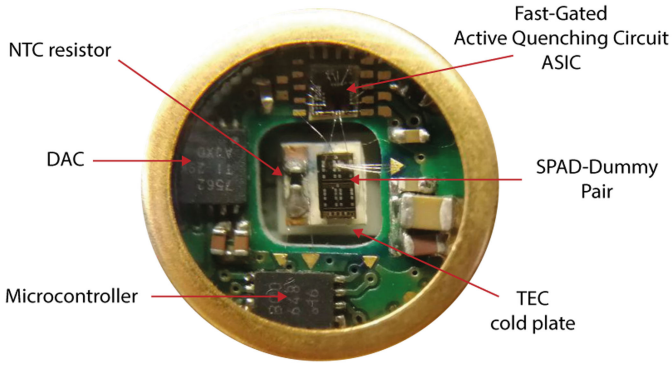
IEEE Photonics Journal
An IEEE Photonics Society Publication

High Detection Rate Fast-Gated CMOS Single-Photon Avalanche Diode Module

Volume 12, Number 5, October 2020

Marco Renna
Alessandro Ruggeri
Mirko Sanzaro
Federica Villa, ***Member, IEEE***
Franco Zappa, ***Senior Member, IEEE***
Alberto Tosi, ***Member, IEEE***



DOI: 10.1109/JPHOT.2020.3017092

IEEE photonics SOCIETY

IEEE

High Detection Rate Fast-Gated CMOS Single-Photon Avalanche Diode Module

Marco Renna, Alessandro Ruggeri, Mirko Sanzaro, Federica Villa, *Member, IEEE*, Franco Zappa, *Senior Member, IEEE*, and Alberto Tosi, *Member, IEEE*

Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano 20121, Italy

DOI:10.1109/JPHOT.2020.3017092

Manuscript received June 28, 2020; accepted July 30, 2020. Date of publication August 17, 2020; date of current version August 31, 2020. This work was supported by the EraNET Cofund Initiatives QuantERA within the European Union's Horizon 2020 research and innovation program under Grant agreement 731473 (project SQUARE). Corresponding author: Marco Renna (e-mail: marco.renna@polimi.it).

Abstract: We present a novel instrument for fast-gated operation of a 50 μm CMOS SPAD (Complementary Metal-Oxide-Semiconductor Single-Photon Avalanche Diode), driven by an integrated fast-gated active quenching circuit with transition times faster than 300 ps (20–80%). The instrument is based on a custom system-in-package where the SPAD and its driving electronics are housed in a TO-8 package. The detector can be operated at repetition rates up to 160 MHz, with gate on-times as short as 500 ps, always guaranteeing a temporal response with 60 ps (FWHM) timing jitter and short exponential decay (53 ps time-constant). A dark-count rate as low as 1 cps is achieved operating the CMOS SPAD at 5 V above breakdown at a temperature of 263 K, still keeping the afterpulsing probability lower than 2%, with only 50 ns hold-off time, thanks to the fast-gating driving electronics. The instrument is housed in a compact $5 \times 4 \times 8$ cm^3 case and can be triggered by either an external or internal source. A USB link allows to adjust measurement parameters, SPAD bias voltage and operating temperature. The high re-configurability of the instrument and its state-of-the-art performance make it suitable for applications where high detection rates and low timing jitter are required.

Index Terms: Single-photon, fast-gated, SPAD, afterpulsing, quantum key distribution, quantum communication.

1. Introduction

Measurement techniques based on single-photon counting are today successfully adopted in all the applications where the detection of fast and faint optical signals is required. Aside from high detection efficiency and low timing jitter, an ever-increasing number of applications is also requiring the possibility to time-filter optical signals, allowing to discard strong unwanted waveforms preceding (or following) the signal of interest, thus increasing measurement Dynamic Range (DR). Most relevant applications requiring such feature are Non-Line-of-Sight (NLOS) imaging [1], Time-Resolved Near InfraRed Spectroscopy (TR NIRS) [2], [3] and Quantum Key Distribution (QKD) [4].

Single-Photon Avalanche Diodes (SPADs) [5] represent one of the most valuable options for the above-mentioned applications, because of their high Photon Detection Efficiency (PDE), good

Single-Photon Timing Resolution (SPTR) and mostly because they allow for gated-mode operation: the latter consists in keeping the SPAD active only during a predefined ON-time, when it is able to detect photons, while keeping it quiescent during the OFF-time. When the temporal transitions between OFF and ON last less than 1 ns, the SPAD is said to be operating in fast-gating mode [6].

Sub-nanosecond gated operation can be also provided by other types of devices, such as Intensified Charge Coupled Device (I-CCD) cameras [7]. However, measurement acquisition time using CCDs is strictly limited by the long read out time of the sensor, which leads to a maximum repetition rate of few hundreds of Hz. In addition, these instruments are generally expensive, bulky and fragile, as they can be damaged by intense light.

State-of-the-art time-gated single-photon detectors rely on gated operation of either Silicon or InGaAs/InP SPAD commercial photon counters, depending on the wavelength of interest. Instruments based on Silicon SPADs can provide SPTR $<$50 ps (Full-Width at Half Maximum – FWHM), Dark Count Rate (DCR) in the order of few tens of count per second (cps) and PDE consistent with Silicon SPAD technology, with peak efficiency in the order of 50–60% [8], [9]. When high detection rates are required, the main limitation of these instruments is the afterpulsing, which is a signal correlated noise peculiar of SPAD detectors. During every avalanche, large current flows through the detector. Some carriers might get trapped in deep levels and then be released after some time (with time delay dependent on the trap nature), thus triggering a secondary avalanche [10]. The afterpulsing probability can be lowered by the read-out circuit mainly in two ways: i) reducing the junction current either by fast-quenching each avalanche or by lowering the excess bias; ii) holding the detector below breakdown after each avalanche for a certain amount of time, commonly known as hold-off time, allowing traps to release the captured carriers without triggering afterpulses. However, the reduction of the excess bias leads to a degradation of the detector performance and the long hold-off time limits the maximum achievable count rate.

In this paper we present an instrument specifically designed to overcome this limitation: starting from an integrated Fast-Gated Active Quenching Circuit (FG-AQC) developed by Politecnico di Milano and extensively described in [11], we developed a fully custom instrument based on a 50 μm diameter CMOS SPAD driven in fast-gated regime with transition times $<$200 ps, providing optical gates as short as 500 ps (FWHM) and an avalanche quenching time $<$1 ns, allowing to drive the SPAD ON/OFF at a repetition frequency up to 160 MHz, still ensuring negligible afterpulsing contributions. Such high gate repetition frequency and short transition times are achievable thanks to the SPAD-dummy approach, which allows to effectively suppress feed-through spikes at comparator input through the SPAD junction capacitance, while guaranteeing optimal timing jitter performance in avalanche-pulse readout [3].

Since this system is based on a silicon SPAD, it can be directly employed in Non-Line-of-Sight imaging and Time-Resolved Near InfraRed Spectroscopy, while it can be employed in fiber-based Quantum Key Distribution for achieving high count rates when coupled to an up-conversion unit [12].

In the following we describe the system architecture of the presented instrument and its experimental characterization.

2. Instrument Design

The presented fast-gated SPAD module is based on a fully-custom assembly made of:

- a System-in-Package (SiP) based on a standard 12-pin TO-8 package, hosting the SPAD and its electronics;
- an FPGA board, managing all the timing signals needed to correctly drive the SiP, besides handling the USB 2.0 communication with the external computer;
- a power board, managing the module power supplies and the SPAD cooling system;
- a front-end board, connecting both the FPGA and power boards to the SiP.

Fig. 1 shows a simplified block diagram of the instrument. The SiP is hosted on the front-end board, which interconnects the SiP to both the power board and the FPGA board. The power board provides the SiP and FPGA power supplies. The FPGA board generates the SiP timing

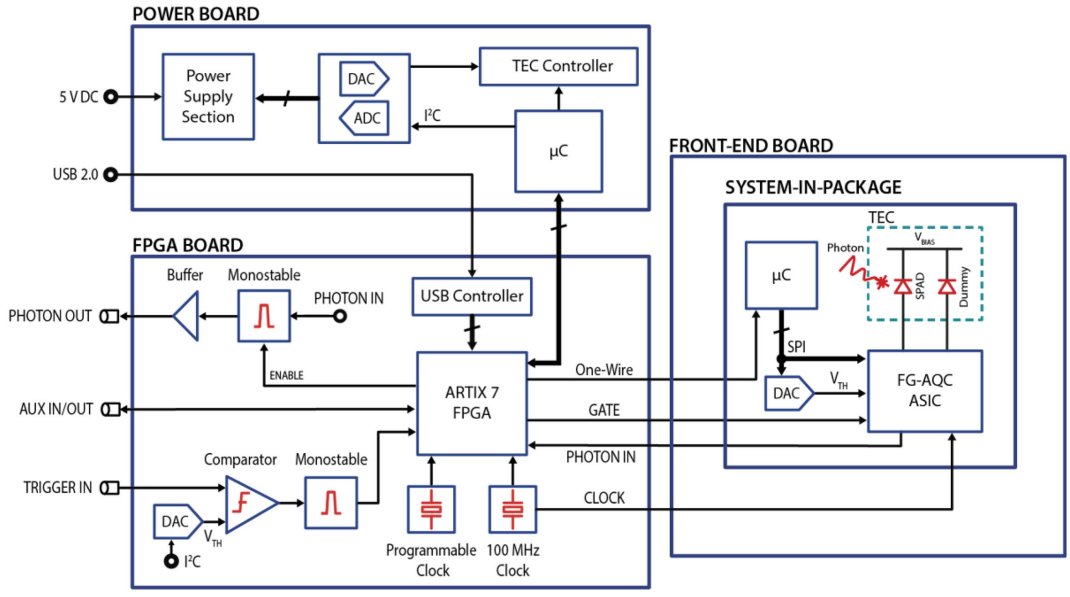


Fig. 1. Simplified block diagram of the module architecture. The System-in-Package (SiP), hosting the SPAD and its driving electronics, is sealed in a nitrogen environment and connected to a front-end board. A power board provides all the voltages needed by the instrument starting from an external 5 V DC supply. The local microcontroller is used to manage the power-on/off routine and drive the instrument I^2C bus. The custom FPGA board, based on an Artix 7 FPGA, manages the instrument timing signals and allows the user to tune measurement and SPAD parameters from an external computer via a USB 2.0 link. It also communicates with the microcontroller within the SiP and the other microcontroller on the power board. TRIGGER IN, PHOTON OUT, and AUX IN/OUT signals are used to interface the detector with standard TCSPC and timing instruments.

signals as well as a one-wire interface for programming the SiP configuration registers. The FPGA and the power board are also mutually interconnected. The communication with a computer host is handled (through a USB 2.0 controller) by the FPGA. Finally, the power board microcontroller, which communicates through a custom interface with the FPGA, drives the I^2C module bus and manages the instrument power up/down sequencing. The instrument is enclosed in a compact $5 \times 4 \times 8$ cm^3 aluminum case, for easy mounting in optical setups, and requires a single external power supply (5 V DC). The overall power consumption ranges from 3.5 to 5 W, depending on SPAD operating temperature and gating frequency.

2.1 System-in-Package

The SiP includes: i) a 50 μm CMOS SPAD [13] paired with a dummy device, mounted on top of a three-stage Thermo-Electric Cooler (TEC); ii) the FG-AQC ASIC [11], which drives SPAD and Dummy devices and detects the avalanche current; iii) a custom annular-shaped Printed Circuit Board (PCB), equipped with a low-pin-count microcontroller, a Digital-to-Analog Converter (DAC) and the FG-AQC. The package is sealed in a nitrogen environment to prevent moisture deposition. A picture of the built SiP is shown in Fig. 2.

The FG-AQC ASIC has been integrated inside the package close to the SPAD-dummy pair in order to minimize distortions of the gate edges (e.g. overshoots and ringing) due to parasitism, such as stray inductances of the TO-8 package pins and bonding-wires. The FG-AQC is wire-bonded directly to an annular-shaped PCB and to the SPAD-dummy pair, allowing to keep the length of critical wire-bonds within few millimeters.

The main drawback of this solution is the high number of connections required to drive the SPAD and the FG-AQC, which cannot be satisfied by standard 12-pin TO-8 packages. To solve

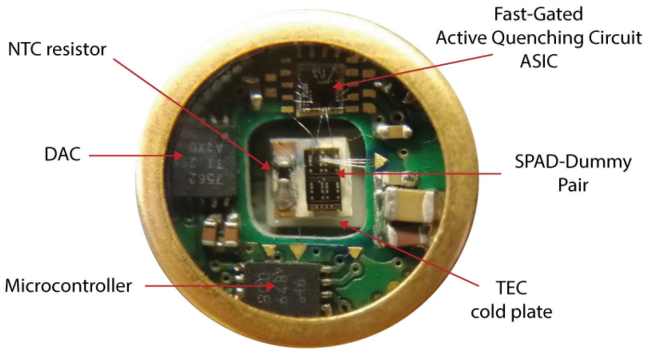


Fig. 2. Picture of the developed System-in-Package. The SPAD-dummy pair and the NTC resistor are mounted on top of the cold plate of a three-stage Peltier cell. The FG-AQC, wire-bonded to the SPAD-dummy pair, a microcontroller and a DAC are mounted on a custom PCB. The package is finally sealed in a nitrogen atmosphere.

this issue, two further ICs have been integrated in the SiP: i) a two-channel DAC is used to set the discriminator threshold voltages; ii) a microcontroller controls both the DAC and the FG-AQC through a local SPI bus according to the commands received from the FPGA, using a custom one-wire interface.

The employed SPAD was designed by Politecnico di Milano and exhibits a PDE equal to 55% at 450 nm and still 10% at 780 nm at 6 V excess bias, together with a DCR in the order of few counts per second when operated at 263 K inside the SiP [13]. The SPAD substrate is connected to ground (i.e. close to the anode voltage) in order to minimize the diffusion tail time-constant and to avoid folding effects in high count-rate applications. The width of the present SPAD depleted region can be modulated by means of the substrate voltage, thus affecting the collection of photo-generated carriers within the neutral region: decreasing the substrate voltage (towards the anode voltage) leads to a widening of the SPAD depleted region across the substrate-cathode junction, thus less photons are absorbed within the underlying neutral region, leading to a reduction of the exponential tail time-constant in the temporal response. On the other hand, this narrows the current path through the cathode neutral region, making it more resistive and leading to a worse SPTR [13], [14].

SPAD and dummy cathodes are connected to a common power supply, with voltage level equal to SPAD breakdown voltage plus the desired excess bias voltage. The anodes are instead actively driven by FG-AQC, in order to bias the SPAD above/below breakdown. The avalanche current is detected by the FG-AQC SiGe front-end comparator with less than 30 ps timing jitter [11]. Thanks to a 17-bit programmable counter in the FG-AQC ASIC, the SPAD hold-off time can be programmed from 50 ns up to 1.3 ms at 10 ns steps, starting from a 100 MHz reference clock (CLOCK in Fig. 1).

2.2 FPGA Board

The core of the FPGA board is an Artix 7 FPGA (Xilinx Inc.), which provides all the synchronization and timing signals required by the system: i) it generates and distributes the GATE signal starting from either internal or external triggering sources; ii) manages the one-wire communication with the SiP and the serial communication with the power board microcontroller; iii) provides the PHOTON OUT signal, synchronous to a photon detection, to the external acquisition electronics; iv) handles the USB 2.0 link through an FTDI interface, thus allowing to tune measurement parameters in real-time. A high precision 100 MHz clock source sets the time-base of the FPGA.

When a photon is detected, the FG-AQC provides at its output a positive voltage pulse with fixed temporal duration of approximately 4 ns. This signal is then buffered through a 1:2 fanout buffer: one copy is fed to the FPGA while the other is connected to a monostable circuit, whose enable input is controlled by the FPGA. The output of the monostable is finally translated into NIM

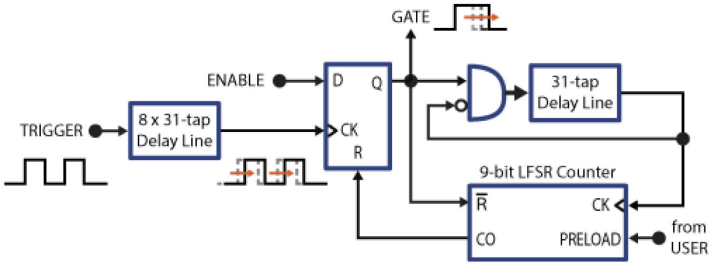


Fig. 3. Simplified schematic of the GATE generation circuit embedded in the FPGA. The 8 x 31-tap delay-line is used to delay the rising-edge of the TRIGGER signal. When a rising-edge is detected at clock input of the D Flip-Flop, its output Q goes high, marking the rising-edge of the GATE signal and enabling a ring-oscillator, whose oscillation period can be varied using a 31-tap delay-line within the ring-oscillator loop. A 9-bit LFSR counter, preloaded to a user-defined word, is used to count ring-oscillator periods. When the Carry Out (CO) signal goes high, it resets the D Flip-Flop, thus marking the falling-edge of the GATE signal, disabling the ring-oscillator and resetting the LFSR counter.

voltage levels and output through a 50 Ω SMA connector (PHOTON OUT in Fig. 1), compatible with most Time-Correlated Single-Photon Counting (TCSPC) commercial instrumentation. As the PHOTON IN signal is also provided to the FPGA, which enables the PHOTON OUT signal chain, this architecture allows for further firmware developments, which can be exploited for example to implement countermeasures against hacker attacks in QKD applications where active selection of "valid-detections" is required at high count rates [15], [16].

The GATE signal fed to the FG-AQC is generated through a specific circuit, based on a ring-oscillator, designed within the FPGA, starting from either an internal or an external signal:

- when the internal trigger is used, a low-jitter programmable crystal oscillator (output frequency from 100 kHz to 250 MHz) sets the GATE repetition frequency;
- when the external trigger configuration is adopted, a wide-bandwidth any-level comparator is used to detect the rising-edge of the signal applied to TRIGGER IN input of the module.

An auxiliary TTL-level input/output line (AUX IN/OUT) is provided through a 50 Ω SMA connector: when configured as an input, it can be used to provide a signal in logic combination with the TRIGGER IN input for specific gate patterns; when configured as an output, it can provide auxiliary signals, such as a copy of either the GATE signal applied to the SPAD, or the PHOTON OUT, etc. Additionally, the auxiliary output can be used for generating a VALID GATE signal that properly traces which gate pulse effectively enabled the SPAD, i.e., which are the pulses that were not blanked during the hold-off time.

2.3 GATE Generation Circuit

The GATE signal is internally generated within the FPGA, thanks to a dedicated circuit, starting from a selectable triggering source. The circuit is implemented in a single bank of the FPGA in order to guarantee low-jitter in GATE generation, both on GATE signal rising-edge and temporal duration. Low-jitter performance on gate rising-edge is of utmost important in many applications. For example, in QKD gate-matching between two or more detectors (operating in parallel) is required for avoiding hacker attacks [17], [18]. The main building block of the circuit is a programmable delay-line primitive of the Artix 7 family, called IDELAYE2 and made of 31 taps, each one introducing a nominal delay equal to 39, 52 or 78 ps, depending on the frequency of their reference clock [19]. In our implementation, the provided reference clock frequency is 400 MHz, obtained through an internal PLL from the FPGA 100 MHz reference clock, thus setting the nominal delay introduced by each tap equal to 39 ps.

A schematic of the GATE generation circuit is shown in Fig. 3: the first block is a cascade of eight delay-line primitives for fine-adjustment of GATE position, allowing to delay TRIGGER signal rising-edge from 0 up to 9.67 ns. The output of this first delay-line is fed to the clock input of a

D Flip-Flop, whose data input is driven by the FPGA: when D is low, Flip-Flop output Q is low independently of rising-edges at clock input, thus no GATE signal is provided to the SiP. Once D is driven high, it enables Q transitions, as a logic "1" can propagate at Flip-Flop output synchronous to the rising-edge of the delayed TRIGGER signal: this marks the rising-edge of the GATE signal. Q also enables a ring-oscillator made of a AND logic gate and a single delay cell primitive. At each rising-edge of ring-oscillator output corresponds an increase in the counts of the following 9-bit Linear Feedback Shift Register (LFSR) counter, starting from its initial value (PRELOAD). When the overflow is reached, the Carry Out (CO) signal resets the Flip-Flop, eventually forcing Q output low. Right after GATE falling-edge, the ring-oscillator is disabled and, thanks to the connection between Q and the reset input, the counter reloads its initial value and circuit operation is restored. An LFSR counter was specifically adopted in this design as the lower number of logic gates required by this architecture with respect to others (e.g. binary, BCD etc.) leads to a lower jitter and higher maximum input frequency.

This circuit allows to set the gate ON-time duration by modifying two separate parameters: on one side the oscillation period of the ring-oscillator (which is tuned by acting on the number of taps of the delay-line, N), while on the other the preload of the 9-bit LFSR counter. The ON-time of the GATE signal can be calculated as $T_{ON} = M{\cdot}2{\cdot}(N{\cdot}39\text{ ps} + T_{AND}) + T_R$, where M is equal to the number of counts of the 9-bit LFSR counter before overflow (thus 2^9–PRELOAD), T_{AND} is the propagation delay of the AND gate and T_R is the reset to output propagation delay of the D Flip-Flop. The minimum achievable GATE ON-time (with $M = N = 0$) is equal to 2.65 ns, which can be increased up to 2.5 μs. Concerning the minimum ON-time, it must be considered that the FG-AQC shortens the temporal duration of the applied GATE signal of about 2.5 ns, leading to a minimum enabling time of the SPAD (i.e. the optical GATE) of 150 ps, whereas SPAD peak efficiency is not reached in such short time.

The maximum GATE repetition frequency ensuring correct and stable operation of the instrument is 160 MHz.

2.4 Power Board

Starting from a single 5 V DC input, the power board provides all the power supplies of the instrument. A microcontroller is employed to guarantee correct system operation, as it manages the turn on/off procedure, communicates with the FPGA and controls peripherals on the I^2C bus: an integrated TEC controller drives the three-stage TEC mounted within the SiP, whereas a high-precision DAC drives a boost DC/DC converter for adjusting the SPAD bias voltage.

The SPAD excess bias voltage can be tuned between 2 and 6 V and the temperature can be adjusted down to 225 K, with the actual detector temperature being read back by a Negative-Temperature Coefficient (NTC) resistor, placed next to the SPAD-dummy pair on the TEC cold plate, and an Analog-to-Digital Converter (ADC).

The FPGA-based architecture and the high re-configurability of the instrument allow to tune the measurement parameters (i.e. measurement repetition frequency, gate ON-time and hold-off time) and SPAD operating conditions (temperature and excess bias) to the specific user application at any time during the measurement.

3. Experimental Characterization

We performed an extensive experimental characterization of the presented instrument: we firstly investigated jitter performance of the GATE generation circuit, both on rising-edge and temporal duration of the signal at FPGA output; DCR and afterpulsing probability at different operating temperatures have been then evaluated. Finally, we performed single-photon timing measurements both at low repetition frequency, in order to quantify detector SPTR and exponential decay time-constant, and at high gate repetition rates ($>$100 MHz) to prove the effectiveness of the presented instrument when high count rates are required.

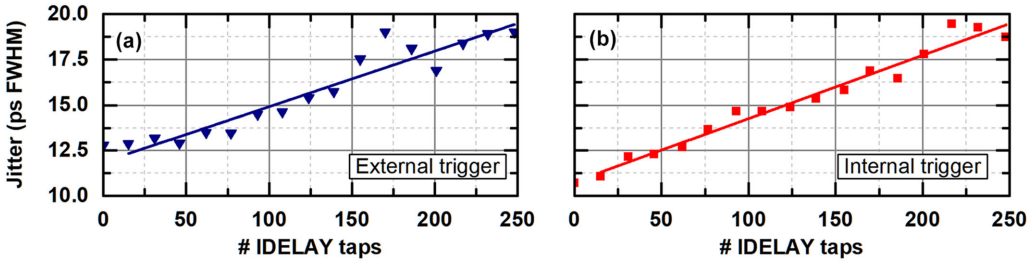


Fig. 4. Rising-edge jitter of the GATE signal generated starting from an external trigger source (a) and internal clock source (b). Nominal tap delay is 39 ps. Data include jitter contributions of both the detector and the TCSPC electronics.

Unless otherwise specified, measurements are performed with 5 V SPAD excess bias voltage and the gate pulse amplitude is 5.5 V. The SPAD breakdown voltage is 22.9 V at 263 K and 24.1 V at 293 K.

3.1 Gate Rising-Edge Jitter

The jitter of the rising-edge of the GATE signal, generated within the FPGA and provided to the FG-AQC, was measured at 100 kHz repetition rate in two different conditions: i) GATE generated starting from the external TRIGGER IN input, ii) GATE generated starting from the internal clock source. Both the measurement setups make use of a TCSPC module featuring a single-shot precision of 7 ps (FWHM). When the internal clock source is used, a copy of the internal programmable clock is provided at the AUX output of the module and used as TCSPC synchronization signal for the measurement, while in the external trigger configuration, the output of a high-precision clock source, with very low jitter, is split and fed to both the SPAD module as TRIGGER input and the TCSPC module for synchronization.

The results shown in Fig. 4 refer to the overall jitter, thus accounting contributions from both the SPAD system and the TSCPC module. When the delay introduced is short, the internal source signal chain (Fig. 4(b)) provides better performance with respect to the external one (Fig. 4(a)), with a contribution to the overall jitter of about 9 ps (FWHM). The slightly higher jitter on the external TRIGGER signal chain is mainly due to the front-end readout electronics, nevertheless its contribution becomes negligible when the delay introduced is longer than 1 ns, i.e. when more than one block of delay-line is used. The maximum overall jitter is lower than 20 ps (FWHM), and removing the contribution of the TCSPC electronics, the maximum jitter introduced by the delay-line is about 17.5 ps (FWHM). The overall jitter on the rising-edge of the GATE signal is therefore negligible for most applications, as it results to be lower than 10% of the GATE signal transition time, i.e. 280 ps (20–80%).

3.2 Gate Width Jitter

The jitter of the GATE width (i.e. the temporal duration of the GATE ON-time) is mainly caused by the jitter introduced by the ring-oscillator and the 9-bit LFSR counter. The GATE width jitter was measured through a TCSPC module by acquiring the temporal distribution between rising and falling edges of the GATE signal generated within the FPGA.

Considering the GATE generation circuit, it can be noticed that a specific ON-time can be obtained with different M and N (except for the extreme values of the ON-time range): for every count of the LFSR counter (M), N taps of the ring-oscillator delay-line are used. In Fig. 5(a) results of the jitter measurements when generating a fixed ON-time in two different conditions are shown: keeping $N = 0$ and $N = 31$. It is clearly visible that when using the full delay-line of the ring-oscillator,

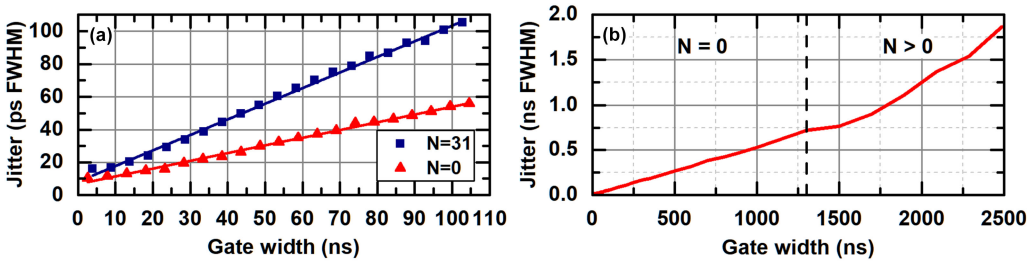


Fig. 5. (a) Gate width jitter comparison: per each GATE ON-time, configurations with $N = 0$ and larger M introduce the lowest jitter; (b) Gate width jitter performance in the 0–2.5 μs range. Data include jitter contributions of both the detector and the TCSPC electronics.

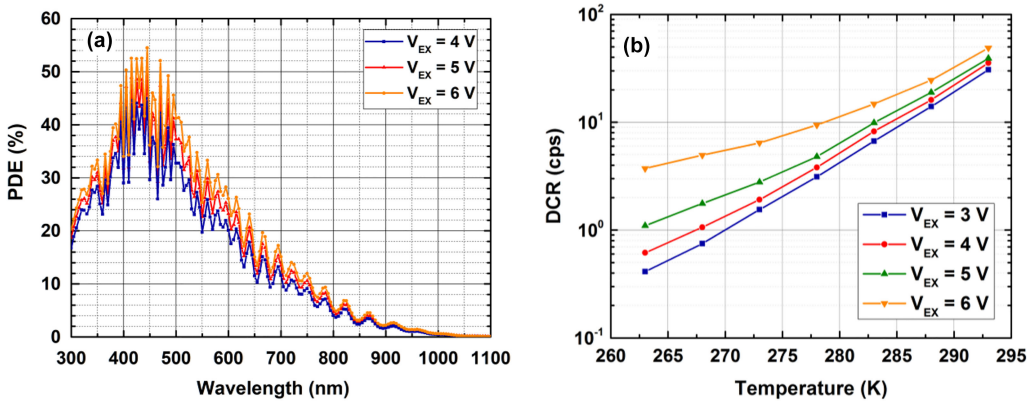


Fig. 6. (a) Detector photon detection efficiency (PDE) as a function of photon wavelength. (b) Detector dark-count rate (DCR) as a function of its operating temperature and excess bias voltage.

jitter contribution is almost double with respect to the case with $N = 0$. These results led to the development of a specific software algorithm used to retrieve the optimal values in order to keep N the lowest possible: jitter performance of the final implementation are shown in Fig. 5(b). It can be seen how for GATE ON-times longer than 1.3 μs, i.e. when the system starts to use delay-line taps within the ring-oscillator loop, jitter contribution starts to increase. It must be noticed that in standard NLOS imaging setups based on silicon SPADs, the GATE ON-time is generally in the order of few tens of nanoseconds, whereas in QKD and TR NIRS it can be even smaller, in the order of just few nanoseconds, thus the contribution of the GATE generation circuit to the ON-time jitter is negligible for most applications, as it results in about 10 ps (FWHM) for 10 ns ON-time and 55 ps for a 100 ns GATE window.

3.3 Photon Detection Efficiency and Dark-Count Rate

The PDE of the employed SPAD is reported in Fig. 6(a) for different excess bias voltages over the whole visible and near-infrared spectral range. A more detailed experimental characterization of this SPAD is reported in [13]. The measurement of primary DCR was carried out in gated-mode operation at 1 MHz gating frequency with a 990 ns ON-time, in order to operate the instrument almost in free-running mode, with a 99% duty cycle. SPAD hold-off time is 1 μs, in order to have a negligible afterpulsing contribution. Dark-count events were acquired through a frequency-counter over 10 minutes integration time per each measurement and properly normalized for duty cycle and acquisition time. Measurement was performed at four different excess bias voltages (3, 4, 5 and 6 V), in a temperature range from 263 to 293 K, at 5 K steps. Results are shown in Fig. 6(b): the

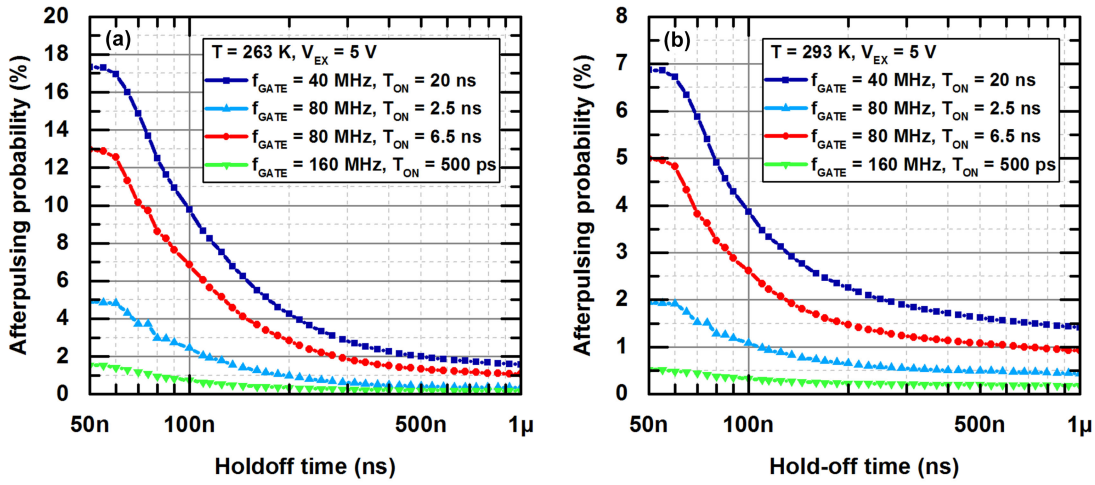


Fig. 7. Afterpulsing probability at two different SPAD operating temperatures: 263 K (a) and 293 K (b). Afterpulsing probability can be kept below 2% even when operating the SPAD at 263 K thanks to the adoption of sub-nanosecond gates.

instrument exhibits a DCR of only few counts per second at all excess bias voltages when operated at 263 K and still 50 cps at 293 K with 6 V excess bias voltage. The higher DCR at high excess bias and low temperature is due to a stronger contribution of the tunneling effect.

3.4 Afterpulsing Probability

We measured the detector afterpulsing probability at two different operating temperatures (263 and 293 K) by means of the Time-Correlated Carrier Counting (TCCC) technique, which consist in collecting a histogram of the inter-arrival times between two subsequent avalanche pulses [20] (results are reported in Fig. 7). Measurements were performed biasing the SPAD 5 V above breakdown in four different operating conditions, i.e. 20 ns gate ON-time at 40 MHz repetition frequency, 6.5 and 2.5 ns ON-time at 80 MHz and, finally, 500 ps ON-time at 160 MHz repetition frequency. Fig. 7(a) shows results obtained operating the SPAD at 263 K: afterpulsing probability is slightly higher than 17% with 20 ns ON-time at 40 MHz with 50 ns hold-off time, but can be reduced to 5% using 2.5 ns gates at 80 MHz and further decreased to just 1.5% with 500 ps gates at 160 MHz. Measurement results acquired at 293 K operating temperature are reported in Fig. 7(b), where afterpulsing probability is lower because the trap time-constant is inversely proportional to SPAD temperature: with 50 ns hold-off time it results lower than 7% with 20 ns gates at 40 MHz and equal to only 0.5% when driving the detector with 500 ps gates at 160 MHz. The obtained results highlight the possibility to adopt this detector in measurement setups where a high count-rate and low noise are required.

3.5 Short ON-Time Gates

Photon-detection uniformity within the optical gate is reported in Fig. 8. Fig. 8(a) reports a 20 ns gate acquired at 10 MHz repetition rate. A flat distribution is achieved after about 2 ns from gate opening, even though initial fluctuations remain lower than $\pm$10% with respect to asymptotic value. In Fig. 8(b) and Fig. 8(c) we report short ON-time gates (1 ns and 500 ps FWHM respectively) acquired at both 40 and 160 MHz repetition rate: same shape at both gating repetition frequency is observed. Despite a flat distribution is not achievable with short gates due to initial fluctuations

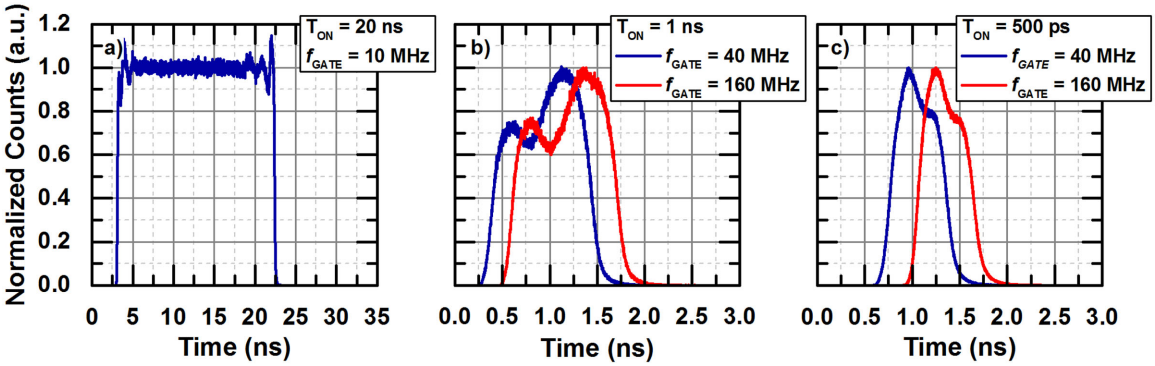


Fig. 8. Photon-counts distribution when the SPAD is illuminated with uncorrelated ambient light. (a) 20 ns ON-time at 10 MHz repetition frequency; (b) and (c) short ON-times both at 40 and 160 MHz GATE repetition rates (1 ns and 500 ps gate ON-time respectively). Red curves are shifted forward for visualization purpose.

(which cannot be completely avoided even with short wire-bondings between the SPAD and the FG-AQC), peak efficiency is achieved with gate ON-times as short as 500 ps (FWHM) [11]. Additionally, short gates are very important in QKD applications, where flat count distributions are not required, but low afterpulsing and high gate repetition frequency are critical. All the measurements reported in Fig. 8 were performed operating the SPAD at 283 K and no significant temperature dependence was observed.

3.6 Single-Photon Timing Resolution

Finally, we evaluated the SPTR of the detector using three different pulsed laser sources: i) a commercial pulsed laser source (MPL-820, Antel Optronics) at 830 nm, emitting 10 ps (FWHM) optical pulses to quantify SPTR, unfortunately limited at 100 kHz repetition rate and exhibiting a second “bump” after the main peak; ii) a commercial laser source (Advanced Laser Diode Systems) at 850 nm that offers 40 ps (FWHM) optical pulses, but with clean Gaussian shape for evaluating the exponential decay time-constant; iii) finally, we used a custom-made laser source at 780 nm providing 60 ps (FWHM) optical pulses up to 150 MHz repetition rate in order to assess timing performance when very short gates are adopted. All the measurements were performed operating the SPAD at 283 K, but same results hold at different temperatures, as SPAD timing performance are mainly affected by its excess bias voltage and by the photon wavelength. Fig. 9(a) shows the acquired waveform when the narrowest laser source is used: the overall timing jitter is equal to 58 ps (FWHM), which leads to a SPTR of the presented detector equal to 57 ps (FWHM), including contribution of both the SPAD (with 5 V excess bias voltage and substrate ground connected) and the FG-AQC. As the “bump” after the main peak is due to the laser source, in Fig. 9(b) we employed a different source emitting clean Gaussian shape laser pulses at 1 MHz repetition frequency, which leads to an estimated exponential decay time constant equal to 53 ps. In both Fig. 9(a) and Fig. 9(b) we set an optical ON-time equal to 15 ns. The worse SPTR with respect to state-of-the-art SPAD-based single-photon detectors (i.e. $<$50 ps FWHM) is mainly due to the increased timing jitter of the SPAD when its substrate is connected at about the same voltage of the anode (for faster tail, as discussed in 2.1) and the intrinsic timing jitter of the FG-ACQ front-end comparator. In Fig. 9(c) a TCSPC measurement at high repetition rates, adopting the 780 nm custom laser, is presented: removing the contribution of the laser source (i.e. 60 ps – FWHM), the SPTR degrades to 63 ps (FWHM) at 150 MHz repetition rate with 1 ns optical gate ON-time due to distortion on gate edges, with no photon accumulation at gate opening [11], differently from single-photon detectors with non-optimized time-gating schemes [21]–[23].

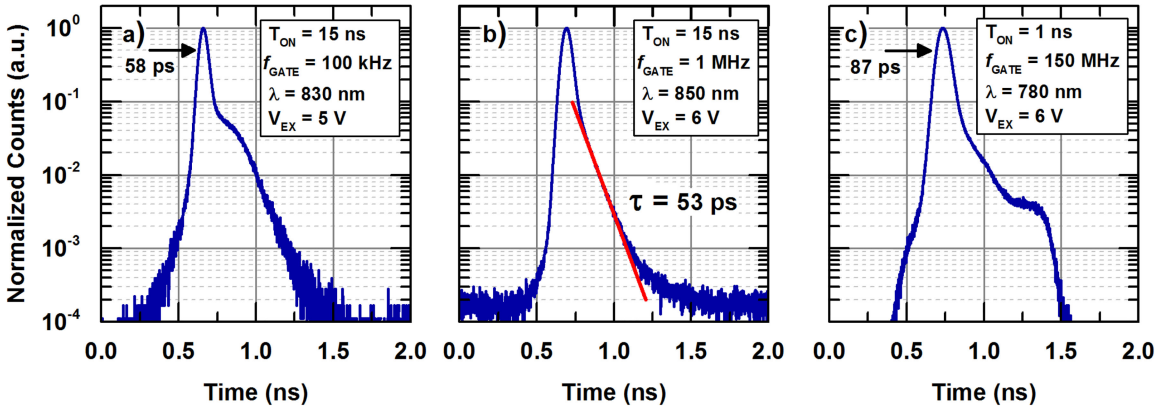


Fig. 9. Single-photon response of the presented detector at three different pulsed laser sources. In (a), a narrow laser pulse (10 ps FWHM) is used to evaluate detector SPTR. In (b), a pulsed laser source with no secondary emission is used to quantify the exponential decay time-constant of the detector, and in (c) a custom laser source emitting at 780 nm is employed to evaluate detector performance at high measurement repetition rates.

4. Conclusion

We presented a novel, fully-custom single-photon detection module based on a fast-gated CMOS SPAD. The instrument can be adopted in NLOS, TR NIRS, and QKD applications where a high detection rate is required, as the SPAD can be enabled with fast rising edges (less than 300 ps) at a repetition frequency up to 160 MHz and with very short ON-times, down to 500 ps. The detector provides a SPTR equal to 57 ps (FWHM), with the SPAD biased 5 V above breakdown, still guaranteeing a DCR equal to 1 cps at 263 K, and afterpulsing probability lower than 2% with only 50 ns hold-off time, thanks to the short-gate capability. Another version hosting an InGaAs/InP SPAD within the System-in-Package is currently under development.

References

[1] M. Buttafava, J. Zeman, A. Tosi, K. Eliceiri, and A. Velten, "Non-line-of-sight imaging using a time-gated single photon avalanche diode," *Opt. Express*, vol. 23, no. 16, pp. 20997–21011, 2015.
[2] S. Del Bianco, F. Martelli, and G. Zaccanti, "Penetration depth of light re-emitted by a diffusive medium: Theoretical and experimentaL investigation," *Phys. Med. Biol.*, vol. 47, no. 23, pp. 4131–4144, 2002.
[3] A. Dalla Mora *et al.*, "Fast-gated single-photon avalanche diode for wide dynamic range near infrared spectroscopy," *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, no. 4, pp. 1023–1030, 2010.
[4] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, "Quantum cryptography," *Rev. Mod. Phys.*, vol. 74, no. 1, pp. 145–195, 2002.
[5] F. Zappa, S. Tisa, A. Tosi, and S. Cova, "Principles and features of single-photon avalanche diode arrays," *Sens. Actuators, A*, vol. 140, pp. 103–112, 2007.
[6] A. Tosi *et al.*, "Fast-gated single-photon counting technique widens dynamic range and speeds up acquisition time in time-resolved measurements," *Opt. Express*, vol. 19, no. 11, 2011, Art. no. 10735.
[7] A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. G. Bawendi, and R. Raskar, "Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging," *Nat. Commun.*, vol. 3, 2012, Art. no. 745.
[8] Aurea Technology, "SPD_A_VIS VIS single photon detector datasheet," [Online]. Available: https://www.aureatechnology.com/images/produits/AUREA_Datasheet_SPD_A_VIS_V1.1_2018_light.pdf, Accessed on: Sep. 5, 2019.
[9] Micro Photon Devices, "FastGated SPAD datasheet," [Online]. Available: https://www.micro-photon-devices.com/MPD/media/Datasheet/FastGatedSPAD.pdf, Accessed on: Sep. 5, 2019.
[10] S. Cova, A. Lacaita, and G. Ripamonti, "Trapping phenomena in avalanche photodiodes on nanosecond scale," *IEEE Electron. Device Lett.*, vol. 12, no. 12, pp. 685–687, Dec. 1991.
[11] A. Ruggeri, P. Ciccarella, F. Villa, F. Zappa, and A. Tosi, "Integrated circuit for subnanosecond gating of InGaAs/InP SPAD," *IEEE J. Quantum Electron.*, vol. 51, no. 7, Jul. 2015, Art. no. 4500107.
[12] M. A. Albota and F. N. C. Wong, "Efficient single-photon counting at 1.55 μm by means of frequency upconversion," *Opt. Lett.*, vol. 29, 1449–1451, 2004.

[13] F. Villa *et al.*, "CMOS SPADs with up to 500 μm diameter and 55% detection efficiency at 420 nm," *J. Mod. Opt.*, vol. 61, no. 2, pp. 102–115, 2014.
[14] M. Buttafava, G. Boso, A. Ruggeri, A. Dalla Mora, and A. Tosi, "Time gated single-photon detection module with 110 ps transition time and up to 80 MHz repetition rate", *Rev. Sci. Instr.*, vol. 85, 2014, Art. no. 083114.
[15] C. Wiechers *et al.*, "After-gate attack on a quantum cryptosystem," *New J. Phys.*, vol. 13, 2011, Art no. 013043.
[16] N. Jain, B. Stiller, I. Khan, D. Elser, C. Marquardt, and G. Leuchs, "Attacks on practical quantum key distribution systems (and how to prevent them)," *Contemp. Phys.*, vol. 57, no. 3, pp. 366–387, Jul. 2016.
[17] B. Qi, C.-H. F. Fung, H.-K. Lo, and X. Ma, "Time-shift attack in practical quantum cryptosystems," *Quantum Inf. Comput.*, vol. 7, no. 1-2, pp. 73–82, Jan. 2007.
[18] Y. Zhao, C.-H. F. Fung, B. Qi, C. Chen, and H.-K. Lo, "Quantum hacking: Experimental demonstration of time-shift attack against practical quantum key distribution systems," *Phys. Rev. A*, vol. 78, no. 4, Oct. 2008, Art. no. 042333.
[19] Xilinx, "7 Series FPGAs SelectIO resources. user guide," [Online]. Available: https://www.xilinx.com/support/documentation/user_guides/ug471_7Series_SelectIO.pdf, Accessed on: Aug. 1, 2019.
[20] S. Cova, A. Lacaita, and G. Ripamonti, "Trapping phenomena in avalanche photodiodes on nanosecond scale," *IEEE Electron. Device Lett.*, vol. 12, no. 12, pp. 685–687, Dec. 1991.
[21] A. Rochas *et al.*, "ASIC for high-speed-gating and free running operation of SPADs," *Proc. SPIE, Photon Counting Appl., Quantum Opt., Quantum Cryptogr*, vol. 6583, pp. 65830F-1–65830F-10, May 2007.
[22] J. Zhang, R. Thew, J.-D. Gautier, N. Gisin, and H. Zbinden, "Comprehensive characterization of InGaAs–InP avalanche photodiodes at 1550 nm with an active quenching ASIC," *IEEE J. Quantum Electron.*, vol. 45, no. 7, pp. 792–799, Jul. 2009.
[23] A. Cominelli, G. Acconcia, I. Labanca, M. Ghioni, and I. Rech. "Accurate non-invasive measurement of the turn-on transition of fast gated single photon avalanche diodes," *Rev. Sci. Instrum.*, vol. 90, 2019, Art. no. 033102.