Actively quenched single-photon avalanche diode for high repetition rate time-gated photon counting

A. Spinelli(a) and L. M. Davis
Center for Laser Applications, University of Tennessee Space Institute, Tullahoma, Tennessee 37388

H. Dautet
EG&G Optoelectronics Canada, 22001 Dumberry Road, Vaudreuil, Quebec J7V 8P7, Canada

(Received 18 September 1995; accepted for publication 7 October 1995)

This paper reports an experimental characterization of the EG&G SPCM-AQ single-photon avalanche diode module with an active quenching bias circuit that gives a dead time of ~35 ns for use in high count rate, fast timing applications. A quantum efficiency of >70% and an optimal timing jitter with a full width at one-thousandth maximum of 1.5 ns is obtained if the light is tightly focused to a spot of <25 μm in the center of the active region, if the signal from the diode before the active quenching circuitry is used as input to the timing electronics, and if an external dead time of ~55 ns is imposed. Under these conditions, the total probability of obtaining an afterpulse is found to be ~2×10⁻³. Limitations of existing time-correlated single-photon counting instrumentation for count rates exceeding 10⁸ s⁻¹ are discussed. © 1996 American Institute of Physics. [S0034-6748(96)03001-9]

I. INTRODUCTION

Single-photon avalanche photodiodes (SPADs) are now in use for an increasingly large range of applications in which light detection with single-photon sensitivity is required. To 2 Compared to photomultiplier tubes, they offer several advantages, including ruggedness, low operating voltage, compact size, and low cost. Although avalanche diodes may also achieve single-photon sensitivity when biased below breakdown, fast timing and high single-photon detection efficiency have only been achieved when the diode is reverse biased above the breakdown voltage. Under these conditions, a single photogenerated electron–hole pair is able to initiate a divergent avalanche current pulse that is strong enough to be detected by the following electronics.

To make possible the detection of subsequent photons, after each avalanche initiation, the current must be quenched by lowering the applied voltage to a value below breakdown and the bias must then be reapplied. This process can be achieved in basically two ways: In a passive quenching arrangement, a large resistor, , is simply placed in series with the SPAD. The voltage drop across due to the avalanche current causes the voltage across the SPAD to decrease below breakdown, thereby quenching the avalanche. The SPAD bias voltage then recharges with a time constant given by the product of and the total (device and stray) capacitance of the SPAD. For a value of sufficiently large to quench the avalanche, the recharge time is typically in the microsecond range. During this time the SPAD has reduced response and thus the maximum available photon count rate is limited. Alternatively, in an active quenching arrangement, an active electronic circuit senses the leading edge of the avalanche current and rapidly lowers the bias. Then, after a fixed time, which may be as short as several nanoseconds, the bias is smoothly switched back on.

Active quenching can considerably reduce the SPAD dead time and is imperative for applications in which high instantaneous count rates must be monitored. In particular, for experiments to rapidly count single molecules in a flowing stream of solution by fluorescence photon burst detection, the dead time of the detector has been shown to significantly limit the signal acquired from each molecule and the rate at which molecules may be detected. In such experiments, subnanosecond time-gated photon detection is used to reject the dominant Raman scattering from the enormous number of solvent molecules surrounding the fluorescent molecule of interest. Also, since the number of fluorescence photons collected is quite low (~10), a detector with high quantum efficiency is needed.

SPADs can provide both high photon detection efficiency and ultrafast timing, but in general there is a trade-off between these parameters. Silicon SPADs with a reach-through design have demonstrated high quantum efficiency throughout the visible spectral range, due in part to the depth of the depletion region, typically ~30 μm. However, the best timing performance is ~150–200 ps. Considerably better timing, down to ~20 ps, has been achieved with shallow junction devices, but in this case the quantum efficiency in the visible is limited to ~40%. The different timing performance of each device geometry is largely due to the different mechanisms for the statistical growth of the avalanche currents. In reach-through devices with a larger active volume, the spreading of the avalanche is thought to be dominated by photon emission from the initial avalanche filament and reabsorption elsewhere in the active region, so that new avalanche filaments are formed around the initiation site. Complete breakdown occurs more slowly and with...
greater statistical fluctuations than in shallow junction devices.

The EG&G/RCA C3092 SPAD, with a reach-through geometry and an active diameter of \(\sim 500 \, \mu\text{m}\), has been extensively used for single-photon detection,\textsuperscript{12,13} although it was not designed specifically for this purpose. EG&G Canada subsequently released a photon counting module employing a new SLIK avalanche photodiode, fabricated in a reach-through geometry from silicon with super-low \(K\), where \(K\) is the ratio of the effective ionization coefficient of electrons to that of holes in the intrinsic region of the diode.\textsuperscript{14} The improved material reduces charge trapping effects and yields a higher probability of avalanche formation from a photogenerated electron–hole pair.

The SLIK diode from a EG&G SPCM-200 module has previously been characterized with a passively quenched circuit, modified for use in fast time-gating applications such as single-molecule detection (SMD).\textsuperscript{8} Considerable improvement over the original SMD experiments\textsuperscript{15} has been achieved by replacing the photomultiplier detector with this passively quenched SPAD.\textsuperscript{6} Monte Carlo simulations of the SMD experiment\textsuperscript{16} have indicated that the 0.8 \(\mu\text{s}\) dead time of the SPAD is the major factor limiting the signal acquisition and rate of molecule detection. More recently, EG&G released their SLIK detector within a new active quenching module, which yields a dead time of only \(\sim 35\) ns. The aim of this work is to accurately characterize the performance of the new EG&G SPCM-AQ active quenching detector, and to select suitable operating conditions for applications such as the SMD experiments.

II. THE DEVICE

The SPCM-AQ module includes the SLIK SPAD, which is mounted with thermistors on a two-stage thermoelectric cooler within a TO-18 can, the cooler driver circuitry, the high voltage converter needed to bias the SPAD above breakdown, and the active quenching circuitry, which triggers a transistor–transistor–logic (TTL) output pulse, all within a compact package. The package requires a single +5 V supply for operation. Preliminary measurements of the timing jitter of the SPCM-AQ module indicated a prompt with a full width at half-maximum [FW(1/2)M] of \(\sim 400\) ps, considerably poorer than that of the passively quenched device,\textsuperscript{8} and moreover, with an additional slow tail. A custom modified module, in which the output signal is taken before the TTL circuitry, was then provided by EG&G. In this case, as shown in the insert of Fig. 1, an operational amplifier is coupled by a capacitor to the load resistor of the SPAD to amplify the output and isolate the active quenching circuitry from the output load.

The shape of the output pulse, as digitized by a boxcar integrator (EGiG 4402) with 25 ps, 50 \(\Omega\) sampling head (Tektronics S4), is shown in Fig. 1. When viewed on a 1 GHz analog oscilloscope (Tektronics 7104) with high viewing intensity, for high photon count rates, additional output pulses can be seen at random times beginning as early as \(\sim 35\) ns after the beginning of the pulses that trigger the oscilloscope, that is, beginning at the point at which the tail of the pulse in Fig. 1 returns to zero voltage and thus being superimposed on the tail. The dead time of \(\sim 35\) ns allows photon counting rates of \(>10^6\) s\(^{-1}\) without appreciable loss of counts, as discussed in Sec. V.

The quantum efficiency of the detector was measured using laser beams of very low power at wavelengths of 585 and 632 nm. Multiple reflections within thick, slightly wedged, uncoated silica substrates at angles close to the Brewster angle are used to reduce the power of the beams from the lasers and a precision power meter (Laser Precision 5720/RKP575) is used to calibrate each split and to measure the initial laser powers. The resulting values of the quantum efficiency are 0.76 at 585 nm and 0.70 at 632 nm. These high values are comparable with those obtained by other experimenters.\textsuperscript{8,17}

III. CHARACTERIZATION OF AFTERPULSES

The phenomenon of afterpulses is well known in avalanche photodiodes operating in the Geiger mode. For the most part, it is a consequence of the presence of trap centers in the depleted region of the device. During current flow, these traps may be filled with free carriers, which are released at later times. The released carriers are able to trigger an avalanche pulse provided that the diode is reverse biased above breakdown at the time. For active quenching, afterpulses may also occur due to reignition of the avalanche by an incompletely quenched discharge current.\textsuperscript{14}

A sensitive method for monitoring afterpulses with time-correlated single-photon counting instrumentation is to connect the output pulses from the SPAD to both the start and stop inputs of the time-to-amplitude converter (TAC), with an additional short delay cable inserted before the start input, as shown in Fig. 2. The TAC performs a conversion only when there are two SPAD pulses in close succession. The first SPAD pulse starts the TAC, but since it had arrived at the stop before the TAC was started, the TAC is stopped on the subsequent SPAD pulse. The resulting timing spectrum represents the probability density function of the separation times between SPAD output pulses. If the SPAD were re-

![Figure 1](image-url)
Corresponding to a light source of constant intensity, the Poisson-distributed photons would have an exponential distribution of separation times, with a decay constant equal to the reciprocal of the count rate. If the mean count rate is kept low, only the beginning part of the exponential is observed, and the timing spectrum should appear to be constant, following an initial interval of zero or reduced counts due to the dead time of the SPAD. If the SPAD generates afterpulses with some small probability, additional counts will be superimposed on the flat spectrum at the characteristic times at which the afterpulses occur.

The result of such a measurement of the afterpulse timing spectrum, obtained using a white light source, a mean count rate of $1.94 \times 10^4$ s$^{-1}$, and a total collection time of 9979 s, is shown in Fig. 3. For clarity, only the first 100 ns of the 250 ns collected spectrum is shown. The major afterpulse contributions are, first, a short-lived component, which is attributed to avalanche reignition, beginning immediately after the dead time of the SPAD and falling with a time constant of $\sim 3$ ns, and, second, an exponential distribution with a time constant of $\sim 98$ ns, which is attributed to charge trapping within the SPAD. However, there are also several additional peaks that are most likely due to a retrigging of the comparator within the active quenching circuit by voltage fluctuations. Thus, most probably the small fluctuations in the output pulse shape seen in Fig. 1 are caused by fluctuations in the reset of the SPAD bias, which in turn are the cause of the additional peaks. Note that the afterpulse spectrum is found to be unaffected by adjustment of the external discriminator level and therefore the output pulse fluctuations themselves are not the origin of the additional afterpulses. Note also that the differential nonlinearity of the time-correlated single-photon counting instrument, which is separately measured by connecting a stable-frequency pulse generator to the TAC stop input, does not account for the additional peaks.

The total afterpulse probability is calculated by integrating the counts superimposed on the flat background of Fig. 3 and dividing this number by the total number of TAC valid start counts collected over the same time ($1.89 \times 10^8$). The afterpulse probability is found to be $\sim 4.1 \times 10^{-3}$. In time-gating applications such as SMD, if the total count rate caused by promptly scattered background light is $\sim 1 \times 10^6$ s$^{-1}$, the rate of nonprompt counts due to afterpulses will be $\sim 4 \times 10^3$ s$^{-1}$, considerably larger than the 180 counts s$^{-1}$ dark noise rate of the SPAD. Furthermore, peaks in the afterpulse spectrum may happen to fall exactly within the signal time window that follows a later laser excitation. In our case, with pulsed laser excitation at a rate of $7.6 \times 10^7$ s$^{-1}$ (i.e., 13.16 ns between laser pulses), afterpulses near $\sim 40$–48 ns fall after the third subsequent laser excitation pulse and thus cause additional bumps to appear in the tail of the detector prompt function. If an external electronic dead time of $\sim 55$ ns is added, as described in Sec. IV, the total afterpulse probability falls to $\sim 2.1 \times 10^{-3}$ and the additional bumps are removed.

**IV. TIMING JITTER**

The dependence of the prompt function or single-photon timing jitter of the detector on the experimental conditions was investigated using the setup shown in Fig. 4. The beam of picosecond pulses from a synchronously pumped dye laser (Coherent 702-1) is attenuated using a series of slightly wedged uncoated silica beamsplitters, calibrated as described in Sec. II, so as to give a SPAD count rate of $\sim 1 \times 10^6$ s$^{-1}$. The attenuated beam is focused onto the active region of the SPAD with a $\times 10$ microscope objective (Newport M-10). Care is taken to correctly aperture weak stray beams, which would otherwise become focused near the edges of the active region and cause the detector timing performance to become degraded. The output pulses from the detector module are passed to a leading edge timing discriminator (a Tennelec 455 quad constant fraction discriminator with a leading edge module), which has a threshold that is variable from $\sim 5$ mV to $\sim 1$ V. The output pulse width of the discriminator, which controls its dead time or blocking width, is adjusted to be $\sim 5$ ns, so that the majority of the detector
afterpulses are blocked and the small additional features in the tail of the prompt, described in Sec. III, are removed.

To measure the prompt, part of the beam from the laser is also directed to a fast PIN photodiode (Hewlett–Packard 4202), which is connected to a constant fraction discriminator (another channel of the Tennelec 455 with 0.2 fraction module). Discriminator pulses triggered by the SPAD and PIN photodiode are used to start and stop the TAC (Ortec 567). The TAC output passes through a biased amplifier (Tennelec 252) to a multichannel analyzer (Ortec ADCAM 921) for acquisition of the single-photon timing spectrum. The overall resolution of the instrument (1.49 ps/channel) is measured using a time calibrator (Ortec 462).

Minimum timing jitter of the detector is expected when the leading edge discriminator threshold is set closest to zero, as triggering of the SPAD output pulse then occurs early in the avalanche process before significant statistical differences in the growth of the avalanche current develop. However, electronic noise on SPAD output may cause false triggering if the absolute value of the threshold is too low. Because the active quenching circuit must switch tens of volts for each detected photon, there is considerable noise on the SPAD output caused by radio frequency pickup of transients within the compact module, particularly when the count rate is large. The noise impairs the use of threshold values above $2 \times 10^{-3}$ mV.

The prompt function of the detector was collected for different values of the discriminator threshold. Figure 5 shows the dependence of the FW(1/2)M, FW(1/100)M, and the FW(1/1000)M on the threshold. For time-gating applications such as SMD the FW(1/1000)M is the best indication of detector performance as it is close to the optimum setting for the time gate. As expected, better results are obtained at lower thresholds. Note however, that the best result, shown in Fig. 6, is considerably poorer than that obtained with a passive quenching circuit. A significant contribution to the timing jitter has been added either by the operational amplifier used to isolate the output pulse from the active quenching circuit or by the active quenching circuit itself. Improvements in the timing jitter may be possible with changes in the design of the electronics.

The dependence of the prompt function on the size of the
spot of light that is imaged onto the SPAD is an important factor to consider in many applications, including the SMD experiments. In most experimental setups, the smallest spot size achievable is limited by the size of the object space from which light is to be collected and the optical magnification that can be attained. To investigate the dependence of the prompt on the region of illumination, two experiments were conducted.

First, the SPAD was translated in a direction transverse to the optical axis while keeping the laser beam tightly focused. Figure 7 shows that the width of the prompt increases as the distance of the illuminated spot from the center is increased. By comparison, the central part of the active region of the SPAD over which the quantum efficiency is constant has a radius of $50 \, \mu m$. Such a degradation in timing could arise if the device were nonuniform, so as to have a lower electric field near the edges of the active region. However, even if uniform, better timing may arise for illumination in the center of the device because of the avalanche spreading mechanism, which is expected to be faster and with less statistical fluctuations if spreading can occur in all directions from a central initiation point.

Second, the SPAD was translated along the optical axis so as to defocus the Gaussian laser beam while keeping it centered on the active region of the SPAD. Beyond $12 \, \mu m$ from the focal plane of the $\times 10$ microscope objective, the half-angle divergence of the laser beam is $125 \, \text{mrad}$ and the size of the spot on the SPAD increases by $25 \, \mu m$ for each $100 \, \mu m$ of translation. From the results in Fig. 8, it can be seen that the prompt remains almost unchanged if the spot size is kept below a diameter of $25 \, \mu m$. However, as the size of the illumination region is increased beyond this value, the timing jitter degrades appreciably. The FW(1/2)M of the prompt becomes greater than that of the tightly focused illumination near the edge (Fig. 7), because for wide area illumination there is a statistical mixture of fast and slow avalanche initiations beginning near the center or the edge. Furthermore, as the Gaussian spot size becomes comparable to the $100 \, \mu m$ active region diameter, the tail of the prompt grows appreciably. This behavior is probably due to the generation of electron–hole pairs outside of the active region, where the field strengths are reduced.

V. PHOTON COUNTING AT HIGH RATES

For most experiments involving photon burst detection, unless photon antibunching occurs, during each short interval of high instantaneous count rate, the photons may be assumed to be Poisson distributed with an exponential distribution of separation times. If the instantaneous photon rate is $R \, \text{counts s}^{-1}$ and the dead time of the detector is $D \, \text{s}$, the fraction of photons that will arrive during the dead time and be missed is $1 - \exp(-RD)$. With the dead time of the SPAD module and external discriminator circuit increased to $55 \, \text{ns}$, instantaneous count rates up to $1.9 \times 10^6 \, \text{s}^{-1}$ can be utilized with only 10% loss. The module can actually be used at photon detection rates considerably higher, up to $2 \times 10^7 \, \text{s}^{-1}$, but the percentage of photons that go undetected increases.

For simple photon counting applications, low cost multichannel scalers can easily count at rates exceeding this value. However, for time-resolved photon detection, it is necessary to also measure the precise arrival time of each photon at high rates. Typically a mode-locked laser operating at $50–100 \, \text{MHz}$ is used to repetitively excite the sample and the nanosecond delay between each detected photon and the prior laser pulse is recorded. If the electronics used for mea-
suring the delay has a dead time that exceeds that of the
detector, then photons that are detected during the elec-
tronic dead time will be lost.

A time-to-digital converter can be used to measure the
delay. However, existing models have a typical dead time of
several microseconds and furthermore are limited in time
resolution to \( \sim 50 \text{ ps/channel} \). More commonly, a TAC is
used together with a biased amplifier and multichannel ana-
lyzer (MCA) to achieve whatever time per channel is desired
and a timing precision of \( \sim 10 \text{ ps} \), limited by the TAC. Com-
mercially available TACs have a dead time of \( \geq 1 \text{ \( \mu \)s} \), in
part, because the output pulse width must be adjusted to be
sufficiently long for the analog-to-digital-converter (ADC)
within the MCA. In turn, most MCAs require an input pulse
width of several microseconds and have a dead time exceeding
10 \( \mu \)s.

In our SMD experiments, a PC-based 12 bit ADC (Na-
tional Instruments EISA-A2000) is used to continually digi-
tize at 2 \( \mu \text{s} \) intervals the amplified output of the TAC. The
TAC is adjusted to produce output pulses of 1.5 \( \mu \text{s} \) width
and is strobed by the ADC clock so that digitization always
occurs at the same time after the beginning of an output
pulse. The resultant dead time is \( \sim 2 \text{ \( \mu \)s} \).

For time-resolved single-photon counting, due to the
dead time of the TAC, the photon detection rate must be
adjusted to be less than \( \sim 1\% \) of the laser excitation rate so
that the Poisson probability of detection of two photons after
each laser pulse is negligible. If two photons were detected,
only the first would be processed by the TAC and the re-
corded time spectrum would become biased, giving rise to
pile-up. Nevertheless, for laser excitation rates of \( \sim 100
\text{ MHz} \), photon detection rates of \( \sim 1 \text{ MHz} \) could be utilized if
a TAC and MCA with dead times of \( < 1 \text{ \( \mu \)s} \) were available.

In contrast, for time-gated photon counting, it is possible
to use a fast coincidence/anticoincidence circuit to process
detector pulses at rates exceeding 1 MHz. For example, the
EG&G Ortec model CO4020 provides nanosecond coinci-
dence/anticoincidence with a dead time of \( \sim 40 \text{ ns} \).

In recent experiments to rapidly detect single molecules in so-
lution using the passively quenched SPCM-200 SPAD, a
locally constructed anticoincidence circuit, with subnano-
second overlap resolution and dead time of \( < 10 \text{ ns} \), was used to
reject promptly scattered photons occurring at a rate of
\( \sim 3 \times 10^5 \text{ s}^{-1} \). During the single-molecule fluorescence
bursts, photons that are delayed from the excitation pulses
by more than \( \sim 1.5 \text{ ns} \) occur at a slower rate of \( \sim 2 \times 10^4 \text{ s}^{-1} \),
and so may still be processed by the TAC and ADC with
significant pile-up or loss. The fraction of photon events lost
due to the 0.8 \( \mu \text{s} \) dead time of the passively quenched de-
tector is calculated to be
\[
1 - \exp(-3.2 \times 10^5 \times 8 \times 10^{-5}) = 0.2.
\]

The new actively quenched detector is expected to en-
able SMD in the presence of an even higher rate of back-
ground from promptly scattered light. By operating at higher
laser excitation powers, it should be possible to achieve SMD
with a higher rate of fluorescence photon detection.

Towards this goal, Fig. 9 shows the single-photon timing
spectrum collected using the subnanosecond anticoincidence
before the TAC and MCA, for a prompt count rate of
\( \sim 1 \times 10^6 \text{ s}^{-1} \). Both the position and the width of the antico-
incidence gate are adjustable. With the exception that there is
no fluorescence, the geometry is essentially the same as that
of a SMD experiment; i.e., light from a 400 \( \mu \text{m} \) pinhole is imaged to a 40 \( \mu \text{m} \) disk in the center of the SPAD active
region with a \( \times 10 \) microscope objective. Although the tim-
ing resolution in Fig. 9 has deteriorated slightly from that in
Fig. 6 due to the larger illumination region, there is negligi-
gible loss of timing caused by the anticoincidence circuit. In
Ref. 22 it was shown that for the passively quenched SPAD,
photons that are detected before the SPAD bias has fully
recharged give rise to slightly delayed output pulses, which
cause an additional count rate dependent tail on the prompt.

With the actively quenched SPAD, there is no such effect,
and the shape of the time spectrum is observed to be inde-
pendent of count rate.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure9.png}
\caption{Time spectrum after anticoincidence gating of prompt photons.}
\end{figure}

\begin{thebibliography}{10}
\bibitem{1} A. D. MacGregor, in \textit{Applications of Photonic Technology}, edited by G. A.
Lampropoulos (Plenum, New York, 1995).
(1983).
(1986).
\bibitem{7} L. M. Davis and L. Q. Li, \textit{Laser Applications to Chemical Analysis}, 1994
Technical Digest Series Vol. 5 (Optical Society of America, Washington,
\bibitem{9} A. Spinelli, Ph.D. dissertation, Politecnico di Milano, Italia, 1995 (in Ital-
ian).
\bibitem{10} A. Lacaita, S. Cova, A. Spinelli, and F. Zappa, Appl. Phys. Lett. \textbf{62}, 606
(1993).
\bibitem{11} A. Lacaita, A. Spinelli, and S. Longhi, Appl. Phys. Lett. (in press).
\bibitem{14} H. Dautet, P. Deschamps, B. Dion, A. MacGregor, D. MacSween, R. J.
\bibitem{15} E. B. Shera, N. K. Seitzinger, L. M. Davis, R. A. Keller, and S. A. Soper,
\end{thebibliography}
Note that a chaotic or pseudochaotic light source has super-Poissonian photon statistics and its distribution of photon separation times has an additional fast exponential component with decay time equal to the temporal coherence. Thus a multimode laser with coherence length of ~1 m would not be a suitable light source for this measurement as the additional fast decay component due to the super-Poissonian photon statistics would overlap the SPAD afterpulse spectrum.

Note that SPCM-AQ modules are available with dark count rates as low as 5 s⁻¹ but the bias voltage and quantum efficiency may be decreased.


L. M. Davis and L. Q. Li, in Ref. 1.