19], while the pixel can only store a single bit. However, this solution requires continuous scanning of the array, thus reducing the throughput and the scalability of the approach. So, to efficiently time-gate and count the relevant photons per-pixel, the analog approach proved to be very effective in the reduction of area needed for the electronics: a first solution found in literature showed a remarkable 25-μm pixel pitch and 20.8% fill-factor, with a minimum time-gating of 1.1 ns [20]. Developing on the same approach, other similar architectures have been realized, even though not in array configuration, achieving down to 530 ps minimum time gating in a 0.15-μm CMOS technology [21]. An extremely small pitch of 9.8 μm, for a SPAD-based pixel, has been achieved in [22] by means of a deep-submicron technology node and the use of analog counting circuitry. In the mentioned implementation, despite time-gating could be potentially achievable, it has not been demonstrated yet; anyway, both analog and 1-bit digital photon counting are possible in a quarter-VGA resolution. All these advantages of the described approach are anyway balanced by the typical weaknesses of analog circuitry with respect to digital. Indeed, the main drawback of all analog-based solutions is the strong non-uniformity of the analog count steps in the pixels, which therefore requires calibration, and the increased sensitivity to noise and disturbances. In other pixel-level applications such as X-ray imaging this has been addressed by designing a non-linear counter with large steps when the non-idealities are more relevant [23].

In order to address the different highlighted requirements of a time-gated imager for FLIM applications, in this work a compact analog pixel based on SPAD for time-gating and photon counting has been implemented, together with a self-referenced analog-to-digital converter addressing the non-uniformity and noise issues [24]. Detailed simulations first, and then

measurements, highlight the effectiveness of the compact pixel architecture in discriminating photons within sharp and stable observation windows. FLIM measurements of known fluorophores validate the sensor.

The paper is structured as follows: in Section II the sensor architecture and circuitry are explained in detail from the top level to the pixel electronics, while the characterization measurements are given in Section III, starting from the electrical and optical test, then continuing with the time-gating measurements and eventually providing the experimental results for a FLIM setup. Section IV summarizes the main results and concludes the paper.

II. TIME-GATED SENSOR ARCHITECTURE

A time-gated imager for FLIM is not a conventional image sensor: in the following the specific requirements will be analyzed and the architecture of the sensor explained, down to the pixel details.

A. FLIM TECHNIQUE AND SENSOR ARCHITECTURE

In a typical time-gated FLIM measurement system a fast pulsed light source repeatedly excites the sample under investigation. The detector operates synchronously with the source, collecting data on photons arrival time statistics by opening short observation windows and counting the detected photons. In order to optimize the acquisition time, this repetition should occur at the maximum rate of the typically employed lasers for excitation, which is about 40-80 MHz. In order to evaluate a single exponential decay $I(t) = k \cdot e^{-t/\tau}$, two acquisition windows are required and good results can be obtained for overlapped windows. As shown in Figure 1, a short laser pulse is used to excite the sample which after few picoseconds starts to emit the fluorescent response. Integrating N_1 and N_2 photons in the two windows of length T_{WIN} and overlapped by half their width, the amplitude and lifetime of the fluorescence can be extracted [25]:

$$I(t) = k \cdot e^{-t/\tau} \begin{cases} k = \dfrac{2N_1^3}{T_{WIN}\left(N_2^2 - N_1^2\right)} ln\left(\dfrac{N_2}{N_1}\right) \\[2ex] \tau = -T_{WIN}\left[ln\left(\dfrac{N_2}{N_1}\right)\right]^{-1} \end{cases}. \tag{1}$$

As proved in [25], the optimal choice for T_{WIN} is of about 1-2 times the expected time constant in order to obtain the best SNR. However, the overlapped windows method has a quite wide plateau which gives some freedom in the setting of T_{WIN}, considering also that the speed of the acquisition, i.e. laser repetition rate, may be limiting the maximum available time. As an additional consideration, it is preferable to keep the window short so as to avoid the capture of DCR events. This means that for typical fluorophores with nanosecond-long lifetime, the gating window should be of that range. Since a single SPAD is available per-pixel, and the dead-time needed to recharge it after an avalanche is in the order of tens of nanoseconds, only a single photon can be recorded per repetition. It is known that to properly record undistorted data, measurement has to happen in low pile-up conditions [26]: this implies that on average, each SPAD should detect one photon every 10-100 time windows. Practically, this is not a constraint since not only the excitation light is converted with a low efficiency by the fluorophore producing a weak intensity, but also the peak power of the laser should not be too high so as to avoid photobleaching, i.e. the deterioration of the fluorescence emission. Therefore, regardless of the SPAD quantum efficiency, the low pile-up condition has to be attained and it poses a design constraint on the pixel for efficient acquisition: it has to maximize the acquisitions between readouts, and therefore to be capable of cycling hundreds or thousand times with a dynamic range requirement that can be one-two orders of magnitude smaller. Then, the acquisition has to be repeated until the total number of collected photons is such that the required precision is achieved, based on the specific optical setup, the power of the exciting laser, and the

concentration of the fluorophore. By applying the error propagation to eq. (1) it is possible to determine that, for example, a number of photons in the range of one thousand is able to return the lifetime with a 10% rms error.

The conceived time-gated sensor architecture is depicted in Figure 2. Besides a conventional addressing scheme for the focal plane array, there is a window generator which has the duty of generating the tight timing waveforms driving the pixel: it requires a clock signal named TRIG which can be the same that is provided to the laser source. Internally, programmable delay lines provide the precise timing of pixel signals that are properly distributed using clock trees to the row decoders, and then delivered to the pixels with a last driving stage. In order to ensure that the delay lines of the window generator are not drifting with temperature or supply voltage variations, a delay locked loop employs the period of an input clock REFCLK as reference delay in order to provide a regulation voltage to the generating logic. Readout can be performed either in analog mode, outputting the analog voltage accumulated in the pixels, or by performing a row-by-row analog-to-digital conversion, delivering directly the counted photons in digital form at a higher speed.

B. ANALOG COUNTING TIME-GATED PIXEL

The novel pixel topology here presented relies on two main concepts: the reuse of transistors by moving part of the complexity from the pixel to the driving electronics in the row decoder, and the use of a "deferred counting" scheme which avoids the need of generating a pulse inside the pixel, like in [20]. The transistor-level NMOS-only schematic is depicted in Figure 3 and the waveforms designed to properly drive it are shown in Figure 4.

The SPAD is connected with its cathode to the high-voltage bias V_S, and the anode to M1. This transistor is operating either as a precharge device or a disabling switch depending on the VEB

state: when on, it charges the SPAD to V_S (VEB = 0V) or it turns off the SPAD by biasing it to V_S-V_{exc} (VEB = V_{exc}, SPAD excess bias, the amount of voltage provided beyond the breakdown). With the digital signal SDRV = L, the SPAD retains its state: when already charged, it may fire a pulse if a photon or a dark count occurs. Afterpulsing (undesired avalanches which can occur up to several microseconds after recharge, due to charge trapping and release) can impact the measurement as an increase of the noise floor. It is known that afterpulsing depends on the capacitance of the detector node [27]: beside the small SPAD size, the capacitance has been minimized by design and with careful layout.

The pixel operation starts first with a hard reset, so as to remove any memory of the previous frame, followed by a soft reset. The reset takes place with the SPAD off and all other signals activated: the hard reset completely discharges node V_{cnt} through M4 and M5, by forcing BIAS = 0V, while the soft-reset operates with BIAS = V_{BH}. Typical operation is with V_{BH} = 3.3V so node V_{cnt} can be precharged to approximately 2.5V, depending on the reset phase duration.

Clocked by the TRIG signal rising edge, the window generator first deactivates all signals except VEB and BIAS: the latter in particular is set to the typical operating voltage V_{BL} = 0.5V. Subsequently, a number of counting cycles can be performed, each composed by counter precharge, SPAD precharge, pulse latching and counting:

- counter precharge: it is performed by means of a pulse of programmable duration T_{PRE}, which is sent to M4 so as to charge the parasitic capacitance C_1 to V_{BL}. In the meantime, signal WIN is activated;

- SPAD precharge: after a programmable delay TDEL from the TRIG rising edge, the gating window is open by bringing the SPAD beyond the breakdown voltage with a short pulse on SDRV. In order to ensure a fast turn-on of the SPAD, the signal SDRV can be

distributed by the row drivers with a pulse width as short as $T_{CHR} \approx 300$ ps, which can also be programmed;

- pulse latching: in this phase, while WIN = H, any SPAD pulse transfers on V_{latch}. Since M1 is now off, there is no recharge of the SPAD and V_{latch} keeps track of the state. The observation window is defined by T_{WIN}, and is concluded by the deactivation of the WIN signal, isolating V_{latch} from the SPAD. Immediately after, the SPAD can be turned off with VEB = V_{exc} and SDRV = H;

- photon counting: eventually, the counting phase takes place using a pulse on signal CNT of typical amplitude $V_{CNTL} = 1.5$V, propagating on V_{pulse} depending on the value stored on V_{latch}. Therefore, if a photon has been recorded, V_{cnt} decreases accordingly due to charge sharing between C_1 and C_2.

All digital signals are 3.3V, except SDRV and WIN, which are increased to 5V in order to speed up the SPAD turn-on and avoiding early clamping of the level propagated to V_{latch}. Indeed, these boosted control signals enable NMOS-only electronics while using a wide range of SPAD excess bias voltages, up to approximately 5V-V_{th2}. Furthermore, the voltage on V_{pulse}, at the gate of M5, is independent from the SPAD pulse level as long as V_{CNTL} is lower than V_{latch}-V_{th3} because M3 is operating as a switch, propagating the CNT levels to the counter. Finally, there is no dependence of the counter step from the time of arrival of the photon, as the pulse on V_{pulse} is always of the same width: it has been observed that a $T_{CNT} > 4$ ns is sufficient to make negligible any propagation different of the CNT signal in the array, ensuring a complete settling. The described deferred counting scheme therefore makes the analog step independent from the SPAD time constant, quenching conditions and breakdown voltage variations, enabling extremely sharp gating window edges. The step size of the analog counting can be tuned by changing the stored

charge in C_1 at the beginning of the cycle with the V_{BL} value of BIAS; its voltage, being kept constant during the counting phase, has no propagation constraints. In order to guarantee shot noise-limited operation, the following condition has to be met:

$$2 \cdot V_{kTC}^2 + N \cdot V_{kTC}^2 + 2 \cdot V_{nRO}^2 < N \cdot V_{step}^2, \tag{2}$$

where the left side includes the kTC noise of the reset operation V_{kTC}, of the charge sharing, and of the readout V_{nRO}, and on the right side the shot noise of photons expressed in terms of the voltage step V_{step}; the factor 2 is required to take into account the double sampling operation of the readout. The step has therefore to satisfy the unequality:

$$V_{step} > \sqrt{\frac{2+N}{N} \cdot V_{kTC}^2 + 2 \cdot V_{nRO}^2} \ . \tag{3}$$

For N = 1, the term in (3) is maximum and sets the design constraint on the step, which can be easily fulfilled with capacitance in the order of several fF and voltage step in the tens of millivolt range.

Additionally, the window generator has an emulation mode which, differently from Figure 4, maintains the SPAD off by forcing SDRV = H and VEB = V_{exc}: in this way, not only SPADs are disabled, but V_{latch} is always high, emulating the arrival of a photon for each cycle. This operating mode is very useful for calibration of the step amplitude and, as will be explained in the following, for the analog-to-digital conversion.

The process, voltage and temperature (PVT) sensitivity of the counting step has been thoroughly analyzed and optimized through simulations, although constrained by the area requirement. The process sensitivity mainly involves the threshold variations of transistors and the capacitance values through the gate oxide variations. Figure 5 shows the MonteCarlo simulations of the pixel output for a single counting step: in the inset the scatter plots of the two variables with the higher

correlation are shown, namely the threshold voltage variation ΔV_{tn} and the oxide thickness variation Δt_{ox}. Voltage variations have been considered to be present only on purely digital signals, while multilevel signals are treated as analog signals, and therefore generated by stable and reliable sources: indeed, the V_{BL} level is intentionally designed to determine the step amplitude. Temperature variations affect the DCR of SPADs, while having a minor impact on the step. Figure 6 depicts the typical and corner simulation of the output voltage in function of the time of arrival of the photon: even though the real window shape will be smoothed by the SPAD jitter, the sharpness of the window edges highlights the precision of the electronic gate. Moreover, temperature corners in the range 0°C-70°C and supply corners of 3.0V-3.6V and 4.5V-5.5V change only marginally the voltage step while not affecting the timing. However, the waveform generation block can be influenced by temperature and therefore a closed-loop delay line has been implemented (see sub-Section II.D).

Nevertheless, although the analog counter has a good reliability in terms of PVT variations, the sensor can also exploit the pixel structure in emulation mode for a self-referenced A/D conversion, in order to remove the remaining non-uniformities as explained in the next paragraph.

C. ANALOG READOUT AND SELF-REFERENCED A/D CONVERSION

After the chosen number of counting cycles, all the SPADs in the array are shut down by setting SDRV = H and VEB = V_{exc}, thus achieving a global shutter operation; then, the imager can be readout either in analog mode or performing a self-referenced analog-to-digital conversion. The employed reconfigurable column amplifier is shown in Figure 7 and the required waveforms in Figure 8: these are repeated for each row of the array in order to perform a complete readout.

The signal V_{line} from the pixel belonging to the selected row can be readout in analog mode by a conventional double sampling operation so as to remove the precharged V_{cnt} offset and the pixel

source follower non-uniformity. Considering the waveforms in Figure 8 in dashed lines, first the pixel output value is stored on C_s with the amplifier in unity-gain configuration ("Sample" phase), precharging at the same time the feedback capacitance C_f to a voltage V_{refCDS} with signal ANALOG; afterwards, the pixel counters of the selected row are locally reset, so the reset value can be subtracted ("Reset" phase) and the resulting output voltage can be serialized to the output. However, this readout mode does not correct any non-linearity and non-uniformity in offset and gain of the analog counters which may arise after several tens or hundreds of accumulation cycles, requiring external calibration for correct reconstruction of the number of detected photons.

Therefore a second operating mode for the column circuitry has been implemented, exploiting the previously explained emulation mode of the pixel in order to generate the reference ramp for the analog-to-digital conversion. Being the ramp generated by the pixel itself, it automatically corrects any first order non-linearity and non-uniformity of the analog counters. The detailed operation, referring to the phases in Figure 8 with waveforms depicted in solid lines, is structured as follows: when the row is selected, the pixel output is sampled on C_s while the amplifier is connected in follower configuration ("Sample" phase). Then, the pixel is reset and the amplifier feedback is open, therefore working as a comparator: being the reset level higher, its output is initially 0V, and the global counter CNT[0:7] is also reset with the active low ADCRN signal ("Reset" phase). The latches of Figure 7 are also reset, and afterwards the selected pixel row is driven with the emulation mode waveforms, while the other rows are instead in a rest state, with all other signals deactivated. Simultaneously, the column ADCs are fed with the global counter output CNT[0:7], clocked by the ADCCK signal which is synchronized with the pixel emulation mode cycles ("Conversion" phase). By so doing, the pixel output is now emulating the detection of a photon for each cycle, therefore generating a decreasing staircase-like voltage: as soon as

V_{line} is crossing the previously stored value in C_s, the comparator switches to VDD thus freezing the count CNT[0:7] state provided so far. Finally, the digital serialization of the converted column values can take place ("Readout" phase).

Transistors connected to digital signals CHBL and SAMPLE are respectively needed in order to precharge, at the beginning of each row selection, the common vertical line to zero for faster settling, and to keep the voltage at the input of the readout channel constant during the transient of the reset operation.

Figure 9 visualizes the behavior of the in-pixel analog counting node V_{cnt}, the input of the comparator V_{cin}, and the output of the OR gate DIS: in order to emulate the presence of pixels with different analog step, the reference voltage V_{BL} has been changed of ±100 mV. It is clearly visible that the comparator switches always at the same moment, thus producing the same digital code.

The use of the self-referenced conversion not only brings the mentioned advantages, but enables higher frame rates: while analog readout is limited by the speed of analog output buffers to 5 MS/s, the digital output can be read at 50 MHz through an 8-bit parallel bus, reducing the output data delivery time from 24 μs to 2.4μs for a 120-pixel row. Even though the conversion process takes some row time, still the digital mode is faster than the analog mode readout.

D. GENERATION OF PIXEL WAVEFORMS

The pixel waveforms are generated by the window generator block of Figure 2 as 3.3 V digital signals and are then propagated through a balanced distribution tree to the two sides of the array. At decoder level, some of the 3.3 V signals are converted to the multi-level control signals for the pixel by using custom-designed logic with pull-up and pull-down transistors connected to the required voltages. The operation of the window generator is based on a state machine and on

programmable delay lines with 6 bit registers. In order to provide stable and reliable timing, the delay elements have a regulated supply voltage which can be generated by a DLL which includes a replica of the same delay lines. The DLL schematic is shown in Figure 10: an input reference clock is fed into the phase detector and into the replica delay line. Being the latter inverting, the phase detector measures zero phase shift when the delay is exactly equal to the reference clock semiperiod: the charge pump, using thick-oxide 5V transistors, closes the loop using an external loop filter. Stability of the DLL is achieved with a single capacitor, without the need of adding zeros in the loop filter: this is because, differently from PLLs, the loop of the DLL exhibits a first-order transfer function. The speed of the DLL tracking has no requirements as its main task is to track variations of process or temperature, so the capacitor value is determined jointly by the charge-pump current and the required ripple on the regulated voltage. However, this scheme has multiple stable points in the large signal domain, in particular for delays being (2N+1) times the semi-period; however, these points result to have very low regulated voltages and it is easy to avoid them by pulling the filter capacitor node towards the 5V supply with a high-valued resistor. The tuning range of the DLL allows to achieve delays with resolution from 190 ps up to 270 ps.

III. EXPERIMENTAL RESULTS

The imager has been fabricated in a 0.35-μm high-voltage CMOS technology: the use of a high-voltage module was motivated by the use of a deep n-well implant for the realization of efficient SPAD devices. Indeed, the SPAD is implemented using a p-plus diffusion inside a deep n-well, with p-well guardrings: this structure allows deep n-well sharing among different SPADs, and therefore NMOS transistors can be placed inside p-wells [20]. The chip micrograph is shown in Figure 11.

A. ELECTRO-OPTICAL CHARACTERIZATION

A first characterization of the chip can be performed without supplying synchronized light to the pixels, i.e. exploiting the SPADs' own DCR or a continuous light source.

When the pixel is exposed to a continuous light of a given intensity, the analog output voltage has the envelope of a Poissonian distribution with standard deviation determined by the shot noise of the photons, where the single steps of the analog counter are visible as small Gaussian distributions having a width dependent on the electrical noise of the counter and readout circuitry. The result is shown in Figure 12, where 10000 cycles with continuous light have been acquired: the peaks due to the analog counter steps are clearly visible, and even though for large number of photons they are merging, the pixel output still remains shot noise-limited. From this measurement it is possible also to evaluate the typical step of the analog counter with the nominal settings, which is 16.5 mV/photon. At the same time, the electronic noise amounts to $V_{nout} = 1.3$ mV$_{rms}$, with a pixel-to-pixel non uniformity of $\sigma_{Vout} = 2.6$ mV$_{rms}$. On the same graph of Figure 12, the expected output distribution is plotted as a solid line: for lower counts, there is a deviation from the expected statistics, while for higher counts the theory fits well with the measurement. The difference with respect to theory is due to the difficulty of maintaining controlled environmental conditions of very low illumination, few photons, during a relatively long acquisition needed to collect a meaningful statistics.

After the step characterization, the DCR map of the imager can be evaluated by letting the chip acquire for a large number of cycles and counting the photons in the total equivalent exposure. The SPAD breakdown voltage resulted to be approximately $V_{BD} = 18.6$ V and the DCR distribution maps have been acquired for several excess bias voltages. The results, shown in Figure 13, show a good DCR performance at the nominal Vexc = 3.0V, and highlight also the

presence of crosstalk at higher bias voltages, mainly optical (generation and absorption of photons in silicon [28]), due to the extremely tight packing of pixels.

Similarly to conventional imagers, it is possible to extract a power responsivity curve, for both analog and digital operating modes. The sensor has been exposed to a stable white lamp through a set of logarithmically-spaced neutral density filters, and both the average analog output voltage and the average digital output code have been acquired. In order to directly compare them, a tradeoff of the bias voltage and timing settings has been made, resulting in an overall lower dynamic range with respect to the optimal points of both modes. However, as shown in Figure 14, this also allowed to evaluate point-to-point the conversion linearity, extracting the integral non-linearity (INL). The differential non-linearity (DNL) is found bound at least to ±1, because it has been verified that there are no missing codes in the converter output population and the response is monotonic.

B. TIME-GATING CHARACTERIZATION

Time gating measurements require the use of a synchronized light source. A 70-ps full-width half-maximum (FWHM) 470-nm pulsed laser shining onto the sensor trough a diffuser has been used for the timing characterization, with the setup of Figure 15. The laser pulse can be delayed with respect to the sensor trigger with minimum steps of 10 ps using a programmable laboratory instrumentation. In order to evaluate the effective operation of the time gating, and the window characteristics such as edges, width, and uniformity, several acquisitions have been taken sweeping the laser pulse in a 5 ns range: the output of the sensor for several window widths is shown in Figure 16 for different settings of the T_{WIN} parameter. It can be seen how the window edges are well defined with rising and fall time of approximately 200 ps: this number is mainly

determined by the SPAD jitter and the laser FWHM. Slight variations of amplitude with TWIN have no impact on performance as they are below the counting step size.

The DLL stabilization has been tested by acquiring the gating window just after turning on the chip, after 30 minutes and after 90 minutes of continuous running. As plotted in Figure 17, the time-gating characteristics of the chip remain constant in time with a non-measurable variation with respect to the other non-idealities, i.e. jitter. Also the spatial uniformity, determined by the process uniformity and signal distribution, has been evaluated and depicted in Figure 18, where it can be seen that the window width variation across the array has a standard deviation of only 80.2 ps. This variation is mainly due to the differences in propagation of the signals SDRV and WIN which are determining the window edges, in particular with respect to the difference between two halves of the array served by different row drivers.

A complete summary of the chip characterization is given in Table I, while in Table II some key parameters are compared to the referenced literature. It can be seen how the presented sensor achieves a good fill-factor and a small pixel pitch even with the use of a low-cost CMOS technology node, without sacrificing performance and keeping a low power consumption: at the same time it is the only time-gated sensor based on analog counting featuring an embedded digital conversion which also compensates for first-order non-uniformities.

C. FLIM MEASUREMENTS

The setup for the FLIM measurements is schematically depicted in Figure 19: the imager has been mounted on the optical output port of a microscope setup (Olympus ix50 with 4x magnification) and the 70-ps laser has been coupled to the sample. The microscope contains a filter wheel to select some specific fluorescence wavelengths, and to remove the laser tail, which may be relevant with respect to fluorescence photons.

In order to provide a reference measurement, a slide containing CdSe/ZnS fluorescent dots has been first measured with a commercial TCSPC system and then evaluated with the developed sensor. The TCSPC lifetime measurement yields 11.18 ns of lifetime, while the time-gated measurement with acquisition of two windows is in good agreement, giving 11.8 ns. The ≈5% discrepancy is due to many factors, in particular to the multiexponential nature of the fluorescence, the signal-to-noise ratio of the measurement, and the spatial non-uniformity of the sample. Figure 20 shows the histogram of selected pixels, discarding those with low detected number of photons, and the map of the measured lifetime in the inset. The dispersion of the measured lifetime is function of the number of detected photons inside the window and can be improved by increasing the number of acquisitions.

Eventually, a FLIM experiment with a standard calibration test slide has been performed. The sample, a rhizome of Convallaria Majalis **Error! Reference source not found.**, has been excited with the 70-ps laser. Figure 21 shows the intensity image of the observed sample obtained without the excitation filter of the microscope and with the gating window overlapping the pulse emission. In order to measure the fluorescence lifetime, two partially overlapped observation windows with $T_{WIN} = 3$ ns have been acquired. Any background light, thanks to its uncorrelated nature and the short gating window, does not influence significantly the measurement, but increase the noise and therefore measurements have been performed in dark conditions. Each acquisition has been repeated to reach an equivalent observation time of 120 ms, acquiring a total of 200 thousand frames. The result is shown in Figure 22, where the fluorescence lifetime is extracted using eq. (1). Expected decay times are 3.8 ns for the Fast Green fluorophore, 1-2 ns for the cell boundary self-fluorescence and 5-7 ns for the chlorophyll in methanol. These decay times become qualitatively clear also observing the images resulting from the single windows:

Figure 23 shows the images obtained from 4 windows with T_{WIN} = 3 ns and spaced by 500 ps. As the delay increases, the finer background details quickly disappear, while the thicker border, besides having a stronger fluorescence amplitude, has a longer persistence and therefore a longer lifetime.

IV.   CONCLUSIONS

A time-gated image sensor based on SPAD and analog counting has been described and validated with FLIM measurements. The compact pixel achieves a 21% fill-factor in a 15μm pitch; the analog counting scheme behavior is independent from the SPAD characteristics and the voltage step can be tuned using external reference voltages, while a self-referenced conversion scheme allows fast and reliable digital readout up to 486 fps with a first-order removal of per-pixel non-uniformities. Programmable-width observation windows in the nanosecond range, down to 750 ps width, can be opened at a maximum frequency of 50 MHz and accumulated over several hundred cycles, with stable operation over time, supply and temperature thanks to the closed loop DLL control. The developed sensor has a power consumption of 20.6 mW in analog mode and 156.7 mW in digital mode, making it suitable for the development of future FLIM portable devices.

References

[1] X.F. Wang, A. Periasamy, B. Herman, and D. M. Coleman, "Fluorescence lifetime imaging microscopy (FLIM): instrumentation and applications," Critical Reviews in Analytical Chemistry, vol.23, no.5, pp.369-395, 1992.

[2] M. Popleteeva, A. Esposito, A. Venkitaraman, "Innovating Solid State Detection Technologies For Biomedical Imaging," Focus on Microscopy 2013, http://www.focusonmicroscopy.org/2013/PDF/358_Popleteeva.pdf (last accessed 22 May 2015).

[3] D. Phillips, R.C. Drake, D.V. O'Connor, and R.L. Christensen, "Time Correlated Single-Photon Counting (TCSPC) Using Laser Excitation," Instrumentation Science & Technology, vol.14, no.3-4, 1985.

[4] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, "Avalanche photodiodes and quenching circuits for single-photon detection," Applied Optics, vol.35, no.12, pp.1956-1976, 1996.

[5] 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 Journal of Solid-State Circuits, vol.43, no.12, pp.2977-2989, Dec. 2008.

[6] M. Gersbach, Y. Maruyama, R. Trimananda, M.W. Fishburn, D. Stoppa, J.A. Richardson, R. Walker, R. Henderson, and E. Charbon, "A time-resolved, low-noise single-photon image sensor fabricated in deep-submicron CMOS technology," IEEE Journal of Solid-State Circuits, vol.47, no.6, pp.1394-1407, June 2012.

[7] C. Veerappan, J. Richardson, R. Walker, D.U. Li, M.W. Fishburn, Y. Maruyama, D. Stoppa, F. Borghetti, M. Gersbach, R. Henderson, and E. Charbon, "A 160×128 single-photon image

sensor with on-pixel 55ps 10b time-to-digital converter," International Solid-State Circuits Conference ISSCC, Digest of Technical Papers, pp.312-314, February 2011.

[8] D. Bronzi, F. Villa, S. Tisa, A. Tosi, F. Zappa, D. Durini, S. Weyers, and W. Brockherde, "100000 Frames/s 64×32 Single-Photon Detector Array for 2-D Imaging and 3-D Ranging," IEEE Journal of Selected Topics in Quantum Electronics, vol.20, no.6, pp.354-363, November-December 2014.

[9] R.M. Field, S. Realov, and K.L. Shepard, "A 100 fps, time-correlated single-photon-counting-based fluorescence-lifetime imager in 130 nm CMOS," IEEE Journal of Solid-State Circuits, vol.49, no.4, pp.867-880, April 2014.

[10] D. Stoppa, F. Borghetti, J. Richardson, R. Walker, L. Grant, R.K. Henderson, M. Gersbach, and E. Charbon, "A 32x32-pixel array with in-pixel photon counting and arrival time measurement in the analog domain," Proceedings of the European Solid-State Circuit Conference, pp.204-207, September 2009.

[11] L. Parmesan, N.A. Dutton, N.J. Calder, A.J. Holmes, L.A. Grant, and R.K. Henderson, "A 9.8 μm sample and hold time to amplitude converter CMOS SPAD pixel," 44th European Solid State Device Research Conference (ESSDERC), pp.290-293, September 2014.

[12] D. Stoppa, N. Massari, L. Pancheri, M. Malfatti, M. Perenzoni, and L. Gonzo, "A range image sensor based on 10-μm lock-in pixels in 0.18-μm CMOS imaging technology," IEEE Journal of Solid-State Circuits, vol.46, no.1, pp.248–258, January 2011.

[13] A. Esposito, T. Oggier, H. C. Gerritsen, F. Lustenberger, and F. S. Wouters, "All-solid-state lock-in imaging for wide-field fluorescence lifetime sensing," Optics Express, vol.13, no.24, pp.9812–9821, November 2005.

[14] Z. Li, S. Kawahito, K. Yasutomi, K. Kagawa, J. Ukon, M. Hashimoto, and H. Niioka, "A time-resolved CMOS image sensor with draining only modulation pixels for fluorescence lifetime imaging," IEEE Transactions on Electron Devices, vol.59, no.10, pp.2715–2722, October 2012.

[15] L.-E. Bonjour, T. Baechler, and M. Kayal, "High-speed general purpose demodulation pixels based on buried photodiodes," in Proc. Int. Image Sensor Workshop, pp. 1–4, 2011.

[16] A.V. Agronskaia, L. Tertoolen, and H.C. Gerritsen. "High frame rate fluorescence lifetime imaging," Journal of Physics D-Applied Physics, vol.36, no.14, pp.1655-1662, 2003.

[17] M.-W. Seo, K. Kagawa, K. Yasutomi, T. Takasawa, Y. Kawata, N. Teranishi, Z. Li, I.A. Halin, and S. Kawahito, "11.2 A 10.8ps-time-resolution 256×512 image sensor with 2-Tap true-CDS lock-in pixels for fluorescence lifetime imaging," IEEE International Solid-State Circuits Conference (ISSCC), 22-26 February 2015.

[18] I. Rech , G. Luo , M. Ghioni , H. Yang , X. Sunney and S. Cova "Photon-timing detector module for single-molecule spectroscopy with 60-ps resolution", IEEE Journal of Selected Topics in Quantum Electronics, vol.10, no.4, pp.788-795, 2004.

[19] Y. Maruyama, and E. Charbon, "A time-gated 128x128 CMOS SPAD array for on-chip fluorescence detection," Proceedings of International Image Sensor Workshop (IISW), 2011.

[20] L. Pancheri, N. Massari, and D. Stoppa, "SPAD Image Sensor With Analog Counting Pixel for Time-Resolved Fluorescence Detection," IEEE Transactions on Electron Devices, vol.60, no.10, pp.3442-3449, October 2013.

[21] L. Pancheri, E. Panina, G.-F. Dalla Betta, L. Gasparini, and D. Stoppa, "Compact analog counting SPAD pixel with 1.9% PRNU and 530ps time gating," Proceedings of the European Solid-State Circuits Conference, pp.295-298, September 2013.

[22] N.A.W. Dutton, L. Parmesan, A.J. Holmes, L.A. Grant, and R.K. Henderson, "320x240 Oversampled Digital Single Photon Counting Image Sensor," IEEE Symposium on VLSI Circuits, June 2014.

[23] B. Dierickx, B. Dupont, A. Defernez, and N. Ahmed, "Indirect X-ray photon-counting image sensor with 27T pixel and 15e−rms accurate threshold," IEEE International Solid-State Circuits Conference (ISSCC), pp.114-116, 20-24 February 2011.

[24] M. Perenzoni, N. Massari, D. Perenzoni, L. Gasparini, and D. Stoppa, "A 160×120-pixel analog-counting single-photon imager with Sub-ns time-gating and self-referenced column-parallel A/D conversion for fluorescence lifetime imaging," IEEE International Solid-State Circuits Conference (ISSCC), 22-26 February 2015.

[25] K.K. Sharman, A. Periasamy, H. Ashworth, and J.N. Demas, "Error analysis of the rapid lifetime determination method for double-exponential decays and new windowing schemes," Analytical Chemistry, vol.71, no.5, pp.947-952, 1999.

[26] H.C. Gerritsen, M.A.H. Asselbergs, A.V. Agronskaia, and W.G.J.H.M. Van Sark, "Fluorescence lifetime imaging in scanning microscopes: acquisition speed, photon economy and lifetime resolution," Journal of Microscopy, vol. 206, no. 3, pp. 218–224. 2002.

[27] G.-F. Dalla Betta, D. Stoppa, J. Richardson, L. Pancheri, and R. Henderson, "Avalanche photodiodes in submicron CMOS technologies for high-sensitivity imaging," in Advances in Photodiodes, IntechOpen, pp. 226-248, 2011.

[28] I. Rech, A. Ingargiola, R. Spinelli, I. Labanca, S. Marangoni, M. Ghioni, and S. Cova, "Optical crosstalk in single photon avalanche diode arrays: a new complete model," Optics Express 16, pp.8381-8394, 2008.

[29] http://www.msmedia.com.au/convallaria-rhizome-training-slide-stained-for-fluorescence.php, last accessed 3 May 2015.
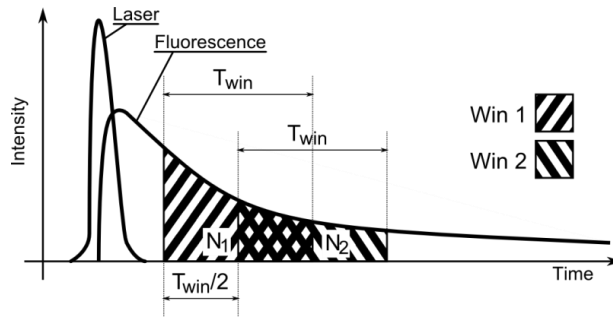
FIGURE CAPTIONS

Figure 1: Principle of operation of the lifetime extraction with two overlapped windows.
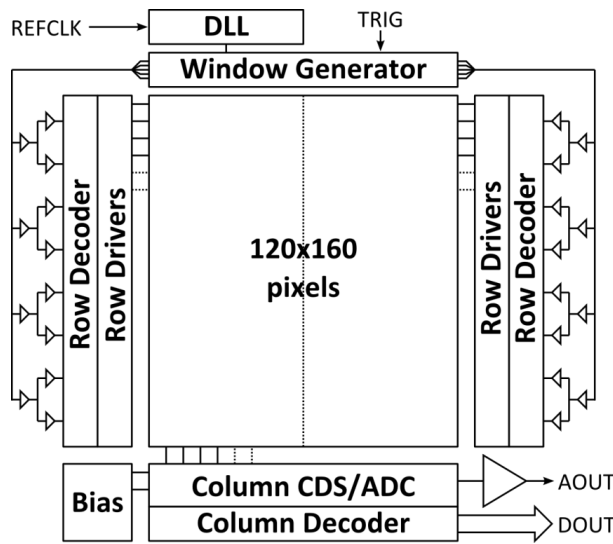
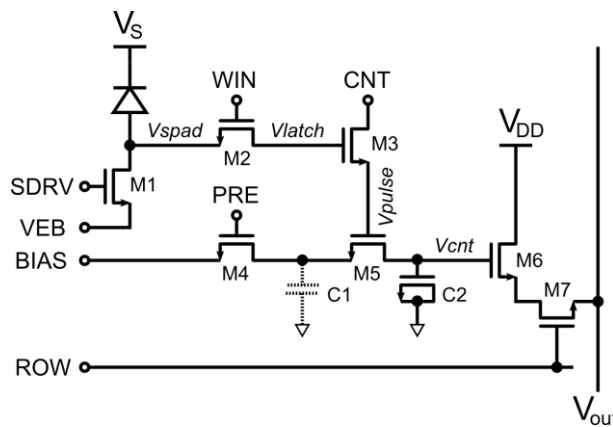Figure 2: High-level architecture of the time-gated SPAD imager.

Figure 3: Schematic of the pixel.

Figure 4: Pixel waveforms with multilevel signals and programmable time parameters.

Figure 5: MonteCarlo simulation of the counter step V_{step}. The inset shows the scatter plot with respect to the process variation of threshold voltage and of the oxide thickness.

Figure 6: Analog counter voltage step variations from a time-sweep simulation of the arrival at time T_{ph} of a photon versus corner conditions of temperature (0°C-70°C) and supply voltage (±10%), showing good robustness and sharp windowing windowing(SPAD jitter not included).

Figure 7: Schematic of the column circuitry for analog readout and analog-to-digital conversion.

Figure 8: Column readout waveforms for the analog mode (dashed line) and the digital mode (solid line).

Figure 9: Simulation of the self-referenced conversion with different counting step by changing the V_{BL} reference voltage.

Figure 10: Schematic of the DLL for the generation of the regulated supply for the delay lines of the window generator block.

Figure 11: Micrograph of the time-gated SPAD imager with main blocks highlighted.

Figure 12: Histogram of the output voltage after 10000 acquisitions (50 frames of 200 cycles) for two different continuous light intensities, with superimposed the expected theoretical Poisson distribution (solid line).

Figure 13: DCR distribution measurements for different excess bias and normalized images of the dark counts, showing relevant crosstalk at the higher V_{exc} value.

Figure 14: Power responsivity of the sensor in analog and digital mode, and analog-to-digital conversion linearity (INL).

Figure 15: Schematic view of the measurement setup for the time-gating characterization.

Figure 16: Measurement of the gating window for 10 different T_{WIN} settings, from 1.4 to 3.1 ns.

Figure 17: Time stability of the gating window, acquired immediately after turn-on, after 30 minutes, and after 90 minutes with the DLL active.

Figure 18: Time window uniformity measurement across the array with $\sigma_{TWIN} = 80.2$ ps: in the inset, the representation of the window FWHM showing a 99.2% operability of the array pixels.

Figure 19: Setup for FLIM measurements with the developed time-gated imager.

Figure 20: Lifetime measurement with CdSe/ZnS fluorescent dots: lifetime measurement with a commercial TCSPC system (left), and with the proposed sensor using the time-gated technique (right).

Figure 21: Intensity image of Convallaria sample, obtained without excitation filter, superimposing the gating window with the laser pulse.

Figure 22: Fluorescence lifetime acquisition of a Convallaria sample, showing the extracted image of the lifetime τ (left), and histograms of the lifetime in two different regions of the sample (right).

Figure 23: Images of 4 subsequent acquisition windows of 3 ns duration, spaced by 500 ps, where it is possible to appreciate the different intensities and lifetimes of the Convallaria sample. Scale is in counted photons.

TABLE CAPTIONS

Table I: Summary of the main characteristics of the sensor.

Table II: Comparison table of SPAD-based image sensors with counting/timing capabilities.

Figure 1: Principle of operation of the lifetime extraction with two overlapped windows.

Figure 2: High-level architecture of the time-gated SPAD imager.
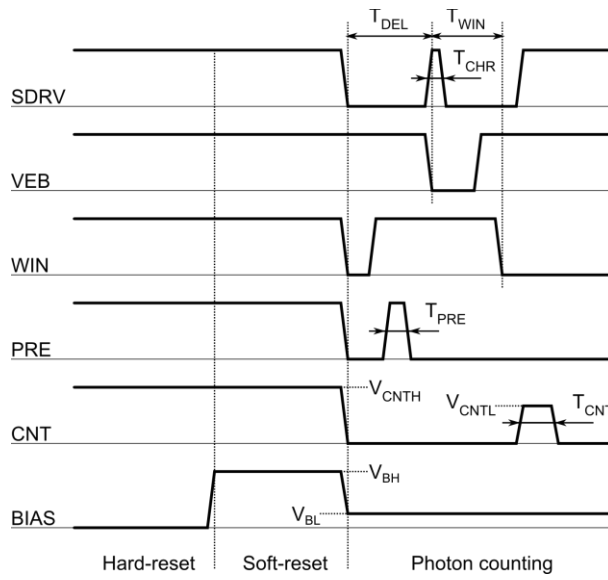

Figure 3: Schematic of the pixel.

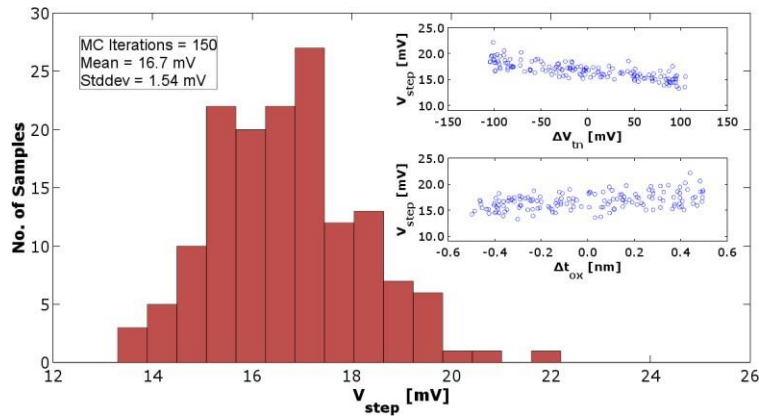Figure 4: Pixel waveforms with multilevel signals and programmable time parameters.

Figure 5: MonteCarlo simulation of the counter step V_{step}. The inset shows the scatter plot with respect to the process variation of threshold voltage and of the oxide thickness.

Figure 6: Analog counter voltage step variations from a time-sweep simulation of the arrival at time T_{ph} of a photon versus corner conditions of temperature (0°C-70°C) and supply voltage (±10%), showing good robustness and sharp windowing windowing(SPAD jitter not included).

Figure 7: Schematic of the column circuitry for analog readout and analog-to-digital conversion.

Figure 8: Column readout waveforms for the analog mode (dashed line) and the digital mode (solid line).

Figure 9: Simulation of the self-referenced conversion with different counting step by changing the V_{BL} reference voltage.

Figure 10: Schematic of the DLL for the generation of the regulated supply for the delay lines of the window generator block.

Figure 11: Micrograph of the time-gated SPAD imager with main blocks highlighted.

Figure 12: Histogram of the output voltage after 10000 acquisitions (50 frames of 200 cycles) for two different continuous light intensities, with superimposed the expected theoretical Poisson distribution (solid line).

Figure 13: DCR distribution measurements for different excess bias and normalized images of the dark counts, showing relevant crosstalk at the higher V_{exc} value.

Figure 14: Power responsivity of the sensor in analog and digital mode, and analog-to-digital conversion linearity (INL).

Figure 15: Schematic view of the measurement setup for the time-gating characterization.
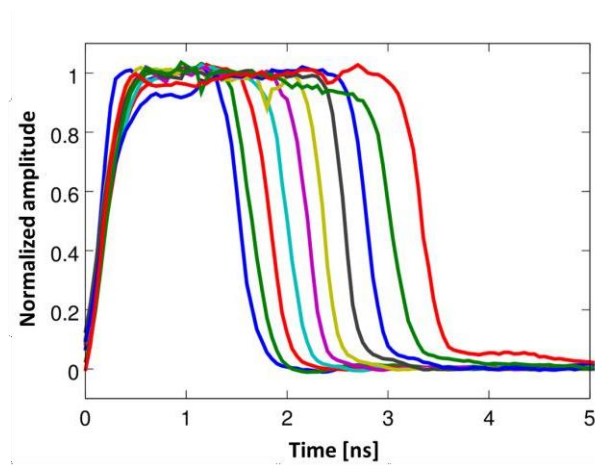
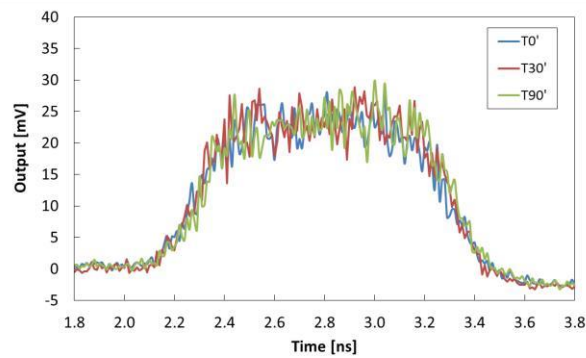Figure 16: Measurement of the gating window for 10 different T$_{WIN}$ settings, from 1.4 to 3.1 ns.

Figure 17: Time stability of the gating window, acquired immediately after turn-on, after 30 minutes, and after 90 minutes with the DLL active.
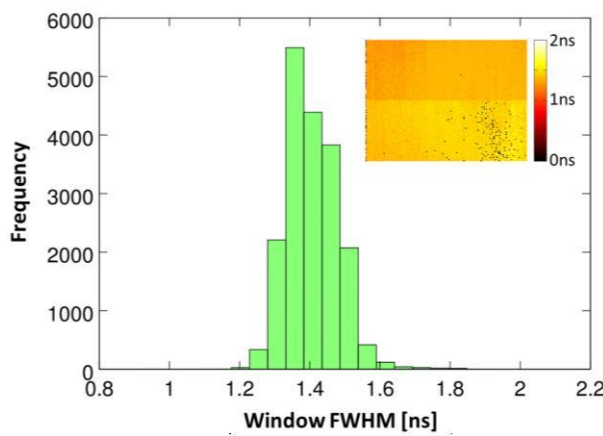

Figure 18: Time window uniformity measurement across the array with σ$_{TWIN}$ = 80.2 ps: in the inset, the representation of the window FWHM showing a 99.2% operability of the array pixels.
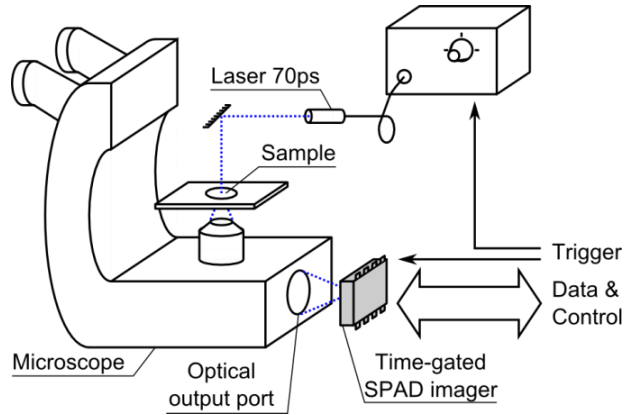
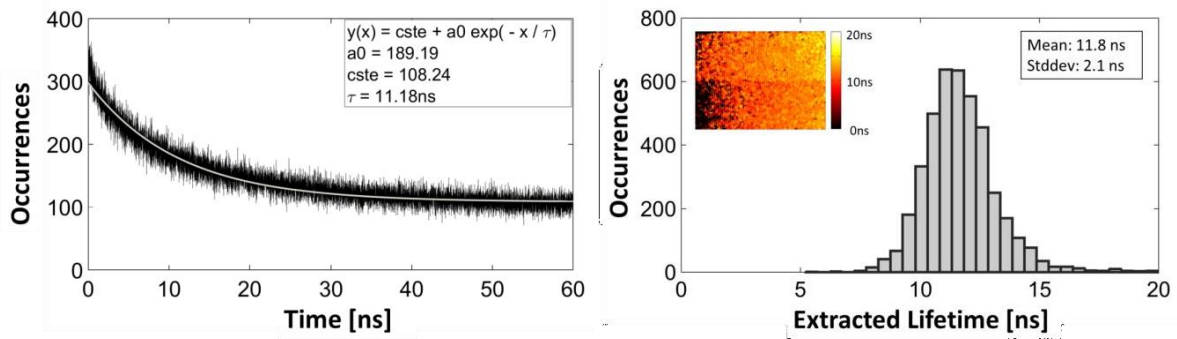Figure 19: Setup for FLIM measurements with the developed time-gated imager.

Figure 20: Lifetime measurement with CdSe/ZnS fluorescent dots: lifetime measurement with a commercial TCSPC system (left), and with the proposed sensor using the time-gated technique (right).
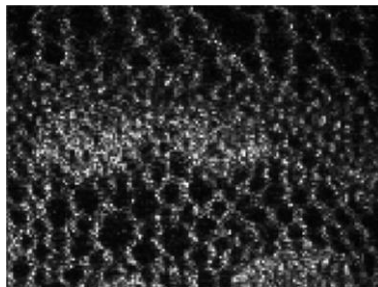

Figure 21: Intensity image of Convallaria sample, obtained without excitation filter, superimposing the gating window with the laser pulse.
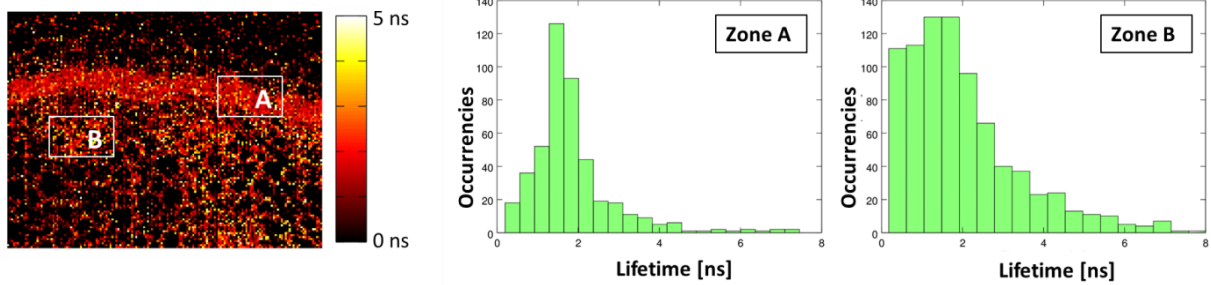
Figure 22: Fluorescence lifetime acquisition of a Convallaria sample, showing the extracted image of the lifetime τ (left), and histograms of the lifetime in two different regions of the sample (right).
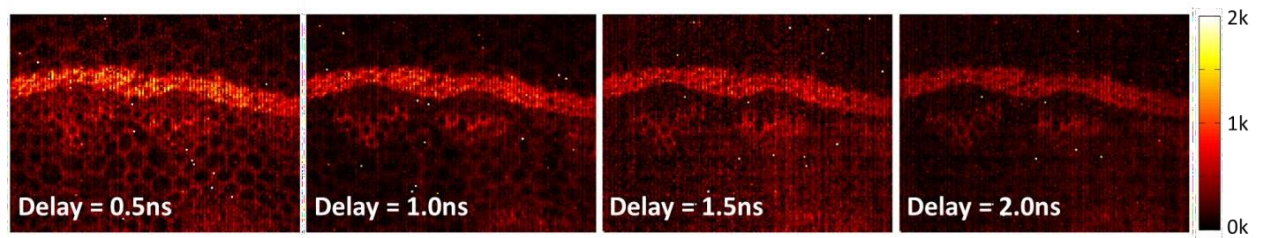

Figure 23: Images of 4 subsequent acquisition windows of 3 ns duration, spaced by 500 ps, where it is possible to appreciate the different intensities and lifetimes of the Convallaria sample. Scale is in counted photons.

Table I: Summary of the main characteristics of the sensor.

Parameter	Notes	Value	Unit
Chip features			
Array resolution		160×120 pixels	
Chip size		3.42×3.55	mm^2
Pixel size		15	μm
Fill-factor		21	%
Transistors/pixel		7T + 1MOSCAP	
SPAD Median DCR	Vexc = 3 V	580	Hz
Time gating			
Delay resolution	Tref =10ns	194	ps
Time gating fmax		50	MHz
Gating windows		200	cycles/frame
Analog output			
Sensitivity		16.5	mV/photon
Pixel noise		1.33	mV
Step uniformity		2.6	mV
Output range		1.15	V
		70	photons
Output rate		5.0	MS/s
Max frame rate		231	Hz
Power	Total	20.6	mW
	Digital 5.0V	4.5	mW
	Digital 3.3V	0.6	mW
	Analog 3.3V	15.5	mW
Digital mode			
Sensitivity		1	LSB/photon
Pixel noise		0.55	LSB
DNL		<1	LSB
INL		<1.01	LSB
Output range		5.4	bit
		41	photons
Output rate		50	MHz
Max frame rate		486	Hz
Power	Total	156.7	mW
	Digital 5.0V	123.2	mW
	Digital 3.3V	18.0	mW
	Analog 3.3V	15.5	mW

Table II: Comparison table of SPAD-based image sensors with counting/timing capabilities.

Reference →	[7]	[20]	[21]	[22]	[9]	[19]	This Work
Technology	0.13μm CIS	0.35μm HV	0.15μm	0.13μm CIS	0.13μm	0.35μm	0.35μm HV
Array resolution	160×128	32×32	n.a.	320×240	64×64	128×128	160×120
Pitch	50μm	25μm	n.a.	8μm	48μm	25μm	15μm
Fill-factor	1%	20.8%	n.a.	26.8%	0.77%	4.5%	21%
Transistor/pixel	>50T	12T + 1C	12T + 2C	8T+1C	>30T	12T	7T + 1C
Power consumption	550 mW@7bit	33 mW	n.a.	69.5 mW^A 69.0 mW^D@1bit	8.79 W	363 mW	20.6 mW^A 157 mW^D@8bit
Photons/frame	128 ph	150 ph	90 ph	400 ph^A 1 ph^D	n.a.	1 ph	70 ph^A 256 ph^D
Method	Counting^D Timestamp	Counting^A Gating	Counting^A Gating	Counting^{A,D}	Counting^D Timestamp	Gating	Counting^{A,D} Gating
Max frame rate	500 kfps	180 fps	n.a.	7 fps^A 5.1 kfps^D 5.1 kcnt/s^D	466 fps	2441 fps	231 fps^A 486 fps^D 124 kcnt/s^D
Min time gating	n.a.	1.1 ns	0.53 ns	n.a.	n.a.	n.d.	0.75 ns
Gate fmax	1 MHz	40 MHz	50 MHz	n.a.	20 MHz	40 MHz	50MHz

Note: ^{A}Analog counting mode, ^{D}Digital counting mode