Avalanche Detector with Ultraclean Response for Time-Resolved Photon Counting

Alessandro Spinelli, Massimo A. Ghioni, Member, IEEE, Sergio D. Cova, Fellow, IEEE, and Lloyd M. Davis

Abstract—Experimental tests have been carried out for characterizing the performance of a new single-photon avalanche diode. The detector is specifically designed for picosecond time-correlated single photon counting, aiming to obtain a time resolution free from the tail effects and/or secondary bumps observed in all other available single-photon detectors. The experimental results confirm that an unprecedented ultraclean response is attained, with full width at half maximum (FWHM) of 35 ps, full width at a thousandth of the peak count of 214 ps, together with a photon detection efficiency of 35% around 580-nm wavelength.

Index Terms—Avalanche photodiodes, photodetectors, photon counting, photon timing.

I. INTRODUCTION

Time-correlated single-photon counting (TCSPC) is widely adopted for measuring transient optical signals, such as multicomponent fluorescence decays and fast luminescence, because it provides a remarkable combination of ultrahigh sensitivity, fast response, [1], [2], and wide dynamic range. The time resolution is usually characterized by the full width at half maximum (FWHM) of the instrument response function (IRF), which is obtained by direct measurement of the excitation pulse. With an ultrafast laser as the excitation source, the IRF is primarily limited by the single-photon detector [2]. After a fluorescence decay profile is recorded, reconvolution analysis techniques may be employed to determine lifetime values that are a fraction of the FWHM of the IRF [1]–[3], but the achievable time resolution is still limited by the IRF. In particular, for accurate reconvolution analysis, it is important to obtain a clean IRF. TCSPC is frequently used to analyze complicated fluorescence decay profiles exhibiting components with disparate timescales. In this case, the fluorescence decay profile must be acquired over a wide dynamic range, and the IRF must maintain a narrow width over this range for accurate determination of all the components. Tailing effects in the IRF can obscure the presence of components with longer lifetimes. Hence, for measuring ultrafast and multicomponent fluorescence decays, for discerning close peaks in optical time-domain reflectometry (OTDR), and for performing fast-gated photon counting [4], it is important that the width of the IRF, typically at one-hundredth (FW(1/100)M) or one-thousandth (FW(1/1000)M) of the peak count, remains small. The development of single-photon detectors with such timing performance is therefore a key factor for the advancement of the TCSPC technique.

II. AVAILABLE SINGLE-PHOTON DETECTORS

Photomultiplier tubes (PMT) have been the traditional detector for use in the TCSPC. With dynode-string PMT’s, the available FWHM has evolved from about 500 ps in the 1960’s to about 200 ps in the late 1970’s [5]; values down to 115 ps have been obtained with monochromatic illumination of a small area using side-on PMT’s [6]. The development of the microchannel-plate (MCP) PMT in the 1980’s brought further progress; the FWHM value has improved from 300 ps down to about 20 ps [5], [7]. However, in all PMT’s and even in the most advanced MCP devices, the shape of the IRF is plagued with secondary bumps and irregular tails, and the FW(1/100)M and FW(1/1000)M values are quite long. Other drawbacks are the moderate or low values of the quantum efficiency (QE) of the available photocathodes, particularly in the red and near-infrared range, and the wide distribution of output pulse amplitudes. Because of the latter, a constant-fraction discriminator (CFD) must be used to obtain a standard timing pulse free of time-walk [1], [2]. Further, in order to obtain the optimum MCP time resolution, the CFD used must be quite sophisticated, and the range of the pulse amplitudes accepted by the CFD must be restricted, thereby further reducing the effective QE.

Also in the 1980’s, the single-photon avalanche diode (SPAD) was introduced as a cost-effective alternative to the MCP for use in TCSPC [8]. In contrast to PMT’s, SPAD’s directly provide an output signal of well-defined amplitude and offer the high QE typical of semiconductor detectors [2]. SPAD’s are basically avalanche photodiodes working above the breakdown voltage in Geiger mode. Active-quenching circuits make it possible to operate these detectors at high counting rates and to obtain optimal timing performance [9], [10]. The IRF typically has a fast main component and a slower tail [11]. The main component has a regular, Gaussian-like shape and is due to photons absorbed in the high electric field region, i.e., in the junction depletion layer. The slow tail is due to photons absorbed in neutral layers, which generate carriers that slowly diffuse and eventually reach the depletion layer with longer delays [12].

Silicon SPAD’s are nowadays well developed. The devices so far reported can be divided into two groups [10], [13], based on the depletion layer of the p-n junction, which can be either...
thick, ~20 μm and more, with a reach-through geometry, or thin, typically ~1 μm, constructed upon a substrate. Thick SPAD’s have fairly wide active areas (up to 500 μm in diameter) and are fabricated in high-resistivity high-quality silicon with special technologies that are not compatible with the standard planar processes of integrated circuits [13], [14]. They have a breakdown voltage, $V_B$, from 200 to 500 V, very high QE, and a fairly good IRF, with FWHM ~170 ps. The earlier reach-through devices have an IRF with a FWHM of 350 ps and a diffusion tail with amplitude two decades lower than the peak and exponential shape with 650 ps decay constant [15], [16]. The newer EG&G SLiK devices [13] have higher QE (better than 50% over the wavelength range from 540 to 850 nm) and improved IRF. FWHM values better than 170 ps have been reported [17], [10]. The diffusion tail is about one decade lower than the peak and has an exponential shape with a decay constant of 160 ps. Values for the FWHM of 650 ps and FWHM values down to 20 ps have been noted. The SLiK devices are commercially available in compact modules, including the bias and quenching circuitry. The model with the active-quenching circuit provides an IRF with 300 ps FWHM, 1000-ps FWHM, and 1500-ps FWHM [18], [19].

Thin SPAD’s have a smaller active area (up to 100 μm in diameter) and are fabricated with current good-quality silicon and planar technologies [20]. They have a breakdown voltage $V_B$ from 10 to 50 V, lower QE than thick SPAD’s, particularly in the red and near-infrared range, and a very good IRF, with FWHM values down to 20 ps [21]. The earlier devices had 70-ps FWHM and a fairly strong diffusion tail, with wavelength-dependent intensity and decay constant [11]. The development of epitaxial device structures in subsequent generations has progressively improved the FWHM value and reduced the tail. In double-epitaxial devices [20], the tail is reduced to an exponential with 270-ps lifetime, independent of wavelength, and an amplitude one decade lower than the peak; the FWHM of 270 ps and FWHM values of 1400 ps have been noted. The SUIC devices are commercially available in compact modules, including the bias and quenching circuitry. The model with the active-quenching circuit provides an IRF with 300 ps FWHM, 1000-ps FWHM, and 1500-ps FWHM [18], [19].

Thin SPAD’s have a smaller active area (up to 100 μm in diameter) and are fabricated with current good-quality silicon and planar technologies [20]. They have a breakdown voltage $V_B$ from 10 to 50 V, lower QE than thick SPAD’s, particularly in the red and near-infrared range, and a very good IRF, with FWHM values down to 20 ps [21]. The earlier devices had 70-ps FWHM and a fairly strong diffusion tail, with wavelength-dependent intensity and decay constant [11]. The development of epitaxial device structures in subsequent generations has progressively improved the FWHM value and reduced the tail. In double-epitaxial devices [20], the tail is reduced to an exponential with 270-ps lifetime, independent of wavelength, and an amplitude one decade lower than the peak; the FWHM of 270 ps and FWHM values of 1400 ps have been noted. The SUIC devices are commercially available in compact modules, including the bias and quenching circuitry. The model with the active-quenching circuit provides an IRF with 300 ps FWHM, 1000-ps FWHM, and 1500-ps FWHM [18], [19].

### III. The Double-Junction SPAD Detector

In order to obtain complete suppression of the diffusion tail, a new planar SPAD device has been specifically designed and fabricated with a patterned double-epitaxial structure [27], which is outlined in Fig. 1. The buried p++ layer, which provided the low resistance path for the avalanche current in the double-epitaxial SPAD [20], is now interrupted in the region immediately below the active area of the diode. Consequently, the depleted region of the junction between the n-substrate and the p-epistrate is further extended into the epitrate, toward the active depletion layer of the n++-p++ junction. The peak electric field is lower and the depletion layer is wider than in the double-epitaxial SPAD, and hence the breakdown voltage is increased to 27 V. As the reverse bias on the n++-p++ junction is increased, the thickness of the p-neutral region sandwiched in between the two junctions becomes reduced. At sufficiently high overvoltage the two depletion layers meet and complete depletion is reached. Under these conditions, it is expected that the tail contribution will disappear and the IRF will be due only to carriers generated in the active depleted region of the upper junction, terminated by the lower substrate junction. The new device is therefore called double-junction single-photon avalanche diode (DJ-SPAD).

Preliminary tests, performed with a 20-ps laser diode, confirmed the validity of the approach, but the inherent skirts in the laser pulse prevented an accurate characterization of the detector performance. In this paper, extensive measurements were carried out using a high-performance laser system in order to accurately determine the detector behavior and evaluate its performance for picosecond TCSPC.

### IV. Experimental Results

A standard TCSPC setup, as shown in Fig. 2, was used to measure the timing jitter of the detector. A mode-locked dye laser (Coherent 702-1) provides pulses at a repetition rate of 76 MHz of duration 4 ps, as measured with a background-free autocorrelator (Femtochrome 103). These pulses are then severely attenuated to yield a SPAD count rate <10^5 s^-1, so that the probability of more than one photon per pulse reaching the SPAD is negligible. To ensure that the timing performance is not degraded by the introduction of collinear ghost beams, attenuation is accomplished by the use of multiple reflections.
the beam. The absence of a glass window covering the SPAD makes it possible to use a small-working-distance microscope objective to tightly focus the attenuated beam onto the active area of the SPAD. (In the results that follow, a Newport ×20 objective was used, although it is possible to use a ×60 objective with 300-μm working distance.)

A custom-designed active quenching circuit (AQC) [9] provides both the necessary quenching of the SPAD avalanche current and a standard NIM-negative output pulse, which is used to start the time-to-amplitude converter (TAC) (Ortec 567). The TAC stop pulses are obtained from the laser pulses by means of a PIN photodiode (Hewlett-Packard HP-4202). When reverse biased at 80 V, this produces negative pulses with a rise time of ~1 ns and amplitude of ~1.2 V into 50 Ω. If laser power fluctuations cause amplitude fluctuations in these reference pulses, a constant-fraction discriminator (CFD) (Tennelec 454) must then be used to avoid time-walk at the 250 mV leading-edge discriminator of the TAC input. However, a slightly sharper IRF is obtained when the laser is operated in light-control and the PIN photodiode pulses are used to directly trigger the TAC stop channel. The TAC is operated in reverse mode at its maximum resolution. The output from the TAC passes through a biased amplifier (Tennelec 252) to a personal-computer-based multichannel analyzer (MCA) (Ortec Adcam 918A). The time calibration of 1.8 ps/channel is obtained with the use of a 10-ns time-calibrator (Ortec 462).

Fig. 3 shows the IRF measured with a SPAD device having an active area diameter of 10 μm, reverse biased at 48 V, i.e., with 21-V excess bias voltage above the breakdown level. To better observe the nature of the tail of the IRF, the experimental data are plotted on a semilogarithmic scale, after subtraction of the small constant background level caused by dark counts (27 counts/channel). It is evident that the detector response is absolutely free from secondary peaks and practically free from the slow tail that exists in other detectors. In fact, the tail is reduced to a small and very fast exponential component with an amplitude two decades lower than the peak and a decay constant of ~30 ps. The latter value is consistent with the computed values for the diffusion of charge carriers photogenerated in the top n+ layer [11], [12], [20]. The thickness of this neutral layer is only 0.3 μm and further significant reduction does not appear attainable, since the n-doping density is already very high, more than 10^{19} cm^{-3}.

It was verified in other experiments that this ultraclean response makes it readily possible to perform measurements practically unattainable with other detectors. For instance, it is possible to resolve and measure a weak scattered or reflected pulse, occurring a few hundred picosecond after a main light pulse, and with an intensity three orders of magnitude weaker [27].

The measured parameters from the IRF are reported in Table I. Several different DJ-SPAD devices with similar structures were tested and the IRF was found to be well reproducible and uniform. The ultrafast performance of the DJ-SPAD is better appreciated when compared with other high-performance single-photon detectors. Tests were carried out also with an EG&G SLiK SPAD, using a similar TCSPC setup, but with a fast laser diode source. The pulses from this source [27] had 20-ps FWHM and weak skirts, about two decades lower than the peak intensity, and were therefore not adequate for an accurate characterization of the DJ-SPAD. However, in previous tests performed with the mode-locked dye laser described above [17], the IRF measured with the EG&G SLiK had 170-ps FWHM and a diffusion tail with 160-ps decay time (corresponding to about 0.8-μm thickness of the neutral p+ layer). The diode laser source was therefore considered adequate for testing the EG&G SLiK. A custom-designed active-quenching circuit was employed, which made it possible to operate the EG&G SLiK at a bias voltage of 450 V, i.e., 40 V above breakdown. The IRF parameters obtained in these conditions (see Table I) are significantly better than those previously reported [10], [17], but the timing performance is nevertheless far from that of the DJ-SPAD. Fig. 3 also shows the IRF obtained in tests previously performed [27] with a Hamamatsu R2809U MCP and a mode-locked dye laser with 5-ps pulse duration. As seen in Fig. 3, the Hamamatsu R2809U MCP has an IRF with a slightly shorter FWHM, but it is plagued by irregular bumps and tails, which have a different shape for each particular MCP detector. The ratio of the peak-to-tail amplitude was less than one decade for the MCP sample used in these tests, although better values of this ratio have been reported [7]. The recently introduced Hamamatsu R3809U-50 has a fairly cleaner shape and correspondingly improved values of the FWHM and FW(1/1000)M. Table I gives the best values reported in the literature for the Hamamatsu R3809U MCP.
and of 214 ps are remarkably better than that of previous devices [21], [28] because of the wider depletion layer and lower electric field adopted in this fabrication run. However, this is not an intrinsic limitation of the new device structure.

The QE for detection of single photons by the device was measured in the spectral range from 585 to 605 nm. The QE value is not as high as that of SPAD’s with a thick depletion layer, such as the SLiK, but it is nevertheless remarkably better than that of most PMT’s and MCP’s. Further, the useful spectral response is expected to follow the absorption spectrum of silicon and hence extend to about 1050 nm.

V. CONCLUSION

The experimental tests confirm that the DJ-SPAD meets the goals set in the design of the device structure. In single-photon timing, the new device provides an IRF with ultraclean shape and steep decay over more than four decades. The FWHM of 35 ps is comparable to values obtained with the fastest available single-photon detectors and it may possibly be improved with further design refinement. The FW(1/100)M of 125 ps and FW(1/1000)M of 214 ps are remarkably better than any other previously reported results. Such timing performance is particularly valuable in fast-gated photon counting and in measurements of fast luminescent decays by the TCSPC technique, in particular in experiments for accurately resolving decay components with small intensity in the presence of stronger ones. The QE of 35% is remarkably better than that available with MCP’s and PMT’s in the visible spectral range. The small sensitive area, typical of fast semiconductor detectors, is acceptable in many applications, particularly since a high magnification objective can be used to image light from a larger region onto the detector. In some experiments, it may represent an advantage, since it makes it easier to avoid the detection of stray light. The detector therefore appears ideally suitable for TCSPC applications requiring high timing resolution and high counting rate capability.

REFERENCES


Alessandro Spinelli was born in 1966 in Bergamo, Italy. He received the Laurea and the Ph.D. degrees in electronics engineering from the Politecnico of Milano, Milan, Italy.

He is currently working as a Researcher at the Politecnico di Milano. In 1995, he was a Visiting Scholar at the University of Tennessee Space Institute, Tullahoma, where he worked on single molecule detection in solution. In 1996, he was a consultant for SGS-Thomson Microelectronics, Central R&D Department, Agrate Brianza, working on quantum effects in MOS transistors. His current research interests include modeling of advanced MOS devices, experimental characterization of MOS electrical parameters, and avalanche photodiodes for single-photon detection.

Massimo A. Ghioni (M’91) was born in Monza, Italy, in 1962. He received the Laurea Degree in nuclear engineering from the Politecnico di Milano, Milan, Italy, in 1987.

Since 1990 he has been Assistant Professor of Electronics at the Politecnico di Milano. In 1992, he was Visiting Scientist at the IBM T.J. Watson Research Center, Yorktown Heights, NY, working on silicon photodetectors for integrated optical receivers. His current research interests include the design of avalanche photodiodes for single-photon detection and the development of fast electronic circuits for picosecond timing applications.

Sergio D. Cova (M’71–SM’82–F’92) was born in Rome, Italy, in 1938. He received the doctor degree in nuclear engineering from Politecnico di Milano, Milan, Italy, in 1962.

He has been Full Professor of Electronics at Politecnico di Milano since 1977 and has taught courses in electronics, microelectronics, and physics at Parma University and Bari University, Italy. He has contributed to the fields of physics of semiconductor detectors for optical and ionizing radiations, microelectronics, and electronic and optoelectronic measurement instrumentation and has collaborated in interdisciplinary research in physics, astronomy, cytology, and molecular biology. He brought innovations to spectroscopy amplifiers and fast timing instrumentation in nuclear electronics. Long before the advent of powerful microcomputers, he devised and demonstrated a digital lock-in detection method for analog signals. He pioneered the development of single-photon counting with photomultipliers and with single-photon avalanche diodes (SPAD). He invented and developed the active-quenching circuit for exploiting fully the performance of SPAD’s up to a high counting rate. He designed new silicon SPAD structures for picosecond photon timing and experimented Germanium and III–V compound semiconductor SPAD’s for extending photon counting in the infrared spectral range. He demonstrated optical time-domain reflectometry in fibers with millimeter resolution and high sensitivity (Rayleigh scattering detection). He has published over 120 papers in international refereed journals and conferences and is the author of two international patents (USA and Europe).

Dr. Cova has served on committees of the IEEE North Italy Section and on the Board of the Italian Electrical Engineering Society Milano Section.

Lloyd M. Davis was born in Suva, Fiji, in 1957. He received the Ph.D. degree in physics from the University of Auckland, New Zealand, in 1985.

He is Associate Professor of Physics with the Center for Laser Applications at the University of Tennessee Space Institute, Tullahoma, where he conducts research in experimental physics and laser technology, specializing in the areas of single molecule detection and ultrafast spectroscopy, quantum, nonlinear, and applied optics, and ultrafast phenomena. He has been a visiting researcher and collaborator at Los Alamos and Oak Ridge National Laboratories and the U.S. Army Laboratory in Huntsville, AL.

Dr. Davis was a member of a team to receive a “Research and Development 100 Award” in 1991 for the first demonstration of the detection of single chromophore molecules in solution.