AIP | Review of Scientific Instruments

Time-gated single-photon detection module with 110 ps transition time and up to 80 MHz repetition rate

Mauro Buttafava, Gianluca Boso, Alessandro Ruggeri, Alberto Dalla Mora, and Alberto Tosi

Citation: Review of Scientific Instruments **85**, 083114 (2014); doi: 10.1063/1.4893385

View online: http://dx.doi.org/10.1063/1.4893385

View Table of Contents: http://scitation.aip.org/content/aip/journal/rsi/85/8?ver=pdfcov

Published by the AIP Publishing

Time-gated single-photon detection module with 110 ps transition time and up to 80 MHz repetition rate

Mauro Buttafava,[1,a)] Gianluca Boso,[1] Alessandro Ruggeri,[1] Alberto Dalla Mora,[2] and Alberto Tosi[1]
[1]*Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria, Piazza Leonardo Da Vinci 32, 20133 Milano, Italy*
[2]*Politecnico di Milano, Dipartimento di Fisica, Piazza Leonardo Da Vinci 32, 20133 Milano, Italy*

(Received 14 July 2014; accepted 6 August 2014; published online 22 August 2014)

We present the design and characterization of a complete single-photon counting module capable of time-gating a silicon single-photon avalanche diode with ON and OFF transition times down to 110 ps, at repetition rates up to 80 MHz. Thanks to this sharp temporal filtering of incoming photons, it is possible to reject undesired strong light pulses preceding (or following) the signal of interest, allowing to increase the dynamic range of optical acquisitions up to 7 decades. A complete experimental characterization of the module highlights its very flat temporal response, with a time resolution of the order of 30 ps. The instrument is fully user-configurable via a PC interface and can be easily integrated in any optical setup, thanks to its small and compact form factor. © *2014 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4893385]

I. INTRODUCTION

In the last years an increasing number of applications require the use of photodetectors able to acquire light signals down to single-photon level with high temporal resolution, extremely low noise and extended Dynamic Range (DR). Some examples are Time-Correlated Single-Photon Counting (TCSPC) applications, like optical time-resolved spectroscopy for functional brain imaging[1–3] and mammography,[4] or fluorescence life-time microscopy,[5,6] ultra-fast time-of-flight imaging,[7] molecular imaging,[8] and quantum information.[9]

Single-Photon Avalanche Diodes (SPADs)[10] are becoming increasingly widespread for the detection of faint light signals, due to their high Photon Detection Efficiency (PDE), low noise, high timing resolution, and good reliability. A SPAD is basically a p-n junction reverse biased above its breakdown voltage (by an amount called excess bias), in which the electric field within the active area is sufficiently high to allow the triggering of a self-sustaining impact ionization process (avalanche), even starting from a single photo-generated charge carrier.[10]

The result is a macroscopic current pulse, whose leading edge marks the arrival time of any absorbed photon, with tens of picoseconds precision. Thin-junction silicon SPADs exhibit PDE of about 60% at 550 nm, 20% at 800 nm, and still 4% at 950 nm, with timing jitter on the photon arrival time of less than 40 ps.[11] Carrier thermal generation can also trigger an avalanche pulse without any absorbed photon and this effect is quantified by the Dark Count Rate (DCR), in terms of counts-per-second. DCR is the main noise contribution in a silicon SPAD detector and for a typical 100 μm active-area diameter device can be of the order of few kcounts/s, decreasing with temperature and applied excess bias voltage.[10]

In addition, carriers trapped in deep levels of the semiconductor during an avalanche current pulse can be released after a certain time delay giving rise to other spurious ignitions, which are also correlated with the signal itself. This phenomenon is known in literature as afterpulsing effect.[10] To lower the probability of an afterpulse event, the SPAD must be kept OFF for a certain amount of time after each avalanche event (hold-off condition). For silicon SPADs, the typical hold-off time needed to keep the afterpulsing probability lower than 1% is in the range from 50 ns to 150 ns.[11]

After every avalanche event (either due to a photon absorption or to a thermally generated carrier) the self-sustaining multiplication process must be quenched by lowering the bias voltage below its breakdown value (quenching phase) and after that, the initial detector conditions must be restored (reset phase) in order to detect the following photons.[10]

A further advantage of a SPAD detector is the possibility of turning it ON and OFF quickly and efficiently, by modulating its bias voltage from below to above the breakdown threshold. A photon impinging on the detector during its OFF time is not able to trigger any avalanche. This gated-mode operation technique is useful to perform a time filtering of incoming photons, for example, in cases where a filtering based on either photon wavelength or polarization is not possible. The precision achievable in terms of time-selection of the incoming light signal strongly depends on the sharpness of the transitions between OFF and ON conditions of the detector and on the uniformity and linearity of the response just after the rising edge. A gated-mode operated SPAD with transition times shorter than 200 ps is known in literature as a fast-gated SPAD.[12] Different improvements arise from this kind of technique: not only it becomes feasible to detect faint and fast light signals just few hundreds of picoseconds after a stronger unwanted light pulse, but also the achieved Dynamic Range (DR) of the measurement can be increased considerably, up to 7 orders of magnitude, while keeping the same acquisition

a)Electronic mail: mauro.buttafava@polimi.it

time.[12] Important applications of fast-gated SPAD detectors are, for example, those in the field of time-resolved diffuse spectroscopy of biological tissues (i.e., photon migration) at small source-detector separation,[13,14] where it is required to detect few so-called "late" photons scattered from deep layers of the sample (carrying all the information about the diffusing medium), while rejecting "early" ones, scattered from upper layers. In addition, the gated-mode operation of a SPAD can also be useful to reduce the effect of DCR and afterpulsing of the detector, improving (further) the Signal-to-Noise Ratio (SNR) of the measurement.

Time-gated SPAD modules for visible and near-infrared wavelengths (from 400 nm to 1000 nm) are reported in literature and are now commercially-available. The "Fast-Gated SPAD" module,[15] developed at Politecnico di Milano (Italy), is characterized by sharp transition times (of the order of 200 ps) and good time resolution (30 ps, Full-Width at Half Maximum—FWHM), but it can operate only at gate repetition rates up to 40 MHz with a maximum gate width of 10 ns, it needs an external trigger source and cannot work in free-running mode. The id110 module,[16] developed by ID Quantique SA (Switzerland), has a higher degree of flexibility regarding the gate repetition frequency and triggering options, but has drawbacks like poorer time resolution (FWHM is about 200 ps), lower quantum efficiency (24% at 530 nm), and limited ON-time duration range (up to 25 ns). Other types of devices are also available for sub-nanosecond time-gated single photon detection, for example, Intensified Charge-Coupled Device (I-CCD) cameras, but they typically are expensive, bulky, and fragile instruments (they can be damaged by strong light pulses).

In this paper we present the design and characterization of a complete and stand-alone single-photon counting module capable of turning ON and OFF a SPAD detector with gating transitions down to 110 ps, repetition rates up to 80 MHz, extended gate width range (up to 500 ns), and low timing jitter (FWHM $\sim$ 30 ps). This instrument can operate SPADs of different active area diameters (from 20 μm to 200 μm) built in different technologies (both custom and Complementary Metal-Oxide-Semiconductor (CMOS) silicon ones) and is also capable of free-running operation mode. It features a complete programmability of the detector parameters (like excess bias voltage, hold-off time, temperature, etc.) for the best flexibility in every application. It is characterized by a compact form-factor (the detection head dimensions are 60 mm $\times$ 60 mm $\times$ 120 mm), high reliability and ease of use.

II. INSTRUMENT DESIGN

The module is made of two separate blocks, connected via a 2-m wideband cable: a control unit and a detection head. The control unit is based on the architecture previously developed for an InGaAs/InP single-photon counting module.[17] It contains the electronic circuits needed to generate the clock signals for the system and to control the gate pulse duration, furthermore it is used to generate the external trigger signal, manage the power supply, the communication with the remote PC and the user interface. For the present project, the detection head has been completely redesigned for the fast-gating operation of silicon SPADs.

Figure 1 shows the schematic block diagram of the module, in particular the structure of the detection head. The clock signal for the SPAD gating can be internally generated, in the range from 1 kHz to 80 MHz, or it can be derived from an external source (for example, the trigger signal of a pulsed laser source) fed to the "*TRIGGER IN*" input. When the internal clock source is used, a synchronization signal is provided on the "*TRIGGER OUT*" output of the module to be used as a reference signal by a TCSPC board. Starting from the clock signal, the control unit modulates the duty cycle to generate the enabling signal for the SPAD, that is, the "*GATE IN*" input for the detection head. This pulse width modulation is implemented with a combination of fast Emitter-Coupled Logic (ECL) and CMOS logic circuits, to guarantee a time duration adjustable from 2 ns to 500 ns. In the detection head, a Complex Programmable Logic Device (CPLD) manages the hold-off and reset conditions of the SPAD. The "*GATE OUT*" signal is fed to a fast pulser, that directly drives the SPAD cathode, while the anode is kept at a bias voltage slightly lower than the SPAD breakdown voltage (the difference between the SPAD bias and the breakdown voltage is called under-voltage, V_{UV}). This is needed in order to avoid

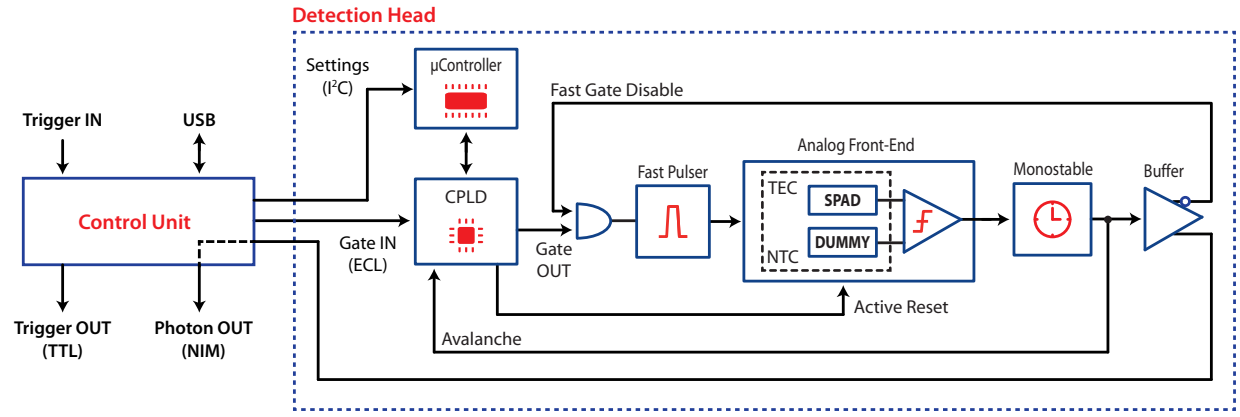

FIG. 1. Block diagram of the fast-gating single-photon counting module. The detection head contains SPAD, fast pulse generator, and front-end electronics. It is characterized by a small form-factor (60 mm $\times$ 60 mm $\times$ 120 mm) for easier integration in any optical setup. The control unit provides all the timing and trigger inputs/outputs, the USB link, and it is connected to the detection head via a multipolar wide-bandwidth cable.

avalanche events during the OFF time of the detector. Given the fast pulser peak-to-peak voltage V_{PP}, it follows that the effective excess bias of the SPAD is equal to $V_{EX} = V_{PP} - V_{UV}$. The excess bias V_{EX} can be selected by the user between 2.5 V and 7 V.

In the event of a photon detection, the avalanche current pulse is detected using a fast ECL comparator, while a monostable circuit generates a signal with a fixed time duration. This "*AVALANCHE*" signal, synchronous to the photon arrival time, is then buffered and sent to the control unit as the "*PHOTON OUT*" output signal of the module. The "*AVALANCHE*" signal is also input to the CPLD to start the hold-off period. Due to the long propagation delay of the CPLD itself (of the order of 4–5 ns), a fast turn OFF path is also needed to lower immediately the SPAD bias voltage below the breakdown value. This fast turn OFF path is implemented using the ECL AND gate at the input of the pulser, controlled by an inverted replica of the "*AVALANCHE*" signal itself. The hold-off time set by the CPLD can be selected by the user, in the range from 25 ns to 1.5 μs. During the hold-off time, the CPLD also controls the reset phase of the SPAD, sending the "*ACTIVE RESET*" pulse signal to the SPAD front-end.

The operating temperature of the SPAD can be lowered down to about $-10\,^{\circ}$C, in order to decrease its DCR noise contribution (for a typical silicon SPAD, the DCR lowers by about an order of magnitude for a drop in temperature of 15 $^{\circ}$C), using an integrated Thermo-Electric Cooler (TEC).

Finally, an 8-bit microcontroller is used to set and monitor all the operating parameters of the detection head. The communication between PC and the control unit is carried out via a USB link, while the detection head is programmed by the control unit using an I^2C serial interface. The end user can change the measurement parameters using a simple interface developed using National Instrument LabVIEW environment.

A. Fast pulse generator

The pulse generator is the most critical sub-circuit of the detection head. The amplitude of the output voltage pulse has to be programmable in a wide range (from 3 V to 7.5 V) and the transition times have to be extremely fast (of the order of 200 ps, for 20%–80% transitions). Since the PDE of the SPAD is directly proportional to its excess bias voltage V_{EX}, to obtain a uniform response of the detector for the whole ON time, the amplitude of the pulse has to be as flat and constant as possible, with no oscillations or ringing after the sharp rising edge. In addition, the output pulses must be DC-coupled to the cathode of the SPAD in order to always keep constant its bias voltage with respect to changes in the gate signal frequency and duty-cycle, or in case of non-periodic synchronization signals. This constraint does not allow the use of Monolithic Microwave Integrated Circuits (MMIC) for the implementation of the pulser (as proposed in Ref. 15).

Figure 2 shows a simplified circuit diagram of the fast pulse generator. The main component is a High Electron-Mobility Transistor (HEMT) with 4 GHz cut-off frequency, acting as a common-source switching stage. When the HEMT

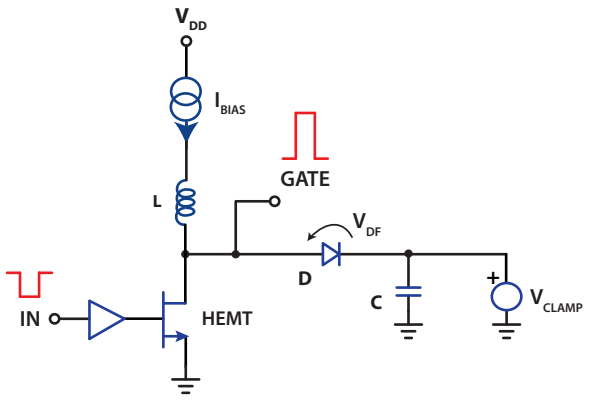

FIG. 2. Simplified schematic diagram of the pulse generator, used for SPAD ultra-fast turn ON and OFF. It is based on a common source switching stage with a clamped output voltage, implemented using a High Electron-Mobility Transistor (HEMT). Voltage V_{CLAMP} can be adjusted to change the SPAD excess bias between 2.5 V and 7 V. The output pulses exhibit very sharp transitions, less than 200 ps rise and fall times (20%–80%).

is ON the output voltage is determined by the transistor dropout voltage (of the order of 100 mV). Turning OFF the HEMT, the bias current (I_{BIAS}) flows into the diode and the voltage generator (V_{CLAMP}). The voltage generator is a sink/source capable linear regulator that clamps the output voltage of the "*GATE*" node at the value $V_{PP} = V_{CLAMP} + V_{DF}$, where V_{DF} is the forward voltage drop on the diode. In this way, a digital potentiometer changes the value of V_{CLAMP} to adjust the excess bias voltage of the SPAD. The capacitor C is actually an array of low Equivalent Series-Resistance (ESR) capacitors to obtain a low impedance path for the high-frequency components of the pulse, keeping the value of V_{CLAMP} as constant as possible during commutations. Special care was used in the design of the inductor L, whose function is to isolate the current generator from the switching node over an extended frequency range (from few kHz to 4 GHz). A broadband equivalent inductor is obtained by connecting in series different discrete components interleaved by passive dumping networks. In particular, a 2 μH conical inductor is placed close to the HEMT. The sharpness of the rising and falling edges of the output pulse can be traded against the flatness of the pulse itself, by changing the value of the current I_{BIAS}. High current values mean lower transition times but higher ringing; for this reason also the I_{BIAS} value can be programmed by the user and can be adjusted depending on the SPAD and on the specific trade-off required by the application. Finally, a HEMT driver is designed using fast ECL components, achieving negligible jitter (less than 3 ps RMS) and extremely low transition times.

B. Detection head design

Figure 3 shows the simplified circuit diagram of the front-end electronics: the avalanche signal is sensed by a differential circuit including the SPAD itself and a second identical device, called "dummy" detector, which is not able to detect any photon since it is biased below its breakdown voltage. This differential topology is needed to discriminate the avalanche signal despite the strong spurious voltage spikes at

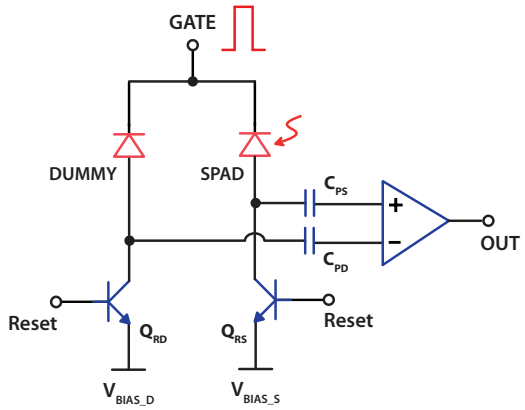

FIG. 3. Simplified circuit diagram of the front-end electronics. ON and OFF pulses are applied to the SPAD cathode, while the avalanche output signal is sensed at the SPAD anode. A differential circuit with a "dummy" detector is used to compensate the capacitive-coupled voltage spikes at the anodes, due to the enabling "*GATE*" pulses. Two wide-bandwidth BJTs act as passive-quenching and active-reset devices, restoring the SPAD bias after each avalanche.

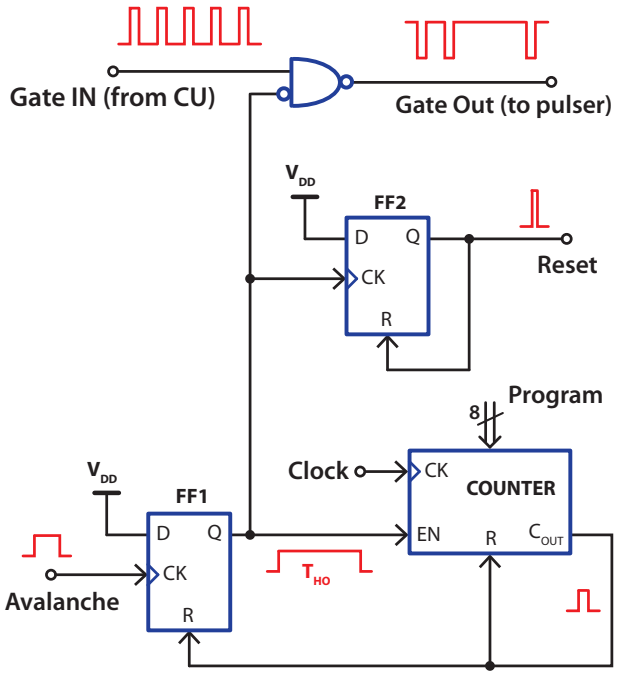

FIG. 4. Logic circuit implemented into the CPLD. The CPLD is in charge of both the hold-off and reset phases of the SPAD. Without any avalanche, the "*GATE*" signal is directly fed to the pulse generator; when an avalanche occurs, "*GATE*" is inhibited and counter is enabled, starting the hold-off phase. At the end of the hold-off time period, a "*RESET*" pulse is sent to the front-end BJTs.

the anodes due to the capacitive coupling of the "*GATE*" signal, as described in Ref. 13. The spurious spikes appear as a common-mode signal at comparator inputs, but the avalanche signal is not balanced by the "dummy" path and can trigger the comparator when a photon is absorbed. The threshold of the comparator can be adjusted changing the bias point of the two inputs, in order to neglect effects of noise and mismatches on the differential path, which translate in a non-perfect canceling of the coupling spikes. The lower is this threshold, the higher is the time resolution of the system and for this reason the entire Printed Circuit Board (PCB) is designed with the highest degree of symmetry between active and "dummy" paths.

The SPAD is operated in the so-called passive-quenching active-reset regime.[13] Bipolar Junction Transistors (BJT) Q_{RD} and Q_{RS} are involved both during quenching and reset phases: they are normally OFF and act as two quenching resistors (of the order of 1 MΩ), but during the reset phase they are turned ON for few nanoseconds, restoring the bias voltage at the anodes. Capacitors C_{PS} and C_{PD} are needed to isolate the comparator inputs from the high-voltage at the SPAD anodes. The comparator is an ultrafast SiGe ECL comparator (Analog Devices Inc.) with 8 GHz equivalent input rise-time bandwidth, 40 ps output rise/fall times, and less than 1 ps timing jitter (RMS). The output pulse from the comparator is fed to an ECL monostable circuit (see Figure 1) that generates a pulse with a fixed time duration of 25 ns for each avalanche, masking other spurious glitches, for example, due to the following reset phase.

After the avalanche detection, the SPAD has to be kept OFF for a hold-off time (few tens of nanoseconds), till when the reset pulse is sent to the front-end BJTs (Q_{RS} and Q_{RD}). This circuit is implemented within a CPLD (see Figure 4), whose advantages over a discrete logic circuit are mainly related to reconfigurability, minimum components number, and lower power dissipation. The chosen CPLD (from MAX family by Altera Corp.) has a maximum operating frequency of 200 MHz, with 4.5 ns propagation delay and a measured jitter that is less than 6 ps (RMS). Referring to Figure 4 and assuming no detected avalanches, the GATE signal from the control unit is inverted and sent to the pulse generator (the NOT operation compensates the further inversion introduced by the HEMT stage). When an avalanche event occurs, the flip-flop FF1 is triggered (by the rising edge of the "*AVALANCHE*" signal), the "*GATE*" signal is thus inhibited via the NAND gate and the counter is enabled. The down-counter was previously pre-loaded by the microcontroller with a digital word proportional to the hold-off duration (the final user can change the hold-off duration using the control interface) and the clock signal is provided by the control unit, with a fixed frequency of 167 MHz. At the beginning of the hold-off phase, the monostable built using FF2 generates also the "*RESET*" pulse with a fixed duration of 3 ns. When the counter overflows, the output C_{OUT} resets FF1 and the counter itself and the hold-off phase ends (its duration T_{HO} is determined by the clock period multiplied by the digital value pre-loaded into the 8-bit counter, giving the range from 25 ns to 1.5 μs). Additional gates (not shown) are also implemented to make sure that the hold-off phase does not end during the ON time of the gate signal. On the contrary, in the free-running operation mode (in which the SPAD is always ON and photons can be detected at any time), the "*GATE IN*" signal is always kept high and the detector is automatically re-enabled after each hold-off phase.

Figure 5 shows the most significant voltage waveforms during the operation of the detection head, highlighting its behavior after a photon detection. Waveform (a) is the output trigger signal, generated using the internal clock and used

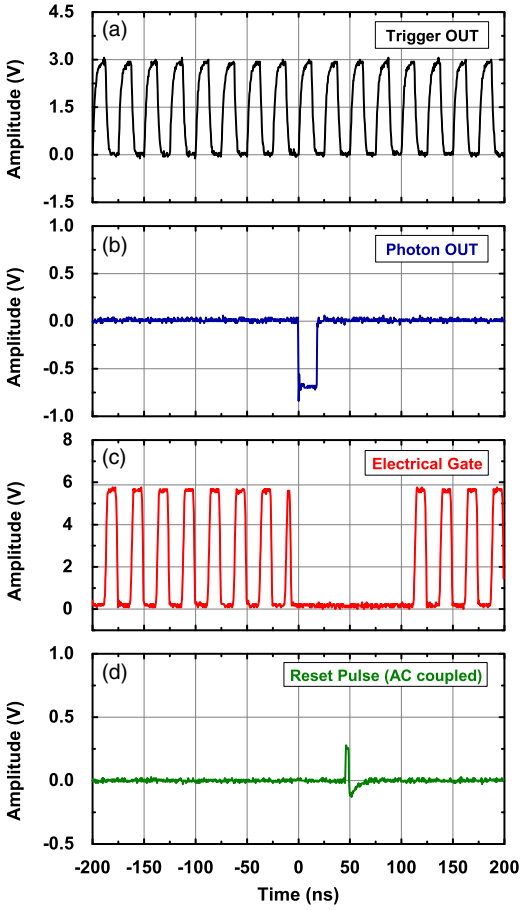

FIG. 5. Voltage waveforms during detection head operation. (a) "*TRIGGER OUT*" output signal from the control unit: it comes from the PLL-based clock generator and is used for the synchronization of the entire system. (b) "*PHOTON OUT*" signal: the falling edge marks the arrival time of detected photons. (c) Pulse generator output: it reflects the SPAD excess bias voltage and it is quickly turned OFF after an avalanche and during the following hold-off phase. The ON time in this sample measurement is set to $T_{ON} = 10$ ns. (d) "*ACTIVE RESET*" pulse: it is given to the front-end BJTs during the hold-off time interval.

as time-base for gate synchronization. This signal is a 50% duty-cycle CMOS waveform and the measurement reported in Figure 5 was taken with the repetition rate set at 40 MHz. Waveform (b) shows the "*PHOTON OUT*" signal (a Nuclear Instrumentation Module—NIM compatible pulse with a duration of 25 ns) whose falling edge marks the arrival time of the detected photon. The hold-off phase can be observed in waveform (c) that shows the output of the fast pulse generator (which is the actual excess bias voltage of the SPAD, apart from the under-voltage). When no avalanche is triggered, the SPAD is turned ON and OFF synchronously with the trigger signal. When a photon is absorbed, the excess bias is immediately lowered by the fast gate-disable path and the hold-off phase starts, as it can be seen in waveform (c), where the gate pulse in which the photon is detected has a shorter duration and the following ones are skipped during the hold-off time (100 ns in this example). The hold-off time interval ends when the first entire gate pulse starts. Finally, in waveform (d), the 3 ns "*RESET*" pulse is shown, as measured at the BJTs base terminals after AC-coupling.

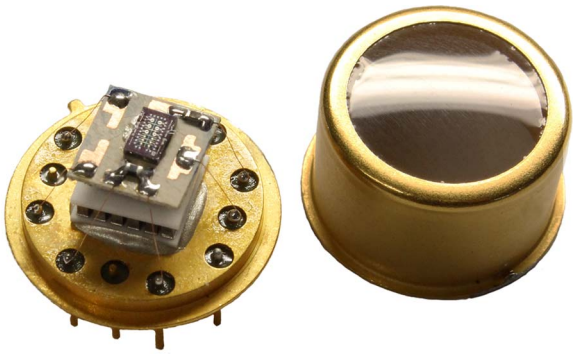

FIG. 6. Windowed TO-8 package assembly, containing both active and "dummy" SPADs glued on a ceramic chip carrier. A two-stage TEC is used to cool the detector down to −10 °C. The package is sealed in nitrogen in order to avoid moisture deposition on the cold side of TEC and on silicon die. Inductive parasitism of leads and wirings had to be reduced in order to avoid oscillations on the edges of the "*GATE*" signal.

C. SPAD cooling system

A significant improvement on the SNR of the measurements is achieved by cooling down the SPAD, in order to reduce its dark-counts noise. Furthermore, lowering the temperature impacts also on the SPAD afterpulsing probability, since release lifetimes of the semiconductor traps increase and more carriers can be released after the hold-off time. The actual operating temperature is chosen as a trade-off between DCR and afterpulsing probability (at a given hold-off time) and ranges between 0 °C and −10 °C, depending on the SPAD fabrication technology and active area diameter.[11] Temperature control is performed using a two-stage TEC and a Negative Temperature Coefficient (NTC) resistor, both in a thermal feedback loop. The actual SPAD temperature of operation is user selectable, with a resolution of 0.1 °C. For an efficient cooling, the detector is sealed into a windowed TO-8 package with a dry nitrogen atmosphere. Due to the tight matching requirements of the differential front-end, also the "dummy" SPAD must be included into the same package with the active SPAD.

Figure 6 is a picture of the package assembly with active and "dummy" SPADs integrated on the same silicon die, placed on a ceramic chip-carrier (where bonding pads are placed) and mounted on the cold-side of the TEC. A 0.1% precision 10 kΩ NTC resistor is also soldered onto the chip carrier. The main limitation in the use of a package (with respect to the chip-on-board solution adapted in Ref. 13) arises from the higher inductive parasitic effects of the leads and the chip carrier wirings. A big challenge was the minimization of these effects, keeping connections and bonding wires as short and symmetrical as possible.

III. EXPERIMENTAL CHARACTERIZATION

Module characterization started from the fast pulse generator. Figure 7 shows the rising edge of the "*GATE*" pulse signal for different peak-to-peak voltages (V_{PP}). The repetition frequency is 10 MHz with a pulse width of 10 ns; waveforms are acquired using a 20 GHz equivalent-bandwidth digital sampling scope (Tektronix 11801C with SD-26 sampling head). As it can be seen, the SPAD enabling signal exhibits

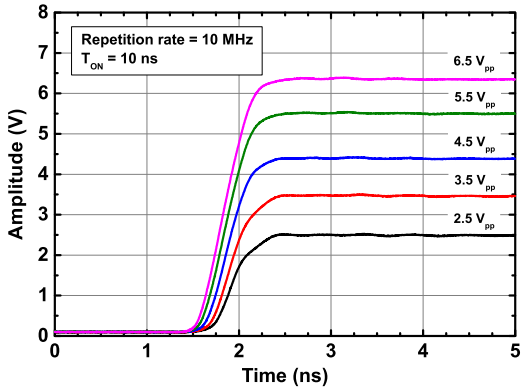

FIG. 7. Rising edges of the SPAD "*GATE*" signal for different peak-to-peak voltages V_{PP} (the actual SPAD excess bias is equal to $V_{EX} = V_{PP} - V_{UV}$). Repetition rate of the measurement is 10 MHz, with $T_{ON} = 10$ ns. Transition times are less than 200 ps (20%–80%) on the whole output voltage swing, without any oscillation or ringing during the ON time.

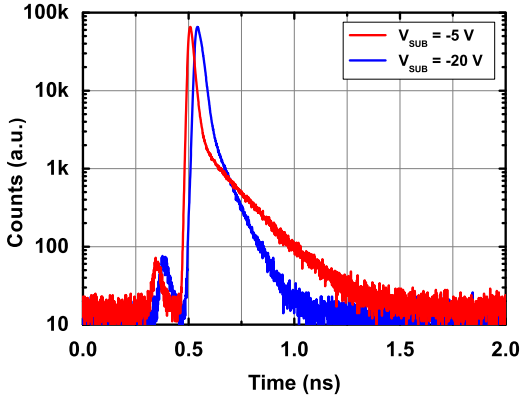

FIG. 9. Time-response of the fast-gated module to a pulsed laser source at 780 nm wavelength, with less than 1 ps pulse-width at two different SPAD substrate bias voltages V_{SUB}. Decreasing V_{SUB} increases time jitter on the photon arrival time, but also reduces the exponential tail time-constant, that is beneficial for measurements of fast light signals. Laser repetition rate is 80 MHz, and the SPAD excess bias is $V_{EX} = 7$ V.

very sharp edges, of the order of 200 ps rise-time (20%–80%) with a flat ON-time shape, both at low and high output voltages.

The SPAD exploited in following characterization is a 20 μm active-area diameter CMOS SPAD,[18] with an higher breakdown voltage "dummy" detector integrated on the same die. Peak PDE is more than 45% at 450 nm and still 5% at 800 nm, with less than 1 dark count-per-second at 0 °C. The response uniformity of the module within the ON-time window is characterized by illuminating the detector with attenuated ambient light and acquiring the temporal distribution of the incoming photons by means of the TCSPC technique (based on SPC-630 TCSPC card, from Becker & Hickl GmbH). Since the detection efficiency of the SPAD is proportional to its excess bias V_{EX}, even small ringing or overshooting of the electrical "*GATE*" signal can be directly observed in the counts distribution. The resulting response is shown in Figure 8, using an excess bias of 5 V, a 12 ns gate-ON time window at 40 MHz repetition rate, and a hold-off time (T_{HO}) of 80 ns. The shape of the gating window is flat with clean

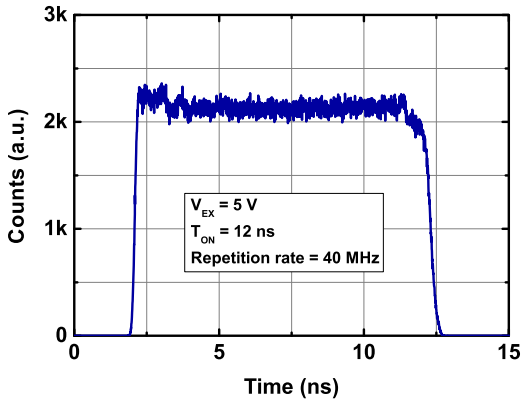

FIG. 8. Photon counts distribution within a 12 ns gate-ON time window, when the SPAD is gated at 40 MHz with an excess bias $V_{EX} = 5$ V during the ON-time. The transition time of the rising edge is 110 ps (20%–80%). Both the sharp edges and the flatness of counts distribution are key parameters in a measurement, providing a good time-filtering of incoming photons and preventing any distortion of acquired waveforms.

and fast rise and fall transitions and does not show any relevant distortion that would be detrimental for the reconstruction of the optical waveform. The obtained rise time of 110 ps is an outstanding result, since the fast-gating technique relies on really sharp transitions of the gate window for promptly rejecting the unwanted "early" photons. This value represents a factor-of-two improvement with respect to state of the art.[15]

Temporal response of the system to a laser pulse is shown in Figure 9. A typical SPAD impulse response function is composed by two contributions: a sharp main peak and a following slower exponential tail.[11] Photons absorbed into the depleted region produce a fast avalanche ignition, with a time-jitter due to the statistical fluctuation of the avalanche build-up, and give rise to the main peak. Instead, photon absorbed in neutral region (below the depleted one) generates carriers that slowly diffuse and, in some cases, reach the depleted region, originating the slower tail. The FWHM of the main peak is the intrinsic time resolution of the detector, while the exponential tail is the main limiting factor in applications where it is needed to acquire fast optical signals overwhelmed by a large amount of "early" arriving photons.[12] Figure 9 shows two impulse response functions of the fast-gated SPAD module, acquired using a mode-locked laser source with less than 1 ps pulse-width (FWHM) at 780 nm, running at a frequency of 80 MHz (Menlo Systems, TC-1550). The SPAD excess bias is 7 V, while its substrate is biased at two different voltages, giving rise to different shapes of the response function due to different collections of the carriers photo-generated in the neutral region:[18] decreasing the substrate bias voltage, the width of depleted region across the substrate-cathode junction increases, hence less photons are absorbed in neutral regions and the exponential tail time-constant decreases. At the same time, the current path through the cathode neutral region is more narrow and thus resistive, and the avalanche build-up is slower and with a higher time jitter (the main peak becomes wider). Therefore, the substrate bias voltage V_{SUB} is a trade-off parameter between time resolution and diffusion tail time-constant. From Figure 9, with $V_{SUB} = -5$ V the time resolution is 28 ps (FWHM) and the tail time-constant is $\tau = 340$ ps,

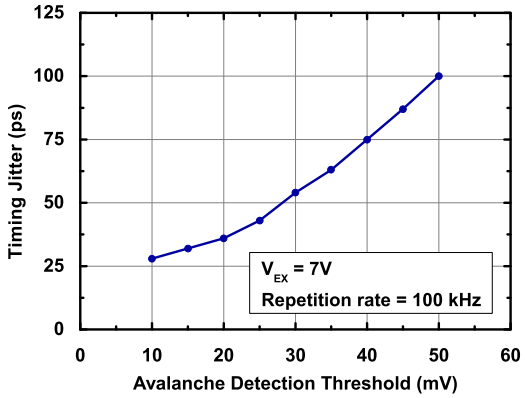

FIG. 10. Timing jitter of the fast-gated module as a function of threshold voltage V_{TH} of the front-end fast comparator. Pulsed laser source is 820 nm wavelength, with a pulse-width of 10 ps FWHM and a repetition rate of 100 kHz. SPAD excess bias voltage is 7 V. The outstanding time resolution of 28 ps FWHM is achieved with a threshold of $V_{TH} = 10$ mV.

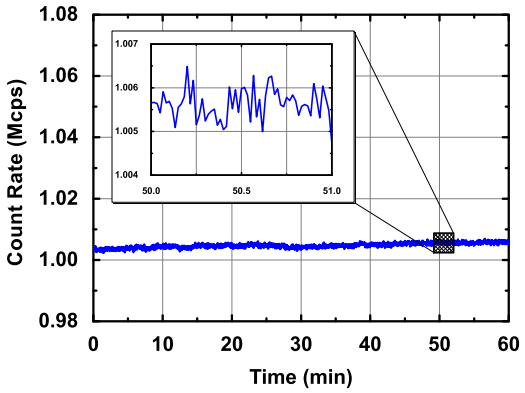

FIG. 11. Stability of the module count-rate over a time range of 60 min. The SPAD was illuminated with a continuous-wave laser source. Inset: short-term stability relative to a 60 s time interval. The overall stability is excellent, with a total drift less than 0.5% over the entire hour and dispersion of less than 0.08%.

while for $V_{SUB} = -20$ V time resolution becomes 37 ps (FWHM) but the tail time-constant lowers to $\tau = 160$ ps. The small peak leading the main one in both waveforms is due to a secondary optical output pulse of the laser at 520 nm, not completely suppressed by the optical filtering stages.

Time resolution of the system strongly depends also on the threshold voltage V_{TH} of the front-end fast comparator (Figure 3). Due to statistical fluctuations on avalanche build-up, the earlier is its detection, the lower is the introduced temporal uncertainty. Figure 10 shows the achieved time resolution of the system (represented as the peak width of the impulse response function) at different threshold voltages V_{TH}. The pulsed laser source has a wavelength of 820 nm, with less than 10 ps FWHM pulse-width and the SPAD excess bias voltage is 7 V. As it can be seen, lower values of V_{TH} give better time resolution, reaching an outstanding value of 28 ps (FWHM) for $V_{TH} = 10$ mV. The limiting factor on the minimum value of the threshold is essentially the mismatch between active and "dummy" paths of the circuit: better performance is then obtained using small active-area SPADs (which guarantee smaller feed-through spurious spikes) with specially designed "dummy" structures. However, also using a 100 μm active-area diameter CMOS SPAD,[18] it is possible to work with a threshold of about 20 mV, achieving an excellent time resolution of 40 ps (FWHM).

As mentioned above, in the fast-gating technique implementation, it is of extreme importance a precise time-selection of incoming photons. This is not only related to the sharpness of the gate ON-time rising edge, but also to the time position of the ON-time window with respect to the system main time-base. A characterization of timing jitter between triggering signals and gate-window opening time is reported in Table I, where the time jitter of the window duration itself is also reported. The gate-width generation circuit is implemented in the control unit by means of both ECL and CMOS delay lines, depending on the selected width range (see Ref. 17). This translates into different jitter performance depending on the selected gate width, as it can be seen from Table I. Using, for example, the internal clock source, the RMS jitter between "*TRIGGER OUT*" output and the ON-time window rising edge is less than 8 ps (RMS) for a gate width shorter than 10 ns. This value becomes 12 ps (RMS) if an external trigger source is used (here time jitter is measured between "*TRIGGER IN*" signal and the gate window rising edge).

TABLE I. Time jitter between trigger input/output and SPAD ON-time window (rising edge), relative to the two different gate-width (T_{ON}) generation circuits (ECL logic for $0 < T_{ON} < 10$ ns and CMOS logic for 11 ns $< T_{ON} < 500$ ns).[17]

Gate-width generation circuit	Internal trigger out (RMS) (ps)	External trigger in (RMS) (ps)	Gate-width jitter (RMS) (ps)
ECL	7.6	11.9	14
CMOS	15.8	16.8	16

Finally, Figure 11 shows the long-term stability of the module count-rate over a time period of 60 min. The photon rate was kept constant at about 10^6 counts/s using a continuous-wave laser source with the module running at ambient temperature (after a warm-up time of 30 min). The inset shows the short-term stability relative to a 60 s time interval, where the fluctuations are essentially due to poissonian noise. The overall drift is less than 0.5% over the entire 60 min, with a dispersion of less than 0.08%. These values are comparable with the intrinsic stability of the laser source, highlighting the excellent stability of the module.

IV. CONCLUSIONS

We presented the design and characterization of a gated-mode single-photon counting module, able to quickly turn ON and OFF a silicon single-photon avalanche diode with transition times down to 110 ps. The SPAD ON-time can be adjusted between 2 ns and 500 ns, with a repetition rate up to 80 MHz. A full experimental characterization of the module was carried out, testing both its electrical ad optical performances. A very flat temporal response of the detector during its ON-time is achieved, with an overall time resolution of

less than 30 ps, thanks to the differential front-end circuit implementation. The instrument can be successfully exploited in applications like optical time-resolved spectroscopy, fluorescence life-time, and time-of-flight imaging, achieving dynamic range improvement and measurement-time reduction, thanks to its extremely fast time-gating capability. The module features a compact form-factor and a complete configurability of all its operating parameters, for an easy integration in any experimental setup.

ACKNOWLEDGMENTS

The authors wish to thank Franco Zappa, Adriano Della Frera, and Carmelo Scarcella for the valuable discussions and Nick Bertone for his support.

[1]J. Steinbrink, H. Wabnitz, H. Obrig, A. Villringer, and H. Rinneberg, Phys. Med. Biol. **46**(3), 879–896 (2001).
[2]S. Del Bianco, F. Martelli, and G. Zaccanti, Phys. Med. Biol. **47**, 4131–4144 (2002).
[3]J. Selb, J. J. Stott, M. A. Franceschini, A. G. Sorensen, and D. A. Boas, J. Biomed. Opt. **10**(1), 011013 (2005).
[4]A. Pifferi, A. Farina, A. Torricelli, G. Quarto, R. Cubeddu, and P. Taroni, J. Near Infrared Spectrosc. **20**(1), 223–235 (2012).
[5]W. Becker, A. Bergmann, M. A. Hink, K. Konig, K. Benndorf, and C. Biskup, Microsc. Res. Technol. **63**, 58 (2004).
[6]G. Vicidomini, G. Moneron, K. Y. Han, V. Westphal, H. Ta, M. Reuss, J. Engelhardt, C. Eggeling, and S. W. Hell, Nat. Methods **8**, 571–573 (2011).
[7]A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. G. Bawendi, and R. Raskar, Nat. Commun. **3**, 745 (2012).
[8]K. Vishwanath, B. Pogue, and M. A. Mycek, Phys. Med. Biol. **47**(18), 3387 (2002).
[9]G. Éthier-Majcher, P. St-Jean, G. Boso, A. Tosi, J. F. Klem, and S. Francoeur, Nat. Commun. **5**, 3980 (2014).
[10]F. Zappa, S. Tisa, A. Tosi, and S. Cova, Sens. Actuators. A **140**(1), 103–112 (2007).
[11]M. Ghioni, A. Gulinatti, I. Rech, F. Zappa, and S. Cova, IEEE J. Sel. Top. Quantum Electron. **13**(4), 852 (2007).
[12]A. Tosi, A. Dalla Mora, F. Zappa, A. Gulinatti, D. Contini, A. Pifferi, L. Spinelli, A. Torricelli, and R. Cubeddu, Opt. Express **19**(11), 10735–10746 (2011).
[13]A. Dalla Mora, A. Tosi, F. Zappa, S. Cova, D. Contini, A. Pifferi, L. Spinelli, A. Torricelli, and R. Cubeddu, IEEE J. Sel. Top. Quantum Electron. **16**(4), 1023–1030 (2010).
[14]A. Pifferi, A. Torricelli, L. Spinelli, D. Contini, R. Cubeddu, F. Martelli, G. Zaccanti, A. Tosi, A. Dalla Mora, F. Zappa, and S. Cova, Phys. Rev. Lett. **100**(13), 138101 (2008).
[15]G. Boso, A. Dalla Mora, A. Della Frera, and A. Tosi, Sens. Actuators A **191**, 61–67 (2013).
[16]See http://www.idquantique.com for ID Quantique SA, id110 data-sheet.
[17]A. Tosi, A. Della Frera, A. Bahgat Shehata, and C. Scarcella, Rev. Sci. Instrum. **83**, 013104 (2012).
[18]F. Villa, D. Bronzi, Y. Zou, C. Scarcella, G. Boso, S. Tisa, A. Tosi, F. Zappa, D. Durini, S. Weyers, W. Brockherde, and U. Paschen, J. Mod. Opt. **61**(2), 102–115 (2014).