A 1 × 400 Backside-Illuminated SPAD Sensor With 49.7 ps Resolution, 30 pJ/Sample TDCs Fabricated in 3D CMOS Technology for Near-Infrared Optical Tomography

Juan Mata Pavia, Mario Scandini, Scott Lindner, Martin Wolf, and Edoardo Charbon, Senior Member, IEEE

Abstract—A 1 × 400 array of backside-illuminated SPADs fabricated in 130 nm 3D IC CMOS technology is presented. Sensing is performed in the top tier substrate and time-to-digital conversion in the bottom tier. Clusters of eight pixels are connected to a winner-take-all circuit with collision detection capabilities to realise an efficient sharing of the time-to-digital converter (TDC). The sensor's 100 TDCs are based on a dual-frequency architecture enabling 30 pJ per conversion at a rate of 13.3 ms/s per TDC. The resolution (1 LSB) of the TDCs is 49.7 ps with a standard deviation of 0.8 ps across the entire array; the mean DNL is ± 0.44 LSB and the mean INL is ± 0.47 . The chip was designed for use in near-infrared optical tomography (NIROT) systems for brain imaging and diagnostics. Measurements performed on a silicon phantom proved its suitability for NIROT applications.

Index Terms—Near-infrared optical tomography (NIROT), near-infrared spectroscopy (NIRS), optical tomography (OT), single-photon avalanche diode (SPAD), single-photon imaging, time correlated single photon counting (TCSPC), time-of-flight imaging, time-resolved imaging, time-to-digital converter (TDC).

Manuscript received March 11, 2015; revised June 03, 2015; accepted August 02, 2015. Date of publication September 10, 2015; date of current version September 24, 2015. This paper was approved by Associate Editor Vivek De. M. Wolf and E. Charbon shared last co-authorship. J. Mata Pavia and M. Scandini contributed equally to this work. This work was supported in part by the KFSP Molecular Imaging and the KFSP Tumor Oxygenation of the University of Zurich, and in part by the National Competence Center for Biomedical Imaging.

J. M. Pavia is with the Quantum Architecture Group (AQUA), Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland, also with the Biomedical Optics Research Laboratory, Division of Neonatology, University Hospital Zurich (USZ), 8091 Zurich, Switzerland, and also with the University of Zurich, 8006 Zurich, Switzerland (e-mail: juan.matapavia@epfl.ch).

M. Scandini is with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology (TU Delft), Mekelweg 4, NL-2628CD Delft, The Netherlands (e-mail: m.scandiuzzo@tudelft.nl).

S. Lindner is with the Quantum Architecture Group (AQUA), Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland, also with the Biomedical Optics Research Laboratory, Division of Neonatology, University Hospital Zurich (USZ), 8091 Zurich, Switzerland, and also with the University of Zurich, 8006 Zurich, Switzerland (e-mail: scott.lindner@usz.ch).

M. Wolf is with the Biomedical Optics Research Laboratory, Division of Neonatology, University Hospital Zurich (USZ), CH 8091 Zurich, Switzerland, and also with the University of Zurich, 8006 Zurich, Switzerland (e-mail: martin.wolf@usz.ch).

E. Charbon is with the Quantum Architecture Group (AQUA), Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland, and also with the Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology (TU Delft), NL-2628CD Delft, The Netherlands (e-mail: e.charbon@tudelft.nl).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSSC.2015.2467170

I. INTRODUCTION

N EAR-INFRARED OPTICAL TOMOGRAPHY (NIROT) is a novel technique capable of determining oxygenation values in tissue by measuring its optical properties (absorption and scattering) as a function of wavelength, between 650 and 850 nm [1]-[9]. Recently, it has been demonstrated how single-photon avalanche diode (SPAD) image sensors capable of time-resolved measurements can be applied to NIROT, paving the way for new contactless setups and optimized reconstruction algorithms [10]. The ideal sensor for NIROT requires a high photon detection probability (PDP) in the near-infrared range and a large array of pixels where each SPAD pixel is coupled with and is adjacent to a time-to-digital converter (TDC), so as to guarantee low skew and high parallelism in the determination of the time-of-flight of photons. With the availability of SPADs in deep-submicron CMOS technologies, such image sensors have become possible; however with very low fill factor [11]–[13]. A possible solution is resource sharing; an example of this approach was achieved in SPADnet, where 720 SPADs share one TDC [14], and in EndoTOFPET-US, where 416 SPADs share 48 TDCs [15]. However, at over 600 and 800 μ m, respectively, the resulting pixel pitch is very large and unsuitable for NIROT.

Up to now, most of the SPAD structures implemented in large arrays had a PDP not higher than 10% in the near-infrared range [11], [12], [14], [16]. It has been demonstrated that non-isolated SPADs employing the bulk substrate as a common anode perform better in the NIROT wavelengths of interest [17], [18]. However, they require the interaction between high and low voltage circuits, making it necessary to use capacitive coupling to interface both [19]. Capacitive coupling requires potentially large passive elements that may reduce the fill factor, therefore they are not well suited for low light applications where the photon detection efficiency (PDE) is one of the most critical factors.

At present the I/O activity in large arrays of TDCs is the most dominant factor in terms of power consumption [20]. However, on chip histogramming [21] shifts this paradigm, and the TDC power consumption becomes a critical factor. Therefore, it is of great interest to develop new TDC architectures that address this problem.

0018-9200 © 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications standards/publications/rights/index.html for more information.



Fig. 1. Cross-section of the Tezzaron 3D wafer stacking process. The wafers are bonded face to face by the top metallization through DBIs. The tier 1 substrate wafer is thinned to $4.2 \mu m$. TSVs enable the interconnection between the metal layers of tier 1, and the bond pads at its back side. The SPAD pixels are placed in tier 1 with the minimum number of transistors necessary for operation to maximize the fill factor. The rest of the circuitry is placed in tier 2.

In this contribution, we propose a novel backside-illuminated (BSI) SPAD array integrated in a 3D multi-wafer stacking CMOS process that enables superimposition of the pixel detector to the ancillary circuitry. In the sensor, denominated 3DAPS, a new resource sharing scheme based on a winner-take-all (WTA) circuit is introduced to enable the integration of a large number of pixels, along with a new TDC architecture for low power operation. The SPAD structures, which employ the bulk substrate as a shared anode, are implemented in a separate substrate from the CMOS circuitry. This enables the direct connection of the SPADs outputs with the CMOS circuitry in non-isolated SPAD structures.

II. CMOS 3D WAFER STACKING PROCESS

The image sensor was implemented in the 3D two-tier Tezzaron's FaStack process, which is based on a CMOS 130 nm technology. In this technology two standard CMOS wafers are bonded face-to-face utilizing copper to copper bonds called direct bond interfaces (DBIs) (see Fig. 1), which are small metal features (0.6 μ m diameter). The top tier implements through-silicon-vias (TSVs), i.e., 6 μ m deep pillars made of tungsten, that are connected to the metal 1 layer of the top tier. After waferto-wafer bonding, the substrate of the top tier is thinned to 4.2 μ m, leaving the TSVs exposed; metallization is added to create the bond pads. The DBIs have a 4 μ m pitch and TSVs are approximately 1.2 μ m in diameter.¹ Fig. 1 shows a representation of the cross-section of this technology.

Both tiers are designed following a standard 2D design flow (schematic capture, layout, parasitic extraction). However, there is a 3D design rule check (DCR) and a 3D layout versus schematic (LVS) check [22], [23] in order to perform a verification of the complete design.

This process enables the integration of SPADs in the top tier with a minimum amount of circuitry for its operation, while the rest of the circuits based on standard CMOS are placed on the bottom tier. This architecture offers a number of benefits with respect to traditional designs with only one substrate. Since most of the circuits are located in the bottom tier, SPAD arrays with very high fill factor can be implemented in the top tier. Moreover, the fact that the pixels are BSI increases the fill factor since there are no metallization layers in the substrate that could block or reflect light. BSI SPADs are particularly interesting in applications employing near-infrared light, since their multiplication region can be engineered to be a few micrometers deep. A deeper multiplication region leads to a higher sensitivity of the SPADs in the NIR range.

III. DESIGN

A. Image Sensor Architecture

Fig. 2 shows a block diagram of the image sensor. It consists of two arrays of 1×400 SPAD pixels implemented in tier 1 connected to an array of 1×100 processing blocks located in tier 2. Two different SPAD structures were implemented, one in each row. In this contribution we report in detail the results of one of these rows. Each processing block includes a WTA circuit, a TDC, and a local 2-port static random access memory (SRAM) block. The SPADs are clustered in groups of 2×4 pixels which are connected to the same processing block. Upon detection of a photon, the TDC in the corresponding processing block calculates the timestamp associated with the photon. This timestamp is locally stored in the processing block's local SRAM until it



Fig. 2. Block diagram of the 3DAPS sensor. Every four pixels in rows A and B form a cluster that is connected to a WTA+TDC+SRAM memory located in tier 2.



Fig. 3. Cluster of pixels diagram. Each cluster has eight SPAD pixels, four from row A and four from row B, connected to the WTA circuit which triggers the conversion in the TDC and writes its result in the local SRAM memory.

is read out. All memory positions in the processing blocks are accessible through the readout multiplexer.

In Fig. 3 the schematic of a processing block is shown. Each block integrates a full TDC along with a redistribution mechanism based on a WTA architecture with a collision detection event-driven backend. This circuitry is used to connect the outputs of 8 SPADs to one TDC while preserving the identity of the triggering SPAD. The present image sensor was built to target time correlated single-photon counting (TCSPC) applications in which the SPADs work in photon-starved mode. In these applications the pixels have a maximum activity rate, i.e., the number of detected photons per clock cycle, of about 1%. Due to the low pixel detection rates in this type of applications, long acquisition times (from hundreds of milliseconds to seconds) are required. In addition, measurement time is a critical aspect in NIROT, because the optical properties of tissue vary over time and more importantly, patients might move, especially during long measurements. Therefore, sequentially measuring different groups of pixels [24] has a severe negative impact since the measurement time increases tremendously (e.g., over 30 min). To overcome this problem, TDCs have been integrated in each pixel to maximize the operational speed of SPAD image sensors [16]. However this approach dramatically reduced the sensor's fill factor and consequently the PDE, making the sensor unsuitable for very low light intensity applications. Since the pixel operation rate is low, sharing a TDC between several pixels enables to concurrently measure all pixels, while reducing the area necessary for TDCs. The only disadvantage of this approach is that, if two photons impinge during the same clock cycle, only one of them will be processed. In the presented design, one TDC was shared among eight pixels; as shown in the following sections,



Fig. 4. SPAD structure cross section. It consist of a N-lightly doped drain (NLDD) cathode, whereas the substrate act as a anode shared among all the SPADs in the image sensor.

this level of sharing has a negligible impact on the overall performance of the image sensor for NIROT applications.

Upon a TDC conversion, the resulting timestamp is stored in a SRAM location. The location in the SRAM memory block will be determined by the address coded in the WTA circuit. When two or more pixels fire simultaneously, the generated code in the WTA circuit will be a non-valid address, thus the result will be discarded and not stored in the SRAM memory. Each pixel has a dedicated position in the SRAM memory that is flagged as unread after a new conversion is stored in it. After the contents of a SRAM location are read out, its unread flag is automatically deasserted. Each memory location in the SRAM memory can be simultaneously accessed by the TDC for writing and by the readout multiplexer for reading operations.

B. Pixel Design

The SPAD structure employed in the line sensor (see Fig. 4) consists of a circular n-type low doped drain (NLDD)/p-well junction, where the high-field multiplication region is generated. This is surrounded by a n-well/p-well junction, which has a higher breakdown voltage, acting as the guard ring that prevents premature edge breakdown in the SPAD. This structure has been reported in a similar technology, albeit in frontside illumination (FSI) [17], yielding PDP values above 30% for 700 nm.

The implemented pixel had a round geometry with an active area of 28 μ m², a 1.1 μ m thick guard ring, and 11 μ m pitch, resulting in a fill factor of 23.3%. These rather conservative parameters were chosen, since this was our first attempt of producing a SPAD in wafer stacking technology and thus the primary objective was obtaining functional devices rather than maximizing fill factor. Based on the achieved results, SPAD structures with square geometry and thinner guard rings could be produced in the near future, that would yield fill factors higher than 70%.

Fig. 5 shows the schematic of the pixel circuit. The quenching circuitry is implemented in tier 1 with a PMOS transistor working in linear mode. The transistor has to be biased in such way that the equivalent ohmic resistance is approximately 100 k Ω to promptly quench the avalanche upon photon detection [25]. Tier 1 substrate is negatively biased to the SPAD's breakdown voltage, while tier 2 substrate is biased to ground. Since the transistors in tier 1 are isolated from their substrate, they can operate at similar supply voltages as the circuits in tier 2. Thus, no level shifting or capacitive coupling is required



Fig. 5. (a) SPAD pixel circuit including quenching transistor and buffer. The number of transistors in tier 1 has been kept low to maximize the space available for SPADs and increase fill factor. The SPAD is passively quenched and recharged with a PMOS; the SPAD recovery time can be adjusted by changing V_{QCH} . The SPAD output is not directly connected to the buffer in tier 2. Instead, a PMOS transistor drives the buffer in tier 2, so as to reduce the parasitic loading of the SPAD cathode, which is dominated by the DBI (the contact between the tiers). (b) A waveform depicting the operation of the circuit is also presented.



Fig. 6. Schematic of the WTA circuit. The collisions in the bus, in the event of two or more pixels firing simultaneously, generate a non-valid code.

to connect the SPAD output to the standard CMOS circuits. Along with the quenching mechanism, a follower drives the metal connecting the two tiers. Although it would be possible to implement the quenching circuitry in the bottom tier, this would increase the parasitic capacitance at the cathode and therefore the current flowing through the diode during an avalanche event. This would negatively affect the afterpulsing, timing resolution, and power consumption [26]. With the reduction of DBI parasitics though, this solution will be a viable one in the future.

C. Winner-Take-All Circuit

A WTA circuit (see Fig. 6) similar to the one introduced in [27] was implemented in each cluster to share the TDC between eight pixels, enabling time-resolved measurements in each pixel while making a efficient use of silicon area. In the WTA circuit the output of each pixel is connected to a digital bus with five address lines through pull-down transistors. A dedicated extra line *TDC_start* is used to propagate the digital pulses from the pixels to the TDC, where a time-to-digital conversion will be started. Each pixel generates a unique code in the bus that is latched at the end of the bus by the *TDC_start* signal.

In the original WTA implementation presented in [27], in the event of two pixels firing simultaneously, the address generated when using a binary code produced an incorrect address, e.g., if

TABLE I Collision Detection Code Table for Eight Pixels

SPAD	Binary code	Collision dection code		
0	000	00111		
1	001	11100		
2	010	11010		
3	011	11001		
4	100	10110		
5	101	10101		
6	110	10011		
7	111	01110		

pixels 1 and 2, with digital codes "001" and "010", respectively, detected a photon at the same time, the generated code in the bus would be "011" which corresponds to pixel 3, thus providing invalid data for pixel 3. We call this event collision.

In order to detect collisions between two or more pixels, we present a new address coding that identifies these collisions in the address bus. With this collision detection coding scheme, invalid addresses are generated in the event of two or more pixels firing simultaneously. The collision detection code is based on the fact that each code has three "1"s and two "0"s. Table I presents the codes for each pixel in the present WTA implementation. Since the bus is implemented in a pull-up manner, when an address is written in the bus some lines will be pulled down by the pixel that detected a photon. In the event of two or more photons being detected, several pixels write their respective addresses in the bus. Since each address pulls down a combination of two out of five lines, in this case the resulting amount of lines that are being pulled down will be higher than two. Therefore the generated code at the WTA output is identified as non-valid. In the event of an invalid code in the address bus, the corresponding TDC result will not be stored in the SRAM memory. The probability of two or more photons being detected in the same clock period can be calculated with a Poisson cumulative probability function. In the proposed WTA implementation where eight pixels are connected to the same bus, for a pixel detection rate of 1%, the probability that during



Fig. 7. TDC general schematic showing the main building blocks. The TDC is based on a dual-speed ring oscillator that minimizes the power consumed during a conversion.



Fig. 8. TDC timing diagram. Upon a photon detection, the ring oscillator starts oscillating at low-speed (T1). At the next rising edge of the *stop* signal, the *stop_reg* signal is asserted, where a synchronizer is used to avoid metastability (T2). The *stop_reg* signal is re-synchronized with *RO_clock* to generate the *boost* signal, which activates the ring oscillator high-speed operation mode (T3). The *stop_del* signal stops the ring oscillator (T4). After writing the data in the local SRAM memory, the TDC is reset and is ready to perform the next conversion (T5).

the same clock period two photons are detected is 0.3% which accounts for 4% of the total number of events in the bus. An increase in the pixel detection rate will result in a higher rate of collisions in the bus, i.e., for a pixel detection rate of 10% the amount of collisions in the bus will account for 35% of the total number of events in the bus. However in time-correlated TCSPC applications such as NIROT it is necessary to keep the pixel activity rate < 1% in order to avoid the pile-up effect.

This collision detection coding can be expanded to a bus with any number of address lines, providing each address has equal numbers of "1"s and "0"s. Consequently, the number of codes per number of address lines is fewer than achieved when using standard binary coding. The maximum number of codes in an address bus with n lines is given by the number of k element combinations of a set of n elements, where k is the number of "0"s in each address code. The total number of possible codes for a bus with n lines is given by (1), where k has to be the integer closest to n/2 in order to obtain the highest possible number of collision detection codes

$$\# \text{codes} = \frac{n!}{k!(n-k)!}.$$
 (1)

For a bus width of 12 bits, n = 12, and k = 6, a total of 924 different collision detection codes can be generated. The equivalent binary coding would require 10 bits to encode the same amount of codes, resulting in a 2 bit overhead due to the collision detection mechanism. In the present design, 5 address lines are employed to connect eight SPADs to the same bus, making an efficient coding of the pixels while being able to detect collisions in the bus.

The size of the cluster, and therefore the WTA, was dimensioned to decrease the TDC footprint per pixel while keeping a high data throughput. Clusters with a high number of pixels will suffer from a considerable data loss since many photons will arrive while the TDC is working, e.g., in a cluster with 32 pixels working at a 1% detection rate, 16% of the detected photons will arrive while the TDC is busy. The cluster size is a tradeoff between the available area for placing the TDCs and the amount of data loss that is acceptable. In TCSPC applications data loss is not critical, however since the total measurement time is important in NIROT, it was decided to implement a cluster of eight pixels since it offered a data loss of only 4% of the total detected photons.



Fig. 9. Ring oscillator is based on a differential buffer whose propagation delay can be adjusted with the *boost* signal. (a) Ring oscillator diagram and (b) schematic of a single delay stage.



Fig. 10. Schematic of the synchronizer circuit implemented in the TDC. The synchronizer is made out of three buffer stages and its purpose is to deliver a "1" at its output synchronized with CK2 when D becomes "1". Once this happens, Q is latched to "1" and it is necessary to reset it to be operational again. Only the first stage includes a latch to minimize the area and to regenerate the integrity of the signal in the internal nodes in case of metastability.

D. Time-to-Digital Converter

The power consumption in SPAD image sensors is strongly dependent on the number of time conversions that are performed per unit of time. For high pixel activity rates, the TDC power consumption is one of the main contributors, if not the largest, to the total image sensor power consumption [11], [13]. In order to save power, TDCs which operate in reverse start-stop configuration have been implemented in most TDC arrays, i.e., the TDC starts the conversion upon the impingement of a photon and it is stopped by the laser's reference clock signal [28], i.e., the laser's internal clock, which represents the emission time of the light pulses. TDC implementations based on ring oscillators (RO) required operation at frequencies of a few gigahertz during the conversion period, drastically increasing the total power consumption of the chip when many of them were active [13]. We propose a new TDC based on a dual-speed ring oscillator that only requires operation at gigahertz frequencies during a small fraction of the conversion time. Upon the start of a conversion, the start reg signal is asserted and the RO oscillates at a frequency of approximately 246 MHz, clocking the low-speed counter whose result will determine the five most

significant bits of the measurement. At the next falling edge of *RO clock*, after the assertion of the laser's reference clock signal stop, the boost signal is asserted and the RO's operating frequency is boosted to 2.52 GHz. At the same time the low-speed counter is deactivated and the high-speed counter is enabled. A delayed version of the stop signal, stop_del, eventually stops the RO. The delay between *stop* and *stop del* has to be larger than a period of RO clock in low-speed operation, otherwise the RO might not start oscillating in high-speed mode. The result of the high-speed counter together with the phase of the RO provides the other 7 bits of the measurement, resulting in a total of 12 bit TDC resolution. This mechanism effectively measures the time difference between *start* and *stop* with a coarse resolution, and subsequently the time difference between the last falling edge of RO clock in low-speed operation and stop del with high accuracy. Therefore, in the worst case, the TDC only operates at 2.52 GHz between the rising edge of *stop* and the rising edge of stop del, which in this case is approximately 5 ns. Fig. 7 shows the block diagram of the TDC, and Fig. 8 the timing diagram associated with a conversion.

The internal structure of the dual-speed RO is shown in Fig. 9. A differential inverter is employed as basic delay cell in order



Fig. 11. Photomicrograph of the image sensor. In the inset, the layout view of a pixel cluster in both tiers is shown. The pixel pitch is $11.75 \,\mu$ m.



Fig. 12. PDP as a function of wavelength for 1.5 V excess bias voltage.

to equalize the delay in the transitions from "0" to "1" and from "1" to "0". The differential stages were designed to have wide voltage swing to reduce the effects of jitter induced by power supply noise. The oscillation frequency is controlled through the tail transistors, which generate the bias current, and the PMOS transistors, which injects current in the differential pair.

In the TDC block diagram presented in Fig. 7, two synchronizers are employed to avoid metastability when signals travel between different clock domains. In the two cases only the transition from "0" to "1" needs to be synchronized, simplifying the design of the synchronizer. Fig. 10 shows the schematic of the synchronizer that is employed to control the boosting of the RO and the latching of the *stop_reg* signal.

E. 2-Port Static RAM Memory Block

A 2-port static SRAM memory is implemented to locally store the TDC conversions. The SRAM memory consists of eight words, one for each pixel, of 13 bits size: 12 bits to store the result of the TDC conversion, and an extra bit to



Fig. 13. DCR cumulative distribution function at 1.5 V excess bias voltage.



Fig. 14. TDC response obtained when varying the arrival time of the *start* signal with a programmable delay. As the low-speed period of the ring oscillator is not a multiple of 2 of the high-speed period, certain codes are unreachable (dashed line) and therefore some steps are observable in the response function. Therefore the TDC output needs to be normalized with a lookup table, which leads to a linear relationship (continuous line).

indicate whether the result has been already read out or is new. The write operations are controlled by the WTA circuit, whereas the SRAM block can be directly accessed from external pads for read operations. The two-port architecture of the SRAM memory allows simultaneous asynchronous reading and writing. This is of great benefit in cases where the reference clock is much slower than the maximum readout frequency and it eases the integration of the image sensor in test setups since clock synchronicity is not required during read out operations.

IV. RESULTS

A. Sensor Implementation

The image sensor was fabricated in the 3D two-tier Tezzaron's FaStack process described in Section II. Fig. 11 shows a microphotograph of the chip. Fig. 11 also shows a layout view of a pixel cluster in the top and bottom tiers. The total area was only $64 \times 47 \ \mu m^2$, most of which is occupied by the 2-port SRAM memory. The cluster was equipped with different



Fig. 15. LSB cumulative distribution function.



Fig. 16. Single-shot accuracy. Results from all delay values employed in the range from 0 to 40 ns are normalized and plotted together. No significant changes were found in the time accuracy for different delay values.

design-for-test capabilities in order to be able to test each block separately. For this reason a readout multiplexer was added to have direct access to the outputs of the main building blocks: WTA, TDC, and SRAM.

B. Experimental Results

The measurements presented in this section were performed under room temperature conditions. All the results hereby presented are from pixels or TDCs integrated in the array, that is DCR, LSB, single-shot accuracy, and linearity, were concurrently obtained for all pixels and TDCs. All the TDCs were active during these measurements.

The SPAD's PDP was measured to characterize the sensitivity of the pixels at different wavelengths. The measured breakdown voltage at room temperature was 16.5 V. Fig. 12 shows the PDP as a function of light wavelength for 1.5 V excess bias voltage. Thanks to its deep multiplication region, the peak performance is achieved around 700 nm. The PDP is above 12% between 650 and 800 nm, performing better than previous CMOS SPAD arrays in the NIR range [11], [13], [14], [16].



Fig. 17. (a) Typical DNL and (b) INL response measured with a 25 ns period reference clock.

The dark count rate (DCR) distribution was measured over the whole array. Fig. 13 shows the DCR cumulative distribution in the array. Although the measured DCR values are higher than in previous SPAD image sensors, it has to be taken into consideration that this is the first BSI SPAD structure that has ever been integrated in an array. Further investigation in new SPAD structures employing this new 3D CMOS technology is expected to yield structures with improved PDP and DCR responses. In fact, a standalone SPAD structure fabricated in the same 3D CMOS technology with similar PDP and an order of magnitude lower DCR has already been reported [29].

A daughterboard with a zero insertion force socket and different power supplies and bias voltages generators was designed for the test and evaluation of the image sensor. The daughterboard is connected to a Virtex-6 FPGA ML605 evaluation kit motherboard, which programs the different voltages necessary for the operation of the image sensor and controls communication with the sensor. A 40 MHz signal was generated in the FPGA and then fed to the image sensor as the reference clock. The Virtex-6 FPGA has a programmable delay line with 17.9 ps taps which is employed to generate the *start* signal that triggers the conversion in the TDC. Fig. 14 shows the conversion results for different delay values in the start signal. Since the low-speed period of the RO is not a multiple of two of the high-speed period, and the minimum count of the high-speed counter is always higher than one (the RO has to be running during a certain period of time before it is stopped to ensure stability in the output frequency), there are certain codes that will never be reached. Thus, it is necessary to use a lookup table that translates the resulting codes of the TDC into a normalized response. In Fig. 14 the normalized response of the TDC with a lookup table is also presented. The nominal resolution of the TDC (1 LSB) is 49.7 ps.



Fig. 18. (a) DNL and (b) INL cumulative distribution function in the image sensor.



Fig. 19. (a) Histograms of the response of the BSI SPAD at different wavelengths for an excess bias voltage of 1.5 V. (b) FWHM jitter as a function of the excess bias voltage for a 750 nm wavelength excitation.

Fig. 15 shows the resolution variations across the array. The observed LSB standard deviation across the whole array was 0.8 ps.

Time accuracy tests were performed on the TDC to measure the conversion variations for a fixed delay between the start and stop signal. For each delay value in the programmable delay line 50 time-to-digital conversions were performed. The conversions obtained for each delay value were normalized to its median value and its standard deviation was calculated, leading to the single-shot RMS jitter. No direct correlation between the delay value and the measured single-shot jitter was observed, indicating that the main source of noise contributing to the jitter has a bandwidth narrower or close to the stop signal frequency. Had the noise a bandwidth much higher than the *stop* signal frequency, the jitter would accumulate during its operation in the ring oscillator and higher variations in the conversion values would be observed for longer delay values. Since the TDC is not compensated for process, voltage and temperature (PVT) variations, power supply coupled noise is most probably the main factor contributing to the time uncertainty. Though no substantial variations were found between different TDCs in terms of operation frequency during the measurements, implying that the TDC seems to be quite stable against voltage and temperature changes, process variations could impact performance robustness. Thus, we plan to implement a PVT compensation mechanism in the circuit, where the current flowing through the oscillator will be controlled by a tail transistor biased with a voltage, which is regulated by an external clock reference. This solution is similar to previous PVT compensated TDC designs based on ring oscillators [13]. Fig. 16 shows the normalized conversion variations for different delay values covering the complete range of the TDC, i.e., 0 to 40 ns. The measured RMS jitter was $\sigma_{jitter} = 1.22$ LSBs.

In order to measure the integral non-linearity (INL) and differential non-linearity (DNL) of the TDC we used a technique known as density test; the SPAD pixels are employed to generate the *start* signal that triggers the conversion in the TDC. Under uncorrelated light conditions, the time of arrival of the photons detected by the SPAD should be completely random. If only one or less photons arrive during a TDC conversion period, the re-

		This work	[16]	[13]	[14]	[11]
SPAD	PDP @800nm [%]	12.2 @1.5Ve	6.5 @1.4Ve	n.a.	2.8 @1.5Ve	7.5 @5Ve
	DCR [kHz]	35 @1.5Ve	0.005 @1.4Ve	n.a.	14 @1.5Ve	20 @5Ve
	Jitter (FWHM) [ps]	260	150	n.a.	150	185
TDC	Resolution (1 LSB) [ps]	49.7	55	52	65	119
	Range [ns]	200	55	53	262	100
	DNL [LSB]	-0.44/+0.44	-0.3/+0.3	-0.4/+0.4	-0.24/+0.28	-0.4/+0.4
	INL [LSB]	-0.47/0.47	-2/+2	-1.4/1.4	-3.9/+2.3	-1.2/+1.2
	Dynamic power consumption [µW @ 500 kS/s]	15	n.a.	38	94	48
Sensor	Process technology	CMOS 3D 130nm	CMOS 1P4M 0.13µm	CMOS 1P4M 0.13µm	CMOS 1P4M 0.13µm	CIS 130nm
	SPAD array size	1× 400	160× 128	32× 32	192× 480	32× 32
	Chip size [mm ²]	0.77× 5	12.3× 11	n.a.	9.85× 5.452	4.8× 3.2
	Output data rate [Gbps]	1.04	51.2	5.12	1.6	10.24
	Clock frequency [MHz]	40	16	40	100	280
	Static power consumption under darkness conditions [mW]	7 [†]	550 [‡]	22 [†]	200	51 [†]

 TABLE II

 SPAD Array Performance Comparison With the State-of-the-Art

[†] Core only.

[‡] Consumption under typical conditions.

sulting histogram showing the distribution of times of arrival should be flat. However in reality some codes show higher occurrences than others. If enough photons are acquired, the DNL can be calculated by dividing the number of occurrences in each code by the average number of occurrences per code. The INL can be calculated by integrating the DNL. Fig. 17 shows typical DNL and INL results obtained from a TDC density test. Fig. 18 shows the peak-to-peak INL and DNL cumulative distribution in the image sensor. These figures show a high homogeneity in the DNL and INL values despite the fact that the TDCs were not compensated for PVT variations.

The sensor's power consumption was measured for different pixel activity rates to determine the amount of energy necessary to perform a conversion. A linear current increase with the pixel activity was observed. From these measurements, an average energy required for a conversion of 30 pJ/Sample was calculated. This results in a power consumption of 15 μ W per TDC at a conversion rate of 500 kS/s. Thanks to the dual-speed ring oscillator mechanism in the TDC, the power consumed by the TDCs is substantially lower than previous TDC designs based on ring oscillators [11], [13], [14], [16], enabling the simultaneous operation of a higher number of TDC on chip and thus higher data throughput.

The timing characterization of the SPAD-TDC subsystem was obtained by illuminating the sensor with a supercontinuum laser (SuperK Extreme, NKT Photonics). The emitted light had a pulse width of 20 ps full width at half maximum (FWHM), while the optical power was adjusted to avoid pile-up during TCSPC measurements. Fig. 19(a) shows the time response



Fig. 20. Experimental setup for determining the absorption and reduced scattering coefficients of a silicon phantom employing time-resolved measurements.

obtained from a SPAD pixel for different wavelengths. The impulse response function is plotted for different NIR wavelengths showing a similar behavior for all of them. The FWHM jitter of the BSI SPADs measured using the proposed TDC is plotted as a function of excess bias voltage. Since the SPADs are not electrically isolated from the substrate, carriers generated after a photon absorption at a certain depth from the multiplication region may eventually reach it and start an avalanche. This translates to a higher time uncertainty in the photons arrival time and therefore in wider time responses, as shown in Fig. 19(b), than in SPAD structures that are isolated from the substrate. The time uncertainty obtained with the implemented SPAD structure is similar to the one obtained

Martin N

Measured values

Simulated values

Mean time of flight [ps] ntensity [a.u.] 800 700 600 500 10 400 2 2.5 4 4 3 3.5 2.5 3 3.5 Source detector distance [cm] Source detector distance [cm] (b)(a)

1300

1200

1100

1000

900

easured values

- Simulated values

Fig. 21. (a) Measured intensity with the SPAD image sensor at the surface of the phantom for different source-detector distances, the simulated intensity values are also plotted. (b) The measured and simulated mean time-of-flights of the detected photons at different locations on the phantom are also shown.

from similar structures in FSI [17]. Table II summarizes the performance of the chip and presents a comparison with the state-of-the-art.

One IC sample of the produced image sensor was integrated in a NIROT setup in order to assess its suitability for this application. A silicon phantom with absorption and reduced scattering coefficients similar to those of the head of a neonate, $\mu_a = 0.06 \text{ cm}^{-1}$ and $\mu'_s = 5.6 \text{ cm}^{-1}$, respectively, was illuminated with a 750 nm picosecond laser with 17 mW optical power. The laser beam hits the phantom a few centimeters outside the field of view of the camera. By measuring the intensity at different distances from the light source, together with the mean time of arrival of photons at those positions, it is possible to retrieve the absorption and reduced scattering coefficients of the target under study [30], [31]. Fig. 20 shows a representation of the experimental setup. In order to minimize the impact of noise on the measurements, the DCR for each pixel was measured. This value was later subtracted from the measured intensity obtained in the experiment. Fig. 21 shows the measured intensity and mean time of flight values for source-detector distances ranging from 2 to 4 cm. The simulated values were calculated with the solution of the diffusion equation for a semi-infinite medium [30], [31]. The measured values have a good correspondence with the expected values indicating the suitability of the sensor for NIROT applications. For distances larger than 3.5 cm, the detected time-of-flight deviates from the simulated values. This divergence is due to the fact that the backscattered light at these distances is very weak. Additionally, light from the shorter source-detector distances is also internally reflected in the camera lens creating lens flare, i.e., light is internally reflected in the lens generating a haze in the image sensor. Since this internally reflected light has a different time-of-flight than the backscattered light from the phantom, it introduces an error when both signals have similar intensities.

V. CONCLUSION

The proposed sensor is, to the best of our knowledge, the first SPAD image sensor with integrated TDCs implemented in a deep-submicron 3D multi-wafer-stacking IC technology. It is also the first CMOS SPAD array with BSI pixels. The power consumption per TDC is the lowest ever reported for any SPAD-TDC. The resource sharing scheme adopted in this image sensor together with the low power consumption of the presented new TDC architecture pave the way towards the implementation of larger SPAD-based sensors for TCSPC applications.

ACKNOWLEDGMENT

The authors would like to thank M. Hohmann for his assistance in the test of the devices. The authors are grateful to Xilinx Inc., for the generous donations of FPGAs under the University Program and to NKT Photonics for their technical support.

References

- [1] S. R. Arridge, "Optical tomography in medical imaging," Inverse Problems, vol. 15, no. 2, p. R41, 1999
- [2] F. Scholkmann, S. Kleiser, A. J. Metz, R. Zimmermann, J. M. Pavia, U. Wolf, and M. Wolf, "A review on continuous wave functional nearinfrared spectroscopy and imaging instrumentation and methodology,' NeuroImage, vol. 85, no. 0, pt. 1, pp. 6-27, 2014.
- [3] S. R. Arridge and J. C. Schotland, "Optical tomography: Forward and inverse problems," Inverse Problems, vol. 25, no. 12, p. 123010, 2009.
- [4] S. D. Konecky, G. Y. Panasyuk, K. Lee, V. Markel, A. G. Yodh, and J. C. Schotland, "Imaging complex structures with diffuse light," Opt. Express, vol. 16, no. 7, pp. 5048-5060, Mar. 2008.
- [5] S. D. Konecky, A. Mazhar, D. Cuccia, A. J. Durkin, J. C. Schotland, and B. J. Tromberg, "Quantitative optical tomography of sub-surface heterogeneities using spatially modulated structured light," Opt. Express, vol. 17, no. 17, pp. 14 780-14 790, Aug. 2009.
- [6] H. Dehghani, M. E. Eames, P. K. Yalavarthy, S. C. Davis, S. Srinivasan, C. M. Carpenter, B. W. Pogue, and K. D. Paulsen, "Near infrared optical tomography using NIRFAST: Algorithm for numerical model and image reconstruction," Commun. Numer. Methods Eng., vol. 25, no. 6, pp. 711-732, 2009
- [7] H. Dehghani, B. R. White, B. W. Zeff, A. Tizzard, and J. P. Culver, "Depth sensitivity and image reconstruction analysis of dense imaging arrays for mapping brain function with diffuse optical tomography,' Appl. Opt., vol. 48, no. 10, pp. D137-D143, Apr. 2009.
- [8] D. J. Cuccia, F. Bevilacqua, A. J. Durkin, F. R. Ayers, and B. J. Tromberg, "Quantitation and mapping of tissue optical properties using modulated imaging," J. Biomed. Opt., vol. 14, no. 2, pp. 024 012-024 012-13, 2009.
- [9] D. Grosenick, K. T. Moesta, H. Wabnitz, J. Mucke, C. Stroszczynski, R. Macdonald, P. M. Schlag, and H. Rinneberg, "Time-domain optical mammography: Initial clinical results on detection and characterization of breast tumors," Appl. Opt., vol. 42, no. 16, pp. 3170-3186, Jun. 2003.

10⁰

10

- [10] J. Mata Pavia, M. Wolf, and E. Charbon, "Single-photon avalanche diode imagers applied to near-infrared imaging," *IEEE J. Sel. Topics Quantum Electron.*, vol. 20, no. 6, pp. 291–298, Nov. 2014.
- [11] M. Gersbach, Y. Maruyama, R. Trimananda, M. Fishburn, D. Stoppa, J. Richardson, R. Walker, R. Henderson, and E. Charbon, "A time-resolved, low-noise single-photon image sensor fabricated in deep-submicron CMOS technology," *IEEE J. Solid-State Circuits*, vol. 47, no. 6, pp. 1394–1407, Jun. 2012.
- [12] D. Bronzi, F. Villa, S. Tisa, A. Tosi, F. Zappa, D. Durini, S. Weyers, and W. Brockherde, "100 000 frames 64 × 32 single-photon detector array for 2-D imaging and 3-D ranging," *IEEE J. Sel. Topics Quantum Electron.*, vol. 20, no. 6, pp. 1–10, Nov. 2014.
- [13] J. Richardson, R. Walker, L. Grant, D. Stoppa, F. Borghetti, E. Charbon, M. Gersbach, and R. Henderson, "A 32 × 32 50 ps resolution 10 bit time to digital converter array in 130 nm CMOS for time correlated imaging," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, Sep. 2009, pp. 77–80.
- [14] L. Braga, L. Gasparini, L. Grant, R. Henderson, N. Massari, M. Perenzoni, D. Stoppa, and R. Walker, "A fully digital 8 × 16 SiPM array for PET applications with per-pixel TDCs and real-time energy output," *IEEE J. Solid-State Circuits*, vol. 49, no. 1, pp. 301–314, Jan. 2014.
- [15] A. J. Carimatto, S. Mandai, E. Venialgo, T. Gong, G. Borghi, D. Schaart, and E. Charbon, "A 67,392 SPAD PVTB-compensated multi-channel digital SiPM with 432 column-parallel 48 ps 17-bit TDCs for endoscopic time-of-flight PET," presented at the IEEE Int. Solid-State Circuits Conf. Dig. Techn. Papers (ISSCC), San Francisco, CA, USA, 2015.
- [16] C. Veerappan et al., "A 160 × 128 single-photon image sensor with on-pixel 55 ps 10 b time-to-digital converter," in *IEEE Int. Solid-State Circuits Conf. Dig. Techn. Papers*, 2011, pp. 312–314.
- [17] S. Mandai, M. W. Fishburn, Y. Maruyama, and E. Charbon, "A wide spectral range single-photon avalanche diode fabricated in an advanced 180 nm CMOS technology," *Opt. Express*, vol. 20, no. 6, pp. 5849–5857, Mar. 2012.
- [18] E. A. G. Webster, J. A. Richardson, L. A. Grant, D. Renshaw, and R. K. Henderson, "A single-photon avalanche diode in 90-nm CMOS imaging technology with 44% photon detection efficiency at 690 nm," *IEEE Electron Device Lett.*, vol. 33, no. 5, pp. 694–696, May 2012.
- [19] E. Webster, R. Walker, R. Henderson, and L. Grant, "A silicon photomultiplier with > 30% detection efficiency from 450–750 nm and 11.6 μm pitch NMOS-only pixel with 21.6% fill factor in 130 nm CMOS," in *Proc. Solid-State Device Res. Conf. (ESSDERC)*, Sep. 2012, pp. 238–241.
- [20] R. Field, S. Realov, and K. Shepard, "A 100 fps, time-correlated single-photon-counting-based fluorescence-lifetime imager in 130 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 49, no. 4, pp. 867–880, Apr. 2014.
- [21] C. Niclass, M. Soga, H. Matsubara, M. Ogawa, and M. Kagami, "A 0.18-µm CMOS SoC for a 100-M-range 10-frame/s 200 × 96-pixel time-of-flight depth sensor," *IEEE J. Solid-State Circuits*, vol. 49, no. 1, pp. 315–330, Jan. 2014.
- [22] K. Torki, "3D-IC integration," presented at the CMP Annu. Users Meet., Paris, France, Jan. 2013.
- [23] S. Lipa, T. Thorolfsson, and P. Franzon, "The NCSU Tezzaron design kit," in *Proc. IEEE Int. 3D Syst. Integr. Conf. (3DIC)*, Nov. 2010, pp. 1–15.
- [24] C. Niclass, C. Favi, T. Kluter, M. Gersbach, and E. Charbon, "A 128 × 128 single-photon image sensor with column-level 10-bit time-todigital converter array," *IEEE J. Solid-State Circuits*, vol. 43, no. 12, pp. 2977–2989, Dec. 2009.
- [25] M. Ghioni, A. Gulinatti, I. Rech, F. Zappa, and S. Cova, "Progress in silicon single-photon avalanche diodes," *IEEE J. Sel. Topics Quantum Electron.*, vol. 13, no. 4, pp. 852–862, Jul. 2007.
- [26] C. Niclass, "Single-photon image sensors in CMOS: Picosecond resolution for three-dimensional imaging," Ph.D. dissertation, Informatics Commun. Dept., Swiss Federal Inst. Technol. Lausanne, Lausanne, France, 2008.
- [27] C. Niclass, M. Sergio, and E. Charbon, "A CMOS 64 × 48 single photon avalanche diode array with event-driven readout," in *Proc.* 32nd Eur. Solid-State Circuits Conf. (ESSCIRC), Sep. 2006, pp. 556–559.
- [28] J. Nissinen, P. Palojarvi, and J. Palojarvi, "A CMOS receiver for a pulsed time-of-flight laser rangefinder," in *Proc. 29th Eur. Solid-State Circuits Conf.*, Sep. 2003, pp. 325–328.
- [29] E. Charbon, M. Scandini, J. Mata Pavia, and M. Wolf, "A dual backside-illuminated 800-cell multi-channel digital SiPM with 100 TDCs in 130 nm 3D IC technology," presented at the IEEE Nucl. Sci. Symp. Med. Imag. Conf. (NSS/MIC), Seattle, WA, USA, 2014.

- [30] R. C. Haskell, L. O. Svaasand, T.-T. Tsay, T.-C. Feng, M. S. McAdams, and B. J. Tromberg, "Boundary conditions for the diffusion equation in radiative transfer," *J. Opt. Soc. Amer. A*, vol. 11, no. 10, pp. 2727–2741, Oct. 1994.
- [31] A. Liebert, H. Wabnitz, D. Grosenick, M. Möller, R. Macdonald, and H. Rinneberg, "Evaluation of optical properties of highly scattering media by moments of distributions of times of flight of photons," *Appl. Opt.*, vol. 42, no. 28, pp. 5785–5792, Oct. 2003.



Juan Mata Pavia received the M.S. degree in electrical engineering from the Royal Institute of Technology (KTH), Stockholm, Sweden, and the telecommunications engineering diploma from the Polytechnic University of Valencia (UPV), Valencia, Spain, in 2004, and the Ph.D. degree from the Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, France, in 2015.

He is currently with the University Hospital Zurich, Zurich, Switzerland, where he is investigating the design of new single-photon detectors

and their application in biomedical imaging. In 2005, he joined Philips Semiconductors, where he was involved in the design of application processors for smartphones. In 2009, he joined the Swiss Federal Institute of Technology Lausanne (EPFL), where he focused on the design of single-photon detectors for near-infrared applications.

Mario Scandini received the degree in electronic engineering from University of Padua, Padua, Italy, in 2002. He is currently pursuing the Ph.D. in electrical engineering from the Delft University of Technology, Delft, The Netherlands.

Since 2002, he has been working in the field of optical imaging. His research activities have been focused on designing optical devices based on CMOS technology. His research is focused on the design of SPAD-based CMOS image sensors for night vision and biomedical applications. He is particularly interested in the field of back side illuminated sensors and three dimensional silicon integration.



Scott Lindner received the M.Eng. degree in electronics and electrical engineering from the University of Edinburgh, Edinburgh, U.K., in 2011.

From 2011 to 2013, he was a trainee with the microelectronics section of the European Space Research and Technology Centre, Noordwijk, The Netherlands, where he worked on analog IC design for space applications. Since 2013, he has been pursuing the Ph.D. degree from the Swiss Federal Institute of Technology Lausanne (EPFL) in collaboration with the Biomedical Optics Research

Laboratory (BORL) at the University Hospital Zürich. His research is focused on the design of SPAD-based CMOS image sensors for use in biomedical applications, e.g., near-infrared optical tomography (NIROT).



Martin Wolf received the M.S. degree in electrical engineering and the Ph.D. degree in biomedical and power engineering from the Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, in 1990 and 1997, respectively.

Since 2002, he heads the Biomedical Optics Research Laboratory, Division of Neonatology, University Hospital Zurich, where he currently expands the research in biomedical optics to many clinical fields. In 2004 he became lecturer and in 2013 professor of the University Zurich. During his Ph.D., he special-

ized in biomedical optics, i.e., near infrared spectroscopy to investigate tissue oxygenation. As a postdoctoral research associate he worked with the Division for Neonatology, University Hospital Zurich until 1999, when he joined the Laboratory for Fluorescence Dynamics, University of Illinois at Urbana-Champaign. He developed imaging methods to non-invasively study hemodynamics and oxygenation of the brain and muscle.



Edoardo Charbon (SM'00) received the Diploma degree from ETH Zurich, Zurich, Switzerland, the M.S. degree from the University of California at San Diego, San Diego, CA, USA, and the Ph.D. degree from the University of California at Berkeley, Berkeley, CA, USA, in 1988, 1991, and 1995, respectively, all in electrical engineering and electrical engineering and computer science.

He is a consultant for numerous organizations, including Bosch, Texas Instruments, Agilent, and the Carlyle Group. He was with Cadence Design Sys-

tems, from 1995 to 2000, where he was an Architect of the company's initiative on information hiding for intellectual property protection. In 2000, he joined Canesta Inc., as the Chief Architect, where he led the development of wireless 3-D CMOS image sensors. From 2002 to 2008, he was a Faculty Member with the École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, where he was involved in research on CMOS sensors, biophotonics, and ultra low-power wireless embedded systems. In 2008, he joined Delft University of Technology, Delft, The Netherlands, as a Full Professor of VLSI design, succeeding Patrick Dewilde. He has authored or co-authored over 250 papers in journals, conference proceedings, magazines, and two books, and he holds 15 patents. His current research interests include 3-D imaging, advanced bio- and medical imaging, quantum integrated circuits, and space-based detection.

Dr. Charbon was a Guest Editor of the IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS and the IEEE JOURNAL OF SOLID-STATE CIRCUITS and a member of the Chair of Technical Committee of ESSCIRC, ICECS, ISLPED, VLSI-SOC, and IEDM. He is a Distinguished Visiting Scholar with the W. M. Keck Institute for Space, California Institute of Technology, Pasadena, CA.