Spatially Resolved Event-Driven 24 × 24 Pixels SPAD Imager With 100% Duty Cycle for Low Optical Power Quantum Entanglement Detection

Fabio Severini[®], Iris Cusini[®], Francesca Madonini[®], Davide Brescia, Robin Camphausen[®], Álvaro Cuevas[®], Simone Tisa[®], and Federica Villa[®], *Member, IEEE*

Abstract-Quantum microscopy requires efficient detectors able to identify temporal correlations among photons. Photon coincidences are usually detected by postprocessing their timestamps measured by means of time-to-digital converters (TDCs), through a time and power-consuming procedure, which impairs the overall system performance. In this article, we propose an innovative single-photon sensitive imager based on single-photon avalanche diodes (SPADs), able to signal coincident photon pairs along with their position through a TDC-free, event-driven architecture. The result is a highly efficient detector (25.8%) with a 100% duty cycle and minimized data throughput. The modular architecture and the 330 ns readout time, independent of pixel number, pave the way to large format imagers based on the same paradigm. The detector enabled quantum imaging at extremely low, microwatt-level optical pump powers, four orders of magnitude lower than previous experiments with similar optical setups.

Index Terms—(On-chip) photon coincidence detection (CD), entangled photons, event-driven readout, quantum microscopy, single-photon avalanche diode (SPAD) array.

I. INTRODUCTION

THE most important performance parameters in optical imaging are sensitivity (i.e., the minimum measurable variation of the quantity under investigation), spatial resolution (i.e., the minimum distance at which two points can be distinguished), and, in a few applications, temporal resolution (i.e.,

Manuscript received 3 October 2022; revised 16 January 2023; accepted 12 February 2023. This article was approved by Associate Editor David Stoppa. This work was supported in part by the Ministerio de Ciencias e Innovación de España (MICIIN) through the European Union Next Generation EU under Grant PRTR-C17.11; in part by the Generalitat de Catalunya; in part by the Fundació Cellex, the Fundació Mir-Puig, and the Generalitat de Catalunya through Centres de Recerca de Catalunya (CERCA) under Grant CEX2019-000910-S [MCIN/ AEI/10.13039/501100011033]; in part by the European Union's Horizon 2020 Research and Innovation Programme under Agreement 801060 [Quantum-enhanced on-chip interference microscopy (Q-MIC)]; in part by the Marie Sklodowska-Curie Grant under Agreement 713729 (ICFOstepstone 2); in part by the Agència de Gestió d'Ajuts Universitaris i de Recerca under Grant 2021 SGR 01458; and in part by the Project Superficies sintonizables para imaging óptico (TUNA-SURF) under Grant PID2019-106892RB-100. (Corresponding author: Federica Villa.)

Fabio Severini, Iris Cusini, Francesca Madonini, Davide Brescia, and Federica Villa are with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, 20133 Milan, Italy (e-mail: fabio.severini@polimi.it).

Robin Camphausen and Álvaro Cuevas are with the Institut de Ciències Fotòniques, 08860 Barcelona, Spain.

Simone Tisa is with Micro Photon Devices srl, 39100 Bolzano, Italy.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSSC.2023.3249122.

Digital Object Identifier 10.1109/JSSC.2023.3249122

time domain analogous to spatial resolution). When using classical light, they are limited by the Poissonian nature of light (shot noise limit) and by diffraction (Abbe limit). To outmatch sensitivity and spatial resolution limits, quantum light sources can be exploited, taking advantage of entangled N00N states and photon correlations to boost sensitivity to the Heisenberg limit and to improve resolution by a factor of k (k being the order of the measured correlation function) [1], [2], [3]. Similarly, temporal resolution can be improved by detecting correlated multiphoton events, which are much better defined in time than classical light [4]. A number of quantum-enabled imaging techniques have been proposed, which include spatial super-resolution [2], [5], enhanced signal-to-noise ratio in intensity and phase imaging [6], [7], and quantum illumination to reject stray light [8]. All of them require the detection of coincidences between temporally correlated photons, thus requiring the development of high-fidelity quantum entanglement imaging detectors.

Thanks to their single-photon sensitivity, single-photon avalanche diode (SPAD) imagers are the forefront solid-state sensors for quantum imaging, showing many other advantages such as CMOS fabrication compatibility, reliability, ease of operation, photon timing resolution, fast readout, and immunity to readout noise.

Photon coincidence detection (CD) can be performed either by counting incident photons in well-defined temporal windows (gated counting) [9] or by timestamping the photons' arrival times with time-to-digital converters (TDCs) and then discriminating coincidences in post-processing [10]. When working with continuous-wave (CW) quantum entanglement sources, these methods (typically relying on frame-based readout), however, limit the observation time per frame and (when working in photon-starved regimes) generate a large overhead of needless data. Conversely, a detector architecture with on-chip CD and event-driven readout, even without timestamping the photons, avoids these issues since postprocessing is no longer needed, readout time gets shortened, and only useful data is saved.

While silicon photomultipliers (SiPMs) with locally generated current pulses are able to detect the presence of coincident photons but are not spatially resolved [11], several SPAD arrays [12], [13], [14], [15], [16] show useful features, although not specifically tailored for quantum imaging applications. None of them embeds all desired features for

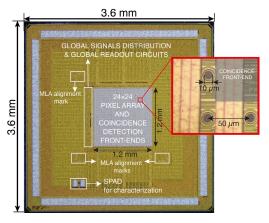


Fig. 1. Microphotograph of the chip, with pixels zoom-in.

entanglement detection, i.e., combining on-chip CD within subnanoseconds coincidence windows and event-driven readout with high photon detection probability (PDP), fill factor, pixel count, spatial resolution at pixel level, and scalability as thoroughly described in [17].

We present an innovative SPAD imager for quantum entanglement detection with on-chip CD, event-driven fast readout, 24 × 24 pixels with spatial resolution, and easily scalable architecture. We characterize its electrical and optical performance in terms of dark count rate (DCR), PDP, microlens concentration factor (CF), and crosstalk and validate it by quantum imaging a space-momentum entangled two-photon state, at record low optical pump power, compared to similar experimental setups.

II. CHIP ARCHITECTURE

We designed a 24×24 pixels SPAD imager in a 160 nm Bipolar-CMOS-DMOS (BCD) technology [18]. The chip (see Fig. 1) has a 3.6×3.6 mm² dimension with 1.44 mm² SPAD active area and a 3.14% fill factor, enhanced through a microlens array (MLA). The breakthrough of this design lies in its ability to detect photon coincidences across the entire array directly on-chip and to provide the addresses of triggered pixels only when a coincidence is detected, thanks to a novel event-driven logic.

Fig. 2 shows the simplified architecture: each pixel includes a SPAD with its sensing/quenching/recharge front-end circuitry, also able to generate a well-defined quantized (in intensity and width) current pulse upon each photon detection, and the readout logic, able to communicate the pixel address through a shared line. An adder node sums all pixels' output currents, then a CD circuitry (CDC) distinguishes if no photons, one photon, or more than one photon have been detected within a 2.5 ns coincidence window. The coincidence window duration is set equal to the pulse-width of the current generators as a compromise between false coincidence rejection (the shorter, the better) and reliable detection (e.g., against mismatches). In order to limit the capacitive load at the CDC input, the 24 × 24 pixels array is divided into elementary subarrays constituted by 12 × 12 pixels, each subarray, including the CDC (the small blue square in Fig. 2), in place of one pixel at each subarray center.

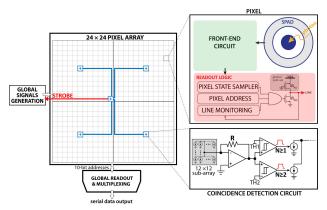


Fig. 2. Simplified architecture of the developed imager, with pixels subarrays and CD circuits (shown as small blue squares).

The CDC consists of a transimpedance amplifier (TIA), two comparators (with equivalent thresholds of 1 and 2 photons, respectively), and two current generators (see Fig. 2 right). The currents of the four 12×12 pixels subarrays (connected through a skew-free H-tree) are summed into another current adder. The same CDC is replicated at the center of the 24 \times 24 array and provides the STROBE signal when its 2-photons threshold is exceeded, independently of the position of the coincidence event (i.e., both in case of two photons within a subarray or across the whole array). The number of subarrays trade-off low capacitive load at each adder node and the short time delay between coincidence event and STROBE signal generation. The dimension of the subarray has been chosen to limit the impact of both time skews and current variability. Indeed, time skews among furthest pixels have been simulated to be 15 ps, and NMOS transistors implementing in-pixel current generators have been set not minimal so that the current variation due to process and mismatches is negligible within a 12×12 pixels area. As it will be thoroughly described in Section IV-C, time skews and current variations within the 12×12 subarrays do not significantly impact the coincidence time window duration. When STROBE is set, each triggered pixel is enabled to transmit its address through a controller area network (CAN)-bus-like communication protocol. The chip-sensitive area and pixel count can be seamlessly scaled up to larger formats, thanks to the conceived modular architecture.

A. SPAD Front-End Circuit

The SPAD front-end is essential to properly sense the avalanche, quench it, and restore the initial bias to the detector. We modified the variable load quenching circuit (VLQC) [19] for free-running mode operation, i.e., to automatically rearm the SPAD after hold-off), as required by a CW laser source. Fig. 3 shows the two main building blocks: the actual sensing, quenching and recharge part (top) and the hold-off logic (bottom). The former makes use of 5 V transistors (shown with thicker gate), needed to withstand 5 V excess bias ($V_{\rm EX}$) so to provide the best trade-off among detection, timing, and noise performance, and 1.8 V transistors to improve signal detection and speed.

The VLQC generates an *EVENT* signal every time an avalanche is triggered, keeps the SPAD quenched for a fixed

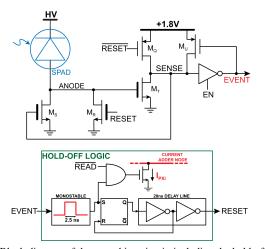


Fig. 3. Block diagram of the quenching circuit, including the hold-off logics.

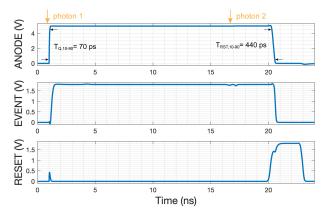


Fig. 4. Postlayout simulation of main signals timings of the VLQC, highlighting free running operation.

hold-off time (about 20 ns), and then restores the initial conditions by biasing the SPAD above breakdown. In the rest condition, the SPAD anode is kept to ground (i.e., with full high-voltage HV applied) through M_S, a 5 V NMOS operating in an ohmic regime, ready to sense an avalanche current. When an avalanche occurs, the current through M_S increases the anode voltage, thus switching M_T on and discharging the SENSE node to the ground. Consequently, M_S switches off, thus maximizing the quenching resistance and speeding up the quenching process (anode reaches $V_{\rm EX} = 5$ V in $T_{O,10-90} =$ 70 ps), thanks to positive feedback. As soon as the SENSE node is pushed to the ground, EVENT is set, thus activating the hold-off logic to keep the SPAD quenched for a fixed hold-off of about 20 ns. After that, RESET is generated to restore the initial conditions, so to switch M_R on and rapidly bring the anode to the ground, and rearm the SPAD. M_O is switched on by \overline{RESET} and recharges the SENSE node, thus quickly discharging the anode voltage ($T_{RST,10-90} = 440 \text{ ps}$). The main timings postlayout simulations in Fig. 4 highlight the free-running operation: upon photon detection, EVENT is set for 20 ns and during such hold-off time, photons cannot be detected.

Each pixel embeds a D flip-flop to selectively enable or disable (*EN*) the corresponding SPAD for disabling hot pixels, i.e., SPADs with a DCR much higher than the median value, which would cause many spurious single-photon detections.

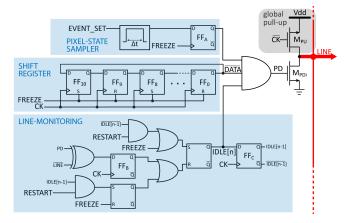


Fig. 5. Block diagram of the readout logic.

The 20 ns hold-off time ($T_{\rm HOFF}$) is a trade-off between afterpulsing probability (about 0.12% [18]) and achievable photon rate (roughly equal to $1/T_{\rm HOFF}$), so the detector is not blinded by a single-photon rate estimated to be below 5 Mcps (considering both DCR and laser source) [20]. The hold-off logic consists of a delay line that propagates the output of an SR-latch, set by the *EVENT* signal (shortened to 2.5 ns), and then generates the *RESET* signal to rearm the SPAD. The hold-off logic embeds the pixel output current generator (a single NMOS transistor), connected to the common adder node, and a monostable circuit sets its 2.5 ns duration. The current pulse is generated just in case the SPAD gets triggered, but the array is not in its readout phase (i.e., the global electronics set \overline{READ} high).

B. Addresses Readout Circuit

A common address event representation (AER) readout method [21] is valid just to discern single non-coincidence photons, being without a dedicated coincidence event arbitration logic. Thus, a dedicated in-pixel addresses readout logic was purposely designed. The readout circuit aims to provide the addresses of the triggered pixels every time the STROBE signal is generated by the highest-level CDC. The in-pixel readout logic, shown in Fig. 2, is inspired by the CAN bus communication ("0" is the dominant state, while "1" is the recessive state), shared among all array pixels. Each triggered pixel sends its address-bits at each clock cycle through an open-drain line (able to impose only low logic values), while simultaneously monitoring the data-line common to all pixels, so to check if the written bit differs from the sent one (meaning that the line is busy with another data-transfer) so to stop communication sending the address again in a next transfer.

The readout logic is implemented through the finite state machine shown in Fig. 5. Communication is initiated by the *FREEZE* signal, which is the *STROBE* distributed back to all pixels through a balanced tree of buffers. In fact, *FREEZE* samples the status of the pixels through flip-flop FF_A; if the pixel got triggered (status is *high*), data transfer is enabled. *FREEZE* also forces the parallel load of the pixel address into the 11-bit shift register, sent as *DATA*. The open-drain enable signal *PD* comes from the AND-ing of three signals: the sampled pixel status, the inverted address-bit (*DATA*), and the

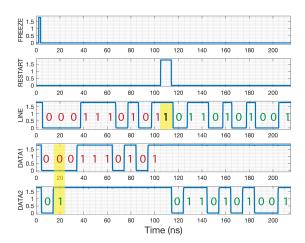


Fig. 6. Postlayout simulation of the readout circuit when two pixels in a 24×24 subarray try to send their addresses.

IDLE signal (which signals whether the line is free from other transfers). The IDLE signal is used to define when a pixel can transfer its address, i.e., after receiving the FREEZE signal and until the EX-OR detects any eventual incongruence between the DATA bit written on LINE (bus collision) and pixel address bit. In case the pixel address communication is interrupted because of bus collision, the IDLE signal is reasserted by the RESTART signal, and the address communication is restarted. For every clock rising edge, for half clock period, a global pull-up transistor M_{PU} resets the LINE to a high state, while for the second half period, M_{PDi} provides the address-bit, one by one. Thanks to the dynamic pull-up operation, transitions operate faster, operating, in simulations, with a 100 MHz reference clock.

The master logic in the global electronics, outside the pixels, provides the transfer and readout of three addresses because the chip is conceived to detect the two-photon coincidence; hence a third valid address would signal a spurious event to be discarded in postprocessing since three (or more) pixels got triggered (either by photons or by DCR noise). Moreover, this can allow modeling the 1-photon statistics and implement accidentals removal as done in recent articles [7], [9].

The transient simulation in Fig. 6 shows two triggered pixels sending their addresses, which are written hard in each pixel (0001110101 and 0110101001, respectively) at the same time. At the second bit, pixel 2 reads a "0" on the *LINE* (as written by pixel 1), which differs from the second bit of its address (which is a "1," instead). Thus, pixel 2 communication stops until the *RESTART* signal resets pixel 1 and enables pixel 2 again. At the end of the communication, both pixels will have successfully transmitted their addresses.

C. Coincidence Detection Circuit

The CDC aims at signaling if no photons, one photon, or more than one photon have been detected within a coincidence window and is hierarchically organized to assure array scalability. The CDC includes a TIA and two comparators and is purposely replicated to eventually propagate the CD information of each subarray up at the upper hierarchical level, eventually generating the global STROBE signal. Efforts have

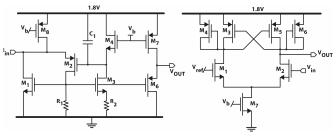


Fig. 7. Schematic of the TIA stage (left). Schematic of the comparator stage core (right).

been put into matching the pixel pitch, so the CDC can be put in place of just one pixel, thus wasting the minimum number of pixels for each subarray. Another critical constraint is the signal propagation delay through the CDC since it impacts the delay between photon detection and pixel status sampling, hence the actual spurious events probability.

For the TIA, we redesigned the one presented in [22], providing high stability of input virtual ground, in order to improve stability and sufficient phase margin (87°) even in the presence of large input stray capacitance. Indeed, in the worst-case condition, the TIA could be fed by 12 × 12 pixels current generators, corresponding to about 2.4 pF a stray capacitance. A stable virtual ground node allows stabilization of biasing conditions of the input current generators and shortens the response time (i.e., avoiding charge/discharge of the adder node parasitic capacitance). Furthermore, we set the TIA gain to provide a relatively high voltage pulse of about 100 mV corresponding to one photon, so to relax low-noise and high-speed constraints of the following comparator's stage.

The TIA schematic is shown in Fig. 7, left. The TIA features a very low-impedance input stage, thanks to one negative feedback current sink and a second positive feedback loop, followed by a simple source follower output stage. The input stage transistors M_1 and M_2 and resistor R_1 are arranged in a negative feedback loop, which makes most of the $I_{\rm IN}$ current flow through M₁, and its gate voltage moves accordingly to be buffered to the output node. To further reduce the input impedance, the second feedback loop consisting of transistors M_3 and M_4 and resistor R_2 lowers the M_2 gate voltage and keeps the input voltage constant. This loop has positive feedback, although lower than one, to prevent instability. Capacitor C_1 is added on purpose to compensate for the stage. In this way, the current pulse generated by a triggered pixel returns quickly to zero, and so does the TIA, with no oscillations that otherwise would widen the coincidence time window. Hence, the double feedback structure helps in solving the low input impedance and high-gain trade-off $(Z_{in}(0) = 100 \Omega, V_{out}/I_{in}(0))$ = 56 dB Ω , simulated), keeping under control area occupation. As a matter of fact, note that to simplify the biasing, transistors M_4 , M_7 , and M_8 are driven by the same voltage V_b generated by a current mirror structure. The output stage is a basic source follower, which buffers the M₁ gate voltage to the output node. The resulting 420 ps propagation delay within the CDC contributes to the overall delay between photon detection and STROBE generation. The input-referred current noise of 1.62 μ A (rms) is much lower than the minimum input current amplitude (about 10 μ A).

The TIA is followed by two voltage comparators able to discriminate the number of detected photons, with 1-photon and 2-photon equivalent thresholds, respectively. The core of the comparator (Fig. 7, right) is based on a long-tailed differential input with an active load consisting of a low-gain semi-latch to speed up commutation. The input stage (M₁ and M₂) matches the TIA common mode output voltage, while the load consists of four transistors with different purposes. Two transistors (M₃ and M₅) introduce positive feedback to completely unbalance the output in the presence of very small input signals, while two transdiodes (M₄ and M₆) reduce the positive feedback gain of the semi-latch, which otherwise would lead to hysteresis, making it impossible to restore the comparator to operation. This differential stage feeds an inverter chain to regenerate a digital output voltage and a monostable to set the 2.5 ns output pulse duration, matching the current pulsewidth generated by triggered pixels. The $V_{\rm b}$ bias is obtained through a current mirror shared among the two comparators to minimize area occupation. $V_{\rm in}$ is the output of the sensing TIA, while V_{ref} is connected to the input of a second floating TIA, acting as a reference to guarantee $V_{\rm in} = V_{\rm ref}$ condition when no photon is detected and to balance the comparator in case of voltage offset. Therefore, the two comparators, having different thresholds, have a different unbalancing factor between the input transistors but have been designed to introduce the same propagation delay (1 ns, simulated), not to lose photon coincidence information. The resulting input-referred voltage noise is 1.65 mV (rms) for the 1-photon threshold comparator and 175 μ V (rms) for the 2-photon threshold comparator, in both cases much lower than the minimum input voltage (about 100 mV). The input-referred noise is dominated by the 1/f noise contribution (inversely proportional to the transistors area), thus justifying the one order of magnitude variation between the two input-referred noise sources, as the two comparators areas greatly differ, in order to obtain different input voltage thresholds.

D. Global Electronics

The global electronics outside the pixels consists of two main blocks. The first one generates the global signals (FREZE, \overline{READ} , RESTART) derived from the STROBE and used by the in-pixel readout logic and current generators, starting from a 100 MHz reference clock. FREEZE is the STROBE signal distributed back to all pixels to sample the status of each pixel and eventually enable the address transfer. \overline{READ} accounts for an ongoing readout and disables pixel current generators. A counter is used to feed up to 33 clock edges (CK) to the readout circuit, 11 clocks for each of the three 10-bit addresses, and to provide the RESTART signal every 11 clocks to begin a new address communication. All signals are distributed through a balanced tree structure.

The second global electronics block is based on a pipelined chain of Shift Registers. Considering a 100 MHz reference clock, the transfer of each address triplet to the final memory bank (from where data is readout) takes about 330 ns. This is the readout dead time, as during this period, the array is idle: all SPADs are off and are not able to detect any



Fig. 8. Camera module based on the 24×24 SPAD imager.

incoming photon. Note that the array dead time is independent of the number of pixels, but it depends just on the number of addresses to be read per each coincident event. As a comparison, considering a SPAD array with frame-based readout, the scanned readout of all pixels of a 24 \times 24 array at 100 MHz reference clock would require 5.76 μ s. Moreover, the required data bandwidth is much lower than in frame-based arrays, especially with the typical low count rate experienced in quantum imaging setups. For instance, the arrays described in [12] and [13] generate about 100 and 200 MB/s, respectively, independent of photon flux, whereas our chip generates just 2.5 kB/s throughput in similar quantum experiments (as the one described in Section V), with an improvement of about five orders of magnitude.

The proposed event-driven architecture also has clear advantages in terms of detection efficiency when compared with more conventional frame-based gated-counting or photon-timing arrays. In fact, in gated-counting arrays, only one gate (i.e., coincidence window) per frame can be open to deterministically identify coincidences, whereas in the photon-timing array, the TDC full-scale range (FSR) limits the detection window per each frame. For instance, the gated-counting array described in [14] and employed in quantum imaging in [9] works at 100 kfps with 10 ns gate windows, resulting in a duty cycle of only 0.1%, whereas the photon timing array described in [13] and used in [10] works at 850 kfps with 50 ns FSR, resulting in a duty cycle of 5%. Instead, our approach provides 100% duty cycle and introduces a dead time only when a photon pair is detected and has to be readout.

III. SPAD CAMERA

In order to manage the SPAD array, we developed a threetier stacked system based on a commercial module (XEM7310 by Opal~Kelly) with a Xilinx~Artix-7 FPGA. The assembly is shown in Fig. 8. The top layer is the chip carrier board hosting the chip, power supplies decoupling capacitors, and signal test points. The main intermediate board generates all power supplies and the readout clock and controls the FPGA interaction with both the chip (through header connectors) and the external measurement setup (through 50 Ω reconfigurable SMA connectors).

IV. EXPERIMENTAL RESULTS

We characterized both pixel and overall sensor chip performance, and we validated the CD in actual quantum imaging

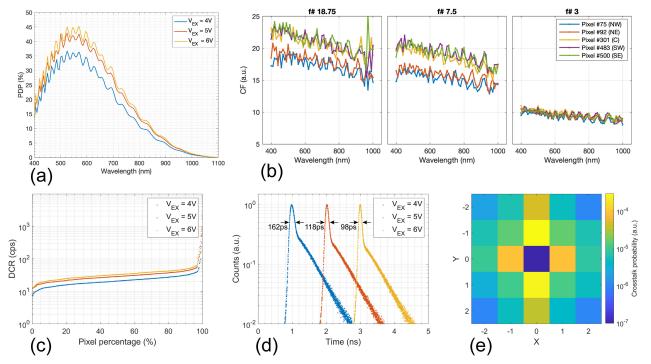


Fig. 9. (a) Measured PDP of SPADs in the array at different excess biases (VEX). (b) Measured MLA CF at different F-numbers and for different pixels at the extreme boundaries of the array. (c) Measured DCR cumulative distribution function of the SPADs in the 24 × 24 array. (d) Timing response of SPADs when illuminated by a 45 ps FWHM pulsed laser, at different excess bias. (e) Crosstalk probability map in logarithmic scale, only crosstalk between nearest neighbors is plotted as it becomes not measurable at longer distances from the "aggressor."

setups. We experimentally verified noise contributions (DCR and crosstalk), SPAD efficiency, pixel time response, and width of the time-coincidence window. All measurements have been carried out with no cooling and at 5 V excess bias.

A. SPAD Performance

The measured SPAD performance, in terms of PDP and DCR, is consistent with those reported in [18]. These parameters were characterized with the imager programed to have a single pixel activated and by turning off the clock to minimize power consumption and self-heating. The PDP peaks to about 43% at 535 nm and 5 V excess bias [see Fig. 9(a)], corresponding to a photodetection efficiency (PDE) of about 1.35% due to a native fill factor of 3.14%. The MLA can greatly improve this figure; thus, we performed a detailed characterization of the MLA CF, defined as $CF = (Counts_{\mu lens}/Counts_{bare})$, i.e., the ratio between the number of photons detected by a chip with and without MLA. As shown in Fig. 9(b), CF reaches about 20 for F-number up to f/7.5, still being about 10 at f/3, meaning that the MLA increases fill factor up to about 60% and peak PDE to about 26%.

The measured median DCR at room temperature and 5 V excess bias is 28 cps with a hot pixel percentage of less than 3% [see Fig. 9(c)]. A low DCR is key in quantum microscopy applications since the typical signal count rate is a few tens of kcps so thermal events can easily cause false coincidences.

B. Pixel Performance

In order to better characterize the VLQC of the proposed pixel, we measured the pixel time response, even if the array was not meant for photon-timing applications but for time CDs. We exposed the array to a pulsed laser source, and we accumulated a histogram of the photon arrival times so as to extract the jitter introduced by the SPAD and its front end: the full width at half maximum (FWHM) of the histogram combines all time-jitter contributions of laser, pixel, and external instrumentation. The pulsed diode laser at 850 nm had a pulsewidth of 45 ps FWHM; the distribution of time delays between the laser sync and the *STROBE* signal was acquired through an 8 ps FWHM TCSPC board (model SPC-630 by Becker and Hickl GmbH). Fig. 9(d) shows the timing response at different excess bias voltages: the FWHM values are 162, 118, and 98 ps at 4, 5, and 6 V excess bias, respectively, well below the 2.5 ns coincidence window.

Optical crosstalk is caused by photons emitted by hot carriers during an avalanche event that triggers other SPADs. It is one of the most significant noise contributions in quantum microscopy measurements since it introduces temporally correlated events that can mask useful coincidence events. According only to Poisson statistics, the theoretical rate of coincident events (no crosstalk) within a time window $t_{\rm cw}$ is

$$R_{\text{coinc},th} = 2R_A R_B t_{\text{CW}} \tag{1}$$

where R_A and R_B are the counting rates of pixel Px_A and Px_B . To characterize optical crosstalk, the camera has been kept in the dark, enabling just selected couples of pixels. Therefore, R_A and R_B correspond to the DCR rates of the two enabled pixels. Note that (1) is valid, provided that $R_i t_{CW} \ll 1$. Since the measured coincidence rate $R_{\text{coinc,meas}}$, is orders of magnitude higher than the theoretical coincidences $R_{\text{coinc,th}}$ computed with (1), all the measured coincidence events can be

considered due to crosstalk. Once the DCR of the two pixels under observation is known, by recording their coincidence rate in the dark, the crosstalk probability between Px_A and Px_B can be computed as

$$X_{AB} = \frac{R_{\text{coinc,meas}}}{R_A + R_B}.$$
 (2)

The measured crosstalk probability is 1.14×10^{-4} for two neighboring pixels; it decreases to 3.26×10^{-5} at a one-pixel distance and to 3.13×10^{-6} at a two-pixel distance, while for diagonal pixels is 3.03×10^{-6} . Fig. 9(e) shows the crosstalk probability as a function of pixel position with respect to the "aggressor" (central) one. Results are in agreement with what is obtained in SPAD arrays with identical diameter-to-pixel pitch ratios [23].

C. Coincidence Time Window

In quantum microscopy measurements, entangled photons are expected to reach the detector within a time lag shorter than 500 fs. Having a detector with a short and well-known coincidence time window reduces the risk of false detection and allows better noise compensation in postprocessing. This figure of merit can be inferred through statistical considerations on the data distribution, so as to set a given degree of uncertainty. The idea is to enable just two pixels, namely Px_A and Px_B , ensuring that their photon events (i.e., counts) are independent Poissonian processes. By rearranging (1), the coincidence time window can be found as

$$t_{\rm CW} = \frac{R_{\rm coinc,th}}{2R_A R_B}.$$
 (3)

Since the SPAD array can work either in CD mode or in *Single Count* mode, the rates R_A , R_B and $R_{\text{coinc,th}}$ must be evaluated in distinct observation windows T_A , T_B , $T_{\text{coinc,th}}$.

The two independent Poisson processes were generated by a constant incoherent light source (a current-controlled infrared LED) impinging on two non-hot and distant pixels, so as to have negligible crosstalk and DCR when compared to the high photon rate. R_A and R_B were chosen in the order of 120 kcps, so as to have $R_i t_{\rm CW} \ll 1$, obtaining a $R_{\rm coinc,th}$ in the order of 50 cps. The chosen observation times of T_A , $T_B = 85$ s, and $T_{\rm coinc} = 170$ s are short enough to ensure a constant temperature across measurements (to have the same triggering probability during the three phases) and keep the uncertainty of the measured $t_{\rm cw}$ below 1%.

The chip has been characterized in different scenarios: two pixels belonging to different 12×12 sub-array (to evaluate the coincidence window related to different CDCs) resulting in an average $t_{\rm cw} = 2.43$ ns (7 ps rms variability), and two pixels belonging to the same subarray (to evaluate the coincidence window due to the same front-end circuit) resulting in an average $t_{\rm cw} = 2.26$ ns (54 ps rms variability). It can be noticed that for two pixels belonging to different subarrays, the average $t_{\rm CW}$ is longer than for pixels in the same subarray. This is due to the chip architecture since $t_{\rm CW}$ is set by two different voltage-controlled current generators in the two cases. In the former, it is controlled by a 2.5 ns (as simulated) voltage pulse generated by the CDC monostable; in the latter, it is controlled

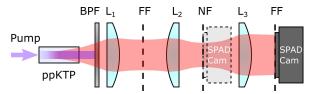


Fig. 10. Setup for imaging space-momentum entangled photon pairs, with periodically poled ppKTP, bandpass filter (BPF), lens (L), FF, and NF images.

by a voltage pulse of 2.3 ns duration (as simulated), generated by the VLQC in-pixel circuit. Moreover, the 143 in-pixel current generators (one less than 12×12 , occupied by the CDC) are spread over a larger $(0.6 \times 0.6 \ \mu\text{m}^2)$ area, with respect to the eight current generators at the input of the global CDC, lumped within a 50 μ m pitch. The resulting limited variability for the same subarray case, however, proves that time skews and process variations within the 12×12 subarrays do not significantly impact the coincidence time window duration.

V. QUANTUM EXPERIMENTS

The current workhorse technology for quantum imaging is spontaneous parametric down-conversion (SPDC), where a nonlinear crystal, illuminated with a pump light source, converts with low probability a λ_p wavelength photon into a pair of lower energy photons with $\lambda_{SPDC} = 2\lambda_p$. This interaction is governed by momentum conservation, which ensures correlations between the two generated photons. These correlations in space and/or momentum degrees of freedom allow exploring high-dimensional entanglement [24] and are required for almost all quantum imaging schemes published to date [2], [6], [7], [8]. CD with SPAD arrays enables the detection of these SPDC correlations and has reduced experiment times by orders of magnitude compared to other cameras [9].

Here, we use our event-driven SPAD array to image photon pair correlations from SPDC in the near-field (NF) and far-field (FF). We show that the measured correlation strengths together violate the so-called "Einstein-Podolsky-Rosen" (EPR) criterion, which proves space-momentum entanglement. Fig. 10 shows the setup with a laser at 405 nm pumping a nonlinear periodically poled potassium titanyl phosphate (ppKTP) crystal, generating entangled SPDC photon pairs at 810 nm. After blocking the pump, the NF is imaged onto the SPAD camera sensor using two lenses (L_1, L_2) with focal lengths $f_1 = 300$ mm and $f_2 = 2500$ mm, respectively) in 4fconfiguration. Using an additional lens (L_3 , $f_3 = 500$ mm), another 4f system images the FF plane after L_1 . FF and NF coincidences were acquired using the event-driven CD SPAD chip, over 120 s in both cases, with pump laser power of 1.2 μ W. Formally, a state is entangled in space and momentum if it violates the EPR condition

$$\Delta \mathbf{r} \cdot \Delta \mathbf{k} > \frac{1}{2}.\tag{4}$$

Here, $\Delta r = \Delta (r_i - r_j)$ and $\Delta k = \Delta (k_i + k_j)$ represent the correlation strengths in position r = (x, y) and transverse wave vector k = (kx, ky), respectively, between the two photons (i, j) in SPDC pairs [9], [10]. The wave vector relates

TABLE I
PERFORMANCE COMPARISON BETWEEN STATE-OF-THE-ART SPAD CHIPSFOR QUANTUM ENTANGLEMENT DETECTION

Parameter	This work	[12]	[13]	[14]	[15][16]	
Year	2022	2019	2018	2014	2014	
Process (nm)	160	180	150	350	130	
Туре	SPAD array	SPAD array	SPAD array	SPAD array	SiPM array	
Architecture	Coincidence detection	Gated counting	TDC	Gated counting	TDC	
Readout time (µs)	0.33	10.2	1.25	10	2	
Pixel number	24×24	512×512	32×32	32×64	8×16 ¹	
Sensitive area (mm²)	1.2×1.2	8.39×8.39	1.43×1.43	4.8×9.6	0.61×0.57	
Power consumption (mW)	40	700	11.1 or 4.8 ⁴	50	N.A.	
Pixel pitch (μm)	50	16.38	44.64	150	N.A.	
SPAD size (μm)	10	6	19.8	30	16.87	
Pixel fill factor (%)	3.14 ² , 60 ³	10.5	19.48	3.14	42.6	
SPAD peak PDP (% @ λ nm)	43 @ 545 12 @ 810	50 @ 520	N.A.	50 @ 400	19 @ 450	
SPAD DCR, median (cps/µm²)	0.36	0.18	0.2	0.21	N.A.	
¹ 720 SPADs per pixel, ² native, ³ with Micro Lens Array, ⁴ depending on frame skipping or row skipping operation mode						

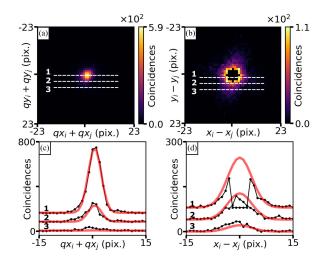


Fig. 11. SPDC coincidences (a) in the FF, projected into sum coordinates, and (b) in the NF, projected into difference coordinates. (c) and (d) Cross sections along dashed lines in (a) and (b), respectively, showing experimental values (black dots) and cross sections of 2-D Gaussian fittings (red solid). Cross-section curves are offset vertically for clarity.

to the transverse momentum operator p through $k = p/\hbar$. When measuring the SPDC FF, k maps onto the FF spatial coordinates q = (qx, qy) according to $k = (2\pi/\lambda_{SPDC} f_{FF})q$, where $f_{\rm FF} = f_1 \cdot f_3/f_2$ is the effective focal length of the FF projection. We, therefore, measure Δk by projecting the coincidences measured in this FF into the sum coordinates $q_i + q_j$ (shown in Fig. 11(a) and finding the width of a 2-D Gaussian fit to the coincidence peak [shown in Fig. 11(c)]. Given the pixel pitch of 50 μ m, we thus obtain Δk = (11.76 ± 0.04) mm⁻¹. Similarly, Δr is measured by projecting NF coincidences into the difference coordinates $r_i - r_j$ [shown in Fig. 11(b)] and fitting a 2-D Gaussian. Note that in the NF we ignore all coincidences for which photons were detected within a 3-pixel radius from each other, to avoid crosstalk. This accounts for the black region in the center of Fig. 11(b); however, it does not prevent an accurate 2-D Gaussian fitting,

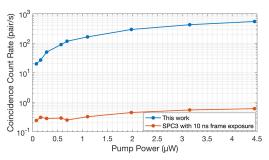


Fig. 12. Comparison between coincidences measurements with the presented event-driven CD camera and a gated camera [13], in the same optical setup, at low optical pump power.

as seen in Fig. 11(d). Taking into account the optical magnification factor $M = f_2/f_1$, we find $\Delta r = (20.4 \pm 0.4) \mu \text{m}$. Therefore, we obtain $\Delta r \cdot \Delta k = 0.240 \pm 0.004 < 1/2$, which violates the EPR condition in (4), thus proving the measurement of a space-momentum entangled quantum state.

The microwatt-level optical pump power used to achieve these results is more than four orders of magnitude lower than comparable results in the literature [9], [10], which used 30 and 50 mW pump laser power, respectively, and was enabled by an almost unity duty cycle of the camera. Our SPAD array, therefore, extends quantum imaging experimental capabilities to inexpensive low-power pump light sources such as light-emitting diodes.

To demonstrate this, the couple rate measured with our event-driven camera is compared with the count rate measured with a gated camera [13], in the same setup at a different low laser pump power (see Fig. 12). Because of some issues in the fabricated chip, the maximum detectable couple rate demonstrated in quantum experiments with the present SPAD array is around 600 pair/s, still comparable with the 4.25 kpair/s demonstrated in [10], which used four orders of magnitude higher laser power. In [9], 100 kpair/s has been achieved, but exploiting a statistical approach to

distinguish photon coincidences, which introduces unavoidable errors and constraints on the minimum number of frames to be analyzed.

The problem is related to the masking logic of the in-pixel current generation signal that is wrongly activated even during the readout phase. Moreover, there is an electrical crosstalk (due to a lack of proper signal shielding) between the readout clock and the STROBE signal. All in all, output data is corrupted if a photon event occurs during the readout. For this reason, the operating coincidence rate (600 pair/s) is much lower than the maximum that can be achieved by the design of around 3 Mpair/s (i.e., the inverse of the readout dead time). We expect to solve this issue in a second implementation of the chip so that the theoretical saturation level for the presented architecture will be reached.

VI. CONCLUSION

Table I shows a comparison between state-of-the-art SPAD arrays previously employed in quantum imaging measurements and the presented chip. In our case, the readout time is the lowest and is independent of the pixel number, so the same advantage holds even when scaling up this architecture to larger array formats. The PDE is the highest, thanks to the very effective microlenses employed, which recover a 60% fill factor. The chip power consumption has been measured to be 40 mW at the maximum output coincidence rate of 600 pair/s. One of the main drawbacks with respect to other arrays is the limited number of pixels. Readout time and coincidence time window duration are independent of the number of pixels, and the main challenge in scaling up the array is represented by the increased probability of detecting a third spurious event after a coincidence has been detected, as the time needed for the array to be blinded to other photon events scales with the number of comparation steps employed. Eventually, power consumption could become the bottleneck for the maximum number of pixels achievable, as in this architecture, it is dominated by the static power consumption of the CDCs, making power consumption increase by a factor N+1, being N the pixel number upscaling factor. A suitable low-power design should be then performed.

Overall performance can be improved by implementing the array in a more scaled technology node. Power consumption of logic circuits reduces thanks to lower supply voltages. Then, within the same area, more processing logic can be implemented (e.g., for trimming and calibration). Overall, area occupation can be reduced thanks to pixel pitch reduction. In planar technologies, SPAD size should scale down accordingly in order to prevent crosstalk among adjacent pixels. On the other hand, considering high-cost 3-D stacking technologies, the detecting performance greatly improves, as the pixel density improves without raising the crosstalk probability, thanks to top-tiers specifically tailored for SPADs. All in all, 3-D-stacking represents the best-in-class technology for quantum detection.

In conclusion, we proved that the reported SPAD imager is a valuable and innovative detector, capable of measuring quantum correlations between SPDC photon pairs at extremely low microwatt-level optical pump powers and short (2 min)

measurement time, and we verified the generation of spacemomentum entanglement. This capability directly enables low-cost implementations of quantum imaging protocols, such as using photon correlations to reject image noise [8], as well as applications in other domains, such as high-dimensional quantum key distribution [25].

REFERENCES

- S. Pirandola, B. R. Bardhan, T. Gehring, C. Weedbrook, and S. Lloyd, "Advances in photonic quantum sensing," *Nature Photon.*, vol. 12, pp. 724–733, Nov. 2018.
- [2] M. Unternährer, B. Bessire, L. Gasparini, M. Perenzoni, and A. Stefanov, "Super-resolution quantum imaging at the Heisenberg Limit," *Optica*, vol. 5, no. 9, p. 1150, 2018.
- [3] O. Schwartz, J. M. Levitt, R. Tenne, S. Itzhakov, Z. Deutsch, and D. Oron, "Superresolution microscopy with quantum emitters," *Nano Lett.*, vol. 13, no. 12, pp. 5832–5836, Dec. 2013.
- [4] C. A. Santivanez et al., "Quantum enhanced lidar resolution with multi-spatial-mode phase sensitive amplification," *Proc. SPIE*, vol. 8163, Sep. 2011, Art. no. 81630Z.
- [5] D. G. Monticone et al., "Beating the abbe diffraction limit in confocal microscopy via nonclassical photon statistics," *Phys. Rev. Lett.*, vol. 113, no. 14, Sep. 2014, Art. no. 143602.
- [6] N. Samantaray, I. Ruo-Berchera, A. Meda, and M. Genovese, "Realization of the first sub-shot-noise wide field microscope," *Light, Sci. Appl.*, vol. 6, no. 7, Jan. 2017, Art. no. e17005.
- [7] R. Camphausen et al., "A quantum-enhanced wide-field phase imager," Sci. Adv., vol. 7, no. 47, Nov. 2021, Art. no. eabj2155.
- [8] T. Gregory, P.-A. Moreau, E. Toninelli, and M. J. Padgett, "Imaging through noise with quantum illumination," *Sci. Adv.*, vol. 6, no. 6, Feb. 2020, Art. no. eaay2652.
- [9] B. Ndagano et al., "Imaging and certifying high-dimensional entanglement with a single-photon avalanche diode camera," NPJ Quantum Inf., vol. 6, no. 1, p. 94, Dec. 2020.
- [10] B. Eckmann, B. Bessire, M. Unternährer, L. Gasparini, M. Perenzoni, and A. Stefanov, "Characterization of space-momentum entangled photons with a time resolving CMOS SPAD array," *Opt. Exp.*, vol. 28, no. 21, 2020, Art. no. 31553.
- [11] F. Nolet, V. P. Rhéaume, S. Parent, S. A. Charlebois, R. Fontaine, and J.-F. Pratte, "A 2D proof of principle towards a 3D digital SiPM in HV CMOS with low output capacitance," *IEEE Trans. Nucl. Sci.*, vol. 63, no. 4, pp. 2293–2299, Aug. 2016, doi: 10.1109/TNS.2016.2582686.
- [12] A. C. Ulku et al., "A 512 × 512 SPAD image sensor with integrated gating for widefield FLIM," *IEEE J. Sel. Topics Quantum Electron.*, vol. 25, no. 1, pp. 1–12, Jan. 2019.
- [13] L. Gasparini et al., "A 32×32-pixel time-resolved single-photon image sensor with 44.64μm pitch and 19.48% fill-factor with on-chip row/frame skipping features reaching 800 kHz observation rate for quantum physics applications," in *IEEE Int. Solid-State Circuits Conf.* (*ISSCC*) Dig. Tech. Papers, Feb. 2018, pp. 98–100.
- [14] D. Bronzi et al., "100 000 frames/s 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. 354–363, Nov. 2014.
- [15] M. Unternährer, B. Bessire, L. Gasparini, D. Stoppa, and A. Stefanov, "Coincidence detection of spatially correlated photon pairs with a monolithic time-resolving detector array," *Opt. Exp.*, vol. 24, no. 25, 2016, Art. no. 28829.
- [16] L. H. C. Braga et al., "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*, vol. 49, no. 1, pp. 301–314, Jan. 2014.
- [17] F. Madonini, F. Severini, F. Zappa, and F. Villa, "Single photon avalanche diode arrays for quantum imaging and microscopy," Adv. Quantum Technol., vol. 4, no. 7, Jul. 2021, Art. no. 2100005.
- [18] M. Sanzaro, P. Gattari, F. Villa, A. Tosi, G. Croce, and F. Zappa, "Single-photon avalanche diodes in a 0.16 μm BCD technology with sharp timing response and red-enhanced sensitivity," *IEEE J. Sel. Topics Quantum Electron.*, vol. 24, no. 2, pp. 1–9, Mar. 2018.
- [19] F. Severini, I. Cusini, D. Berretta, K. Pasquinelli, A. Incoronato, and F. Villa, "SPAD pixel with sub-NS dead-time for high-count rate applications," *IEEE J. Sel. Topics Quantum Electron.*, vol. 28, no. 2, pp. 1–8, Mar. 2022.

- [20] F. Madonini, F. Severini, F. Zappa, and F. Villa, "Performance assessment of SPAD arrays for coincidence detection in quantum-enhanced imaging," *Proc. SPIE*, vol. 11835, Aug. 2021, Art. no. 1183505.
- [21] R. Berner, "Building-blocks for event-based vision sensors," Ph.D. dissertation, ETH Zürich, Switzerland, 2011.
- [22] M. Crotti, I. Rech, G. Acconcia, A. Gulinatti, and M. Ghioni, "A 2-GHz bandwidth, integrated transimpedance amplifier for singlephoton timing applications," *IEEE Trans. Very Large Scale Integr. (VLSI)* Syst., vol. 23, no. 12, pp. 2819–2828, Dec. 2015.
- [23] E. Conca, I. Cusini, F. Severini, R. Lussana, F. Zappa, and F. Villa, "Gated SPAD arrays for single-photon time-resolved imaging and spectroscopy," *IEEE Photon. J.*, vol. 11, no. 6, pp. 1–10, Dec. 2019.
- [24] M. P. Edgar et al., "Imaging high-dimensional spatial entanglement with a camera," *Nature Commun.*, vol. 3, no. 1, p. 984, Aug. 2012.
- [25] S. Etcheverry et al., "Quantum key distribution session with 16dimensional photonic states," Sci. Rep., vol. 3, no. 1, pp. 1–5, Jul. 2013.



Fabio Severini received the M.Sc. degree in electronics engineering and the Ph.D. degree in information technology from the Politecnico di Milano, Milan, Italy, in 2018 and 2022, respectively.

During his Ph.D. degree, his research interests focused on the design of integrated single-photon detectors in CMOS and Bipolar-CMOS-DMOS (BCD) technologies, targeting quantum imaging, and quantum communication applications. In 2022, he joined STMicroelecronics, Milan, as an Analog and Mixed-Signals Designer, centering his activity

on the design of high voltage drivers and analog front-ends for MEMS mirrors.



Iris Cusini received the B.Sc. degree in automation engineering from the Politecnico di Milano, Milan, Italy, and Tongji University, Shanghai, China, in 2016, within a double degree project, and the M.Sc. degree (summa cum laude) in electronics engineering from the Politecnico di Milano, in 2019, where she is currently pursuing the Ph.D. degree in information technology.

Her research interests include the design, development, and testing of systems with single-photon avalanche diodes (SPADs) and digital silicon photomultipliers (dSiPMs).



Francesca Madonini received the M.Sc. degree (summa cum laude) in electronics engineering from the Politecnico di Milano, Milan, Italy, in 2019. She was double degree awarded at the Politecnico di Torino, Turin, Italy, as a part of the Alta Scuola Politecnica Program. She is currently pursuing the Ph.D. degree in information technology (area electronics) with the Politecnico di Milano.

Her current research focuses on the design of microelectronic circuits for single-photon avalanche diode (SPAD) arrays, specifically tailored for photon

coincidences measurements in the field of quantum-enhanced microscopy.



Davide Brescia was born in Milan, Italy, in 1997. He received the B.Sc. degree in engineering physics and the M.Sc. degree in electronics engineering from the Politecnico di Milano, Milan, in 2019 and 2021, respectively.

He is currently working as a Firmware and Hardware Engineer for Leonardo, Edinburgh, U.K., where his activities are directed toward digital electronics and specifically FPGA programming.



Robin Camphausen was born in Berlin, Germany, in 1992. He received the B.Sc. degree (Hons.) majoring in physics from The University of Sydney, Camperdown, NSW, Australia, in 2017. He is currently pursuing the Ph.D. degree with the ICFO—The Institute of Photonic Sciences, Castelldefels, Spain working on quantum technology, applying quantum optics to imaging and sensing, quantum communication, and quantum random number generation.



Álvaro Cuevas received the Ph.D. degree in physics from the Sapienza Università di Roma, Rome, Italy, in 2018, where he studied a variety of information propagation phenomena, assisted by photon entanglement.

He is currently an Experimental Physicist in the field of quantum information and photonics, holding a Postdoctoral position in the Optoelectronics Group led by Prof. Valerio Pruneri at the ICFO-The Institute of Photonic Sciences, Castelldefels, Spain, where he develops technologies

for quantum-enhanced imaging, sensing and ranging, as well as entangled photon sources, entanglement distribution protocols, and discrete variable quantum key distribution implementations. He is the coauthor in more than 20 publications, holding a research H-index is 10.



Simone Tisa was born in Milan, Italy, in 1977. He received the M.Sc. degree (summa cum laude) in electronics engineering and the Ph.D. degree in electronics from the Politecnico di Milano, Milan, in 2001 and 2006, respectively.

Since 2007, he has been a Senior Engineer with Micro Photon Devices, Bolzano, Italy, since 2007. His research interests are connected with the use of single-photon detectors and active-quenching circuits for ultrafast imaging applications in astrophysics and with the design of bidimensional arrays

of single-photon avalanche diodes with integrated electronics for the acquisition of time-resolved images at the single-photon level.



Federica Villa (Member, IEEE) was born in Milan, Italy, in 1986.

Since 2015, she has been an Assistant Professor in electronics engineering with the Politecnico di Milano, Milan. Her academic research mainly focuses on the design of microelectronic integrated sensors and circuits in CMOS and bipolar-CMOS-DMOS (BCD) technologies for the development of single-photon detectors, such as single-photon avalanche diodes (SPADs) and silicon photomultipliers (SiPMs), photon-counting array imagers with

single-photon sensitivity, and high-resolution and linearity time-to-digital converters (TDCs).