Architecture and applications of a high resolution gated SPAD image sensor

Samuel Burri,[1,*] Yuki Maruyama,[2] Xavier Michalet,[3]
Francesco Regazzoni,[4] Claudio Bruschini,[1] and Edoardo Charbon,[2]

[1]*School of Engineering, Swiss Federal Institute of Technology, Lausanne, Switzerland*
[2]*TU Delft, The Netherlands*
[3]*Department of Chemistry & Biochemistry, UCLA, Los Angeles, USA*
[4]*ALaRI, USI, Lugano, Switzerland*

[*]*samuel.burri@epfl.ch*

Abstract: We present the architecture and three applications of the largest resolution image sensor based on single-photon avalanche diodes (SPADs) published to date. The sensor, fabricated in a high-voltage CMOS process, has a resolution of 512 x 128 pixels and a pitch of 24 μm. The fill-factor of 5% can be increased to 30% with the use of microlenses. For precise control of the exposure and for time-resolved imaging, we use fast global gating signals to define exposure windows as small as 4 ns. The uniformity of the gate edges location is ~140 ps (FWHM) over the whole array, while in-pixel digital counting enables frame rates as high as 156 kfps.

Currently, our camera is used as a highly sensitive sensor with high temporal resolution, for applications ranging from fluorescence lifetime measurements to fluorescence correlation spectroscopy and generation of true random numbers.

OCIS codes: (030.5260) Photon counting; (040.0040) Detectors; (040.1240) Arrays; (100.0118) Imaging ultrafast phenomena; (110.0110) Imaging systems; (180.2520) Fluorescence microscopy; (230.5160) Photodetectors.

References and links

1. R. H. Haitz, A. Goetzberger, R. M. Scarlett, and W. Shockley, "Avalanche effects in silicon p-n junctions. I. Localized photomultiplication studies on microplasmas," J. Appl. Phys. **34**, 1581 (1963).
2. A. Goetzberger, R. M. Scarlett, R. H. Haitz, and B. Mcdonald, "Avalanche effects in silicon p-n junctions. II. Structurally perfect junctions," J. Appl. Phys. **34**, 1591 (1963).
3. S. Cova, A. Longoni, and A. Andreoni, "Towards picosecond resolution with single-photon avalanche diodes," Rev. Sci. Instrum. **52**, 408–412 (1981).
4. R. J. McIntyre, "Recent developments in silicon avalanche photodiodes," Measurement **3**, 146–152 (1985).
5. A. Rochas, M. Gösch, A. Serov, P. A. Besse, R. S. Popovic, T. Lasser, and R. Rigler, "First fully integrated 2-D array of single-photon detectors in standard CMOS technology," IEEE Photonics Technol. Lett. **15** (2003).
6. E. Charbon and S. Donati, "SPAD sensors come of age," Opt. Photonics News **21**, 34–41 (2010).
7. M. Gersbach, R. Trimananda, Y. Maruyama, M. W. Fishburn, D. Stoppa, J. Richardson, R. Walker, R. K. Henderson, and E. Charbon, "High frame-rate TCSPC-FLIM using a novel SPAD-based image sensor," SPIE Optics+Photonics, Single Photon Imaging Conference (OP111), SPIE Paper 7780C-58 (2010).
8. A. P. Singh, J. W. Krieger, J. Buchholz, E. Charbon, J. Langowski, and T. Wohland, "The performance of 2D array detectors for light sheet based fluorescence correlation spectroscopy," Opt. Express **21**, 8652–8668 (2013).
9. S. Bellisai, F. Villa, S. Tisa, D. Bronzi, and F. Zappa, "Indirect time-of-flight 3D ranging based on SPADs," Proc. SPIE **8268**, 82681C–82681C–8 (2012).
10. C. Niclass, K. Ito, M. Soga, H. Matsubara, I. Aoyagi, S. Kato, and M. Kagami, "Design and characterization of a 256x64-pixel single-photon imager in CMOS for a MEMS-based laser scanning time-of-flight sensor," Opt. Express **20**, 11863–11881 (2012).

11. L. Braga, L. Gasparini, L. Grant, R. Henderson, N. Massari, M. Perenzoni, D. Stoppa, and R. Walker, "A fully digital 8 x 16 SiPM array for PET applications with per-pixel TDCs and real-time energy output," IEEE J. Solid-State Circuits **49**, 301–314 (2014).
12. J. Blacksberg, Y. Maruyama, E. Charbon, and G. R. Rossman, "Fast single-photon avalanche diode arrays for laser raman spectroscopy," Opt. Lett. **36**, 3672–3674 (2011).
13. R. A. Colyer, G. Scalia, F. A. Villa, F. Guerrieri, S. Tisa, F. Zappa, S. Cova, S. Weiss, and X. Michalet, "Ultra high-throughput single molecule spectroscopy with a 1024 pixel SPAD," Proc. SPIE **7905**, 790503–790503–8 (2011).
14. E. S. Harmon, M. Naydenkov, and J. T. Hyland, "Compound semiconductor SPAD arrays," Proc. SPIE **8727**, 87270N–87270N–13 (2013).
15. E. Charbon and M. W. Fishburn, *Monolithic Single-Photon Avalanche Diodes: SPADs* (Springer, 2011), chap. 7, pp. 123–156.
16. Y. Maruyama and E. Charbon, "A time-gated 128x128 CMOS SPAD array for on-chip fluorescence detection," Proc. Intl. Image Sensor Workshop (IISW) (2011).
17. R. J. Walker, E. A. G. Webster, J. Li, N. Massari, and R. Henderson, "High fill factor digital silicon photomultiplier structures in 130nm CMOS imaging technology," IEEE NSS/MIC (2012).
18. S. Donati, G. Martini, and M. Norgia, "Microconcentrators to recover fill-factor in image photodetectors with pixel on-board processing circuits," Opt. Express **15**, 18066–18075 (2007).
19. C. Veerappan, J. Richardson, R. Walker, D.-U. Li, M. W. Fishburn, Y. Maruyama, D. Stoppa, F. Borghetti, M. Gersbach, R. K. Henderson, and E. Charbon, "A 160x128 single-photon image sensor with on-pixel 55ps 10b time-to-digital converter," IEEE Intl. Solid-State Circuits Conference (ISSCC) (2011).
20. J. M. Pavia, M. Wolf, and E. Charbon, "Measurement and modeling of microlenses fabricated on single-photon avalanche diode arrays for fill factor recovery," Opt. Express **22**, 4202–4213 (2014).
21. M. Motoyoshi, "Through-silicon via (TSV)," Proc. IEEE **97**, 43–48 (2009).
22. S. J. Koester, A. M. Young, R. R. Yu, S. Purushothaman, K.-N. Chen, J. D. C. La Tulipe, N. Rana, L. Shi, M. R. Wordeman, and E. J. Sprogis, "Wafer-level 3D integration technology," IBM J. Res. Dev. **52**, 583–597 (2008).
23. M. Entwistle, M. A. Itzler, J. Chen, M. Owens, K. Patel, X. Jiang, K. Slomkowski, and S. Rangwala, "Geiger-mode APD camera system for single-photon 3D LADAR imaging," Proc. SPIE **8375**, 83750D–83750D–12 (2012).
24. R. M. Clegg, *Fluorescence Imaging Spectroscopy and Microscopy* (John Wiley & Sons, 1996).
25. K. Suhling, "Fluorescence lifetime imaging," in "Cell Imaging," , D. Stephens, ed. (Scion Publishing, Bloxham, 2006).
26. M. Y. Berezin and S. Achilefu, "Fluorescence lifetime measurements and biological imaging," Chem. Rev. **110**, 2641–2684 (2010).
27. R. A. Colyer, O. H. W. Siegmund, A. S. Tremsin, J. V. Vallerga, S. Weiss, and X. Michalet, "Phasor imaging with a widefield photon-counting detector," J. Biomed. Opt. **17**, 016008 (2012).
28. J. R. Lakowicz, *Principles of Fluorescence Spectroscopy* (Springer, 2006), p. 954.
29. I. Kanter, Y. Aviad, I. Reidler, E. Cohen, and M. Rosenbluh, "An optical ultrafast random bit generator," Nat. Photonics **4**, 58–61 (2010).
30. W. Wei, G. Xie, A. Dang, and H. Guo, "High-speed and bias-free optical random number generator," IEEE Photonics Technol. Lett. **24**, 437–439 (2012).
31. NIST, "A statistical test suite for the validation of random number generators and pseudo random number generators for cryptographic applications," Pub 800-22 rev1a (2010).

1. Introduction

Solid-state photon counting is a powerful technique enabling the detection of light at the highest level of sensitivity: a single photon. Available since the 1960s, avalanche photodiodes (APDs) operating in proportional and in Geiger mode have opened the way to photon-counting at the microscopic level, allowing time-resolved imaging with high resolution. When operating in Geiger mode, APDs are called single-photon avalanche diodes (SPADs), and their structure, modelling, and performance characteristics have been described in numerous publications [1–4].

Until recently, SPADs were discrete, fragile devices fabricated in special processes usually incompatible with mass-production. The development of SPADs in standard CMOS is a recent innovation [5] and a breakthrough that enabled their wide adoption [6].

A CMOS process, optimized for the creation of classical electronic circuits, needs to be sufficiently mature to enable the integration of SPADs with good performance. The main cri-

teria are the cleanliness of the process and the availability of the necessary layers to create the SPAD. Since most manufacturers do not disclose the internal details of their processes and are prone to tune the process parameters even at the later stages of the process life-cycle, the whole fabrication of CMOS SPADs remains challenging.

In comparison with classic image-sensors based on charge accumulation (CCD and CMOS), SPAD-based sensors enable photon-counting at the pixel level. As photons impinge at different instants, they can be identified and assigned a timestamp. With a large number of pixels, each of them capable of detecting single photons and measuring their arrival time, it is possible to explore phenomena which would be otherwise difficult to study. Among the different approaches that can take advantage of this capability, time-correlated single-photon counting (TCSPC) techniques, which is used for fluorescence lifetime imaging microscopy (FLIM), figure prominently and have been demonstrated using SPAD-based imaging systems [7, 8].

Other applications for SPAD array sensors include time-of-flight 3D imaging [9, 10] and positron emission tomography [11], different kinds of spectroscopy [12, 13] and imaging in high radiation environments [14].

In this paper, we introduce the largest SPAD-based image sensor to date, SwissSPAD featuring a resolution of 512 x 128 pixels and a global gating circuit for accurate timing information.

After describing the architecture of our sensor system (Section 2) and evaluating its performance (Section 3), we demonstrate the benefits of microlenses (Section 4) before presenting three applications in the next three sections: high-speed movies (Section 5), FLIM (Section 6), and true random number generation (Section 7).

2. Imager architecture

We describe the architecture of our sensor system following a bottom-up approach. We first describe the CMOS SPAD structure, then present the pixel circuit and the whole chip architecture. Finally we describe the full system, including the FPGA firmware controlling the sensor and providing the computer interface.

2.1. CMOS SPAD

Any p-n junction designed in a CMOS fabrication process can, in principle, be used as a SPAD. For most combinations, however, the electric field at the edges of the implants in a planar process tends to exceed that of the center, thus causing premature edge breakdown (PEB) and preventing operation in Geiger mode, i.e. above breakdown where optical gain is virtually infinite. As a result PEB has to be avoided when implementing SPADs. This is done for instance by using a guard ring with different doping concentration around the multiplication region. In [15], Charbon et al. provide more details on techniques to prevent PEB.

SwissSPAD is fabricated in a standard high-voltage 0.35 μm CMOS process. The SPAD sub-circuit, which has a cross-section shown in Fig. 1, was derived from the structure presented by Maruyama et al. in [16]. A deep n-well in a p-type silicon substrate is used as a cathode inside which a p+ anode is implanted. The p+ to deep n-well junction forms the active region of our SPAD structure and is circular with a diameter of approximately 6 μm. A p-well implantation is used as a guard ring to prevent PEB. Using this approach, anode and cathode are decoupled from ground and from other operating voltages.

Figure 2(a) shows the photon detection probability (PDP) and efficiency (PDE) of the SPAD structure for different excess bias voltages. Excess bias is the voltage above breakdown at which the SPAD operates. PDP measures the probability of detection for a photon impinging the active region; PDE includes the fill-factor of 5% and denotes the global detection probability for an array.

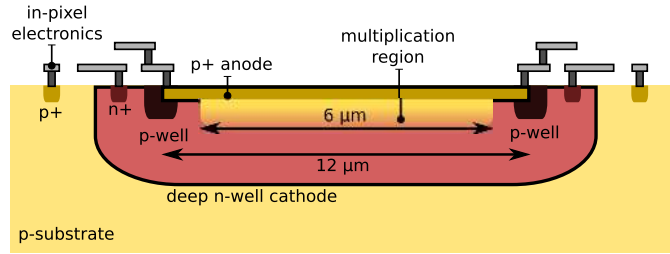Figure 2(b) shows the distribution of the dark count rate (DCR) as measured in SwissSPAD

Fig. 1. Cross-section of the SPAD structure fabricated in a 0.35 μm high-voltage CMOS process. A p+ - deep n-well junction is used to create the multiplication region with a p-well guard ring to prevent premature edge breakdown.

for different excess bias voltages. DCR is the number of events registered in the absence of photon flux. The typical DCR distribution has most pixels around the median value with a few percent non-working and very noisy pixels.

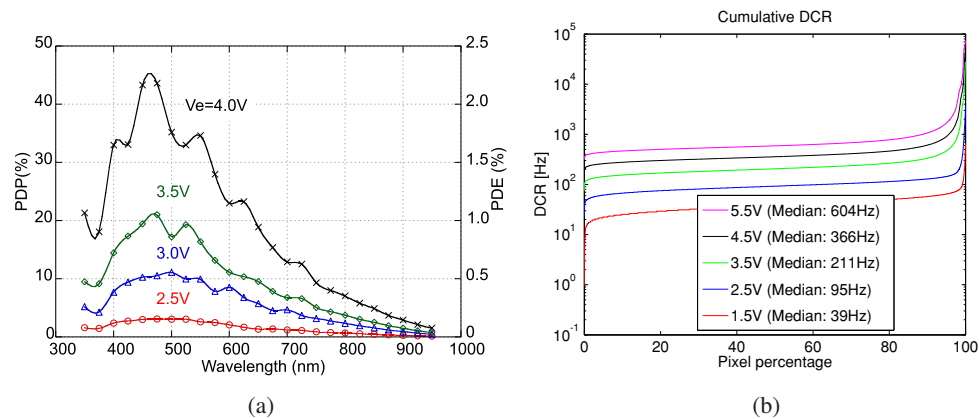Table 1 lists all parameters characterizing the performance of SwissSPAD.

(a) (b)

Fig. 2. (a) Pixel photon detection probability (PDP) and array photon detection efficiency (PDE) in the range of 350 nm to 950 nm for various excess bias voltage. (b) SPAD dark count rate (DCR) distribution as a function of excess bias voltage at room temperature. The hottest 1% of the pixels (reaching 2 MHz) were removed from this plot.

2.2. Pixel circuit

When designing an integrated pixel around a SPAD the main trade-off is between the fill-factor and pixel functionality. In order to have a large active area per pixel the electronics inside a single pixel has to be as small as possible. On the other hand the electronics has to provide the necessary performance in terms of timing resolution and dynamic range and at the same time support the creation of a pixel array. Our goal was to achieve a high spatial resolution with a reasonable fill-factor which lead to the design of a pixel with a 1-bit memory and global shutter.

The circuit of the pixel used in SwissSPAD is composed of 11 transistors arranged around the SPAD structure in a square of 24 x 24 μm^2. The entire pixel circuit is fabricated using NMOS logic exclusively to achieve tighter spacing thanks to the lack of n-wells necessary for PMOS transistors. The circuit is presented in Fig. 3.

The pixel comprises three parts. The photo-sensitive front-end is composed of the SPAD and
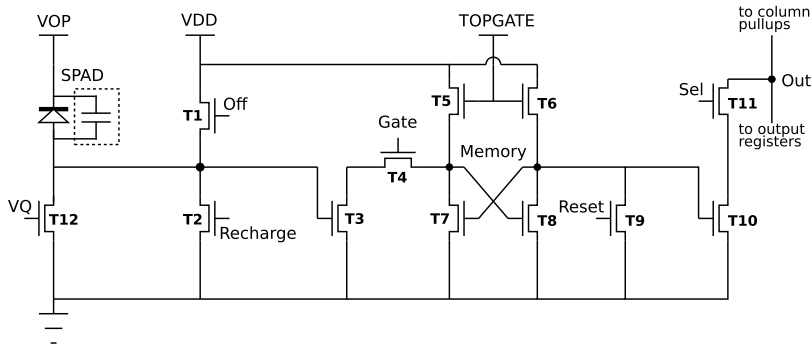
Fig. 3. Transistor level schematic of the pixel circuit. The SPAD is shown together with its junction capacitance. T12 can be used for passive quenching and separates the SPAD from ground. Transistors T1 and T2 control the SPAD bias and are used to switch the SPAD on and off. T4 controls the access to the NMOS-latch formed by T7 and T8, loaded by T5 and T6. T9 is used to reset the storage latch, previously set by T3. Finally T10 is used to transfer the memory value to the output line through the row select transistor T11.

the transistors controlling its anode voltage. The back-end is formed by the memory reset, value transfer and row access transistors. Between these two parts there is an NMOS latch used to memorize photon events.

Transistors T1, T2 and T4 form the gating circuit. They are used to define the time window during which the SPAD is photo-sensitive as well as to shut it off when it is not used. When T1 is turned on the SPAD is deactivated by removing the excess bias voltage from the SPAD anode. T2 is pulsed to restore the SPAD to operating condition and T4 prevents registration of false events from turning the SPAD on and off. These three transistors are controlled by three global high-speed signals. Their relative timing is critical for the operation of the sensor and it is essential that signal skew be minimized among the pixels.

The avalanche triggered by a photon hitting an active SPAD structure is quenched naturally due to the rising anode voltage. This voltage is sensed at the gate of T3 and the 1-bit memory cell formed by T5-T8 is flipped when T3 turns on. The value of the memory is read out through T11 and reset through T9.

2.3. Image sensor chip architecture

An array of 512 x 128 pixels is formed by connecting the select signals along the rows and the output signals along the columns. Reset generators are inserted on the rows, providing an automatic reset pulse when the row is deselected after readout. The output line going to the data registers is connected to an active pull-up circuit for faster operation. When a new row is selected, the effective value of the pull-up resistor is lowered to achieve a faster pull-down by the small in-pixel transistors. A full line is registered before being multiplexed on the output signals to ensure synchronous readout and to permit the selection of the next row during data transfer. The chip operates as a three stage pipeline with the critical path from the row select register through the pixel to the data register.

To achieve a high timing precision in the global shutter we use matched signal trees with large drivers on the long side of the array. Thus we can distribute the three shutter signals with a low skew of a few hundred picoseconds across the full array. The shutter signals operate independently from the readout and memory reset circuitry and they are the same for all pixels. The buffers driving these signals are powered by an independent supply bus.

To read a 1-bit image from the chip the control circuit first generates the shutter pattern to

detect photons and then iterates row-by-row over the array reading the data. During this readout the memory is reset and the process can begin anew. Shutter and readout can be overlapped to detect photons between successive readouts of the rows.

Figure 4 shows the image sensor chip micrograph. The pixel array measuring 12.3 mm by 3 mm is surrounded on three sides by control and readout logic. As the array touches one edge of the chip, two chips can be mounted side by side on a PCB with a gap smaller than 6 pixels as seen in Fig. 5. Figure 6 illustrates the image sensor chip architecture in a block diagram including the external readout system.

512 x 128
SPAD pixel array

24um

12.3mm

Column registers, multiplexers, signal trees
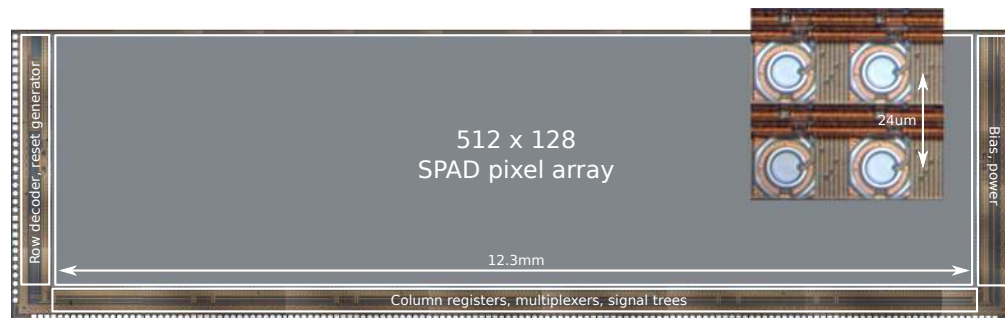
Row decoder, reset generator

Bias, power

Fig. 4. SwissSPAD die micrograph with the SPAD-array in the center and logic on three sides. Two sensors can be abutted with a gap less than 6 pixels wide for a combined resolution of 1/8 Megapixel. The inset is a detail of the SPAD cells.

2.4. System architecture

To achieve a useful dynamic range from a digital pixel one needs to accumulate a high number of measurements. A Xilinx Virtex 4 FPGA is used to generate the control signals and receive the column outputs. The FPGA firmware architecture is separated in three independent parts. The first handles gating or illumination control, by generating the gate signals for pixel transistors T1, T2 and T4. These signals specify when the chip is sensitive to photons. The second part controls the readout and generates row addresses and register clocks. Finally, the third part handles the double buffer image accumulation, accumulating and reordering data, before storing the final images in the off-chip memory. Figure 6 shows the entire image sensor system and the relations between the firmware parts.

To build our imaging system the PCB, shown in Fig. 5, is connected with four high density connectors to the larger PCB carrying the FPGA.

The readout part of SwissSPAD is operated at 80 MHz generating a data rate of 10.2 Gbps over the 128 output lines. The USB2 interface, with a top speed of 480 Mbps, cannot sustain this data rate. The data is therefore accumulated in the FPGA in the form of grayscale images and then written in large SO-DIMM buffer DRAMs. A double-buffering scheme is employed such that the full speed can be sustained at all times. The intensity resolution of images can be configured between 1 and 16 bits in order to obtain the optimal trade-off between memory space and time resolution.

3. Performance evaluation

The timing precision reached by the gating circuitry determines the performance and flexibility of the SwissSPAD imager. Three input signals: Spadoff, Recharge and Gate, control the photon detection time-window. These signals are generated by the FPGA and distributed on-chip to all
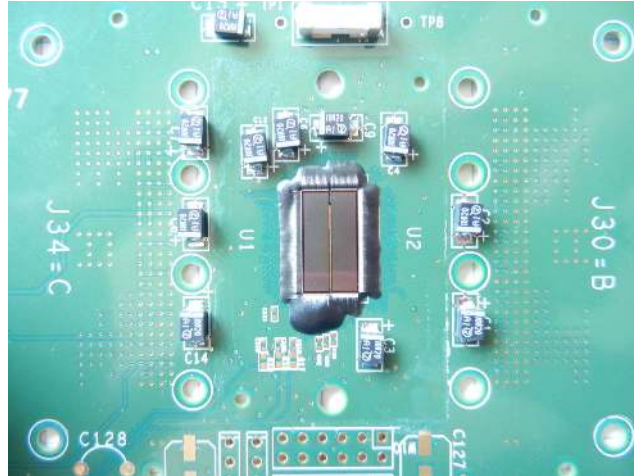
Fig. 5. Chip carrier PCB with two chips bonded side by side for simultaneous operation at doubled resolution. The gap is less than 6 pixels wide.
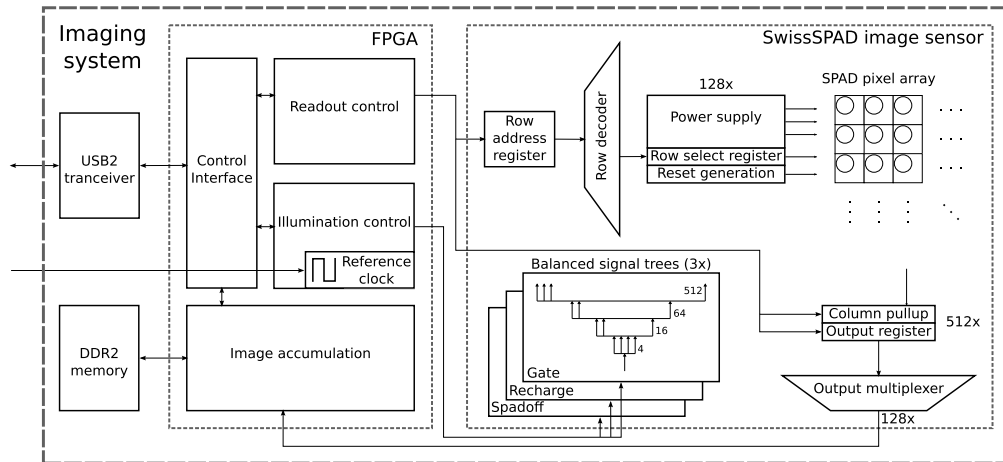

Fig. 6. A block-level representation of our imaging system. The part on the right depicts the interior of the SwissSPAD chip which is built around the central array of 128 x 512 pixels. The power supplies and gating signals are shared among all pixels, while the selection and reset signals are row-by-row. The pixel outputs are connected column-by-column with pullup circuits on one end and data registers and output multiplexers on the other end. An FPGA, depicted on the left, is used to generate the control signals and receive the data generated by the pixels.

pixels using a balanced signal tree on the long side of the array to minimize skew. A detailed skew analysis will be presented later.

The three gating signals are driven from the same clock signal which imposes a fixed phase relationship and restricts their timing granularity. This clock signal, however, can be generated from an external reference signal and the phase offset between this reference signal and the clock of the gating signals can be programmed in increments of approximately 20 ps. This allows defining a fixed-duration signal pattern, which can then be shifted in time with respect to a reference signal.

For characterization of the timing performance, we used a 637 nm picosecond laser with 40 MHz repetition rate and 35 ps FWHM pulse (Advanced Laser Diode Systems A. L. S. GmbH). The laser Sync signal was used to generate the reference clock, so that in practice, we evaluated the gating precision of the whole system comprised of a laser, the FPGA and the SwissSPAD chip. The chip was continuously illuminated by the pulsed laser, while the gating pattern was chosen such as to produce a photon detection window shorter than the laser repetition period. After data accumulation with a given window, the pattern was then shifted with respect to the laser reference and data was accumulated for the same duration. This procedure was repeated until the whole laser repetition period was covered. Figure 7 shows the signals and relationships involved in the procedure.
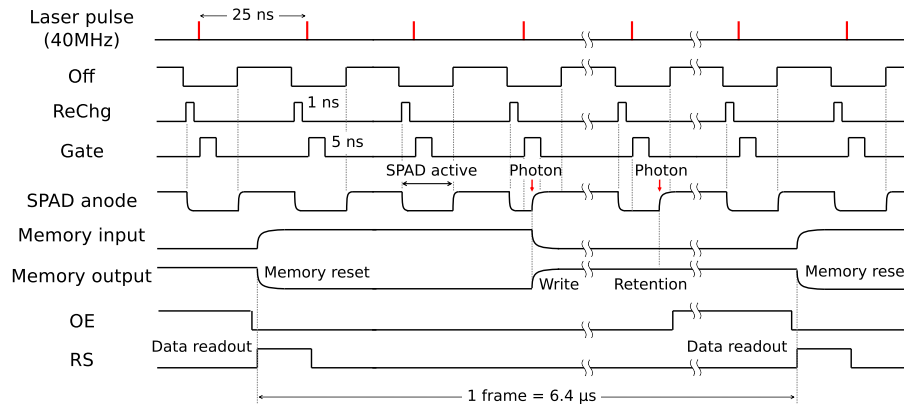

Fig. 7. Timing diagram for pulsed illumination imaging. The gating signals (Off, ReChg, GATE) are derived from the reference clock supplied by the illumination system (here a picosecond laser with 40 MHz repetition rate). Output enable (OE) and reset (RS) signals are used to control the chip readout.

For performance characterisation, 255 readouts were performed for each of the 1280 pattern positions possible in one laser period. We obtain 8-bit values covering the 25 ns clock period in steps of approximately 20 ps. The result is represented for one pixel in Fig. 8 and can be seen as a convolution of the picosecond laser pulse and the sensitivity window defined by our gating mechanism.

Figure 8(a) shows that the rising edge is very fast, while the falling edge is comparatively slower, exhibiting a tail. This can be explained by the fact that the fast edge is the one corresponding to the end of the sensitive window, marked by the turning off of the gate transistor T4, blocking access to the memory. This happens quickly and it is completely independent from photon flux and SPAD bias voltage. The slow edge on the other hand marks the beginning of the sensitive window, when the SPAD bias voltage is restored to its nominal value through the recharge transistor T2. How fast this voltage is restored in response to the recharge pulse on the gate of T2 is not only dependent on the SPAD bias prior to the attempted recharge, but also on photon flux occurring during the recharge time. An avalanche can be triggered as soon as the SPAD's bias is above breakdown and a recharge pulse might fail to bring the anode voltage to ground when it coincides with the arrival of a photon. Whether the photon is recorded in for these cases depends on the SPAD anode voltage during the next time period where gate is active. Thus the turn-on time for a SPAD is always limited by the discharge time of the SPAD anode parasitic capacitance for this circuit. Figure 8(b) shows the distribution of the rise and fall time across the array.

We evaluated the characteristics of the sensitive window over the whole chip, in order to
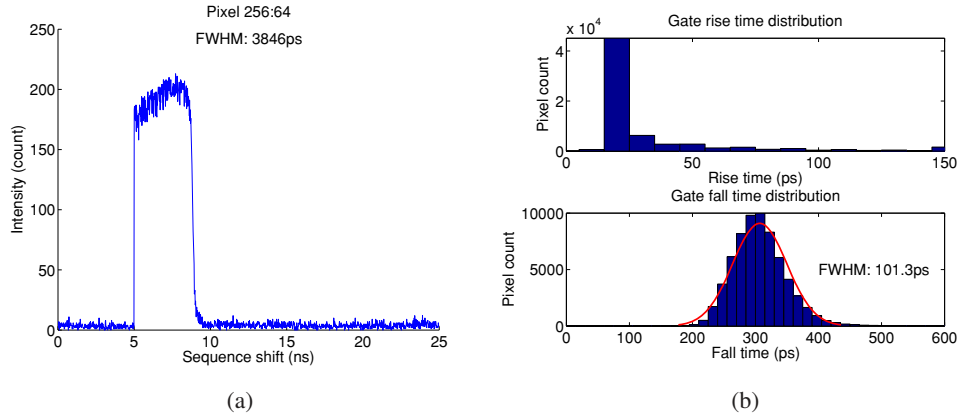
Fig. 8. (a) Typical pixel intensity response obtained by sliding the detection gate over the whole pulsed laser period. In this representation the falling (second) edge corresponds to the start of the detection where the SPAD is turned on. This response is analysed as bilevel waveform according to *The IEEE Standard on Transitions, Pulses, and Related Waveforms, Std-181-2003*. (b) Distribution (for all pixels) of the rise and fall times of the waveform shown in a. While the rising edge is very short (~20 ps), the falling edge is distributed around 300 ps. There is no apparent pattern for the distribution over the array.

assess the uniformity and matching of the gating circuitry consisting of the signal trees and in-pixel transistors.
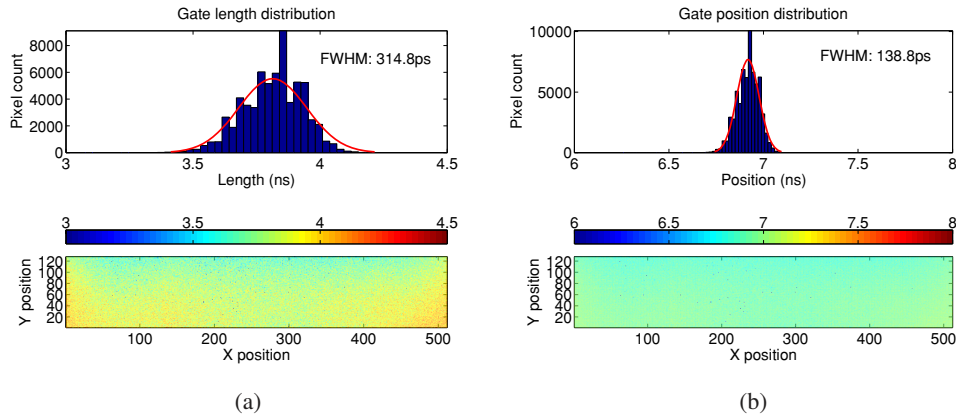

Fig. 9. (a) Distribution of the gate length for the pulse specification used in the measurement of Fig. 8 over the full array. The gate length increases from top to bottom and from the center to the left and right edges in a systematic way. The total variation is about 500 ps for the pattern of gating signals used in these measurements. (b) Distribution of the gate position over the full array.

Figure 9 shows the distribution of gate length and position over the full pixel array. The spatial distribution indicates a systematic skew leading to shorter gate values at the top center and larger values at the bottom and short edges. While the gradient from bottom to top can be explained by the placement of the signal trees on the long side of the chip, the origin of the edge gradient is less obvious. Power distribution, which is placed on the short sides might account

for most of it.

The performance figures for our imaging system are summarized in Table 1.

Table 1. SwissSPAD performance parameters.

Parameter	min.	typ.	max.	Unit
Process		$0.35\mu m$		HV CMOS
Chip size		13.5 x 3.5		mm^2
Format		512 x 128	512 x 256	pixel
Pixel pitch		24		μm
Dead time		100		ns
Nominal fill-factor		5		%
Microlens median concentration factor		6		
Microlens resulting fill-factor		30		%
Photon detection probability $(V_e = 3.5V, \lambda = 490nm)$		22		%
Photon detection probability $(V_e = 4V, \lambda = 490nm)$		46		%
Spectral range $(V_e = 3.5V, PDP > 5\%)$	350		850	nm
Median dark count rate $(V_e = 2.5V)$		95		Hz
Median dark count rate $(V_e = 4.5V)$		366		Hz
Electrical cross-talk		negligible		
Optical cross-talk[*] $(V_e = 2.5V @ 24\mu m)$		<3.5		%
Gating duration (FWHM)	4	20	6400	ns
Clock frequency		80		MHz
I/O bandwidth (80 MHz)		10.24		Gbps
Sensor power consumption (idle)		660		mW
Sensor power consumption (active)		1650		mW

[*]Extracted from noise analysis of neighbouring pixels. [16]

4. Fill-factor improvement using microlens arrays

One disadvantage of SPADs implemented in standard CMOS technology is the limited fill-factor due to the need for guard rings and the placement of in-pixel electronics. Two main approaches have been proposed to mitigate the problem. The first, and comparatively the simplest one, uses larger and rectangular shaped SPADs to increase the active area to pixel size ratio and moves the electronics outside of the pixel array [17]. The downside of this approach is increased pixel noise (DCR) and reduced spatial and temporal resolution. The approach also does not scale well for a large number of pixels. The second solution, which was adopted for SwissSPAD, uses a microlens array [18]. Each pixel has its own micro-optical concentrator in order to collect photons that would otherwise hit nonsensitive areas of the pixel. The downside

of this solution is increased fabrication complexity and cost, especially for large format sensors such as SwissSPAD.

Whereas in other chips with more complex pixel level electronics, the fill-factor is extremely low ($\leq 1\%$, [19]), SwissSPAD has relatively simple pixels, resulting in a small pixel size and a fill-factor of approximately 5%. This is still very low for low-light level applications, and calls for an improvement. To improve the fill-factor a microlens array, shown in Fig. 10, was formed directly on the chip surface. The array was fabricated using a quartz mold to imprint the lenses in a sol-gel polymer on top of the chip. Upon fabrication the polymer is UV-cured and stabilized through thermal cycling. In [20], Mata Pavia et al. provide details on the fabrication process and lens performance analysis.
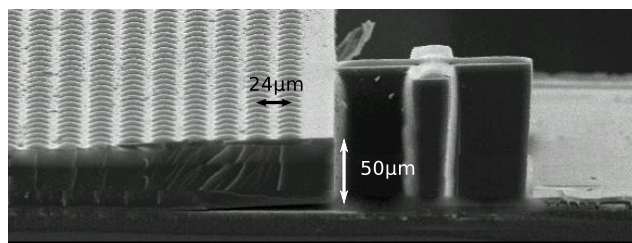

Fig. 10. Scanning electron microscope image of the microlens array with alignment mark deposited on SwissSPAD. Square lenses optimized for collimated light were deposited by CSEM Muttenz, Switzerland.

Another promising solution, to be implemented in future CMOS SPADs, uses 3D fabrication techniques. In this approach, the SPAD structure with the guard rings only, is put on one substrate, while all the pixel-level and chip-level electronics are put on an additional substrate connected to the SPAD chip using, for example, through-silicon-vias (TSVs) [21] or chip to chip bonding [22]. While first systems employing these techniques are being demonstrated, such as the one in [23] with InGaAsP SPADs, the techniques are not yet integrated in standard design flows.

4.1. Measurements

The sensitivity increase due to microlenses is measured by means of a parameter known as 'concentration factor', which is defined as the ratio of light intensity detected with and without microlenses. To perform an accurate measurement accounting for potential minute pixel-to-pixel sensitivity variations, we used the fact that our microlenses are designed for use with collimated light as found in many modern microscopy applications. Since the lenses do not focus uncollimated light, the intensity for this case is the same whether lenses are present or not. Using a standard 16 mm focal length TV lens with a f-number (focal ratio equal to the focal length divided by the lens diameter) ranging from 1.8 to 16 we collect reflected, uncollimated light hitting the lens at a fixed intensity while increasing the f-number from f/1.8 to f/16. With increasing f-number less light is hitting the sensor while at the same time it is more collimated. As expected, the concentration factor increases, as the f-number increases; it reaches a maximum of approximately 6 at f/16. We compared the recorded light intensity between chips with and without microlenses and we confirmed the concentration factor values measured above. Figure 11 shows the concentration factor, defined as intensity ratio, as a function of the f-number. The ratio was calculated over an evenly lit area of 20 x 20 pixels.
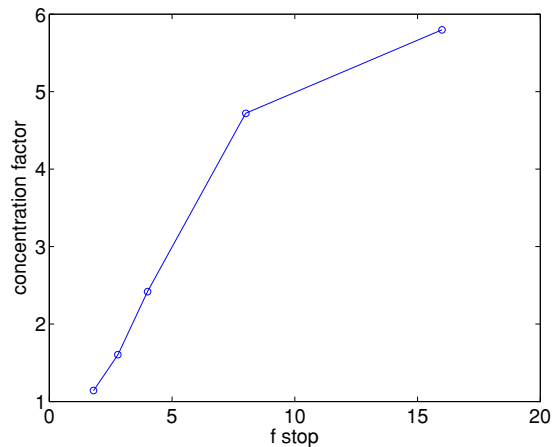
Fig. 11. Concentration factor as ratio of light intensity of a chip with microlenses compared to a chip without microlenses. The objective lens is illuminated by uncollimated white light at a fixed intensity and measurements for increasing f-numbers are performed. As the f-number increases, the resulting concentration factor increases as well, confirming the effectiveness of our microlens array.

5. Application 1: High frame-rate movies

SwissSPAD can be used in a standard wide-field optical setup as high frame rate single photon sensitive camera. Image series with configurable intensity resolution can be captured in two fundamental modes. First, the global gating/shutter can be run in sequence with the readout for a classical global shutter, full-frame transfer mode of operation. In this mode the maximum frame rate that can be achieved is determined by the illumination time plus the fixed readout time of 6.4 μs since no light is captured during readout. Second, the shutter can be run independently from the pixel memory readout. This results in a rolling shutter mode of operation and the maximum frame rate of 156,250 1-bit frames per second can be achieved. The shutter in this case is active for a fraction of the frame time and the illumination time for a row is determined by the photo sensitive time between two successive readouts.

The FPGA accumulates the sensor output at a rate of 10.2 Gbps and sends the frames to the attached memory for intermediate storage. As demonstrated in Fig. 12, single frames with different bit depths can be extracted from a high-speed movie shot with SwissSPAD. Even at a resolution of 1-bit per pixel, the sine wave traced by the electron beam impacting the oscilloscope's phosphorescent screen (as well as the screen's persistence) can be clearly distinguished. The maximum recording length is only limited by the size of fast storage memory connected to the FPGA. For special applications, domain-specific data compression and triggering could be implemented using the reconfigurable control logic in order to increase recording length.

Figure 13 gives the relationship between the maximum frame rate and the number of bits per pixel in an image. The frame-rate is quickly reduced as the number of bits per pixel is increased. At the same time, the resulting data rate is reduced accordingly, allowing recording longer sequences.

6. Application 2: FLIM measurements

Fluorescence lifetime imaging is a powerful technique used for in vitro, live cell or live animal (in vivo) measurements [24–26]. In contrast to other fluorescence-based microscopy ap-
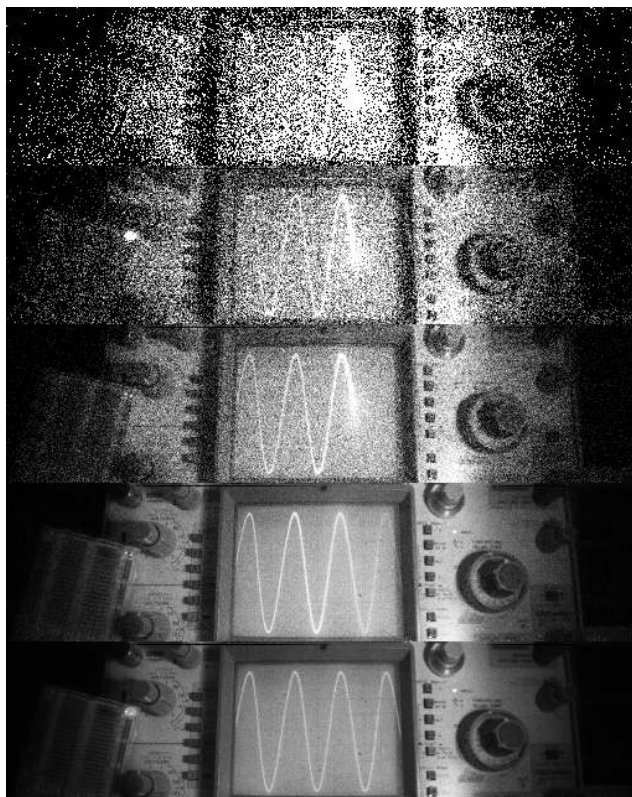
Fig. 12. Five intensity images of a scene acquired with 1-, 2-, 4-, 8- and 16-bit intensity resolution. The electron beam tracing the sine wave can be clearly distinguished at the highest frame-rate of 156 kfps with 1-bit intensity resolution. The video file (Media 1) shows image sequences from 16-bit to 1-bit.

proaches, FLIM is not merely concerned with fluorescence intensity or spectrum, but also in the fluorescence decay time scale after excitation (fluorescence lifetime). This additional information can report on the type of molecules detected at different locations, based on the known lifetimes of individual fluorophores. It can also report on the different environments in which a known fluorophore is located, if its fluorescence lifetime is affected by the chemical species surrounding it. In particular, distance-dependent fluorescence resonant energy transfer (FRET) between fluorophores bound to different molecular species, can be detected by a reduced lifetime of the "donor" species compared to its lifetime in the absence of nearby "acceptor" molecules [24–26].

Fluorescence lifetime measurements require using either pulsed excitation with sophisticated TCSPC techniques or very fast-modulated excitation with parallel modulation of a detector intensifier (frequency-modulation approach). TCSPC-based FLIM typically uses a confocal approach, where a single illumination spot is raster-scanned through the sample and fluorescence is collected by a point-like photon-counting detector with fast timing capabilities (PMT, SPAD or equivalent). The process is slow (several minutes per image) and can lead to sample damage due to the high laser intensity needed. Parallel acquisition alternatives using wide-field time-gated (or modulated) intensified CCD cameras or wide-field time-resolved photon-counting detectors based on PMT technology are costly and fragile, and come with their own sets of
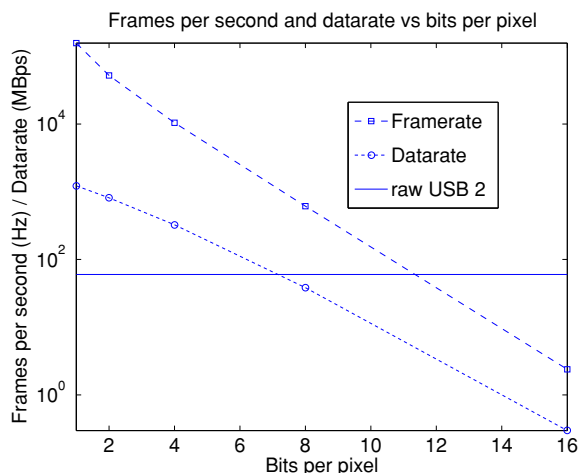
Fig. 13. Relation between maximum frame rate and bits per final frame. The maximum frame-rate of 156,250 fps is attained for 1-bit images. The processed data rate for different bitdepths and the fixed data rate of USB 2.0 are added to underline the need for high-bandwidth connections to the ultimate consumer.

limitations [27].

FLIM would clearly benefit from better detectors. An ideal detector would be (i) wide-field, (ii) capable of time-stamping individual photons with a resolution better than 1 ns, (iii) necessitating only low voltage, (iv) robust, (v) photon efficient and, if possible, (vi) with a cost comparable to standard scientific cameras. As will be demonstrated below, SwissSPAD fits most of this bill.

To calculate the fluorescence lifetime of molecules, the time of arrival of each detected photon with respect to the exciting laser pulse needs to be known. One way to achieve this goal is to incorporate one or more time-to-digital converters (TDC) in the camera. The limiting case is a detector where each SPAD pixel has its own TDC [19], but this is impractical for very large SPAD arrays and in fact unnecessary as long as at most a few SPAD pixels detect a photon during each laser excitation period. In this case, several SPAD pixels can share a single TDC.

An alternative approach used in SwissSPAD, consists of including a global shutter mechanism. This solution is advantageous as it scales well with large numbers of pixels without aversely affecting the fill-factor. However, as for any time-gating approach, it is photon-inefficent since it is blind to photons arriving when the pixels are not on. In the case of SwissSPAD the detection efficiency is bounded by the ratio between the gate length and the pulse period of the laser. The repetition period of the laser together with the desired time resolution dictate the number of gating positions necessary for the coverage of the fluorescence response period. For the FLIM measurements detailed in this section we used a gate length of approximately 6 ns on a period of 14.7 ns resulting in a detection efficiency of about 40%. It is desirable for future systems to be photon sensitive for the full period duration, which can be achieved for example by adding a second memory to the pixel.

Fluorescent decays were acquired using the same procedure as for the characterisation of the gating circuit (Section 3). Instead of illuminating the chip directly with the excitation laser, the laser (PicoTrain, High Q Laser: 532 nm wavelength, 68 MHz repetion rate, 8 ps pulse) was focused into the back focal plane of a microscope objective lens via a dichroic mirror in order to excite the fluorescence of various solutions of fluorophores characterized by different lifetimes. Fluorescence emitted by the molecules was redirected through the dichroic mirror and

a band pass filter to a camera port where the SwissSPAD was attached. The detected signal was recorded for each position of the detection gate within the excitation period, thereby yielding a signal equal to the convolution of the excitation laser pulse, the molecules' fluorescent decay and the detection window. The recorded waveform was then fitted by a single exponential decay model convolved with the detection window profile, acquired separately with the same optical system and a scattering sample (Ludox CL-C solution of colloidal silica particles).

Figure 14 shows typical waveforms recorded for different fluorophore solutions, as well as the instrument response function (IRF) obtained with the Ludox solution and the best fit with a mono-exponential decay characterized by a lifetime τ indicated in each panel.

Table 2 summarizes the results obtained for different fluorophores. Some of the measured lifetimes are shorter than the values quoted in the literature due to the highly concentrated solutions used in our experiments, resulting in partial fluorescence quenching [28].
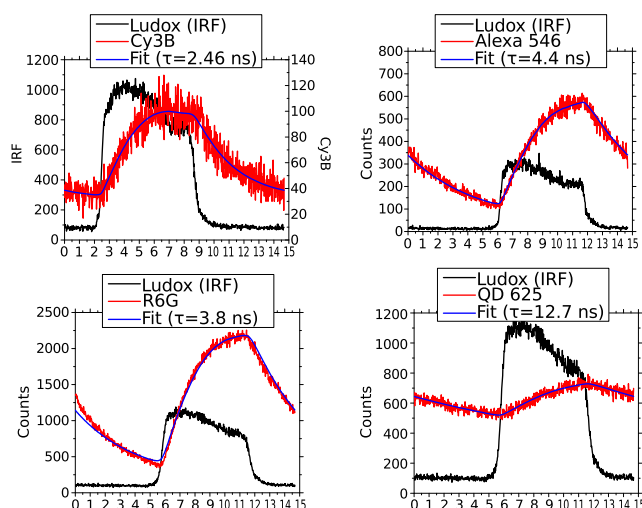

Fig. 14. FLIM recording of different fluorescent samples using a 68 MHz (period: 14.6 ns), 532 nm wavelength laser with a 8 ps pulse duration.

Table 2. FLIM lifetime data.

Fluorophore	Reference[*] lifetime [ns]	Extracted lifetime [ns]
Rhodamine 6G	4.08	3.8
CY3B	2.8	2.4-2.5
Alexa Fluor 546	4.1	4.2-4.4
QD 625	N/A	12.7

[*]ISS database: `http://www.iss.com/resources/reference/data_tables/LifetimeDataFluorophores.html`

7. Application 3: SwissSPAD as True Random Number Generator

SwissSPAD can be used for the generation of high quality random numbers. Past works already used the quantum nature of photons as entropy source and coupled it with digital circuits for building a complete True Random Number Generator (TRNG) [29, 30]. However, the performance and the energy consumption of the TRNGs proposed so far were not satisfactory.

Also, there was no experimental evidence which demonstrated the possibility of increasing the throughput of the TRNG by increasing the number of SPADs in the design.

We addressed these issues using SwissSPAD. In order to build a complete TRNG, the architecture previously described was completed by a LED placed on top of the two matrices of pixels, and the firmware was adapted to control the LED as well as to post process the bit stream produced by the arrays. The data were de-biased using a simple von Neumann filter, which discards the bit pairs whose elements are identical.

We evaluated the quality of the produced stream at different temperatures, bias voltages, and length of the LED pulse. The quality was evaluated by applying to the produced streams the battery of tests proposed by the NIST test suite for the validation of random number generators [31]. We applied the test to both raw data as well as de-biased ones, and we repeated the test applying to several sequences having different length. In particular, we applied the tests to sequences having lengths ranging form 10.000 to 1.000.000 of bits. As required by the NIST procedure, the test parameters were adjusted accordingly to the sequence length.

Our results show that all the de-biased sequences are passing the tests. As example, the test results of one sequence are reported in Table 3. Also the raw sequences performed well on the NIST tests, as most of them passed before any post-processing. Overall, we demonstrate that coupling two SwissSPAD matrices we can reach up to 5 Gbit/s, proving the scalability and performance for random number generators based on SPADs. Based on the chip power consumption for the data acquisition, we estimated the power needed to produce one random bit at 25 pJ/bit, the lowest power consumption to date.

Table 3. Results of the NIST tests applied to a sequence generated with a LED pulse length of 100 ns and an excess bias voltage of 2.8 V. The tests were run on the data from the de-biasing filter.

Test	Accept Threshold	Von Neumann	Pass / No Pass
Frequency	0.951464	0.9833	Y
BlockFrequency	0.951464	0.9833	Y
CumulativeSum	0.951464	0.9833	Y
Runs	0.951464	1.0000	Y
LongestRun	0.951464	1.0000	Y
Rank	0.951464	1.0000	Y
FFT	0.951464	0.9833	Y
NonOverlappingTemplate	0.951464	0.9667	Y
Universal	0.951464	1.0000	Y
ApproximateEntropy	0.951464	1.0000	Y
RandomExcursion	0.951464	0.9744	Y
RandomExcursionVariant	0.942202	0.9744	Y
Serial	0.951464	1.0000	Y
LinearComplexity	0.951464	1.0000	Y

8. Conclusion

SwissSPAD is a large format, highly sensitive SPAD based image sensor with fast and precise global shutter circuitry. In this article we have described the image sensor design, discussing the trade-offs and solutions adopted in the architecture. We illustrated the control and interface of

the chip and have shown the characterisation of the gating circuitry and some of the problems associated with this type of circuits in SPADs. Furthermore, we have shown how low fill-factor in large resolution SPAD sensors can be recovered using an array of microlenses to concentrate incoming light on the sensitive areas of the chip. Lastly, three very different applications were demonstrated using SwissSPAD. We believe that SPADs, implemented in standard CMOS, offer distinct advantages to warrant their evaluation in applications where traditional cameras are currently at their limits. The image sensor chip presented in this paper has exposed some of these advantages and has helped quantify the potential of this technology.

Acknowledgments