InGaAs/InP SPAD With Monolithically Integrated Zinc-Diffused Resistor

Mirko Sanzaro, *Student Member, IEEE*, Niccolò Calandri, *Student Member, IEEE*, Alessandro Ruggeri, and Alberto Tosi, *Member, IEEE*

***Abstract*— Afterpulsing and optical crosstalk are significant performance limitations for applications employing near-infrared single-photon avalanche diodes (SPADs). In this paper, we describe an InGaAs/InP SPAD with monolithically integrated resistor that is fully compatible with the planar fabrication process and provides a significant reduction of the avalanche charge and, thus, of afterpulsing and optical crosstalk. In order to have a fast SPAD reset (<50 ns), we fabricated quenching resistors ranging from 10 to 200 kΩ, smaller than what is available in the literature. The resistor, fabricated with the zinc diffusions already used for avoiding premature edge-breakdown, promptly reduces the avalanche current to a low value (∼100 μA) in less than 1 ns, while an active circuit completes the quenching and enforces a well-defined hold-off. The proposed mixed-quenching approach guarantees an avalanche charge reduction of more than 20 times compared with similar plain SPADs, enough to reduce the hold-off time down to 1 μs. Finally, a compact single-photon counting module based on this detector and featuring 70-ps photon-timing jitter is presented.**

***Index Terms*— Single-photon avalanche diode, InGaAs/InP, single photon, photon counting, near-infrared detector, afterpulsing, NFAD.**

I. Introduction

InGaAs/InP Single-Photon Avalanche Diodes (SPADs) today are employed in high-performance, compact and reliable instruments for single-photon counting applications in the near-infrared range (up to 1.7 μm). In the last years, many efforts were made to improve SPAD material quality for reducing noise [1]. However, afterpulsing, i.e. false counts arising from the emission of carriers that were trapped in deep levels during previous avalanche events [2], still limits the maximum count rate of such devices [3], a key feature for many applications, like quantum-key distribution (QKD) [4].

So far, the most effective approach for reducing afterpulsing is to limit the avalanche charge flow [5]. This strategy leads to less carrier trapping and has been successfully applied to gated mode operation by employing either fast quenching circuitry or sub-ns gate duration. Novel integrated quenching circuits demonstrated faster quenching time and, thus, lower avalanche charge compared to discrete components solutions [6], [7], consequently permitting to achieve reduced afterpulsing probability with shorter hold-off time. Additionally, gigahertz sinusoidal gating approaches have been widely exploited [8]–[10] to further reduce afterpulsing probability, enabling count rates of more than 100 Mcps. However, optical signals have to be synchronized to the gate peak and have a short duration (∼100 ps), because of the sub-ns gate-on time and the non-flat sensitivity. While this is compatible with QKD, in many other fields it is mandatory to reconstruct optical waveforms with much longer decay constants, e.g. in time-resolved spectroscopy of highly diffusing media [11].

An attractive solution to reduce the avalanche charge flow without the complexity of short gating techniques is to employ a monolithically integrated resistor in series with the SPAD (what is usually called integrated passive quenching). Concerning InGaAs detectors, this approach has been introduced by Itzler *et al.* for fabricating NFAD (negative-feedback avalanche diode) devices, by depositing high-value thin film resistors on top of a SPAD. In order to effectively achieve avalanche self-quenching, resistors in the MΩ range have been fabricated at the expense of a long (hundreds of ns) exponential recovery time after each avalanche [12]–[14]. These NFAD devices have been exploited in quantum key distribution [15] and singlet oxygen detection [16] applications, where they demonstrated low afterpulsing probability.

In this paper, we propose a novel implementation of quenching resistors, monolithically integrated with the InGaAs/InP SPADs. Neither additional process steps nor new masks are added to the standard fabrication process flow we used for our plain SPADs [17]. We designed and fabricated "lower-value" resistors, ranging from 10 to 200 kΩ, in order to have a faster recovery compared to NFADs, thus making gated-mode operation feasible (in addition to free-running), and limiting the effects of the non-constant detection efficiency and poorer timing resolution during recovery [12], [18]. We demonstrated that such low value resistor reduces most of the avalanche current, but is not enough for a complete avalanche quenching. Therefore, we introduced an external active circuit to completely quench the avalanche. Thanks to the sub-pF parasitic capacitance, the proposed mixed-quenching approach eventually results in a significant avalanche charge reduction compared to similar plain SPADs operated with external quenching circuits. Moreover, thanks

This work was supported by the Project EMPIR 14IND05 MIQC2 (the EMPIR initiative is Co-Funded by the European Union's Horizon 2020 Research and Innovation Programme and the EMPIR Participating States).

M. Sanzaro, N. Calandri, and A. Tosi are with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan 20133, Italy (e-mail: mirko.sanzaro@polimi.it; niccolo.calandri@polimi.it; alberto.tosi@polimi.it).

A. Ruggeri was with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan 20133, Italy. He is now with Micro Photon Devices, Bolzano 39100, Italy (e-mail: alessan-dro.ruggeri@polimi.it).

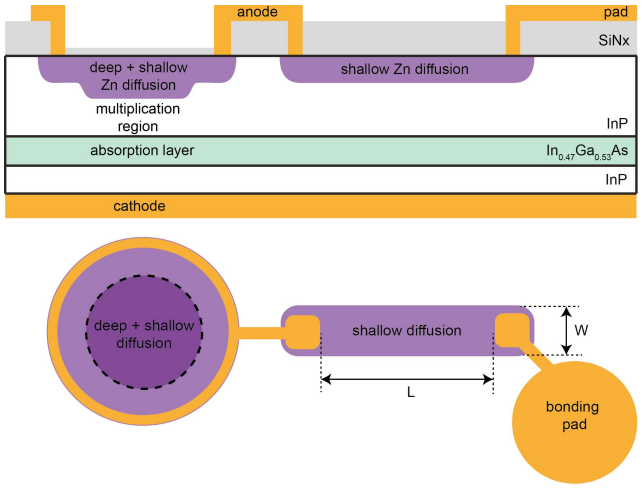

Fig. 1. Schematic representation of a SPAD connected to a diffused quenching resistor: cross-section (top) and top view (bottom). The connection between the SPAD anode and the resistor can be made by a metal strip (as shown here) or by just extending the SPAD shallow diffusion. The SPAD active area is highlighted by the dashed line.

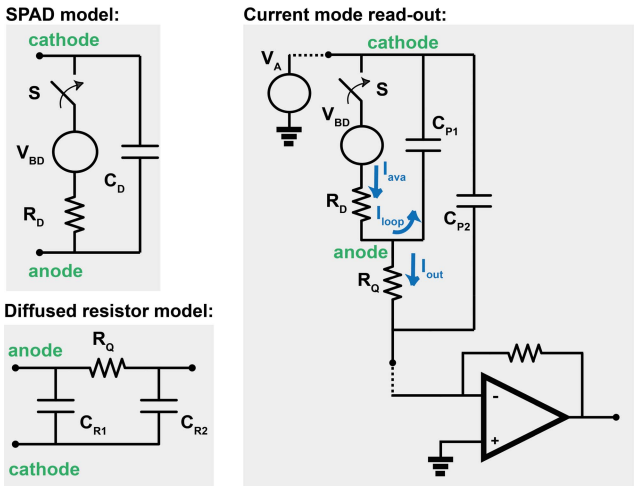

Fig. 2. Simplified model for the InGaAs/InP SPAD with its integrated quenching resistor. $C_{P1} = C_D + C_{R1}$ is the capacitance between anode and cathode and $C_{P2} = C_{R2} + C_{PAD}$ is the capacitance between bonding pad and cathode. The switch models the avalanche triggering and quenching, while the voltage source V_{BD} (breakdown voltage) and the dynamic resistance R_D (diode resistance) represent the avalanche current limitation due to space charge effects and series resistance.

to the external active circuit, a selectable well-defined hold-off can be enforced. Finally, in order to bring these new detectors out of the laboratories and exploit them in practical applications, a compact and reliable single-photon detection module has been developed. Except where otherwise stated, the developed module has been employed for the experimental characterizations presented in the following sections.

II. Device Structure and Modeling

In our work, passive quenching is applied through a diffused resistor, fully compatible with the fabrication process we use for our plain front-illuminated InGaAs/InP SPAD. In particular, in order to fabricate junction-isolated diffused resistors, we exploited the same shallow Zinc diffusion that avoids premature edge-breakdown effects. Fig. 1 shows a schematic illustration of the device cross-section and top-view. The diffused resistor is connected to the SPAD anode through a metal strip, while the other resistor terminal is connected to a bonding pad. Linear resistors of width W = 1 μm and aspect ratio L/W (Length/Width) ranging from 10 to 200 have been fabricated.

The overall equivalent circuit results from the combination of the InGaAs/InP SPAD model and of the diffused resistor model (see Fig. 2). For the SPAD, we used the basic model described in [19], while for the diffused resistor we chose a π-equivalent, which splits equally the distributed depletion layer capacitance in two lumped elements at the resistor ends.

The resistance R_Q between the two endpoints of the shallow diffusion has been measured from stand-alone resistor test structures from the same wafer. The resistance proved to be weakly dependent on its junction isolation depletion layer width, which in turn can be modulated by changing the cathode reverse voltage. In the measurement reported in Fig. 3, the bias at the cathode backside contact was kept 5 V above the breakdown voltage of the SPADs (which is about 60 V at 225 K), thus having the same depletion layer that is present when the SPAD is biased above breakdown. The resistance

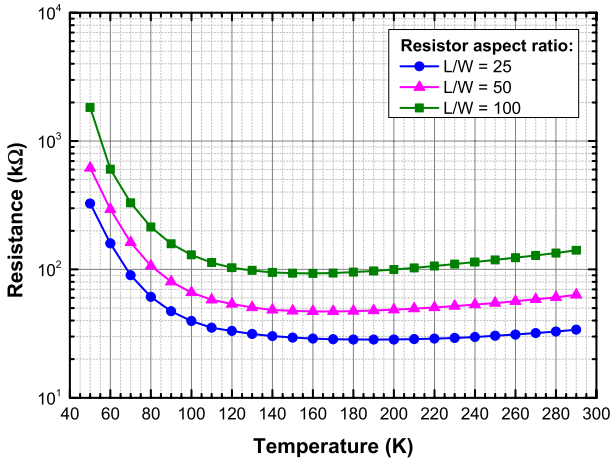

Fig. 3. Resistance as a function of temperature for different resistor aspect ratios (L/W = Length/Width). Measurements have been carried out on standalone resistor test structures from the same wafer. The applied bias at the backside cathode contact of the chip was 5 V higher than the breakdown voltage of the SPAD structure for each temperature.

has been measured in a cryostat from 290 K down to 50 K, taking into account the breakdown voltage dependence on temperature. Fig. 3 shows that the sheet resistance is about 1 kΩ/□ at 225 K and that the temperature coefficient is about 500 ppm/°C. From 140 K to 290 K, the sheet resistance increases with temperature because of the mobility reduction due to the higher interaction with acoustic phonon of the lattice. For temperatures lower than 100 K, freeze-out effects start to take place, increasing the sheet resistance abruptly.

The parasitic capacitance C_{P1}, which accounts for the SPAD depletion layer capacitance C_D, and the additional stray capacitance C_{R1}, of the diffused resistor, has been estimated from the device exponential self-recovery transient, after estimating the associated resistance with the measurements

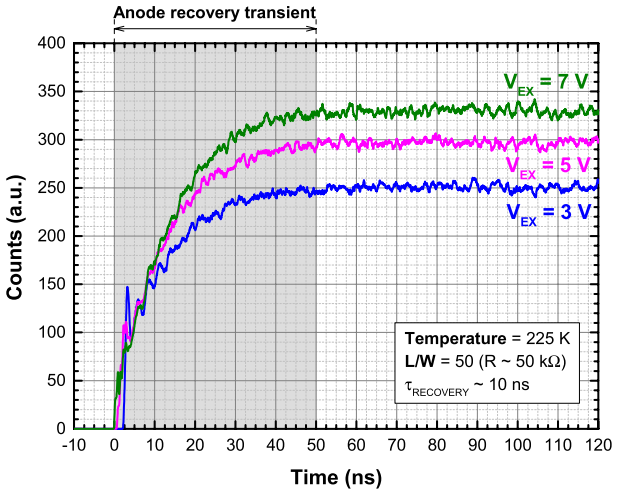

Fig. 4. Anode exponential recovery transient observed with a TCSPC setup. Device under test has been enabled for $T_{ON} = 200$ ns at 1 MHz repetition rate.

described above. Fig. 4 reports a Time-Correlated Single-Photon Counting (TCSPC) acquisition of the time-dependent sensitivity of a detector to a single-photon level CW illumination when gating the cathode voltage (V_A in Fig. 2) with a square-wave signal at 1 MHz and with 200 ns ON-time. The SPAD count rate was kept well below 10% of the gate frequency, so that, according to Poisson distribution, the probability of having more than one count per gate is negligible.

Since the SPAD detection efficiency is proportional to its excess bias, the counts distribution reproduces the voltage signal actually applied to the detector. As a result of the intrinsic low-pass filter made by the resistance R_Q in series with the capacitance C_{P1} (see the SPAD equivalent circuit in Fig. 2), the actual rise and fall times of the gating voltage are slowed down [18]. As shown in Fig. 4, with a series resistor having L/W = 50, the asymptotic excess voltage is reached exponentially with a time constant $\tau_R = R_Q C_{P1} \sim 10$ ns. Therefore, an overall anode stray capacitance $C_{P1} \sim 200$ fF can be estimated. As a comparison, Ref. [12] reports a 55 ns time constant for a 22 μm SPAD with a 600 kΩ series resistor, thus having a parasitic capacitance of $\sim$ 92 fF. Therefore, for NFADs the capacitive load is lower, because the oxide passivation thickness can be increased to reduce the parasitic capacitance of the thin film resistor without degrading other parameters of the underlying detector. In our implementation, instead, the capacitance per unit area of diffused resistors is set by the depletion layer thickness. Besides, guard rings are employed to prevent premature breakdown of the junction surrounding the resistor due to the higher curvature radius of the mask opening at resistor ends. As a result, the actual resistor capacitance is comparable to the SPAD one.

From I-V characteristics, the breakdown voltage of the diffused resistors was measured to be approximately 10 V higher than the one of the SPAD active area, which is defined by the presence of the Zinc deep diffusion (see Fig. 1). Therefore, it is possible to operate the detector with up to 10 V excess bias voltage.

Finally, a "dummy" diode, i.e. a SPAD with about 10 V higher breakdown voltage, has been fabricated in the same chip of devices with aspect ratio L/W = 50. Such a device, whose capacitance is close to the one of the actual detector, mimics the parasitic feedthrough of gate pulses, allowing a low-threshold differential sensing of the avalanche signal.

III. Avalanche Readout

In this work, we employed a current-mode approach for sensing the current flowing from the SPAD into the external circuit and marking an avalanche event. In particular, as shown in Fig. 2, a wideband transimpedance amplifier (TIA) sinks I_{out} and converts it into an amplified voltage signal that can be easily fed to a comparator.

During the initial transient just after the avalanche triggering (i.e. after closing the switch S in the equivalent circuit of Fig. 2), the avalanche current I_{ava} swiftly rises to its peak value, whose amplitude (a few mA) is given by the excess bias voltage $V_{EX} = V_A - V_{BD}$ divided by the SPAD's series resistance R_D [18]. Most of this current, I_{loop}, discharges the parasitic capacitance, C_{P1}. Since $R_D \ll R_Q$, the excess bias voltage exponentially decreases towards zero with a time constant approximately given by $\tau_Q \sim C_{P1} R_D$, i.e. less than 200 ps with $C_{P1} \sim 200$ fF (as estimated in Sect. II) and a SPAD internal resistance $R_D < 1$ kΩ. The current I_{out} flowing through the virtual ground exponentially increases up to V_{EX}/R_Q ($10 \div 100$ μA) with the same time constant τ_Q.

Even if most of the avalanche current pulse flows through C_{P1}, I_{out} can still be exploited for sensing the avalanche with low jitter, because its rise time ($\sim 2.2 \cdot \tau_Q$) is comparable with the typical avalanche-current build-up time [20]. However, the TIA closed loop bandwidth has to be high enough to not introduce a further significant low-pass filtering on the avalanche signal.

IV. Detection Module

In order to assess the performance of these new devices and exploit them in real world applications, a new compact single-photon detection module based on the above approach has been developed. Detector, front-end electronics and cooling system have been housed in a small case (7 cm $\times$ 5 cm $\times$ 4 cm), which, together with a low power dissipation ($\sim$ 5 W), makes it suitable for many applications. Moreover, all the operating parameters settings can be adjusted via USB interface.

A simplified block diagram of this module is sketched in Fig. 5. The detector (L/W = 50) and "dummy" device are housed in a hermetically sealed TO-8 package together with a three-stage thermoelectric cooler and a thermistor. A thermal feedback loop sets the detector operating temperature down to 220 K, thus reducing the dark counts down to few kcounts/s [17]. The module can operate in free-running mode with adjustable dead time, or in gated mode with adjustable gate duration, gate repetition rate and hold-off time. In the latter case, the trigger can be internally generated or taken from an external source.

The analog front-end is built on a couple of matched TIAs, based on two current-feedback amplifiers (CFAs). This differential architecture allows the low-jitter current-mode logic (CML) comparator to sense the avalanche signal with

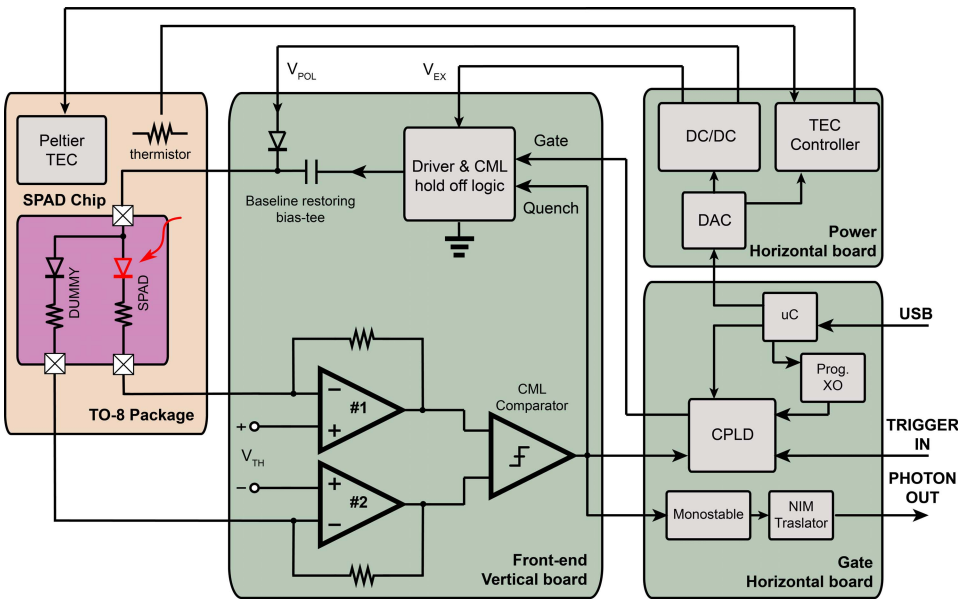

Fig. 5. The module developed for the SPAD with integrated resistor is composed of three boards. The power board hosts DC/DC converters and the TEC controller, which senses the thermistor mounted in the TO-8 package and drives the 3-stage thermoelectric cooler for setting the operating temperature. The gate board hosts the generation of gate, hold-off and photon-out signals, and adjusts all the operating parameters with commands received via the USB interface. The low-jitter front-end electronics and the gate driver are on the front-end board.

few millivolt threshold, while gate pulses are rejected, since they are common mode signals. The comparator threshold is set by offsetting the voltage at the non-inverting inputs of the two CFAs. The comparator output is split on three paths: one generates an output signal (Nuclear Instrumentation Module - NIM compatible pulse), named "photon output"; another one is used to quickly quench the SPAD through a CML AND logic gate that masks the gate signal; the last is fed to a counter (within a Complex Programmable Logic Device - CPLD) in order to generate the hold-off time. For biasing the detector, we employed a bias-tee in which the inductor has been replaced by a diode in order to overcome the drawbacks imposed by linear high-pass filtering [21]: i) the limitation on the maximum pulse duration, due to a variation of the bias voltage during the gating pulse; ii) the limitation on the maximum duty-cycle, due to the superposition of negative tails that build-up a negative baseline offset.

V. Electro-Optical Characterization

A. Avalanche Charge

In order to estimate intensity, shape, duration and its integral (i.e. the avalanche charge) of the avalanche current waveform, we measured the near-infrared electroluminescence emitted by the SPAD during an avalanche. Indeed, as discussed in [20], the SPAD electroluminescence is representative of the current flowing through the junction. Such faint emission is measured by and compared to that of a standard InGaAs/InP SPAD, operated by the module described in [22], optically coupled to the SPAD system under test. As described in [20], both gates of the modules have been synchronized to the same reference signal. The output pulses of the two modules are the START and STOP signals of a time-to-amplitude converter (TAC) that acquires the time between a photon detection by the first module and another one by the second module. The STOP signal was delayed of one gate-ON time in order to be able to acquire also when the STOP signal is generated before the START one.

Both the standard module and the system under test mount 25 μm SPADs from the same wafer [17]. Therefore, the only difference is the quenching circuit, thus making relative comparisons straightforward. Moreover, since the operating conditions are the same in terms of excess-bias voltage and temperature (233 K), the internal electric field, detection efficiency and count rate of the two detectors are very similar, apart from very small across-wafer variations due to process tolerances (breakdown voltage non-uniformity is less than 2 % across the wafer). Therefore, a direct comparison of the acquired waveforms is possible, after subtracting the background due to uncorrelated events. Being the convolution of the uniform dark count distribution within the two square gate windows of the two SPAD, such background has a triangular shape and can be easily fitted and subtracted from raw data.

As shown in Fig. 6, the peak luminescence (and thus the peak current) is reduced by a factor of 3.3, 5.7 and 7.2 for 3, 5 and 7 V excess bias voltages respectively (corresponding to a PDE of 19.5%, 29% and 36% at 1550 nm [17]). Moreover, the avalanche current duration is reduced in the module with integrated resistor, from approximately 5 ns down to 1 ns. The integral of the curves of Fig. 6 is proportional to the avalanche charge and the measured charge ratio between the two modules is 21.3, 45.8 and 47.3 for an excess bias of 3, 5 and 7 V, respectively.

B. Optical Crosstalk

We also investigated the impact of the integrated quenching resistor on the optical cross-talk between neighboring pixels of

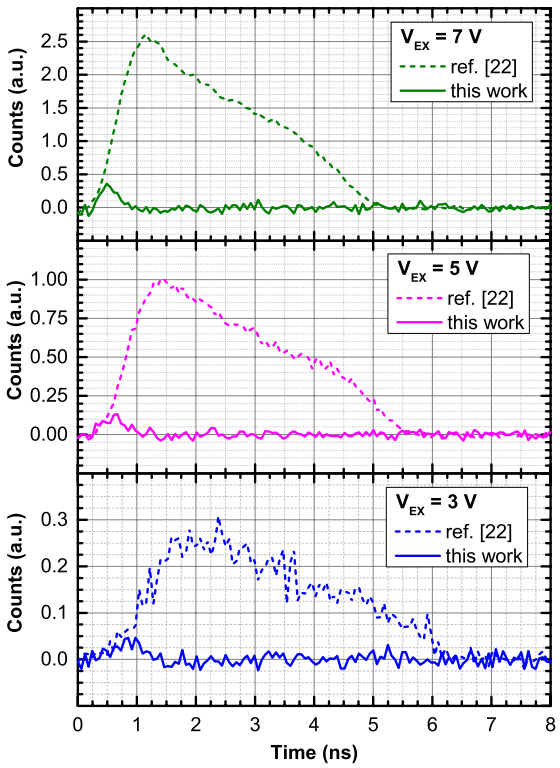

Fig. 6. Electroluminescence waveforms acquired from the modules based on standard SPAD (dashed line) and on the one with integrated resistor (solid line). Measurement time is 12 hours for each curve.

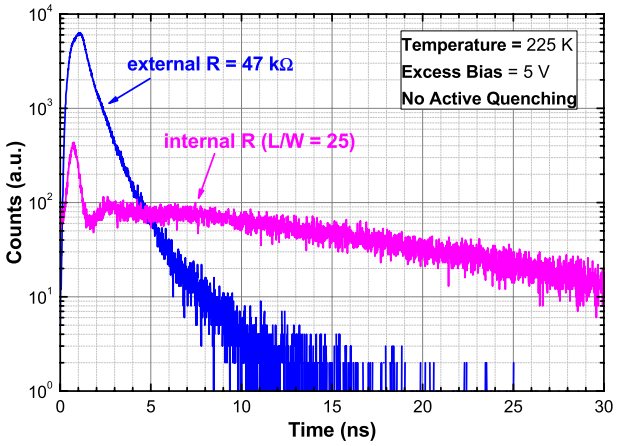

Fig. 7. Distribution in time of optical cross-talk events between neighboring pixels with integrated and external quenching resistor. The two couples of SPADs are both in the same chip at a distance of 120 μm. The integrated series resistance has an aspect ratio L/W = 25 (~30 kΩ at 225 K), while the external one is 47 kΩ. Since this integrated resistor is not high enough to completely quench the avalanche, a slow tail due to the steady-state current electroluminescence can be observed after the fast (~1 ns) initial transient. Uncorrelated events are very few compared to correlated ones.

a linear SPAD array. We compared the optical cross-talk of two standard SPAD pixels (quenched by external resistors) with two SPADs with integrated resistor (same active area diameter of 25 μm, same distance of 120 μm and comparable DCR). For each gate in which both pixels fired, we measured the time interval between the two clicks, eventually reconstructing the histogram of Fig. 7.

Since no active quenching has been employed with arrays and the internal resistor is not sufficient to quench the avalanche, a steady-state current (V_{EX}/R_Q = 167 μA) keeps flowing after each avalanche until the gate trailing edge, as we verified acquiring the output of the TIA. This constant current produces a weak but constant flux of emitted photons collected by the adjacent pixel, which is responsible for the slow "tail" in Fig. 7. The exponential decay after the current peak is typical of passive quenching and is a direct consequence of the exponential quenching transient. As can be seen from Fig. 6 and 7, most of the quenching transient completes in about 1 ns when using internal resistors, while several ns are needed with external quenching.

Since the correlated counts due to the optical crosstalk dominate over the uncorrelated dark counts and the sample-to-sample DCR variation of the four SPADs here employed is small, the integrals of the histograms reported in Fig. 7 can be directly used for calculating the reduction of the optical crosstalk with integrated resistors, which results to be about 4.2 times.

Such cross-talk reduction can be exploited for designing high-density arrays. However, the actual layout (see the top view of Fig. 1) is suitable only for single pixels and linear arrays, for which the resistor on a side is not an issue. In order to fabricate high-density focal plane arrays, the resistor area has to be reduced, e.g. by narrowing the strip while keeping the same number of squares. However, tolerances in the fabrication may set a lower limit to the width of the resistive strip for matching the resistance value of all the pixels. Alternatively, the resistor can be arranged along the SPAD perimeter.

C. Afterpulsing

Afterpulsing probability was measured with a setup based on an FPGA (Field Programmable Gate Array) that time-stamps each avalanche event, when the module is kept in dark conditions. After collecting a large number of avalanche interarrival times, we created a histogram with one gate period quantization. The probability of having an afterpulse in a gate with a delay $\Delta t = n \cdot T_{GATE}$ from the first count can be calculated by: i) dividing the counts of each bin of such histogram by the total number of collected events; ii) subtracting the events due to primary dark counts. This latter is obtained by subtracting the exponential fit of the acquired curve (given the Poisson distribution of dark counts interarrival times), performed for $\Delta t > 200$ μs (i.e. where the afterpulsing probability is negligible). Dividing the result by the effective gate width (in nanoseconds), we obtained the afterpulsing probability density in 1 ns reported in Fig. 8.

With the same procedure, the same operating conditions and two detectors from the same wafer [17], we compared the newly developed module with the system presented in [22]. Fig. 8 shows that the SPAD with integrated resistor allows to reduce the hold-off from 10 μs to 2 μs for the same afterpulsing probability.

Another way to estimate the maximum gate repetition rate is by plotting the dark-count rate as a function of hold-off time. We performed these measurements at three different excess bias voltages (2.5, 5 and 7 V) for both modules: at short hold-off times, afterpulsing increases the DCR (see Fig. 9). The comparison confirms the lower afterpulsing of the new

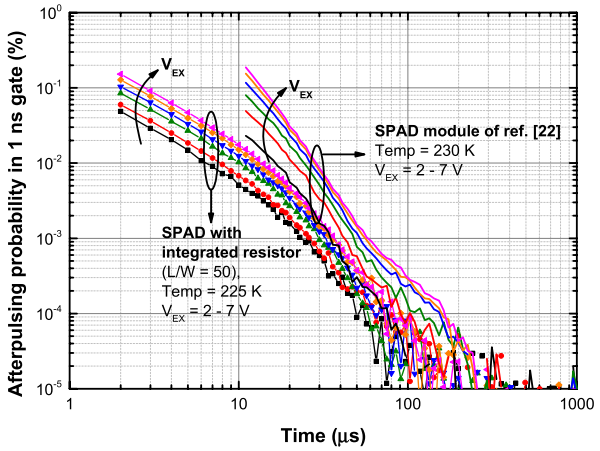

Fig. 8. Dependence of afterpulsing probability in 1 ns on the time after the previous avalanche for the presented module and the discrete component one of [22], at different excess bias values.

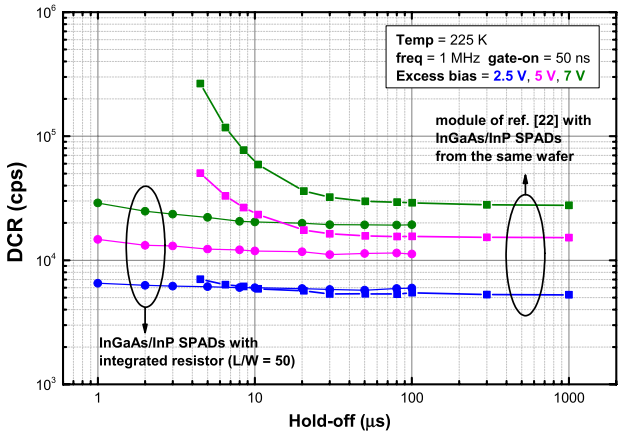

Fig. 9. DCR as a function of the hold-off time for the presented module and the discrete component one of [22] at different excess bias values. The increment of DCR at short hold-off times is due to afterpulsing.

module. In particular, for each excess bias voltage, the hold-off time corresponding to the knee in the curve shifted from tens of μs for the standard plain devices to few μs for the ones with integrated resistors. Moreover, the DCR increase is much less sharp. This further confirms that the integrated fast quenching significantly reduces the afterpulsing effect, thus allowing higher photon counting rates.

D. Timing Jitter

The temporal response of the system has been assessed employing a picosecond pulsed laser at 1550 nm (pulse width is 20 ps FWHM) that uniformly illuminates the active area of the device and the time-correlated single-photon counting (TCSPC) setup already used in [7]. Fig. 10 shows the distribution of photon arrival times when using a SPAD with an integrated resistor having L/W $=$ 50 and excess bias voltages of 3, 5 and 7 V. For $V_{EX} \geq 5$ V the second peak (more than two decades lower than the main one) is not present and the timing response can be fitted as the sum of a sharp Gaussian distribution and a fast exponential tail, as

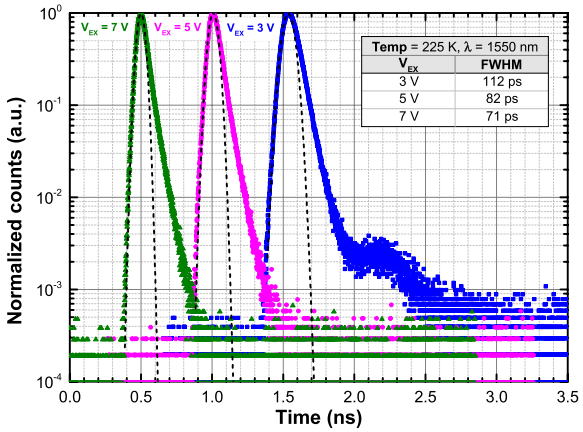

Fig. 10. Temporal response of the presented module based on SPAD with integrated resistor with an excess bias of 3, 5 and 7 V and a resistor with aspect ratio L/W $=$ 50, when illuminated by a narrow ($\sim$20 ps) pulsed laser at 1550 nm. The response closely approaches a Gaussian distribution (standard deviation of 27, 33 and 42 ps at V_{EX} of 7, 5, 3 V), plus an exponential tail (time constant of about 50 ps for all the curves).

reported in [17] for standard InGaAs/InP SPADs from the same wafer. In particular, with an excess bias of 5 V and at a temperature of 225 K, the distribution has a full-width at half maximum (FWHM) well below 100 ps and a full-width at 1/1000 of maximum of about 450 ps. Such characteristics are very important in applications where optical waveforms have to be acquired with high time resolution and with very wide dynamic range.

Despite the low-pass filtered output current flowing in the external readout electronics, these results are fairly consistent with those already reported in [7], where the temporal response of standard InGaAs/InP SPADs from the same wafer have been measured using both a fast-gating AQC chip and the discrete components module of [22].

VI. Conclusions

In this paper, we have discussed about design and characterization of InGaAs/InP SPADs with monolithically integrated quenching resistors.

We pursued a novel implementation of the integrated quenching resistor, which does not require neither additional process steps (like material depositions) nor masks, for guaranteeing easy processing and good yield. In detail, we exploited the shallow Zinc diffusion, already present in the SPAD, to fabricate a diffused resistor. Compared to NFADs, we designed smaller resistor values (tens of kΩ) in order to have shorter recovery times (tens of ns) for making gated mode feasible, despite the inherently higher capacitive load. Given the low-value integrated resistors, an external quenching circuit is needed to enforce avalanche quenching, thus limiting the exploitability of this solution for large focal plane arrays.

For the experimental characterization, we developed a custom front-end electronics, which employs a couple of broadband and matched trans-impedance amplifiers, a low jitter comparator and an active quenching circuit. Detector, front-end electronics and cooling system have been housed in a small case, which is ready for practical applications.

We demonstrated that the proposed mixed quenching approach guarantees a strong avalanche charge reduction. The comparison of the electroluminescence emission of these devices with standard SPADs operated with external quenching pointed out that the avalanche current peak is reduced by more than 3 times, while the total avalanche charge is lowered by more than 20 times. Accordingly, the measured afterpulsing probability is also strongly reduced, permitting gated mode operation with hold-off times in the μs range. Furthermore, the high speed and low jitter comparator makes it possible to achieve timing jitter much lower than 100 ps and the fast exponential tail in the timing response allows extremely wide dynamic ranges in time-correlated single photon counting measurements.

ACKNOWLEDGEMENTS

The authors wish to thank Franco Zappa and Fabio Acerbi for their useful discussions and contributions.

REFERENCES

[1] M. A. Itzler *et al.*, "Advances in InGaAsP-based avalanche diode single photon detectors," *J. Mod. Opt.*, vol. 58, nos. 3–4, pp. 174–200, 2011.
[2] 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.
[3] M. A. Itzler, X. Jiang, and M. Entwistle, "Power law temporal dependence of InGaAs/InP SPAD afterpulsing," *J. Mod. Opt.*, vol. 59, no. 17, pp. 1472–1480, Oct. 2012.
[4] H.-K. Lo, M. Curty, and K. Tamaki, "Secure quantum key distribution," *Nature Photon.*, vol. 8, no. 8, pp. 595–604, Jul. 2014.
[5] C. Hu, X. Zheng, J. C. Campbell, B. M. Onat, X. Jiang, and M. A. Itzler, "Characterization of an InGaAs/InP-based single-photon avalanche diode with gated-passive quenching with active reset circuit," *J. Mod. Opt.*, vol. 58, nos. 3–4, pp. 201–209, 2011.
[6] 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.
[7] 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.
[8] A. Restelli, J. C. Bienfang, and A. L. Migdall, "Single-photon detection efficiency up to 50% at 1310 nm with an InGaAs/InP avalanche diode gated at 1.25 GHz," *Appl. Phys. Lett.*, vol. 102, no. 14, p. 141104, 2013.
[9] N. Namekata, S. Adachi, and S. Inoue, "1.5 GHz single-photon detection at telecommunication wavelengths using sinusoidally gated InGaAs/InP avalanche photodiode," *Opt. Exp.*, vol. 17, no. 8, pp. 6275–6282, 2009.
[10] C. Scarcella, G. Boso, A. Ruggeri, and A. Tosi, "InGaAs/InP single-photon detector gated at 1.3 GHz with 1.5% afterpulsing," *IEEE J. Sel. Topics Quantum Electron.*, vol. 21, no. 3, May/Jun. 2015, Art. no. 3800306.
[11] I. Bargigia *et al.*, "Time-resolved diffuse optical spectroscopy up to 1700 nm using a time-gated InGaAs/InP single-photon avalanche diode," *Proc. SPIE*, vol. 8090, pp. 80900U-1–80900U-5, Nov. 2011.
[12] X. Jiang, M. A. Itzler, K. O'Donnell, M. Entwistle, and K. Slomkowski, "Shortwave infrared negative feedback avalanche diodes and solid-state photomultipliers," *Opt. Eng.*, vol. 53, no. 8, p. 081908, 2014.
[13] X. Jiang *et al.*, "InP-based single-photon detectors and geiger-mode APD arrays for quantum communications applications," *IEEE J. Sel. Topics Quantum Electron.*, vol. 21, no. 3, May/Jun. 2015, Art. no. 3800112.
[14] B. Korzh, T. Lunghi, K. Kuzmenko, G. Boso, and H. Zbinden, "Afterpulsing studies of low-noise InGaAs/InP single-photon negative-feedback avalanche diodes," *J. Mod. Opt.*, vol. 62, no. 14, pp. 1151–1157, 2015.
[15] B. Korzh *et al.*, "Provably secure and practical quantum key distribution over 307 km of optical fibre," *Nature Photon.*, vol. 9, no. 3, pp. 163–168, 2014.
[16] G. Boso, B. Korzh, T. Lunghi, B. Sanguinetti, and H. Zbinden, "Low noise InGaAs/InP single-photon detector for singlet oxygen detection," *Proc. SPIE*, vol. 9370, pp. 93701S-1–93701S-8, Feb. 2015.
[17] A. Tosi, N. Calandri, M. Sanzaro, and F. Acerbi, "Low-noise, low-jitter, high detection efficiency InGaAs/InP single-photon avalanche diode," *IEEE J. Sel. Topics Quantum Electron.*, vol. 20, no. 6, Nov./Dec. 2014, Art. no. 3803406.
[18] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, "Avalanche photodiodes and quenching circuits for single-photon detection," *Appl. Opt.*, vol. 35, no. 12, pp. 1956–1976, 1996.
[19] A. Dalla Mora, A. Tosi, S. Tisa, and F. Zappa, "Single-photon avalanche diode model for circuit simulations," *IEEE Photon. Technol. Lett.*, vol. 19, no. 23, pp. 1922–1924, Dec. 1, 2007.
[20] F. Acerbi, A Della Frera, A. Tosi, and F. Zappa, "Fast active quenching circuit for reducing avalanche charge and afterpulsing in InGaAs/InP single-photon avalanche diode," *IEEE J. Quantum Electron.*, vol. 49, no. 7, pp. 563–569, Jul. 2013.
[21] S. J. Dimler, J. S. Ng, R. C. Tozer, G. J. Rees, and J. P. R. David, "Capacitive quenching measurement circuit for geiger-mode avalanche photodiodes," *IEEE J. Sel. Topics Quantum Electron.*, vol. 13, no. 4, pp. 919–925, Jul. 2007.
[22] A. Tosi, A Della Frera, A. Bahgat Shehata, and C. Scarcella, "Fully programmable single-photon detection module for InGaAs/InP single-photon avalanche diodes with clean and sub-nanosecond gating transitions," *Rev. Sci. Instrum.*, vol. 83, no. 1, p. 013104, Jan. 2012.

Mirko Sanzaro was born in Avola, Italy, in 1988. He received the bachelor's degree in electronics engineering and the M.Sc. (*summa cum laude*) degree in electronics engineering from the Politecnico di Milano, in 2010 and 2013, respectively, where he is currently pursuing the Ph.D degree in electronics engineering. His current research activity aims at the design, development and characterization of visible and near-infrared single-photon avalanche diodes.

Niccolò Calandri was born in Como, Italy, in 1989. He received the B.Sc. degree in biomedical engineering and the M.Sc. (*summa cum laude*) degree in electronics engineering with the Politecnico di Milano, in 2011 and 2013, respectively, where he is currently pursuing the Ph.D. degree in electronics engineering. He is currently involved in the design and characterization of InGaAs/InP single-photon avalanche diodes with innovative structures, exploiting linear and bi-dimensional arrays.

Alessandro Ruggeri was born in Alzano Lombardo, Italy, in 1987. He received the master's degree in electronics engineering and the Ph.D. degree in information technology from the Politecnico di Milano, Milan, Italy, in 2012 and 2015, respectively.

He was with the IBM T. J. Watson Research Center, Yorktown Heights, NY, in 2014, where he was involved in optical testing of ULSI circuits. He is currently a Research and Development Electronic Circuit Designer with Micro Photon Devices, Bolzano, Italy. His main research activity deals with the design and development of electronics for visible and near-infrared single photon avalanche diodes, for biomedical and communications applications.

Alberto Tosi (M'07) was born in Borgomanero, Italy, in 1975. He received the master's degree in electronics engineering and the Ph.D. degree in information technology engineering from the Politecnico di Milano, Milan, Italy, in 2001 and 2005, respectively. He was an Assistant Professor from 2006 to 2014.

He has been an Associate Professor of Electronics with the Politecnico di Milano since 2014. In 2004, he was a Student with the IBM T. J. Watson Research Center, Yorktown Heights, NY, involved in optical testing of CMOS circuits. He is currently involved in silicon and InGaAs/InP single-photon avalanche diodes. His research activity includes arrays of silicon SPADs for 2-D and 3-D applications and time-correlated single-photon counting electronics.