Avalanche photodiodes and quenching circuits for single-photon detection

S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa

Avalanche photodiodes, which operate above the breakdown voltage in Geiger mode connected with avalanche-quenching circuits, can be used to detect single photons and are therefore called single-photon avalanche diodes SPAD's. Circuit configurations suitable for this operation mode are critically analyzed and their relative merits in photon counting and timing applications are assessed. Simple passive-quenching circuits (PQC's), which are useful for SPAD device testing and selection, have fairly limited application. Suitably designed active-quenching circuits (AQC's) make it possible to exploit the best performance of SPAD's. Thick silicon SPAD's that operate at high voltages (250–450 V) have photon detection efficiency higher than 50% from 540- to 850-nm wavelength and still ~3% at 1064 nm. Thin silicon SPAD's that operate at low voltages (10–50 V) have 45% efficiency at 500 nm, declining to 10% at 830 nm and to as little as 0.1% at 1064 nm. The time resolution achieved in photon timing is 20 ps FWHM with thin SPAD's; it ranges from 350 to 150 ps FWHM with thick SPAD's. The achieved minimum counting dead time and maximum counting rate are 40 ns and 10 Mcps with thick silicon SPAD's, 10 ns and 40 Mcps with thin SPAD's. Germanium and III–V compound semiconductor SPAD's extend the range of photon-counting techniques in the near-infrared region to at least 1600-nm wavelength.

Key words: Photon counting, photon timing, avalanche photodiodes, quenching circuits.

1. Introduction: Photon Counting, Single-Photon Avalanche Diodes, and Quenching Circuits

Photon counting and time-correlated photon-counting techniques have been developed over many years by exploiting the remarkable performance of photomultiplier tubes (PMT's).[1–3] In recent years, special semiconductor detectors, called single-photon avalanche diodes (SPAD's), have been developed to detect single optical photons.[4–6] In the literature they have also been referred to as Geiger-mode avalanche photodiodes or triggered avalanche detectors. Significant experimental results have been obtained with these techniques in various fields: basic quantum mechanics[7,8]; cryptography[9]; astronomy[10,11]; single molecule detection[12,13]; luminescence microscopy[14–16]; fluorescent decays and luminescence in physics, chemistry, biology, and material science[17–21]; diode laser characterization[22]; optical fiber testing in communications[23–26] and in sensor applications[27,28]; laser ranging in space applications and in telemetry[29,30]; and photon correlation techniques in laser velocimetry and dynamic light scattering.[4,31,32] Among other advantages with respect to PMT's, remarkable progress has been made for photon detection efficiency, particularly in the red and near-infrared range.

Silicon SPAD's have been extensively investigated and are nowadays well developed; considerable progress has been achieved in design and fabrication techniques, and devices with good characteristics are commercially available.[4,33] The devices so far reported can be divided into two groups, based on the depletion layer of the *p-n* junction, which can be thin, typically 1 μm,[5,6,34,35] or thick, from 20 to 150 μm.[4,31,33] Their main features can be summarized as follows. For thin-junction SPAD's (i) the breakdown voltage V_B is from 10 to 50 V; (ii) the active area is small with a diameter from 5 to 150 μm; (iii) photon detection efficiency is fairly good in the visible range, ~45% at 500 nm, declines to 32% at 630 nm and to 15% at 730 nm and is still useful in the near IR, being ~10% at 830 nm and as little as 0.1% at 1064 nm; (iv) resolution is very high in photon timing, remarkably

The authors are with the Dipartimento di Elettronica e Informazione and Centro di Elettronica Quantistica e Strumentazione Elettronica, Consiglio Nazionale delle Ricerche, Piazza Leonardo Da Vinci 32, Milano 20133, Italy.

Received 20 December 1994; revised manuscript received 26 July 1995.

0003-6935/96/121956-21$06.00/0

better than 100 ps FWHM. In particular, devices with a small active area (~10-μm diameter) attain better than 30 ps FWHM at room temperature and better than 20 ps when cooled to −65 °C.[5,6,34–37] For thick-junction silicon SPAD's (i) the breakdown voltage, V_B is from 200 to 500 V; (ii) the active area is fairly wide, with a diameter from 100 to 500 μm; (iii) photon detection efficiency is very high in the visible region, remarkably better than 50% over all the 540–850-nm wavelength range and declines in the near IR but is still ~3% at 1064 nm; (iv) resolution in photon timing is fairly good, better than 350 ps FWHM for reach-through types[33] and around 150 ps FWHM for devices having a smoother field profile.[4,12] In recent years deeper insight has been gained in the physical phenomena that underlies detector operation, and ultimate limits of the performance in photon timing have been understood for both thin and thick detectors.[6,36–39]

With germanium SPAD's, photon detection efficiency greater than 15% at the 1300-nm wavelength and photon timing with 85-ps FWHM resolution have been experimentally verified.[9,26,40–43] With regard to III–V devices, photon detection efficiency above 10% at the 1550-nm wavelength and photon timing with 250-ps FWHM resolution have been verified for InGaAsP SPAD's.[44,45] In both cases, devices specifically designed to work in the Geiger mode have not yet been reported, and the behavior of commercially available photodiodes is plagued by strong afterpulsing effects because of carrier trapping (see Section 2). The situation bears some similarity to that of silicon devices 25 years ago,[46] meaning that there is much room for improvement in the material technology.

Essentially, SPAD's are p–n junctions that operate biased at voltage V_A above breakdown voltage V_B. At this bias, the electric field is so high that a single charge carrier injected into the depletion layer can trigger a self-sustaining avalanche.[46–49] The current rises swiftly (nanoseconds or subnanosecond rise time) to a macroscopic steady level in the milliampere range. If the primary carrier is photogenerated, the leading edge of the avalanche pulse marks the arrival time of the detected photon. The current continues to flow until the avalanche can be quenched by lowering the bias voltage to V_B or below. The bias voltage is then restored, in order to be able to detect another photon. This operation requires a suitable circuit that must (i) sense the leading edge of the avalanche current, (ii) generate a standard output pulse that is well synchronized to the avalanche rise, (iii) quench the avalanche by lowering the bias to the breakdown voltage, (iv) restore the photodiode voltage to the operating level. This circuit is usually referred to as the quenching circuit. As discussed below, the features of the quenching circuit dramatically affect the operating conditions of the detector and, therefore, its actual performance. Our aim in this paper is to discuss different quenching strategies, comparing simple passive-quenching arrangements and more elaborate active-quenching circuits. We also discuss the suitability of the various quenching circuits for operation with a remote SPAD connected by a coaxial cable, with reference to detectors mounted in receptacles within an apparatus or a cryostat, where the circuit cannot be mounted unless it is integrated with the detector.

A brief review of the main features of SPAD's, which must be taken into account in the design or selection of the quenching circuit, is given in Section 2. The operation of SPAD's in passive-quenching circuits is analyzed in Section 3. The operating principle and the essential features of the active-quenching circuits are dealt with in Section 4. Gated detector operation is analyzed in Section 5 for both passive and active circuits. Circuits with mixed passive–active features are discussed in Section 6. In conclusion, the main advantages offered by SPAD detectors and the role of active and passive circuits in their application and development are highlighted in Section 7. All the experimental data reported have been obtained in our laboratory unless otherwise specifically quoted.

2. Single-Photon Avalanche Diode Operating Conditions and Performance

Bias supply voltage V_A exceeds breakdown voltage V_B of the junction by an amount called excess bias voltage $V_E = (V_A - V_B)$, which has a determining influence on detector performance. It is worth stressing that the value of ratio V_E/V_B matters, not the V_E value alone, because the performance is related to the excess electric field above the breakdown level.[4,6,36–39] Since V_B ranges from 10 to 500 V in the different available SPAD's, the V_E values to be considered are from ~1 to 50 V and more.

A. Photon Detection Efficiency

For a photon to be detected, not only must it be absorbed in the detector active volume and generate a primary carrier (more precisely, an electron–hole pair), it is also necessary that the primary carrier succeeds in triggering an avalanche. The efficiency of photon detection thus increases with excess bias voltage V_E, since a higher electric field enhances the triggering probability.[4,6] Typical data obtained with thin-junction and thick-junction SPAD's are shown in Fig. 1.

B. Time-Resolution

The resolution in single photon timing also improves at a higher electric field and hence at higher V_E,[6,36–39] as illustrated in Fig. 2.

C. Dark-Count Rate

As it happens in PMT's, thermal generation effects produce current pulses even in the absence of illumination, and the Poissonian fluctuation of these dark counts represents the internal noise source of the detector. As illustrated in Fig. 3, the SPAD dark-count rate increases with excess bias voltage.

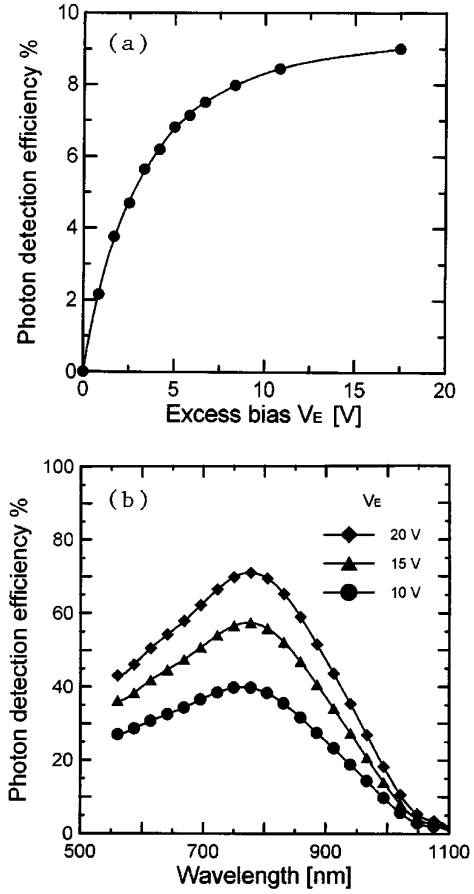


Fig. 1. Dependence of the photon detection efficiency of SPAD's on excess bias voltage V_E: (a) detection efficiency for photons at 830-nm wavelength versus V_E for a thin SPAD developed in our laboratory[6,34] (1-μm junction width, breakdown voltage V_B = 16 V, 10-μm active area diameter); (b) detection efficiency versus wavelength with parameter V_E for a thick SPAD, the EG&G Slik[4] (25-μm junction width, breakdown voltage V_B = 420 V, 250-μm active area diameter). Experimental data are from our laboratory.

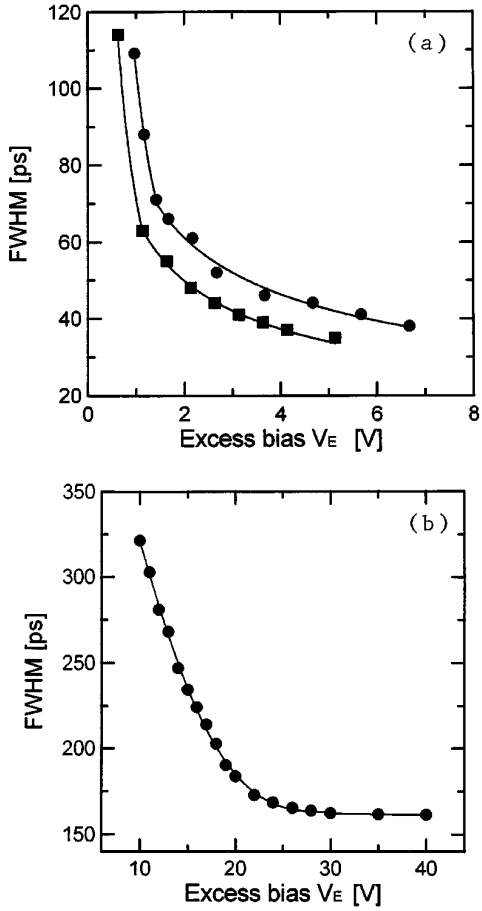


Fig. 2. Dependence of the FWHM resolution in photon timing on excess bias voltage V_E: (a) thin-junction SPAD of Fig. 1(a) at room temperature (filled circles) and cooled to −65 °C (filled squares), (b) thick-junction SPAD of Fig. 1(b) at room temperature. Experimental data are from our laboratory.

The dark-count rate includes primary and secondary pulses.[46] Primary dark pulses are due to carriers thermally generated in the SPAD junction, so that the count rate increases with the temperature as does the dark current in ordinary photodiodes. The rate also increases with V_E because of two effects, namely, field-assisted enhancement of the emission rate from generation centers and an increase of the avalanche triggering probability.

Secondary dark pulses are due to afterpulsing effects that may strongly enhance the total dark-count rate. During the avalanche some carriers are captured by deep levels in the junction depletion layer and subsequently released with a statistically fluctuating delay, whose mean value depends on the deep levels actually involved.[46,47] Released carriers can retrigger the avalanche, generating afterpulses correlated with a previous avalanche pulse. The number of carriers captured during an avalanche pulse increases with the total number of carriers crossing the junction, that is, with the total charge of the avalanche pulse. Therefore afterpulsing increases with the delay of avalanche quenching and with the current intensity, which is proportional to excess bias voltage V_E. The value of V_E is usually dictated by photon detection efficiency or time resolution requirements, or both, so that the trapped charge per pulse first has to be minimized by minimizing the quenching delay. If the trapped charge cannot be reduced to a sufficiently low level, a feature of the quenching circuit can be exploited for reduction of the afterpulsing rate to a negligible or at least an acceptable level. By deliberately maintaining the voltage at the quenching level (see Section 3), during a hold-off time after quenching, the carriers released are prevented from retriggering the avalanche.[47] As shown in Fig. 3(a), for silicon SPAD's at room temperature a few hundred nanoseconds hold off can reduce by orders of magnitude the total dark-count rate at higher excess bias voltage, since it covers most of the release transient and practically eliminates afterpulsing. For SPAD's that work at cryogenic temperatures the method is less effective, since the release transient becomes much slower and the hold-off time required to cover it may be much

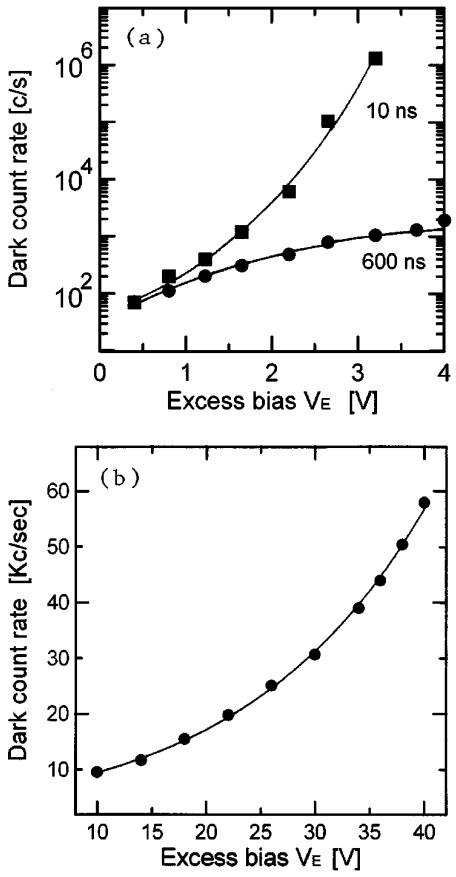


Fig. 3. Dependence of the dark-count rate on excess bias voltage V_E: (a) thin SPAD of Fig. 1(a) at room temperature; the parameter quoted is the hold-off time after each avalanche pulse (see text); (b) thick SPAD of Fig. 1(b) operated at room temperature with 40-ns hold-off time; substantially equal results are obtained with longer hold off, indicating that trapping effects are almost negligible in this device. Experimental data are from our laboratory.

longer and hence seriously limit the dynamic range in photon-counting measurements.[43]

The key factor for attaining a low dark-count rate is detector fabrication technology. In silicon technology, efficient gettering processes minimize both the concentration of generation centers that are responsible for the primary dark-current pulses and of deep levels that act as traps of avalanche carriers. As illustrated in Fig. 3(b), silicon SPAD's have been recently produced with a very low dark-count rate, that is, with an extremely low generation rate and almost negligible trapping (extremely weak trapping with very short release, ~10 ns at room temperature[4,48]).

D. Thermal Effects

Breakdown voltage V_B strongly depends on junction temperature. The thermal coefficient value depends on the SPAD device structure and is fairly high, typically around $0.3\%/\mathrm{K}$.[4,46,49] At constant supply voltage V_A, the increase of V_B causes a decrease of excess bias voltage V_E, which in percentage terms is greater than that of V_B by the factor V_B/V_E. The resulting percent variation of V_E is very strong at a low V_E level, $\sim 30\%/\mathrm{K}$, and fairly high also at a high V_E level, $\sim 3\%/\mathrm{K}$. The effects on device performance are significant. The avalanche current itself dissipates considerable energy in the device: the instantaneous pulse can attain watts of power. The thermal resistance from the diode junction to the heat sink strongly depends on the type of mounting (packaged device, chip on carrier, etc.) and can range from less than 0.1 to 1 °C/mW. At a high counting rate, the mean power dissipation causes a significant temperature increase, particularly in SPAD's with high V_B (see Sections 3 and 4). Remarkable effects are observed in detector performance, particularly in cases in which the photodiode chip is not mounted on an efficient heat sink and the mean count rate of the avalanche pulses varies.[4] It is therefore important to stabilize accurately the junction temperature in working conditions. It is also possible to stabilize V_E directly by increasing the supply voltage V_A as the junction temperature rises. However, this introduces a positive feedback with moderate loop gain, since it slightly increases the power dissipation (see Sections 3 and 4). For SPAD's having high V_B, an upper limit or a coarse stabilization of the temperature must be associated with the bias voltage feedback.

3. Passive-Quenching Circuits

In the experimental setups employed in the early studies on avalanche breakdown in junctions,[46,49] the avalanche current quenched itself simply by developing a voltage drop on a high impedance load. These simple circuits, illustrated in Fig. 4, are still currently employed and have been called[50,51] passive-quenching circuits (PQC's). The SPAD is reverse biased through a high ballast resistor R_L of 100 kΩ or more, C_d is the junction capacitance (typically ~1 pF), and C_s is the stray capacitance (capacitance to ground of the diode terminal connected to R_L, typically a few picofarads). The diode resistance R_d is given by the series of space–charge resistance of the avalanche junction and of the ohmic resistance of the neutral semiconductor crossed by the current. The R_d value depends on the semiconductor device structure: it is lower than 500 Ω for types with a wide area and thick depletion layer [Figs. 1(b), 2(b), and 3(b)] and from a few hundred ohms to various kiloohms for devices with a small area and a thin junction [Figs. 1(a), 2(a), and 3(a)].

Avalanche triggering corresponds to closing the switch in the diode equivalent circuit. Figure 5 shows the typicaly waveforms of diode current I_d and diode voltage V_d, or of the transient excess voltage $V_{\mathrm{ex}} = V_d - V_B$:

$$I_d(t) = \frac{V_d(t) - V_B}{R_d} = \frac{V_{\mathrm{ex}}(t)}{R_d}. \tag{1}$$

A. Quenching Transition

The avalanche current discharges the capacitances so that V_d and I_d exponentially fall toward the

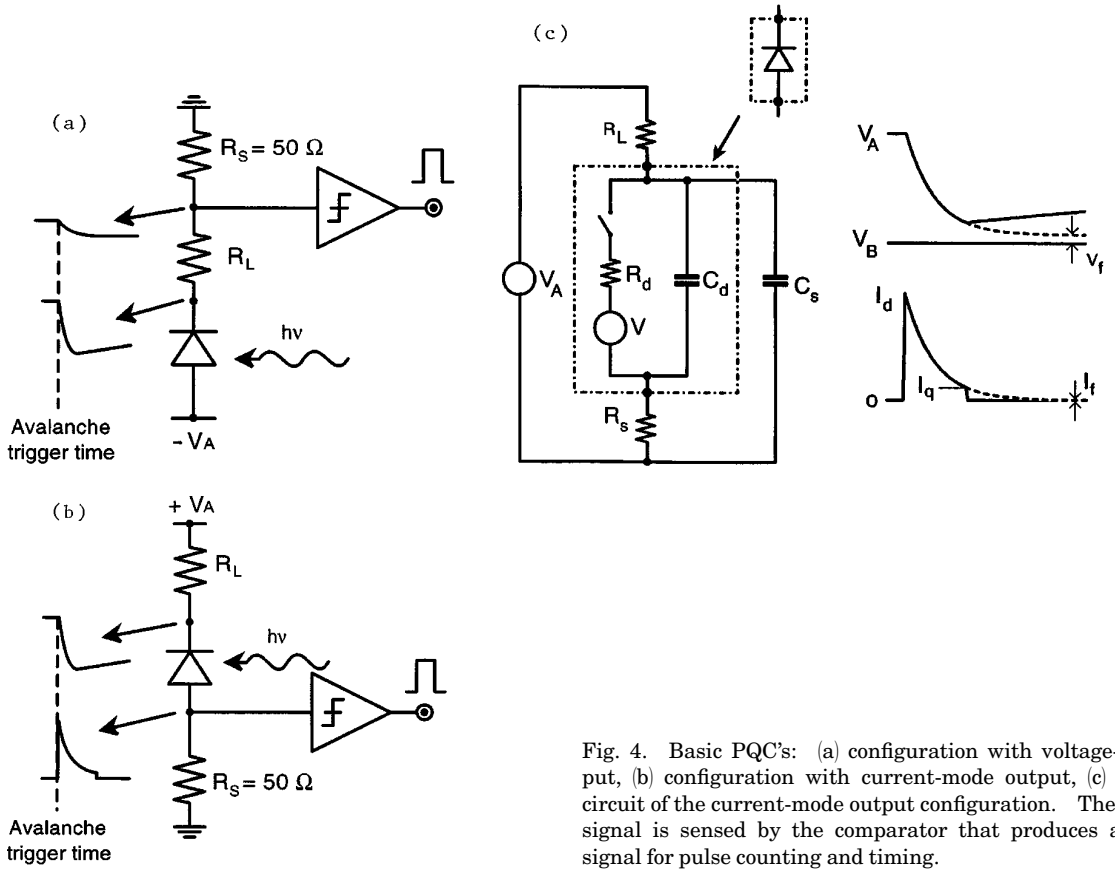


Fig. 4. Basic PQC's: (a) configuration with voltage-mode output, (b) configuration with current-mode output, (c) equivalent circuit of the current-mode output configuration. The avalanche signal is sensed by the comparator that produces a standard signal for pulse counting and timing.

asymptotic steady-state values of V_f and I_f:

$$I_f = \frac{V_A - V_B}{R_d + R_L} \cong \frac{V_E}{R_L}, \tag{2}$$

$$V_f = V_B + R_d I_f. \tag{3}$$

The approximation is justified since it must be $R_L \gg R_d$, as shown in the following. The quenching time constant T_q is set by the total capacitance $C_d + C_s$ and by R_d and R_L in parallel, i.e., in practice simply by R_d,

$$T_q = (C_d + C_s)\frac{R_d R_L}{R_d + R_L} \cong (C_d + C_s)R_d. \tag{4}$$

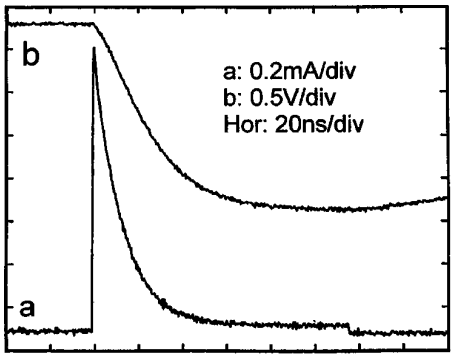


Fig. 5. Pulse waveforms of a SPAD of the type in Fig. 1 that operates in the PQC of Fig. 4(b), displayed on a digital oscilloscope: a, avalanche current I_d; b, diode voltage V_d.

If I_f is very small, V_f is very near to V_B. When the declining voltage $V_d(t)$ approaches V_B, the intensity of $I_d(t)$ becomes low and the number of carriers that traverse the avalanche region is then small. Since the avalanche process is statistical, it can happen that none of the carriers that cross the high field region may impact ionize. The probability of such a fluctuation to zero multiplied carriers becomes significant when the diode current I_d falls below ≈100 μA, and rapidly increases as I_d further decreases.[49] The avalanche is self-sustaining above a latching current level $I_q < 100$ μA and is self-quenching below it. The I_q value is not sharply defined, as is evidenced by a jitter of the quenching time with respect to the avalanche onset and by a corresponding jitter of diode voltage V_q at which quenching occurs. In most computations V_q can be assumed practically equal to V_B although it is slightly higher:

$$V_q = V_B + I_q R_d. \tag{5}$$

The total charge Q_{pc} in the avalanche pulse, an important parameter for evaluating the trapping effects (see Section 1), can thus be evaluated, setting in evidence its relation to asymptotic current I_f and characteristic time constant T_r of the voltage recovery

$$Q_{pc} = (V_A - V_q)(C_d + C_s) \cong V_E(C_d + C_s) \cong I_f T_r, \tag{6}$$

$$T_r = R_L(C_d + C_s). \tag{7}$$

B. Minimum Value of the Load Resistor

If the asymptotic current I_f is set to a value much lower than the latching current level I_q, the behavior of the PQC is the correct one: the declining avalanche current crosses the I_q level with good slope, so that the avalanche is neatly quenched after a well-defined time, with fairly small jitter. However, if the I_f value is raised toward I_q, quenching occurs with a progressively longer delay and wider time jitter.[49] With I_f very close to I_q, quenching still occurs, but with very long and wildly jittering delay. Finally, when I_f is made higher than I_q, the avalanche is no longer quenched and a steady current flows, in a situation just like that of diodes currently used as voltage references in electronic circuits. In this case, if the power dissipation $I_f V_f \cong I_f V_B$ is too high for the actual device mounting, excessive heating may even cause permanent damage to the diode. The experimenter must always check that the asymptotic I_f value is sufficiently lower than the quenching level I_q; that is, that the load resistance R_L is sufficiently high. As a rule of thumb, I_f should not exceed 20 μA, that is, the R_L value should be at least 50 kΩ/V of applied excess bias voltage V_E. Therefore, the minimum values of R_L to be employed in PQC's range from 50 to 500 kΩ for thin-junction SPAD devices [see Figs. 1(a) and 3(a) and Refs. 5, 6, 34, 35, and 37] that work with V_E from 1 to 10 V, from 200 kΩ to 2.5 MΩ for thick-junction SPAD's [such as the EG&G C30902S and Slik; see Figs. 1(b) and 3(b) and Refs. 4 and 31–33], that work with V_E from 4 to 50 V. Lower values of R_L may be used with caution.

C. Recovery Transition and Small-Amplitude Pulses

Avalanche quenching corresponds to opening the switch in the diode equivalent circuit [Fig. 4(c)] so that the capacitances are slowly recharged by the small current in ballast resistor R_L. The diode voltage exponentially recovers toward the bias voltage (curve b of Fig. 6) with time constant T_r, so that it takes $\sim 5T_r$ to recover the correct excess voltage within 1%. Given the typical values of load R_L and of the total capacitance $C_d + C_s$, T_r is typically in the microsecond range. As recovery starts, the diode voltage V_d rises above V_B. A photon that arrives during the first part of the recovery is almost certainly lost, since the avalanche triggering probability is very low. Subsequently, the arriving photons have a progressively higher probability of triggering an avalanche. However, as illustrated in Fig. 6, the diode fires at a voltage lower than V_A. It then operates with lower photon detection efficiency and impaired photon-timing resolution and produces voltage and current pulses having smaller amplitude as shown in Fig. 7.

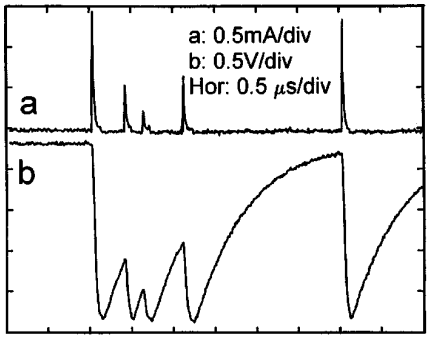


Fig. 6. Retriggering of a SPAD in a PQC (same as in Fig. 5) during the recovery transient after an avalanche pulse, which is the first one displayed on the left-hand side. The a, diode current and b, voltage waveforms are displayed on a fast oscilloscope in a single-sweep mode. Experimental data are from our laboratory.

D. Output Pulse

One can obtain an output pulse from a PQC by inserting a low-value resistor R_s in series on the ground lead of the circuit. A convenient value is $R_s = 50\ \Omega$, since it provides matched termination for a coaxial cable. The sign of the output pulse can be changed by interchanging the diode terminal connections and by changing the polarity of the bias supply voltage.

With R_s on the ground lead of the ballast resistor, as shown in Fig. 4(a), the pulse is an attenuated replica of the diode voltage waveform (see curve b of Fig. 5) and is therefore called voltage-mode output, with peak amplitude V_u of

$$V_u = (V_A - V_B - I_q R_d)\frac{R_s}{R_L + R_s} \cong V_E \frac{R_s}{R_L} \cong I_f R_s. \quad (8)$$

A drawback of the voltage-mode output is that the detector timing performance is not fully exploited, because of the intrinsic low-pass filter with time constant T_q that acts on the fast current pulse to produce the voltage waveform. It has been seen in theoretical analysis and verified by experiments[39,52] that such filtering has a detrimental influence on the photon-timing accuracy, which can be only partly compensated by employing a very low threshold level in the timing circuit. Furthermore, the limitation to I_f (see Subsection 3.B.) causes V_u to be necessarily small. When $R_s = 50\ \Omega$, $V_u = 1$ mV, so that the connection to an external comparator by way of a

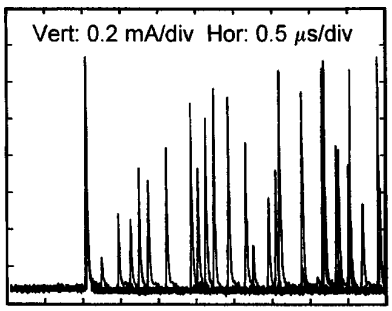


Fig. 7. Avalanche current pulses of a SPAD in a PQC (same as in Fig. 5) that occur at different times during the recovery from a previous pulse, which triggers the oscilloscope scan and is the first one displayed on the left-hand side. Note that the pulse amplitude tracks the recovery diode voltage [compare with Fig. 6(b)]. The waveforms are displayed on a fast oscilloscope in a repeated-sweep mode. Experimental data are from our laboratory.

coaxial cable is not advisable. It is better to employ a higher R_s value, typically 1 kΩ, and mount the comparator (or a voltage buffer) close to the detector.

With R_s on the ground lead of the photodiode as shown in Fig. 4(b), the waveform of the pulse is directly that of the diode current (see curve a of Fig. 5) and is therefore called current-mode output. It is important to realize that, in order to have a significant voltage pulse on R_s, the stray capacitance C_s must be comparable to or greater than the diode capacitance C_d. Otherwise, only a small fraction of the avalanche current will flow through R_s: only the current that discharges C_s flows in the loop including R_s [see Fig. 4(c)], whereas the current discharging C_d flows in the internal loop within the diode. The steplike voltage transition observed on R_s has amplitude V_s:

$$V_s \cong V_E \frac{R_s}{R_d\left(1 + \frac{C_d}{C_s}\right)} \cong I_f R_s \frac{R_L}{R_d\left(1 + \frac{C_d}{C_s}\right)} \cong V_u \frac{R_L}{R_d\left(1 + \frac{C_d}{C_s}\right)}. \quad (9)$$

The waveform has the same fast rise time of the avalanche current (~1 ns or less) and amplitude V_s from tens to hundreds of millivolts with $R_s = 50\ \Omega$. The timing performance of the detector is best exploited, and a coaxial cable can be directly connected to the SPAD terminal. V_s is inherently much higher than V_u because it is generated by the high current V_E/R_d of the avalanche (or most of it, see above) flowing in R_s, whereas V_u is due to the small current $V_E/R_L = I_f$ drawn by load R_L.

The current-mode output [Fig. 4(b)] offers the best performance in high-rate counting and in precision pulse timing and is usually preferred. The voltage-mode output [Fig. 4(a)] is fairly simple and has useful features. For example, since it produces pulses with longer duration, it is easier to monitor an avalanche pulse sequence on a long time scale of the oscilloscope.

E. Small-Pulse Effects in Photon Counting

Pulses having amplitude lower than the threshold of the comparator are not sensed. Figure 7 illustrates the situation: since the comparator threshold level cannot be very low because of electronic noise and temperature drift, each pulse is followed by a quite long dead time T_{pd}. Since the pulse repetition rate $n_t = n_p + n_b$ (sum of the detected photon rate n_p and of dark pulse rate n_b) is random, significant count losses ensue at higher counting rates.

One might consider correcting these count losses by applying the well-known methods developed for counting pulses from nuclear radiation detectors. However, the available correction equations[1,2,53,54] apply to detection systems having behavior either strictly paralyzable (a radiation quantum that arrives during a dead time does not generate an output pulse, but it restarts the dead time) or strictly nonparalyzable (during the dead time the system is completely insensitive: no output pulses are generated, and there is no restarting of the dead time). Furthermore, these equations can be reliably employed only if the value of the dead time is constant and accurately known. None of these conditions is fulfilled in the case of SPAD's in PQC's. After each avalanche pulse, the triggering probability has a continuous evolution, starting from practically nil and finally reaching a steady value. The behavior of the detector is thus quite peculiar: it is paralyzable, but with time-dependent sensitivity to triggering events. The result is a loss of linearity at high counting rates, which may be measured empirically but for which equations for accurate correction of the count losses have yet to be worked out. More important, the precise value of T_{pd} depends on the relative height of the threshold level and of the normal output pulse and is neither well known nor very stable, so that even an empirically measured correction may not be fully reliable. The pulse amplitude may vary [see Eq. (8)] because of variations of excess bias V_E, that is, of the supply voltage V_A or, more likely, of the breakdown voltage V_B. A typical example may better clarify the question. Let us assume that $V_E = 2\text{V}, R_d = 1\ \text{k}\Omega, C_d = C_s = 1\ \text{pF}, R_L = 1\ \text{M}\Omega, R_s = 50\ \Omega$, so that $V_s = 50$ mV and $T_r = 2$ μs. With the discriminator threshold set at 25 mV, the dead time T_{pd} is $\sim T_r/2 = 1$ μs. A shift of 1 mV in the threshold level causes a variation of 20 ns in T_{pd}. The same occurs for a variation of 2 mV in V_s, which may be due to a variation of 80 mV in excess bias V_E. The latter may arise from a variation in the junction temperature of 0.1 K for a thick-junction SPAD with $V_B = 250$ V (e.g., the EG&G C30902S[4]) or 1.7 K for a thin-junction SPAD with $V_B = 16$ V [e.g., the device of Figs. 1(a) and 3(a)].[6,34,35,37]

In summary, with a counting dead time depending on the SPAD voltage recovery, photon-counting measurements can be obtained with an accuracy better than 1% only if the total counting rate n_t is low enough to make count losses negligible and correction unnecessary. This corresponds to keeping a lower than 1% probability P_L of having one or more pulses within the time interval T_{pd} following a counted pulse. From Poisson statistics, we have $P_L = [1 - \exp(-n_t T_{pd})]$ or approximately $P_L = n_t T_{pd}$ for low values of the average number $n_t T_{pd}$ of events in T_{pd}. Accurate counting can therefore be obtained with total pulse counting rate $n_t < 1/100T_{pd}$. In the example considered above, this means $n_t < 1/50T_r$, that is, counting rates lower than 10 kcps. The situation can be improved by inserting a circuit having constant dead time T_{ed} after the comparator with $T_{ed} > T_{pd}$, typically $T_{ed} \geq 2T_r$. This electronic dead time will have a dominant effect on the count losses and, therefore, equations valid for constant dead time will yield fairly accurate corrections to a

moderate count rate, i.e., to $n_t \leq 1/10T_{ed} = 1/20T_r$, that is, to 25 kc/s in the example considered.

F. Small-Pulse Effects in Photon Timing

The time resolution is severely degraded by various effects connected to small-pulse events. First, the intrinsic time resolution of the SPAD is impaired when the diode voltage is reduced.[5,6,33–39,42–44] Second, the reduction of the pulse amplitude causes the triggering time of the comparator to walk along the rising edge of the avalanche pulse. With a pulse rise time of ~1 ns, a 10% reduction in the pulse amplitude causes a delay of ~100 ps in the threshold crossing time. One might try to avoid this walk by employing a constant-fraction-trigger circuit[55] instead of a simple threshold trigger, but this solution turns out to be only partially effective. In fact, a constant-fraction-trigger circuit can only eliminate or strongly reduce the walk time in the case of pulses having varying amplitudes but constant shape, whereas the avalanche pulses have a rise time that becomes progressively slower as the diode voltage is lowered.[6,33,36–39,43,44]

As shown in Fig. 8, a degradation of the resolution is experimentally observed at a higher counting rate n_t because a higher percentage of photons correspond to small-pulse events and, therefore, are timed with intrinsically lower resolution and with additional random delay. The probability of such events is the Poisson probability of having one or more photons over the entire recovery transient, which lasts ~$5T_r$. This recovery is much longer than the dead time T_{pd}, so that a more stringent limitation to the counting rate is set for photon timing than for photon counting, notwithstanding that the allowed percentage of events that occur during the guard interval is somewhat higher in photon timing than in photon counting. In fact, setting the limit to 5%, the corresponding limitation is $n_t < 1/100T_r$, which in the example considered above means $n_t < 5$ kcps.

With regard to the experimental results in Fig. 8, a question may arise about the effect of the conversion time of the time-to-amplitude converters (TAC's) that were used to record the timing resolution. The SPAD pulses are sent to the stop input, whereas a pulse synchronous with the light pulse is sent to the start input. TAC's typically have a conversion time of several microseconds, during which time they do not respond to subsequent start and stop pulses. It might be concluded that, if a photon is detected during a SPAD recovery period, it is usually not recorded because the TAC is usually busy processing the prior photon, which is detected with full bias voltage. However, in a typical high counting rate situation, SPAD pulses time correlated to the light pulse are mixed with a steady rate of uncorrelated pulses because of dark counts and stray light. Uncorrelated SPAD pulses that occur before the start pulse are not accepted, since the TAC does not respond to stop pulses that occur before the start. The TAC then accepts a subsequent time-correlated pulse that may occur during a SPAD voltage recovery caused by one of the uncorrelated events. The situation is also similar in the so-called reverse TAC configuration (in which SPAD pulses are sent to the start input of the TAC and the pulse synchronous with the light signal to the stop input by way of a delay greater than the measured time range), since the fast input gating facility of TAC's is usually exploited to reject pulses that occur before the time interval of interest.

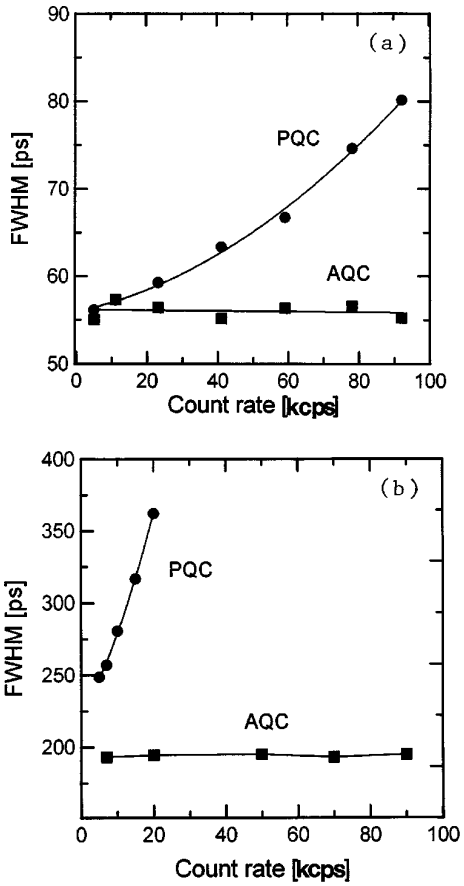


Fig. 8. Effect of the counting rate on the FWHM resolution in photon timing with SPAD's in PQC's. The total counting rate is progressively increased by increasing the steady background light that falls on the detector. For comparison, the performance obtained with the same SPAD operating with an AQC is also reported: (a) thin SPAD device of Fig. 1(a) that operates at room temperature with excess bias $V_E = 2.5$ V in a PQC with recovery time constant $T_r = 500$ ns; (b) thick SPAD of Fig. 1(b) that operates with excess bias $V_E = 20$ V in a PQC with $T_r = 500$ ns and cooled to 0 °C to reduce the dark-count rate.

G. Working at Higher Counting Rates

To extend the working range toward higher counting rates, the recovery time of SPAD's in PQC's must be minimized by minimizing the values of ballast resistor R_L and stray capacitance C_s. However, excessive reduction of R_L will bring the steady current I_f close to the latching current level I_q (see Subsection 3.B.). The turn-off probability is then so low that the duration of the avalanche current (or turn-off time[49]) will contribute significantly to the counting dead time T_{pd} and eventually will dominate it. Further-

more, this duration fluctuates, so that T_{pd} becomes not well defined. For example, with $I_f = 50$ μA, the turn-off probability is $\sim 10^4$ s^{-1},[49] so that the duration of the avalanche current has a 100-μs average value and is affected by fluctuations with 100-μs rms deviation. Even in the most favorable cases, that is, for SPAD's with low capacitance (C_s and $C_d \leq 1$ pF) that operate at low excess bias V_E (between 1 and 2 V) with minimum ballast resistor R_L (from 50 to 100 kΩ), the recovery time constant T_r will be around 200 ns and the entire recovery will take ~1 μs. Therefore, the counting rate limitations for accurate photon counting and photon timing can be improved, but not much beyond 200 and 50 kc/s, respectively.

H. Power Dissipation

The energy E_{pd} dissipated in the SPAD during an avalanche pulse corresponds to the decrease of the energy stored in the capacitance $C_d + C_s$:

$$E_{pd} = \frac{1}{2}(C_d + C_s)(V_B + V_E)^2 - \frac{1}{2}(C_d + C_s)V_B^2$$

$$= (C_d + C_s)V_E\left(V_B + \frac{V_E}{2}\right) \cong (C_d + C_s)V_E V_B,$$

i.e., to the pulse charge Q_{pc} [see Eq. (6)] falling by a voltage slightly higher than V_B:

$$E_{pd} = Q_{pc}\left(V_B + \frac{V_E}{2}\right) \cong Q_{pc}V_B.$$

The dissipation therefore depends not only on excess bias voltage V_E and total capacitance $C_d + C_s$, but also on breakdown voltage V_B. At moderate total counting rate n_t, the mean power dissipation is given by $n_t E_{pd}$ and may cause significant heating, particularly in SPAD's with high V_B. For example, with $V_B = 400$ V, $V_E = 20$ V, and $C_d + C_s = 5$ pF, at $n_t = 100$ kcps the mean power dissipation is 4 mW. With a the thermal resistance of 1 °C/mW, this may increase the junction temperature by 4 °C and V_B by ~5 V. It is worth stressing, however, that PQC's are fairly safe for SPAD's, since they inherently avoid excessive power dissipation. At higher n_t values, the dissipation rise is limited by the increased percentage of small-pulse events; at very high n_t, the limit dissipation corresponds to the product of latching current I_q and breakdown voltage V_B.

I. Operation with a Single-Photon Avalanche Diode Remote from the Circuit

Working with a SPAD mounted in a position remote from the quenching circuit implies long electrical connections between the detector and circuitry. For the ballast resistor R_L, a long connection is unacceptable because it strongly increases the stray capacitance C_s, the avalanche pulse charge [Eq. (6)] and the transition times [Eqs. (4) and (7)], enhancing the associated drawbacks. Luckily, mounting R_L close to the SPAD is not a problem: the physical size of R_L can be very small, since the power dissipation in it is much smaller than in the SPAD. For the signal output, a coaxial cable connected to the remote circuit is perfectly suitable, provided it is terminated in its characteristic resistance: it then represents a purely resistive load and guarantees good transmission of a fast signal. In the configuration with current-mode output [Fig. 4(b)], $R_s = 50$ Ω can be normally employed, so that a PQC, in which only the SPAD and load resistor R_L are remote from the circuit board, can be readily assembled. The PQC configuration with voltage-mode output is not as well suited: a higher R_s must be employed so that at least a voltage buffer must be mounted near the SPAD.

4. Active Quenching

A. Active-Quenching Principle

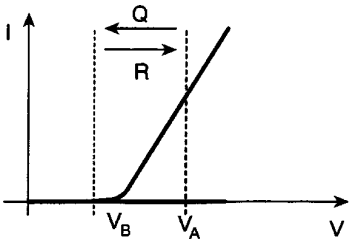
To avoid drawbacks that are due to the slow recovery from avalanche pulses and exploit fully the inherent performance of SPAD's, a new approach was devised and implemented experimentally in our laboratory.[56] The basic idea was simply to sense the rise of the avalanche pulse and **react back on** the SPAD, forcing, with a controlled bias-voltage source, the quenching and reset transitions in short times. A terminology that focuses on the essential features of the circuits was also introduced in our laboratory[50,51] and has been universally adopted: active-quenching circuits (AQC's) are based on the new principle and PQC's are those without a feedback loop. The studies carried out in various laboratories on active or partially active (see Section 6) quenching circuits can be outlined as a sort of family tree. The first AQC configuration (opposite terminal type, see below) dates back to 1975,[56] but it was 1981 before its application to photon timing was attempted,[50] fast gating of the detector was demonstrated,[51] a second basic AQC configuration (opposite terminal type, see below) was introduced,[51] and an AQC based on complementary metal oxide semiconducting integrated circuit blocks was reported.[57] In 1983 the application of SPAD's and AQC's to time-resolved fluorescence measurements was demonstrated.[17] In 1987 a fast AQC was specifically developed for photon correlation and laser Doppler velocimetry.[32] In 1988 an AQC configuration suitable for remote SPAD operation was introduced, patented, and licensed for industrial production.[58] In 1990 the application of AQC's to satellite laser ranging with centimeter resolution was reported.[29] In 1991 a compact AQC module was specifically developed for astronomy.[10] In 1993 application of SPAD's with AQC's to fiber-optic sensors was reported.[27] Various quenching circuits based on fast semiconductor switches [double-diffused metal oxide semiconductor (DMOS) field-effect transistors (FET's)] have also been developed and reported: in 1988 an active-reset circuit[48]; in 1993 two AQC's, one to be employed as a component of detectors for elementary particle physics experiments,[59] the other as the basic ele-

ment of a high-performance general-purpose photon-counting module to be produced industrially[4]; in 1994 a compact AQC for detectors in adaptive-optics telescopes.[11]

Figure 9(a) illustrates the principle of the active-quenching method. The rise of the avalanche pulse is sensed by a fast comparator whose output switches the bias voltage source to breakdown voltage V_B or below. After an accurately controlled hold-off time, the bias voltage is switched back to operating level V_A. A standard pulse synchronous to the avalanche rise is derived from the comparator output to be employed for photon counting and timing. The basic advantages offered by the AQC approach are the fast transitions (from quenched state to operating level and vice versa) and the short and well-defined durations of the avalanche current and of the dead time. The approach is fairly simple and bears some similarity to an approach employed in an original study with true Geiger–Mueller gas detectors for ionizing radiation, but completely new problems arise in its development and application with SPAD's, as discussed below.

B. Basic Active-Quenching Circuit Configurations and Design Problems

An AQC inherently has two connections to the SPAD for sensing the avalanche current and for applying the quenching pulse. Therefore, two basic AQC configurations can be considered, one with a quenching terminal opposite the sensing terminal (Fig. 10),



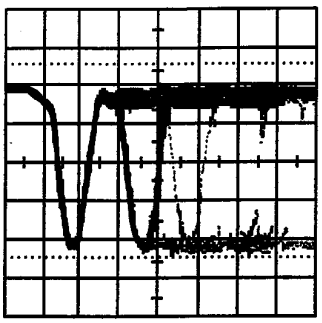


Fig. 9. (a) Principle of active quenching: current–voltage I–V characteristic curve of the SPAD and switching load line (dashed lines) of the AQC controlled voltage source. The Q arrow denotes the quenching transition, the R arrow the reset transition. (b) Output pulses from an AQC designed for minimum dead time that operates with a SPAD of the type in Fig. 1, biased 0.9 V above the breakdown voltage, displayed on a fast oscilloscope at 5 ns/div. Experimental data are from our laboratory.

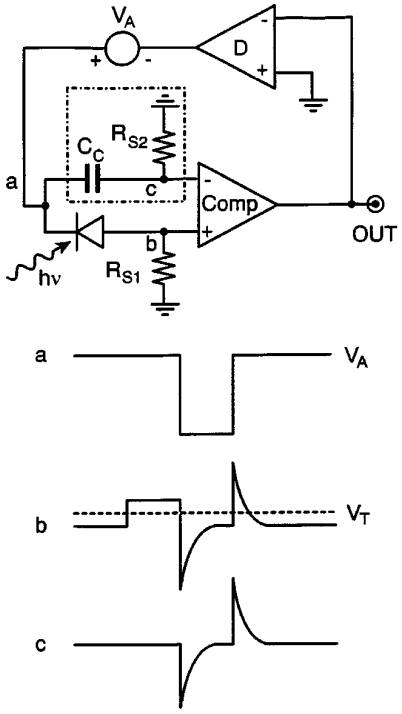


Fig. 10. Simplified diagram of the basic AQC configuration with opposite quenching and sensing terminals of the SPAD. The network in the dotted box compensates the current pulses injected by the quenching pulse through the SPAD capacitance, thus avoiding circuit oscillation. The voltage waveforms drawn correspond to the circuit nodes marked with the same letter.

the other with a coincident quenching and sensing terminal (Fig. 11). In any case, the sensing terminal has a quiescent voltage level at ground potential or not far from it, since it is directly connected to the AQC input.

The quenching and reset driver, labeled D in Figs. 10 and 11, represents circuit means that, when driven from a low-level logic pulse, generate a high-voltage pulse. It can be implemented with either a pulse-booster circuit stage[10,17,27,29,32,50,51,56,57] or with electronic switches[4,11,48,59] connected to two different dc voltage sources that correspond to the operating and quenching voltage levels. For example, such switches can be DMOS FET's capable of withstanding the required voltage and of switching in nanosecond time from a low-series-resistance on state to a high-series-resistance off state and conversely. Both solutions have their relative merits and have been employed in practice. With fast switches the AQC can be simpler, more compact, and have lower power dissipation, since the driver dissipates power only during the transitions. With a pulse-booster circuit the AQC output can better approximate a constant impedance source, as required for remote SPAD operation (see below), and better control and fine adjustment of the pulse waveform is usually obtained.

The amplitude of the quenching pulse should be larger than excess bias V_E. The amplitude margin should be sufficient to overcome possible reignition

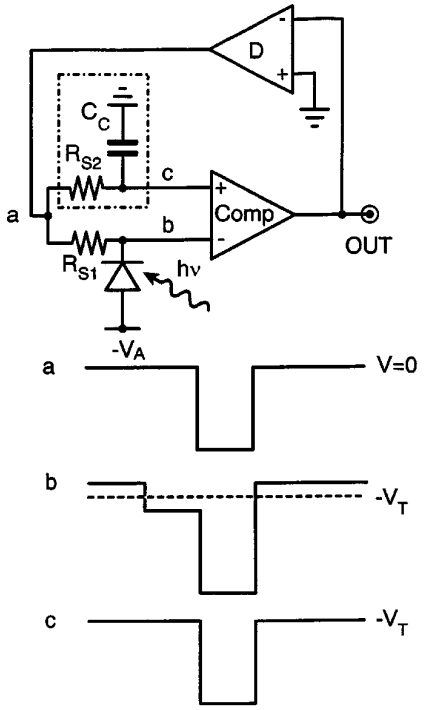


Fig. 11. Simplified diagram of the basic AQC configuration with coincident quenching and sensing terminals of the SPAD. The network in the dotted box is employed to avoid (i) locking of the circuit in the triggered state by the quenching pulse, and (ii) circuit oscillation that is due to small overshoots and ringing of the quenching pulse. The voltage waveforms drawn correspond to the circuit nodes marked with the same letter.

effects that are due to nonuniformities of the breakdown voltage over the detector area. For SPAD's with high V_B, deviations from uniformity can attain several volts[4] and make more severe the requirement of producing large voltage swings with short transition times. Determining suitable electronic components and devising circuit schemes for working with excess bias voltage higher than 20 V are nontrivial tasks for the circuit designer.

1. *Opposite Terminal Configuration*

In the opposite terminal configuration[10,50,51,56] the quenching pulse must be superimposed on the detector dc bias voltage. An ac coupling could be employed for the quenching pulse, but dc coupling is preferred. With ac coupling, if it happens even once that the avalanche is not quenched, for example, because of an interfering electromagnetic disturbance from a spark, the SPAD is locked in a stationary avalanche-on condition, insensitive to subsequent photons and possibly subject to catastrophic end, caused by excessive heating. Even in the absence of such anomalous events, with ac coupling the baseline of the voltage applied to the SPAD suffers a count-rate-dependent mean shift toward a lower value (see Section 5); furthermore, it is affected by random fluctuations because of the random-time distribution of the pulses. On the other hand, the problems met in the design of a dc coupled circuit become increasingly difficult as the bias voltage is increased. In practice, the configuration is not suitable for SPAD's having breakdown voltage higher than 30 V. A further problem has to be faced with any SPAD, which arises from the current pulse that is injected into the AQC input through the detector junction capacitance by the fast quenching pulse transition. The reset transition injects a pulse with polarity that is equal to the avalanche pulse and comparable amplitude and that retriggers the circuit forcing it into steady oscillation. For example, in the case of $C_d = 1$ pF, $V_E = 2$ V, with a 2-ns transition time, the current injected is 1 mA. Specific provisions to avoid such spurious retriggering should be taken. Fast integrated circuit comparators usually have a latch input[60]: by applying to it a pulse covering with sufficient margin the SPAD voltage reset transition, spurious retriggering is inhibited. While the comparator is still latched, however, the voltage on the SPAD recovers and the device can be triggered by incoming photons that are sensed as events that occur at the end of the latch command. In order to discard such incorrectly timed events, the circuit must be further elaborated. An alternative solution, adopted in the first AQC design,[50,56] is to compensate the capacitive pulse by deliberately injecting another pulse through an auxiliary capacitor that has been trimmed to match the detector capacitance. If the compensating pulse is returned to the same input terminal connected to the detector, an inverted quenching pulse has to be applied to the capacitor, as shown in Refs. 50, and 56. If the AQC comparator has a differential input, the compensating capacitor can be connected to the other input terminal and the quenching pulse itself can be employed,[10] as shown in Fig. 10.

2. *Coincident Terminal Configuration*

Coincident terminal configuration[4,11,17,27,29,32,48,51,57,59] has the basic advantage of being suitable for all SPAD's with any breakdown voltage because one of the device terminals is free, not connected to any AQC point, and is available for applying any required dc bias voltage. The quenching pulse is applied to the same terminal and with the same polarity of the avalanche pulse; thus it locks the comparator in the triggered state unless suitable circuit means are provided to avoid it. A monostable circuit that limits the duration of the quenching pulse is a simple solution. If carefully designed, such a circuit produces clean rectangular pulses with fast transitions affected by minimal overshoots and ringings, typically limited from 1 to 3% of the pulse amplitude. However, the pulse amplitude ranges from a few volts to tens of volts: in absolute terms, this means overshoots from tens to hundreds of millivolts applied to the AQC input. Since the circuit must be sensitive to pulses smaller than 50 mV (avalanche pulse of 1 mA or less on input resistance of 50 Ω), the overshoots on the reset transition can retrigger the comparator and drive the circuit into oscillation. The overshoots could be

minimized by slowing down the transition, but at the cost of giving up some of the basic advantages of the AQC. As in the opposite terminal configuration and with the same remarks, the latch input may be employed for inhibiting spurious retriggering. A better solution, illustrated in Fig. 11, was devised for the second generation AQC's.[17,51] A comparator with differential input is employed and the quenching pulse is applied to both terminals (common-mode signal), whereas the avalanche pulse is applied to one side only (differential signal). If the waveforms on the two input sides are identical, the action of the quenching pulse on the comparator is canceled. To equalize the shape of the pulse transitions, one can improve the input symmetry by adding a capacitor in parallel to the second terminal, emulating the detector capacitance. There is some analogy with the compensating capacitor of the previous configuration, but in practice the capacitance matching turns out to be not at all critical in this configuration.

C. Hold-Off Time

The hold-off feature can be introduced in any AQC configuration with simple circuit means. A monostable circuit, triggered by the leading edge of the comparator output and combined with the latter in OR configuration, can be employed to extend the quenching pulse for a controlled time. The hold-off time is effective in reducing the effects of trapped carriers in the dark-count rate (see Section 2). However, it adds to the AQC dead time and is not a convenient solution for cases with a long trap release transient.[43]

D. Remote Detector Operation

In AQC's the ports connected to the detector terminals inherently have a low resistive impedance, so that it is possible to select a 50-Ω value, providing a matched termination to a coaxial cable. In principle, AQC's appear inherently suitable to work with remote SPAD's connected by coaxial cables; in practice, nontrivial problems are met in the design of such circuits. The coaxial cable enhances the problem of avoiding spurious retriggering of the AQC at the end of the reset transition, particularly in cases with high amplitude of the quenching pulse or long cables, or both. Since the cable cannot be terminated at the detector end, the inductive and capacitive mismatches generate there and reflect back to the AQC input overshoots and ringing, corresponding to the transitions of the quenching pulse. The quenching driver is therefore subject to more severe requirements. It must provide a good termination to the cable at the circuit end; therefore, it must provide a constant impedance output, which is usually better obtained with a pulse generator circuit rather than with DMOS FET switches. It must charge not only the stray and detector capacitance, but also the capacitance of the coaxial cable (~100 pF/m of cable); therefore, it has to supply a higher current surge.

In summary, circuits based on the AQC principle and suitable for remote detector operation appear attractive but present severe problems for the circuit designer. In recent years, the practical value and the high-performance level obtainable with such circuits has been demonstrated in many experiments carried out in our laboratory; a reliable circuit of this kind has been developed and patented.[27,58]

E. Active-Quenching Circuit Essential Features

From the previous discussion, the essential features and basic characteristics of AQC's can be highlighted as follows.

(a) Duration T_{ac} of the avalanche pulse is constant and can be very short. It is set by the quenching delay, given by the sum of circulation time T_L in the quenching loop (from detector to sensing circuit and back from quenching circuit to detector), plus rise time T_{aq} of the quenching pulse:

$$T_{ac} = T_L + T_{aq}. \tag{10}$$

The least obtainable T_{ac} value depends on the required excess bias voltage V_E, since the value of T_{aq} increases with the amplitude of the quenching pulse. The minimum T_{ac} values verified range from less than 5 ns with V_E below 1 V [see Fig. 9(b)] to ~10 ns for V_E around 20 V.[4,11] In cases for which one must set the duration of the current at longer values, for example, for studying carrier trapping phenomena,[43,47] this can be simply and accurately obtained by inserting a known additional delay in the loop.

(b) Dead time T_{ad} of an AQC is constant, well defined, and can be accurately controlled. Its minimum value $(T_{ad})_{\min}$ is given by twice the circulation time T_L in the quenching loop plus the sum of rise time T_{aq} and fall time T_{ar} of the quenching pulse:

$$(T_{ad})_{\min} = 2T_L + T_{aq} + T_{ar}. \tag{11}$$

As for T_{ac}, the shortest attainable $(T_{ad})_{\min}$ value increases with excess bias voltage V_E. The minimum value experimentally obtained is 10 ns working at low V_E [below 1 V, see Fig. 9(b)] and 40 ns at high V_E (approximately 20 V, see Refs. 4 and 11).

(c) A hold-off time after avalanche quenching can be easily introduced, with simple circuit means that maintain low voltage for a longer and accurately controlled time.

(d) Since duration T_{ar} of the reset transition is very short, the probability of small-pulse events during recovery is minimized. Avalanche triggering almost always occurs at a well-defined, constant bias voltage condition. Furthermore, the few small-pulse events occur only in correspondence with the well-defined and short reset transition, so that they can be easily recognized and inhibited or discarded by suitable auxiliary circuits.

(e) Thanks to the low resistance of the bias source, practically all the avalanche current contributes to the detector output pulse by flowing in the external circuit (that is, through resistor R_{s1} in Figs. 10 and 11).

(f) The total charge Q_{ac} in the avalanche pulse is constant and well defined. It can be evaluated by assuming that it has a trapezoidal pulse shape with negligible rise time, a constant current at the maximum value $I_{d\max} = V_E/R_d$ during circulation time T_L, and a linearly declining current during quenching transition T_{aq}. We can make reference to pulse width T_w of a square pulse with equal area and equal amplitude,

$$T_w = T_L + \frac{T_{aq}}{2}, \tag{12}$$

and write

$$Q_{ac} = \frac{V_E T_w}{R_d} = I_{d\max} T_w. \tag{13}$$

For a given SPAD device, it is interesting to compare AQC's and PQC's from the standpoint of avalanche charge, which is the source of trapped charge and related afterpulsing effects. From Eqs. (6) and (13), one can see that

$$Q_{ac} > Q_{pc} \quad \text{for } T_w > T_q, \tag{14a}$$

that is,

$$Q_{ac} > Q_{pc} \quad \text{for } \left(T_L + \frac{T_{aq}}{2}\right) > R_d(C_d + C_s). \tag{14b}$$

The comparison depends on excess bias voltage V_E applied, because at higher V_E rise time T_{aq} of the active-quenching pulse becomes longer, increasing pulse width T_w. It is interesting to note that, to minimize the trapped charge, depending on the case, the active- or passive-quenching approach may be advantageous. For SPAD devices having small R_d and C_d and small stray capacitance C_s, passive quenching provides the least pulse charge, because T_q is shorter than the characteristic times in any AQC loop, that is, shorter than any possible T_w. On the other hand, for detectors with higher resistance R_d and larger stray capacitance C_s, active quenching minimizes the charge. As discussed in Section 6 a mixed passive–active-quenching approach may be the most suitable for minimizing the avalanche charge. In any case, it is necessary to bear in mind that it is not the trapped charge alone, but the resulting afterpulsing effect in actual working conditions that does matter. Therefore, an accurate comparison between different circuits should also take into account the presence or absence of the hold-off feature in the quenching circuit.

(g) The energy E_{ad} dissipated in the SPAD during an avalanche pulse depends on breakdown voltage V_B, on excess bias voltage V_E, on SPAD resistance R_d, and on pulse duration T_w, since it is due to pulse charge Q_{ac} [see Eq. (13)] falling by voltage $V_B + V_E$:

$$E_{ad} = Q_{ac}(V_B + V_E) \cong Q_{ac}V_B = V_E V_B T_w/R_d.$$

Q_{ac} (and therefore pulse width T_w) must also be minimized for reducing the dissipation, not only the trapping effects. Therefore, also from the standpoint of power dissipation mixed passive–active-quenching circuits (see Section 6) may be most suitable. The mean power dissipation is given by pulse energy E_{ad} multiplied by the pulse-counting rate. In the operation of SPAD's having high V_B in an AQC with a minimum dead time, the high counting rate capability is essentially limited by thermal effects in the SPAD. Caution is necessary, since at very high counting rates the mean power can be fairly high, not much lower than the instantaneous pulse power $(V_B + V_E)V_E/R_d$.

5. Gated Detector Operation

In various applications, it is necessary or at least advantageous to operate a highly sensitive photodetector under the control of a gate command. Typical cases are laser ranging,[29,30] testing of fibers with optical time-domain reflectometry,[23–28] and measurements of weak fluorescent emissions immediately following an intense light pulse that is due to scattering from an excitation laser.[12–21] With PMT's for gated operation it is necessary to apply voltage pulses of hundreds of volts to a PMT dynode. With SPAD's, the task is much easier: a pulse with amplitude V_g slightly larger than excess bias voltage V_E is sufficient to switch the detector from gated off (at bias voltage $V_A < V_B$) to gated on (at the operating bias voltage $V_A + V_g = V_B + V_E$). Furthermore, gated operation can also be effective in avoiding the dark-count rate enhancement that is due to trapping effects in SPAD's. If the SPAD has been gated off for a sufficiently long time interval (much longer than the trap release time) before bringing it into operation, trap levels are almost all empty and do not interfere with the following measurement.[40–44,47]

The operation and performance of circuits suitable for gated detector operation are analyzed and discussed. The gate command is a rectangular voltage pulse from a low impedance source, with duration T_g and fast rise and fall times. The time interval between a gate pulse and the following is T_a, the repetition rate is $f_a = 1/T_a$, the duty cycle is $w = T_g/T_a = f_a T_g$. In cases for which the repetition rate is not periodic but random, T_a and f_a represent the mean interval and repetition rate.

A. Passive Gated Circuits: Basic Features

As outlined in Fig. 12, the passive gated circuit configurations can have gate input with either dc or ac coupling. In both cases, the pulse actually applied to the SPAD is modified by the filtering action of the circuit.

1. Direct Current Coupled Gate

The gating pulse is added directly in series to dc bias V_A. It can be added at the same terminal where dc bias V_A is applied [Fig. 12(a)], leaving the other SPAD terminal at ground potential free to take the output

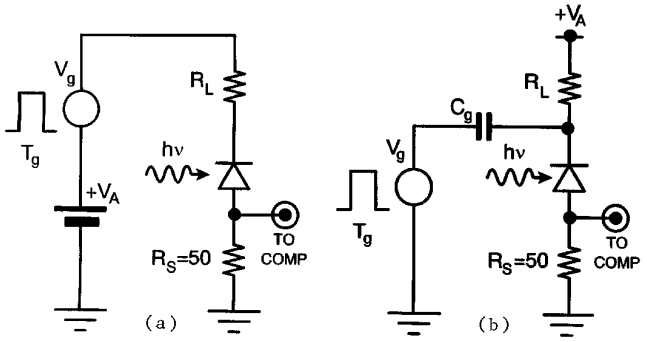


Fig. 12. PQC configurations for gated detector operation: (a) dc coupled gate input, (b) ac coupled gate input. The equivalent circuit of the SPAD [see Fig. 4(c)] must be taken into account in the circuit analysis.

signal. For SPAD's with high V_B, however, this approach implies nontrivial problems in the gate-driver circuit. Alternatively, the gate pulse can be added at the other terminal, but it is then necessary to employ an elaborate electronic output stage that separates the avalanche signal at this terminal from the gate pulse. In any case, the voltage signal actually applied to the detector is modified by the low-pass integrator filter made by resistance R_L in series with capacitance $C_d + C_s$ [see the SPAD equivalent circuit in Fig. 4(c)]. The actual rise and fall times of the gating voltage are thus slowed down to $\sim 2.2T_r$. Of course, during these transitions small-pulse events may occur (see Section 3).

2. *Alternating Current Coupled Gate*

The voltage pulse actually applied to the detector [Fig. 12(b)] is modified by the differentiator filter made by C_g in series with resistance R_L paralleled by capacitance $C_d + C_s$. Two effects have to be considered. First, the actual amplitude V_g' is attenuated with respect to V_g by the capacitive voltage divider:

$$V_g' = V_g \frac{C_g}{(C_g + C_d + C_s)}. \tag{15}$$

If C_g were comparable with $C_d + C_s$, the attenuation would be significant and not well controlled, because the value of $C_d + C_s$ is usually not accurately known. It is advisable to minimize the relative loss in pulse amplitude $(V_g - V_g') \le 0.01V_g$ by employing a sufficiently high C_g:

$$C_g \ge 100(C_d + C_s). \tag{16}$$

Second, the shape of the gating pulse applied to the detector is modified by the differentiator with time constant

$$T_{gr} = R_L(C_g + C_d + C_s) \cong R_L C_g. \tag{17}$$

The top of the gate pulse is no longer flat; it decays almost linearly from the initial V_g to $V_g(1 - T_g/T_{gr})$. In order to keep the decrease within 1%, it is required that

$$g = \frac{T_{gr}}{T_g} \ge 100. \tag{18}$$

The gate pulse is followed by a long negative tail, starting with amplitude $V_g T_g/T_{gr} = V_g/g$ and decaying exponentially with the slow time constant T_{gr}, which corresponds to the slow release of the charge accumulated by C_g in the gate-on interval T_g. With a periodic sequence of gate pulses with duty cycle w, the linear superposition of exponential tails build up a negative baseline offset V_n:

$$V_n = V_g b w, \tag{19}$$

wherre b is a numeric factor that can be easily computed as $b \cong 1$ for $w > 1/g$ and $b < 1$ for $w \le 1/g$. Offset V_n has an effect equivalent to a loss in amplitude V_g and is subject to the same limitation $V_n \le 0.01\ V_g$. Therefore, ac coupling is suitable for operation only with low duty cycle w:

$$w \le \frac{0.01}{b} \cong 0.01. \tag{20}$$

In cases in which the gate pulse sequence is random and not periodic, V_n given by Eq. (19) is the mean baseline shift, but it is advisable to consider an even more stringent limitation because of the random fluctuations around the mean.

It is also important to take into account that the charge of the avalanche pulse is stored in C_g, causing the voltage applied to the SPAD to decay with time constant

$$T_{gq} = (C_g + C_d + C_s)\frac{R_d R_L}{R_d + R_L} \cong C_g \frac{R_d R_L}{R_d + R_L}. \tag{21}$$

The voltage recovery that follows has the time constant T_{gr} of Eq. (17), so that

$$T_{gq} = T_{gr}\frac{R_d}{R_d + R_L} = T_g g \frac{R_d}{R_d + R_L}, \tag{22}$$

and we can define

$$h = \frac{T_{gq}}{T_g} = g\frac{R_d}{R_d + R_L}. \tag{23}$$

A thorough analysis of passive gated circuits has to take into account that the end of the gate pulse will terminate the avalanche current, if still flowing at that time. We therefore have to deal not only with self-quenching passive gated circuits, but also with passive gated circuits with quenching by gate termination.

B. Passive Gated Circuits: Self-Quenching

1. *Direct Current Coupled Gate*

Load resistor R_L must be high if self-quenching is required, particularly in cases for which high excess

bias voltage has to be employed, as discussed in Subsection 3.B. The rise and fall times of gate voltage $2.2T_r$ are thus fairly slow, from hundreds of nanoseconds to microseconds. Therefore, these circuits are unsuitable in most cases; they can work only with fairly long gate durations, at least various microseconds, and suffer significant small-pulse effects.

2. *Alternating Current Coupled Gate*

The avalanche quenches itself only if it fully charges C_g within gate time T_g, allowing the SPAD voltage to decay to V_B. This can occur only if T_{gq} is much shorter than T_g, typically

$$T_{gq} \leq T_g/5, \quad \text{that is, } h \leq 1/5. \tag{24}$$

We see from Eqs. (18) and (23) that the ac coupled circuit is self-quenching only with high load $R_L \gg R_d$, typically

$$R_L \geq 5gR_d \geq 500R_d. \tag{25}$$

It is worth stressing that, with a self-quenching passive circuit having an ac coupled gate, no more than one event per gate pulse can be observed, because the voltage recovery after quenching is much longer than T_g. Furthermore, this recovery sets a more severe limitation on the duty cycle than inequality (20). At the gate end, the SPAD voltage decays to V_B, just above the quiescent bias level V_A. The negative step of the gate voltage produces a voltage undershoot below V_A with almost full amplitude V_g, followed by exponential recovery with time constant T_{gr}. To limit the baseline shift imposed on the next gate pulse, it is necessary to wait until the undershoot amplitude decays from V_g to $V_g/100$, that is, $\sim T_{gr} \log_e 100 \sim 5T_{gr} = 5gT_g \geq 500T_g$. In conclusion, correct operation of a self-quenching passive circuit with ac coupled gate input is possible only with very low duty cycle w, typically

$$w \leq \frac{T_g}{T_{gr} \log_e 100} \cong \frac{1}{5g} \leq 0.002. \tag{26}$$

A definite conclusion can be drawn about self-quenching gated passive circuits: their performance and applications are severely limited. The dc coupled types are practically unsuitable in most cases. The ac coupled types can be employed in fast gating with very low duty cycle for detecting not more than one event per gate pulse.

C. Passive Gates Circuits: Quenching by Gate Termination

1. *Direct Current Coupled Gate*

To keep gate rise and fall times at the nanosecond level, the load resistor R_L value must be 100 Ω or less.[42–44] In correspondence with the fast gate-on and gate-off transitions the gate voltage, through the diode junction capacitance (see Subsection 4.B.1), injects on the output R_s fast current pulses, with polarity equal to the avalanche pulse at the gate opening[45] and opposite at the gate closing. With low R_L values, these spurious pulses have amplitude comparable with the avalanche pulses and complicate the operation of circuits employed for pulse counting or timing. To obtain nanosecond gate duration, the situation must be analyzed with more complex equivalent circuits that better take into account the distribution of resistances, capacitances, and inductances in the actual detector device and mounting. Employing microwave design techniques, even subnanosecond gate durations have been obtained.[61] The following conclusions can be drawn about SPAD's in dc coupled gated circuits with quenching by gate termination:

(a) They can be well exploited for detecting a single photon within a gate pulse of short duration T_g, from subnanosecond to a few microseconds.

(b) The limits to gate duration, repetition rate, and duty cycle are set by the thermal and trapping effects caused in the SPAD's by the avalanche current, which flows until the end of the gating pulse.

(c) They are totally unsuitable for cases in which more than one photon should be detected within one gate interval.

2. *Alternating Current Coupled Gate*

In circuits having a load resistor R_L insufficient for self-quenching, the voltage drop caused by the avalanche is smaller than V_g and quenching occurs at the end of the gate pulse. No more than one event per gate pulse can be observed. When T_{gq} is comparable with T_g, the SPAD voltage decay is still considerable and, therefore, the amplitude of the undershoot at the gate end is still fairly large. It depends on the time of triggering within the gate and can attain $V_g[1 - \exp(-T_g/T_{gq})]$. With respect to the self-quenching type, the time needed for the undershoot to decay to $V_g/100$ and the corresponding limitation to the duty cycle are reduced: in fact, the amplitude ratio in the logarithm is reduced from 100 of inequality (26) to $100[1 - \exp(-T_g/T_{gq})]$.

If $T_{gq} \gg T_g$, the SPAD voltage decay and the undershoot are very small, so that this limitation that is due to the recovery from a single avalanche pulse becomes insignificant. However, the avalanche pulses enhance the limitation that is due to the linear superposition of negative tails of the gate pulses. In fact, when the avalanche is triggered, R_d is added in parallel to R_L and the differentiating time constant is accordingly switched from T_{gr} to T_{gq}. The amplitude of the negative tail increases from V_g/g to $V_gT_g/T_{gq} = (V_g/h) = (1 + R_L/R_d)(V_g/g)$. If P_g is the probability of having an avalanche within gate time T_g, the mean amplitude of the negative tail following a gate pulse is approximately

$$(1 - P_g)\frac{V_g}{g} + P_{gs}\frac{V_g}{g}\left(1 + \frac{R_L}{R_d}\right) = \frac{V_g}{g}\left(1 + P_g\frac{R_L}{R_d}\right). \tag{27}$$

The mean baseline shift V_n is accordingly modified with respect to Eq. (19):

$$V_n = V_g w b\left(1 + P_g \frac{R_L}{R_d}\right), \tag{28}$$

and to have baseline shift $V_n \leq 0.01 V_g$, it is necessary that

$$w \leq \frac{0.01}{b\left(1 + P_g \frac{R_L}{R_d}\right)} \cong \frac{0.01}{\left(1 + P_g \frac{R_L}{R_d}\right)}. \tag{29}$$

Inequality (29) points out that a high factor R_L/R_d can make this limitation significantly more stringent than inequality (20), notwithstanding that P_g is usually quite lower than unity. It is therefore advisable to employ a low value for the load resistor, that is, $R_L \leq R_d$ and a correspondingly higher capacitor C_g to keep the time constant T_{gr} sufficiently long [see Eq. (18)]. These conditions appear the most advisable for ac coupled gate circuits with quenching by gate termination. In such conditions, the ac configuration is an interesting alternative to the dc coupled configuration, particularly valuable for working with SPAD's having high V_B. For very short gate times T_g, resistive load R_L can be advantageously substituted or complemented by an inductive load.[61] The conclusions drawn for the dc coupled configuration with quenching by gate termination also apply in this case, except that duty cycle w is subject to the more severe limitation given by inequality (29).

D. Passive Gated Circuits: Conclusions

Some general conclusions can be drawn about gated passive circuits. The dc or ac coupled configurations with quenching by gate termination are the most interesting for practical applications. They can be employed for detecting not more than one photon per gate time. They are well suited for gate pulses with short duration. The dc coupled type can work with duty cycle w limited only by thermal and trapping effects associated with the avalanche. For the ac coupled type, the circuit behavior sets a further limitation to w. It is worth noting that passive gated circuits, provided they are employed with a sufficient time interval between gate pulses, offer a feature of remarkable interest for investigations of SPAD behavior and also for some practical applications. In fact, the avalanche triggering always occurs without the interfering effects that are due to previous pulses as described in Section 2, because only the first event is detected. Therefore, simple gated passive circuits can be reliably employed to characterize the true timing accuracy of a SPAD device and to obtain uncompromised resolution in some photon-timing measurements, for example, in laser ranging applications.

E. Gated Active-Quenching Circuits

Adapting AQC's to gated operation[51] is straightforward: the external gate-off command and the internally generated quenching command can be applied to the input of an OR circuit whose output goes to the quenching driver. There is no limitation to long gate times T_g and, thanks to the low output resistance of the driver, short transition times can be obtained for the detector voltage. It was indeed verified early[51] that AQC's are almost ideally suited for effective and versatile gated operation of SPAD's. It is worth noting, however, that gated operation makes even more strict the requirement of avoiding any spurious triggering of the AQC because of forced transitions of the SPAD voltage. In fact, the transition from gated off to gated on is equal to the reset transition after quenching and, unless the circuit is properly designed and operated as illustrated in Section 4, it can generate a false output pulse just when the gate is open.

6. Circuits with Mixed Passive–Active Features

The quenching and reset transitions do not necessarily have to be both passive or both active and even a single transition can be managed partly in passive and partly in active ways. Such mixed solutions can be an effective approach to designing simple and compact circuits or for satisfying specific application requirements, or both.

A. Active-Quenching Circuits with Passive Reset

Looking for a simple solution to avoid overshoots and ringings in the recovery of the SPAD voltage may be a motivation for resorting to a passive reset in an AQC. An example is the circuit reported by Brown *et al.*[32] diagrammed in Fig. 13. A fast bipolar n-p-n transistor $Q1$ is switched on by the comparator and quenches the avalanche by pulling current through load R_L. After a delay just longer than the quenching transistion, a p-n-p transistor $Q2$ is switched on and resets the comparator, which then switches $Q1$ off. The fast switching diodes $D1$ and $D2$ are now both off and the avalanche photodiode with load R_L is isolated from the rest of the circuit. A low-resistance load $R_L = 1\ \text{k}\Omega$ is employed for smoothly pulling the voltage to the quiescent level, with a

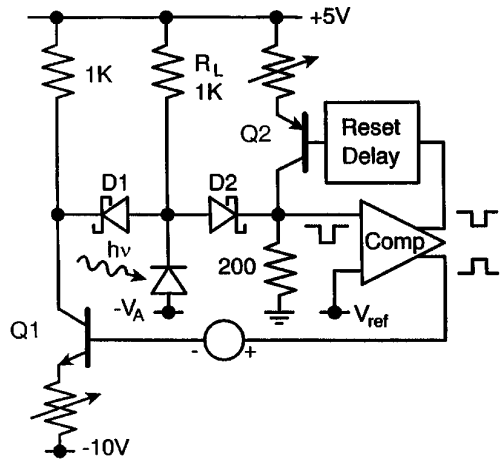


Fig. 13. Simplified diagram of the AQC with the passive reset reported in Ref. 32.

recovery time that can be made shorter than 20 ns by accurately minimizing the stray and detector capacitance. This circuit has been developed for photon correlation measurements to be performed with SPAD's with excess bias voltage V_E to 8 V and operating with a short dead time of less than 50 ns. The circuit is fairly simple and has been successfully employed in the application devised; however, it is not versatile. It is not suitable to work with high excess bias voltage because of the limits to the collector voltage and current of the fast transistor $Q1$. In fact, $Q1$ has to stand quiescently a voltage higher than V_E and must then be used to switch on a current that is high enough to produce a voltage drop at least equal to V_E on a 500-Ω total resistance. For example, with $V_E = 20$ V this means at least 40 mA. Furthermore, the presence of this steady current through $Q1$ in the quenched state complicates the introduction of a hold-off time (Section 2). Gated operation requires nontrivial further elaboration of the circuit design. In the gated-off condition $Q1$ and $Q2$ should be both maintained on and transition to the gated-on condition should be obtained by cutting off first $Q1$ and then $Q2$. The circuit does not appear suitable for remote detector operation, since it does not provide a matched termination to the connecting cables and does not have the capability of discharging and recharging the cable capacitance quickly.

B. Passive Quenching with Active Reset

Starting from the PQC configuration with voltage-mode output [see Fig. 4(a)] the evolution is straightforward. A circuit of this kind has been reported by Lightstone and McIntyre.[48] A fast voltage switch (a FET) is introduced between circuit ground and the node constituted by the SPAD cathode and the low end of ballast resistor R_L. The switch is quiescently open and the R_L value is high enough to guarantee passive quenching (see Subsection 3.B.). After avalanche quenching, when the detector voltage has to be restored, the switch is closed for a short time to recharge the diode and stray capacitance. The reset command is derived from the comparator that senses the avalanche pulse and is applied to the switch with a delay just longer than the time taken by the quenching transition. It is important to employ a FET with low capacitance and minimize all stray capacitances by placing a FET and load resistor close to the SPAD. A hold-off time can be simply enforced by introducing an additional delay after avalanche quenching, before applying the reset command. This compact circuit is not much more complex than the original PQC and provides a remarkably faster voltage recovery. It has, however, various limitations. Like the corresponding PQC configuration, it is not suitable for accurate photon timing since it exploits the voltage-mode output signal. The duration of hold off is limited to a fraction of recovery time constant T_r [Eq. (7)]. In gated operation, the circuit suffers limitations similar to those of PQC's (see Section 5). Since the gate command should be applied through a large coupling capacitor C_g [see Fig. 12(b)], the active reset transition would become much slower because the switch should also discharge C_g. Remote detector operation appears problematic: a FET switch remote from the SPAD must also discharge the large capacitance of the coaxial cable and provide a matched termination.

An actively reset scheme derived from the PQC configuration with current-mode output [Fig. 4(b)] might be interesting for applications that require accurate photon timing at high counting rates. However, the introduction of a reset FET switch in this configuration requires more complex modifications in the circuit. In fact, the FET switch should be connected to the SPAD terminal biased at high voltage. Furthermore, the reset pulse would inject a spurious signal through the SPAD capacitance at the comparator input, with the problems outlined dealing with AQC having opposite terminal configuration. In fact, as far as we know, no circuit of this kind has been reported.

C. Mixed Passive–Active Quenching

Mixed passive–active quenching[4,11] may be a convenient approach for minimizing the charge in the avalanche pulse, particularly for SPAD's having small diode resistance and small stray capacitance [see inequalities (14)]. A passive load R_L can provide a prompt drastic reduction (quasi-quenching) of the avalanche current. An active loop can complete the task after loop delay by forcing the SPAD voltage well below the nominal V_B, with sufficient margin to ensure final quenching and avoid reignition that is due to nonuniformity of V_B over the SPAD active area.[4] The active loop also forces a fast reset transition and makes it possible to introduce a controlled hold-off time. By minimizing the pulse charge, trapping and SPAD power dissipation can be minimized. Furthermore, this mixed approach appears particularly suitable for developing compact circuits. It is worth stressing that it gives flexibility in the choice of the passive load. Since a quasi-quenching action is sufficient, load resistance R_L can be smaller than the minimum value required for passive quenching (see Subsection 3.B.). An inductive load can also be employed to enhance the reduction of the avalanche current for a short time, covering the delay of the active feedback loop. However, to obtain satisfactory results and avoid overshoots and ringing that are due to the unavoidable combination of inductance and capacitances, such a circuit must be accurately analyzed and carefully implemented.

D. Mixed Active–Passive Reset

A mixed active–passive reset can be adopted to avoid overshoots and ringing in the voltage recovery. An active driver can be employed to bring the voltage close to the quiescent level swiftly and can then be switched off, leaving the task of completing the small residual part of the transition to a slower but

smoother passive reset. In our laboratory solutions of this kind have been tested with good results and are currently employed in AQC's.

7. Conclusions

Photon-counting and time-correlated photon-counting techniques provide the ultimate sensitivity and accuracy in measurements of weak and/or fast optical signals. Avalanche photodiodes, called single-photon avalanche diodes (SPAD's), have photon detection efficiency higher than photomultiplier tubes (PMT's), particularly in the red and near-IR spectral range. The spectral region covered with photon detection efficiency higher than 1% extends to ~1000 nm with silicon SPAD's and to at least 1600 nm with germanium and III–V compound semiconductor devices. With the available PMT photocathodes, the long-wavelength cutoff occurs between 600 and 750 nm for ordinary high-sensitivity photocathodes (S-11, S-20, etc.) and around 900 nm for GaAs negative-electron-affinity types. The old S-1 photocathode still has the widest near-IR range, but its photon detection efficiency is very low: from a peak value of ~0.2% at 850 nm it exponentially declines at higher wavelengths, falling to $\sim 10^{-6}$ at 1300 nm.[3] Recently, a new photocathode with photon detection efficiency of ~0.1% to 1600 nm was announced.[62]

In time-correlated photon counting, the neat time response of SPAD's, free from the secondary peaks and ringing that plague PMT's,[5,6,12,22,33–35,43] makes easier and more accurate the reconvolution analysis of measured waveforms. Decay time constants at least five times shorter than the FWHM resolution[14,15,19–21] can thus be accurately measured. The FWHM values obtained with various SPAD types, ranging from 100 to 350 ps, compare favorably with those of fast PMT's. The best FWHM obtained with ultrafast silicon SPAD's[5] is 20 ps, equivalent to or better than that of the most advanced microchannel-plate photomultipliers.[3]

The SPAD sensitive area is much smaller than that of PMT's but is nevertheless sufficient for most of the envisaged applications. In particular, the small area is an advantage for working with optical fibers, since it is perfectly matched to the fiber output and avoids the additional dark-counting rate associated with an unnecessarily wider area. SPAD detectors bring to photon counting the inherent advantages of solid-state devices: ruggedness, low supply voltage, low power consumption, low sensitivity to magnetic fields, and external disturbances in general. Thanks to their tiny size, they can be easily mounted in small receptacles within an apparatus, for example, in the interior of a microscope,[14–16] and can be easily cooled to cryogenic conditions.[26,42–44]

SPAD's must operate in association with quenching circuits. Passive, active, and gated quenching circuit configurations have been critically analyzed, and their relative merits in photon counting and timing applications have been assessed.

For continuous operation of the detector, PQC's are suitable only for low or moderate total counting rates (optical signal plus stray light plus dark counts). For photon counting with 1% accuracy, the counting-rate limit is at best ~200 kc/s. For photon timing that fully exploits the SPAD resolution, the limit is even more stringent, at best ~50 kc/s. The counting-rate limit is often exceeded by the background alone, that is, by dark counts and stray light reaching the detector. It is therefore concluded that simple passive circuits find fairly limited application. However, they still deserve interest for simple experiments and for tests of initial characterization and selection of the SPAD devices. Passive circuits are attractive because they are simple, they inherently limit the average current in the detector device (thus preventing excessive power dissipation and heating), and give easily detectable output pulses. Suitably designed AQC's make it possible to exploit SPAD's fully, which can be useful at performance levels that are less limited by the circuit and approach their intrinsic physical limits. The best-achieved counting dead times are 10 ns for low-voltage SPAD's and 40 ns for high-voltage types and operation at counting rates exceeding 40 and 10 Mc/s, respectively, has been verified. AQC's thus open the way to widespread application of these detectors.

For gated operation of the detector, the range of application of PQC's is broader. Essentially, passive gated circuits provide satisfactory performance in cases for which only the first photon that arrives in the gate interval has to be detected. They are very interesting for cases with very short gate times, in the nanosecond and subnanosecond range. They can also be fairly satisfactory with somewhat longer gate duration, not exceeding a few microseconds. Actively quenched gated circuits are better suited to gated operations in general, and they are absolutely mandatory in applications for which more than one photon has to be detected within a gate time.

In summary, by exploiting the available features of available quenching circuits and particularly AQC's SPAD's are no longer a laboratory curiosity. They are a solid-state alternative to PMT's for measurements of very weak and/or fast light signals and provide a higher sensitivity over a wider spectral range, particularly in the near infrared, and a better time resolution.

Appendix. List of Symbols

b	numeric factor
C_d	junction capacitance
C_g	gate coupling capacitance
C_s	stray capacitance
$D1$	fast switching diode
$D2$	fast switching diode
E_{pd}	energy dissipated in the SPAD during an avalanche pulse
f_a	gate pulse repetition rate
g	ratio between T_{gr} and T_g

h ratio between T_{gq} and T_g
I_d transient diode current
$I_{d\max}$ peak amplitude of the diode current
I_f asymptotic steady-state diode current
I_q latching current level
n_b dark pulse rate
n_p detected photon rate
n_t pulse repetition rate
P_g probability of having an avalanche within gate time T_g
P_L probability of having one or more pulses within time interval T_{pd}
$Q1$ fast bipolar n-p-n transistor
$Q2$ fast bipolar p-n-p transistor
Q_{ac} total charge in the actively quenched avalanche pulse
Q_{pc} total charge in the passively quenched avalanche pulse
R_d diode resistance
R_L ballast resistance
R_s low-value series resistance
T_a time interval between a gate pulse and the following one
T_{ac} avalanche pulse duration
T_{ad} dead time of an active-quenching circuit
T_{aq} quenching pulse rise time
T_{ar} reset transistion duration
T_g gate pulse duration
T_{gq} time constant of the voltage decay across the diode when a gate pulse is applied
T_{gr} time constant of the differentiator made by C_g in series with R_L
T_L circulation time in the quenching loop
T_{pd} dead time of a passive-quenching circuit
T_q quenching time constant
T_r voltage recovery time constant
T_w width of a square pulse approximating the actively quenched avalanche current signal
V_A bias voltage
V_B breakdown voltage
V_d transient diode voltage
V_E excess bias voltage
V_{ex} transient excess diode voltage
V_f asymptotic steady-state diode voltage
V_g gate pulse amplitude
V_g' actual amplitude of the gate pulse
V_n negative baseline offset
V_q peak amplitude of the current-mode output signal
V_u peak amplitude of the voltage-mode output signal
w gate pulse duty cycle.

This paper is based on extensive research that has been carried out for several years in our laboratory with the joint support of the Italian Space Agency, the Italian National Research Council, and the Ministry of University and Scientific and Technological Research. The stimulating discussions, constructive criticism, and help in carrying out experiments given by various colleagues and students, in particular, G. Ripamonti and I. Dal Santo, are gratefully acknowledged.

References and Notes

1. V. O'Connor and D. Phillips, *Time-Correlated Single Photon Counting* (Academic, London, 1984).
2. S. Cova, M. Bertolaccini, and C. Bussolati, "The measurement of luminescence waveforms by single photon techniques," Phys. Status. Solid A **18,** 11–62 (1973).
3. H. Kume, K. Koyama, K. Nakatsugawa, S. Suzuki, and D. Fatlowitz, "Ultrafast microchannel plate photomultipliers," Appl. Opt. **27,** 1170–1178 (1988).
4. H. Dautet, P. Deschamps, B. Dion, A. D. MacGregor, D. MacSween, R. J. McIntyre, C. Trottier, and P. P. Webb, "Photon counting techniques with silicon avalanche photodiodes," Appl. Opt. **32,** 3894–3900 (1993); SPCM-AQ Single-photon Counting Module Data Sheet (EG&G Optoelectronics Canada, Ltd., Vaudreuil, Quebec, Canada, 1994).
5. S. Cova, A. Lacaita, M. Ghioni, G. Ripamonti, and T. A. Louis, "20 ps timing resolution with single-photon avalanche diodes," Rev. Sci. Instrum. **60,** 1104–1110 (1989).
6. A. Lacaita, M. Ghioni, F. Zappa, G. Ripamonti, and S. Cova, "Recent advances in the detection of optical photons with silicon photodiodes," Nucl. Instrum. Methods A **326,** 290–294 (1993).
7. J. G. Rarity and P. R. Tapster, "Experimental violation of Bell's inequality based on phase and momentum," Phys. Rev. Lett. **64,** 2495–2498 (1990).
8. Y. H. Shih and C. O. Alley, "New type of Einstein-Podolsky-Rosen-Bohm experiment using pairs of light quanta produced by optical parametric down conversion," Phys. Rev. Lett. **61,** 2921–2924 (1988).
9. P. D. Townsend, J. G. Rarity, and P. R. Tapster, "Single photon interference in 10 km long optical fibre interferometer," Electron. Lett. **29,** 634–635 (1993).
10. N. S. Nightingale, "A new silicon avalanche photodiode photon counting detector module for astronomy," Exp. Astron. **1,** 407–422 (1991).
11. D. Bonaccini, S. Cova, M. Ghioni, R. Gheser, S. Esposito, and G. Brusa, "Novel avalanche photodiode for adaptive optics," in *Adaptive Optics in Astronomy,* M. Ealey and F. Merkle, eds., Proc. SPIE **2201,** 650–657 (1994).
12. L-Q. Li and L. M. Davis, "Single photon avalanche diode for single molecule detection," Rev. Sci. Instrum. **64,** 1524–1529 (1993).
13. S. A. Soper, Q. L. Mattingly, and P. Vegunta, "Photon burst detection of single near-infrared fluorescent molecules," Anal. Chem. **65,** 740–747 (1993).
14. T. A. Louis, G. Ripamonti, and A. Lacaita, "Photoluminescence lifetime microscope spectrometer based on time-correlated single-photon counting with an avalanche diode detector," Rev. Sci. Instrum. **61,** 11–22 (1990).
15. G. S. Buller, J. S. Massa, and A. C. Walker, "All solid-state microscope-based system for picosecond time-resolved photoluminescence measurements on II–VI semiconductors," Rev. Sci. Instrum. **63,** 2994–2998 (1992).
16. K. P. Ghiggino, M. R. Harris, and P. G. Spizzirri, "Fluorescence lifetime measurements using a novel fiber-optic laser scanning confocal microscope," Rev. Sci. Instrum. **63,** 2999–3002 (1992).
17. S. Cova, A. Longoni, A. Adreoni, and R. Cubeddu, "A semiconductor detector for measuring ultra-weak fluorescence decays with 70 ps FWHM resolution," IEEE J. Quantum Electron. **QE-19,** 630–634 (1983).
18. T. E. Ingerson, R. J. Kearney, and R. L. Coulter, "Photon counting with photodiodes," Appl. Opt. **22,** 2013–2018 (1983).
19. A. Andreoni and R. Cubeddu, "Photophysical properties of

photofrin in different solvents," Chem. Phys. Lett. **108,** 141–144 (1984).
20. A. Andreoni, R. Cubeddu, C. N. Knox, and T. G. Truscott, "Fluorescence lifetimes of angular furocoumarins," Photochem. Photobiol. **46,** 169–173 (1987).
21. T. A. Louis, G. H. Schatz, P. Klein-Bolting, A. R. Holzwarth, G. Ripamonti, and S. Cova, "Performance comparison of a single-photon avalanche diode with a microchannel-plate photomultiplier in time-correlated single-photon counting," Rev. Sci. Instrum. **59,** 1148–1152 (1988).
22. S. Cova, A. Lacaita, M. Ghioni, and G. Ripamonti, "High accuracy picosecond characterization of gain-switched laser diodes," Opt. Lett. **14,** 1341–1343 (1989).
23. B. F. Levine and C. G. Bethea, "Room-temperature optical time domain reflectometer using a photon counting InGaAs/InP avalanche detector," Appl. Phys. Lett. **46,** 333–335 (1985).
24. G. Ripamonti and S. Cova, "Optical time-domain reflectometry with centimetre resolution at 10^{-15} W sensitivity," Electron. Lett. **22,** 818–819 (1986).
25. G. Ripamonti, M. Ghioni, and A. Lacaita, "No dead-space optical time-domain reflectometer," IEEE J. Lightwave Technol. **8,** 1278–1283 (1990).
26. A. Lacaita, P. A. Francese, S. Cova, and G. Ripamonti, "Single-photon optical time-domain reflectometer at 1.3 μm with 5-cm resolution and high sensitivity," Opt. Lett. **18,** 1110–1112 (1993).
27. G. Ripamonti and A. Lacaita, "Single-photon semiconductor photodiodes for distributed optical fiber sensors: state of the art and perspectives," in *Distributed and Multiplexed Fiber Optic Sensors II,* J. P. Dakin and A. D. Kersey, eds., Proc. SPIE **1797,** 38–49 (1993).
28. M. Hoebel and J. Ricka, "Dead-time and afterpulsing correction in multiphoton timing with nonideal detectors," Rev. Sci. Instrum. **65,** 2326–2336 (1994).
29. I. Prochàzka, K. Hamal, and B. Sopko, "Photodiode based detector package for centimeter satellite laser ranging," in *Proceedings of the Seventh International Workshop on Laser Ranging Instrumentation,* C. Veillet, ed. (OCA-CERGA, Grasse, France, 1990), pp. 219–221.
30. F. Zappa, G. Ripamonti, A. Lacaita, S. Cova, and C. Samori, "Tracking capabilities of SPADs for laser ranging," in *Proceedings of the Eighth International Workshop on Laser Ranging Instrumentation,* J. J. Degnan, ed., NASA Conf. Publ. 3214 (NASA, Greenbelt, Md., 1992), pp. 5, 25–30.
31. R. G. Brown, K. D. Ridley, and J. G. Rarity, "Characterization of silicon avalanche photodiodes for photon correlation measurements. 1: Passive quenching," Appl. Opt. **25,** 4122–4126 (1986).
32. R. G. Brown, R. Jones, J. G. Rarity, and K. D. Ridley, "Characterization of silicon avalanche photodiodes for photon correlation measurements. 2: Active quenching," Appl. Opt. **26,** 2383–2389 (1987).
33. M. Ghioni and G. Ripamonti, "Improving the performance of commercially-available Geiger-mode avalanche photodiodes," Rev. Sci. Instrum. **62,** 163–167 (1991).
34. A. Lacaita, M. Ghioni, and S. Cova, "Double epitaxy improves single-photon avalanche diode performance," Electron. Lett. **25,** 841–843 (1989).
35. A. Lacaita, S. Cova, M. Ghioni, and F. Zappa, "Single photon avalanche diodes with ultrafast pulse response free from slow tails," IEEE Electron. Devices Lett. **14**(7), 360–362 (1993).
36. A. Lacaita, M. Mastrapasqua, M. Ghioni, and S. Vanoli, "Observation of avalanche propagation by multiplication assisted diffusion in p–n junction," Appl. Phys. Lett. **57,** 489–491 (1990).
37. A. Lacaita and M. Mastrapasqua, "Strong dependence of time resolution on detector diameter in single photon avalanche diodes," Electron. Lett. **26,** 2053–2054 (1990).
38. A. Lacaita, S. Cova, A. Spinelli, and F. Zappa, "Photon-assisted avalanche spreading in reach-through photodiodes," Appl. Phys. Lett. **62,** 606–608 (1993).
39. A. Lacaita, S. Longhi, and A. Spinelli, "Limits to the timing performance of single photon avalanche diodes," in *Proceedings of the International Conference on Applications of Photonic Technology,* G. A. Lampropulos, J. Chrostowski, and R. M. Measures, eds. (Plenum, London, 1994).
40. B. F. Levine and C. C. Bethea, "10-MHz single-photon counting at 1.3 μm," Appl. Phys. Lett. **44,** 581–582 (1984).
41. B. F. Levine and C. G. Bethea, "Single-photon detection at 1.3 μm using a gated avalanche photodiode," Appl. Phys. Lett. **44,** 553–555 (1984).
42. A. Lacaita, S. Cova, F. Zappa, and P. A. Francese, "Subnanosecond single-photon timing with commercially available germanium photodiodes," Opt. Lett. **18,** 75–77 (1993).
43. A. Lacaita, P. A. Francese, F. Zappa, and S. Cova, "Single-photon detection beyond 1 μm: performance of commercially available germanium photodiodes," Appl. Opt. **33,** 6902–6918 (1994).
44. F. Zappa, A. Lacaita, S. Cova, and P. Webb, "Nanosecond single-photon timing with InGaAs/InP photodiodes," Opt. Lett. **19,** 846–848 (1994).
45. B. F. Levine, C. G. Bethea, and C. G. Campbell, "Near room-temperature single photon counting with an InGaAs avalanche photodiode," Electron. Lett. **20,** 596–598 (1984).
46. R. H. Haitz, "Mechanisms contributing to the noise pulse rate of avalanche diodes," J. Appl. Phys. **36,** 3123–3131 (1965).
47. S. Cova, A. Lacaita, and G. Ripamonti, "Trapping phenomena in avalanche photodiodes on nanosecond scale," IEEE Electron. Devices Lett. **12,** 685–687 (1991).
48. A. W. Lightstone and R. J. McIntyre, "Photon counting silicon avalanche photodiodes for photon correlation spectroscopy," in *Photon Correlation Techniques and Applications,* Vol. 1 of OSA Proceedings Series (Optical Society of America, Washington, D.C., 1988), pp. 183–191.
49. R. H. Haitz, "Model for the electrical behavior of a microplasma," J. Appl. Phys. **35,** 1370–1376 (1964).
50. S. Cova, A. Longoni, and A. Andreoni, "Towards picosecond resolution with single-photon avalanche diodes," Rev. Sci. Instrum. **52,** 408–412 (1981).
51. S. Cova, A. Longoni, and G. Ripamonti, "Active-quenching and gating circuits for single-photon avalanche diodes (SPADs)," IEEE Trans. Nucl. Sci. **NS-29,** 599–601 (1982); presented at the IEEE 1981 Nuclear Science Symposium, San Francisco, Calif., 21–23 October 1981.
52. A. Lacaita, A. Spinelli, and S. Longhi, "Avalanche transients in shallow p–n junctions biased above breakdown," Appl. Phys. Lett. **67,** 2627–2730 (1995).
53. R. D. Evans, *Atomic Nucleus* (McGraw-Hill, New York, 1955), Chap. 28, pp. 785–793.
54. W. Nicholson, *Nuclear Electronics* (Wiley, New York, 1974), Appendix B5, pp. 473–376.
55. Ref. 54, pp. 259–260.
56. P. Antognetti, S. Cova, and A. Longoni, "A study of the operation and performances of an avalanche diode as a single photon detector," in *Proceedings of the Second Ispra Nuclear Electronics Symposium,* EURATOM Publ. EUR 537e (Office for Official Publications of the European Communities, Luxembourg, Belgium, 1975), pp. 453–456.
57. P. A. Ekstrom, "Triggered-avalanche detection of optical photons," J. Appl. Phys. **52,** 6974–6977 (1981).
58. S. Cova, "Active quenching circuit for avalanche photodiodes," U.S. patent 4,963,727 (20 October 1990) (Italian patent 22367A/88); licensed for industrial production to Silena SpA, Milano, Italy.

59. T. O. Regan, H. C. Fenker, J. Thomas, and J. Oliver, "A method to quench and recharge avalanche photodiodes for use in high rate situations," Nucl. Instrum. Methods A **326,** 570–573 (1993); H. C. Fenker, T. O. Regan, J. Thomas, and M. Wright, "Higher efficiency active quenching circuit for avalanche photodiodes," ICFA Instrumentation Bulletin No. 10 (Fermilab, Batvaia, Ill., December 1993), pp. 12–14.
60. Ultrafast comparators AD96685, Linear Product Databook (Analog Devices, Inc., P.O. Box 9106, Norwood, Mass., 1988), pp. 3–17.
61. B. K. Garside, "High resolution OTDR measurements," Photon. Spectra **22**(9), 79–86 (September 1988).
62. "1.7-μm near infrared photomultiplier" (patent pending), preliminary data sheet E(500) (Electron Tube Center, Hamamatsu Photonics KK, Hamamatsu, Japan, March 1994).