OSA Proceedings on
Picosecond Electronics
and Optoelectronics

Volume 4

Edited by
T. C. L. Gerhard Sollner and David M. Bloom

Proceedings of the OSA Topical Meeting,
March 8 – 10, 1989,
Salt Lake City, Utah

This topical meeting was cosponsored
by the Optical Society of America
and by the Lasers and Electro-Optics Society
of the Institute of Electrical and Electronics Engineers

Optical Society of America
1816 Jefferson Place, N.W.
Washington, DC 20036
(202) 223-8130
Abstract

Single Photon Avalanche Diodes (SPADs) are avalanche photodiodes specifically designed for reverse bias operation above the breakdown voltage (Geiger-mode operation), and used for detecting single optical photons. Studies were performed to relate the attainable timing resolution to the device geometry and operating conditions. A new silicon device structure was designed in order to obtain improved timing performance with respect to previous SPADs.

Extensive tests were carried out in order to ascertain the timing resolution of the device in time-correlated photon counting. The SPAD timing resolution, in terms of its full-width at half-maximum (FWHM) contribution to the overall instrumental response width, is 20 ps with the detector cooled to -65 °C, and 28 ps at room temperature. This is the highest resolution so far reported for solid-state single-photon detectors.

Among vacuum tubes, comparable results are obtained only with special microchannel-plate photomultipliers (MCP-PMT). With the excellent timing resolution of the SPAD and the well-known advantages of Time Correlated Photon Counting systems (high sensitivity, linearity etc.), various applications are foreseen in areas so far dominated by streak cameras.

INTRODUCTION

Single Photon Timing techniques have proven to be of considerable interest in various fields. Remarkable performances have been reported in optical fiber characterization by Optical Time-Domain Reflectometry, in laser ranging, in fluorescence and luminescence decay measurements. One of the most interesting properties of the technique is the attainable time resolution, nowadays in the tens of picosecond range. The main limit to the time resolution lies in the detector. In fact, since the advent of the synchronously pumped dye laser, the generation of light pulses of a few picosecond duration is standard. By proper circuit design, the electronic jitter in the measurement set-up can also be reduced to \( \approx 10 \) ps.

On the detector side, time resolutions of the order of some tens of picoseconds were demonstrated with Photomultiplier Tubes (both discrete-dynode and Microchannel-Plate) \([1-3]\) and also with avalanche photodiodes. Geiger-mode single-photon detectors have been demonstrated in silicon, germanium and III-V compounds \([4-7]\).

In this paper we point out the design strategies aimed at minimizing the time jitter of a single-photon avalanche diode (SPAD). The attainable resolution turns out to depend on a number of specifications, often setting conflicting requirements. Other important parameters of the detector, such as the quantum efficiency and the noise (i.e. the dark-count rate) depend on the same quantities. Therefore trade-offs in the device design must be taken into account.

By exploiting these strategies, we were able to design detectors having better than 20 ps full-width at half maximum (FWHM) resolution.

SINGLE-PHOTON AVALANCHE DIODES

Single-Photon Avalanche Diodes (SPADs) are p-n junctions that operate at reverse bias above the breakdown voltage. In this bias condition, a single carrier can trigger a self-sustaining avalanche current. In case of a photogenerated carrier, the leading edge of the avalanche pulse marks the arrival time of the detected photon. The current will continue to flow until an external quenching circuit \([8,9]\) lowers the bias close to or below the breakdown voltage. The
At the breakdown voltage, the multiplying region can be considered as a positive feedback amplifier having unit loop gain. In this situation the average number of carriers in the multiplying region is constant in time. Since we were comparing samples having different BV, the normalization to the breakdown allows the performance comparison in similar multiplication regimes. The choice of the electric field depends on the consideration that the ionization rate is a strong (exponential) function of the electric field. For this reason, most of the ionizations happen in the highest electric field region. Since the samples have very similar geometries of the multiplying region (a shallow diffused junction), the maximum electric field well defines the multiplication process.

From Fig.2 we can deduct the formula:

$$\text{FWHM} = A \cdot (E - E_b)^{-k}$$

where $E_b$ is the maximum electric field at the breakdown voltage; $k$ is 0.5 in shallow diffused junctions; $A$ is a parameter dependent on the device breakdown voltage, that will be discussed in the following [15].

It is evident from Fig.2 that a high excess field, and therefore a bias voltage well above the breakdown, strongly increases the timing resolution. One could also like to increase the voltage as much as possible in order to increase the quantum efficiency.

In a single photon detector, the quantum efficiency is defined as the probability that a photogenerated carrier will succeed in triggering a pulse, that is, a self-sustaining avalanche. Apart from geometrical factors (in particular the depletion region width), the quantum efficiency is known to increase with the electrical field, since the ionization probability increases strongly with the field. The problem was studied by Oldham et al. [6], who found the following relations:

$$\text{d}P_e/\text{d}x = (1 - P_e) \alpha_e [P_e + P_h - P_eP_h]$$

and:

$$\text{d}P_h/\text{d}x = - (1 - P_h) \alpha_h [P_e + P_h - P_eP_h]$$

Here the alphas are the ionization coefficients, whereas $P_e$ and $P_h$ are the probabilities that an electron and an hole respectively will have an infinite number of descendents, i.e. will trigger an avalanche.

We have studied the avalanche trigger probability for our structures. We obtained the ionization coefficients by using Thornber's formula [17] with Grant's coefficients. The results for the average probability of a carrier generated in the depletion layer to trigger an avalanche are shown in Fig. 3. From Fig. 3 we note that a 80 % triggering probability is possible at the highest electric fields.

Two limits to the maximum field are met in the SPAD design. The first one is the occurrence of breakdown in regions outside the active region. This can be due to localized breakdown in the edges of the diffusions, or to the breakdown of the guard rings. Edge breakdown, for example, happened in batch 3 at 34 V, thereby limiting the employable excess bias to 6 V. With a correct device design, in particular by using a shallow guard ring and a metal field plate, this limitation can be avoided. In batch 9, for example, the guard ring breakdown was 40 V, thus allowing 27 V excess bias.

However, a second more fundamental limitation to the excess bias is present. It is in fact well-known that at very high electric fields the generation rate is increased due to a number of physical effects, such as the Frenkel-Poole effect and the phonon-assisted tunnelling. Due to these effects, the dark count rate of the SPAD strongly increases with bias. Since the low-breakdown voltage diodes have an higher breakdown electric field, the dark count rate is much higher for these SPADs. In our case, the maximum applicable voltage was limited by guard ring breakdown in batch 3 and by too high a dark count rate in batch 9.

The maximum tolerable dark count rate depends on various considerations. First, a higher dark count rate implies a lower signal-to-noise ratio in single-photon measurements. Second, there is an increasing probability for an avalanche caused by a photon to happen when the diode is recovering from a previous dark avalanche pulse. The timing resolution is in this case impaired.

We have been able to make single-photon experiments with dark count rates up to 100 kpps, by using an active quenching circuit [9] which we developed in our laboratory. If simpler passive quenching circuits were used, the dark count rate must be lower than some kpps.

In order to reduce the dark count rate to a tolerable value, the depleted region volume must be small. For a given doping profile, this means a smaller sensitive area. We experimented with circular active areas of 10 $\mu$m diameter.

![Figure 3. Calculated triggering probability vs. maximum electric field for two SPAD structures.](image-url)
Note that this value is close to the dimension of a single channel in a microchannel-plate photomultiplier [3].

**Multiplying region depth.**

As shown in Fig. 2, the best timing resolution was achieved with the diodes having the lowest breakdown voltage, that is, the smallest depth of the multiplying region. The reasons why this happens can only be qualitatively understood at this point.

First, we note that the limit resolution must be of the order of the transit time in the avalanching region (this limit would be attained with a multiplication of such high value, that the number of free carriers rise to a detectable value within the transit time of the photogenerated primary carrier). This consideration suggests that the thinner the multiplication region, the better the timing resolution.

The multiplying region in the SPADs has a field that varies almost linearly with the position. The lower the breakdown voltage, the steeper the slope. In relatively high breakdown voltage devices, the multiplying region with lower field extends longer. If the fluctuations in the ionization process in this lower field region have importance to the overall avalanche build-up statistics, the resolution will be reduced. In order to check this hypothesis, we are developing a Monte Carlo simulation of the process.

In Fig. 4 a typical result of the simulation of the avalanche build-up is shown, as obtained with a first version of the program. The program implements the case of an avalanche region with an uniform field. One can thus simulate the different devices only by changing the multiplying region thickness. The timing resolution can be inferred from the spread of the delay from the photogeneration of the primary carrier to the crossing of a given avalanche current level. The primary carriers are generated at random positions within the depletion layer. Even in this case, however, we noted a dependence of the timing resolution on the multiplying region thickness. We are now working on a second version of the program that will take into account the actual field profile.

Two main points are worth stressing. First, the absolute maximum resolution of the SPAD appears to be a few picoseconds. Second, it appears possible to improve the resolution beyond 20 ps FWHM by tailoring the field profile.

**Active region uniformity.**

As previously discussed, there is a trivial limit to the sensitive area, that lies in the dark count rate. In addition, too large a sensitive area could cause other effects that contribute to reduce the timing resolution.

The previously mentioned simulation of the avalanche build-up pointed out that a 10% non-uniformity of the electric field could cause a significant increase of the FWHM resolution. Experiments performed by us with SPADs having different dimensions of the active area confirmed the existence of nonuniformity effects.

Another very important problem, that probably sets a limit in the maximum sensitive area, is the time needed for the propagation of the avalanche pulse over the whole area. This was first found by McIntyre on RCA reach-through avalanche photodiodes [18], and confirmed by us [19]. Whether the horizontal propagation of the avalanche pulse is subbandgap photon assisted, phonon assisted or simply due to the diffusion of the avalanching carriers perpendicular to the electric field, is not yet clear at this point. We are performing experiments to understand this point.

In summary, the sensitive area must be kept small for a number of reasons. A 10 micron diameter was employed in our experiments, and we found no difference of timing resolution on these diodes as compared with devices with even smaller sensitive area that had a lower dark count rate.

**Temperature effects.**

Finally, the operating temperature was observed to influence the observable resolution. In particular, a better resolution was observed for lower temperatures. We carried out experiments in the temperature range -60 C to +20 C by using a controlled temperature chamber. We noted the well-known temperature dependence of the breakdown voltage, and therefore of the breakdown field, and a timing resolution improvement of ~25% between the two temperature extremes.

The reason why this change in the attainable resolution is observed lies probably in the greater efficiency of the ionization process at reduced temperature, as compared with other energy-exchanging processes, such as optical phonon emission.

A low temperature is also known to reduce the thermal carrier generation. However, the reduction in the dark-count rate with temperature does...
not exactly follow the reduction in the generation rate. In fact, deep trap levels release with longer time constants at lower temperature. Therefore, levels filled by an avalanche pulse can release the stored charge at with longer delay time, possibly longer than the hold-off time of the active quenching circuit, thereby increasing the dark count rate [20]. Furthermore, tunnelling effects are not sensitive to temperature, except those assisted by phonons.

EXPERIMENTAL RESULTS

Based on the results of the preceding section, we designed and implemented SPADs having the lowest breakdown voltage compatible with the dark count rate requirement. The aforementioned batch 9 is the result of this design.

The FWHM resolution of the new SPAD is expected to be a few tens of picoseconds. On this time scale, other sources of time dispersion will make significant contributions to the measured overall instrumental FWHM resolution. In order to obtain the true detector resolution, all other contributions must be independently measured and then quadratically subtracted. Errors of the order of a few percent in these terms may well be significant. Possible alternative experimental set-ups were therefore carefully evaluated and results compared, in order to identify individual sources of additional timing jitter and quantify their contribution to the overall instrumental FWHM resolution.

In our standard waveform measurements with a single-photon timing set-up employing a synchronously pumped dye laser, the total additional time dispersion was about 35 to 40 ps FWHM, mostly due to the synchronization jitter in the start channel. By using a gain-switched laser diode we estimated a somewhat higher dispersion, due to the width of the optical pulse.

In order to better exploit the ultrashort pulses of the synchronously pumped dye laser, the contribution due to the synchronism jitter was avoided by using two SPADs for the two branches of the single-photon timing set-up, and thus measuring the autocorrelation of the SPAD resolution. In this case we obtained an instrumental additional FWHM of 25 ps.

In the experiments a cavity dumper reduced the pulse repetition rate to 30 kHz. Neutral density filters were used to reduce the signal intensity to below the single photon level. The signal intensity was limited to produce a photon count rate 1.5 kcps (5% of 30 kHz) in order to prevent pulse pile-up effects. Since the dark-count rate of the SPADs exceeds 100 kcps at room temperature, the high dark count rate caused some complication. Dark pulses in the start branch activate the time to pulse height converter (TPHC), causing an increased data collection time. Therefore, in order to avoid most of the useless starts, the TPHC fast gate facility was exploited. Another split off portion of the laser pulse train was directed to the fast p-i-n photodiode (HP 5082-4220) and constant-fraction discriminator (CFD) normally employed as start branch in the waveform measurements. The CFD output triggered a monostable circuit, that applied to the TPHC gate input a 100 ns square pulse, synchronized with the laser pulse. Start pulses were accepted only if they arrived within this 100 ns window.

Satisfactory histograms were collected in 10 min or less. A 46 ps FWHM autocorrelation curve was obtained, with a data collection time of 600 s. By quadratic deconvolution of the 25 ps contribution of the electronics, the FWHM resolution of the SPAD response at room temperature is therefore demonstrated to be 28 ps.

Extensive experiments were then carried out with fast laser diode pulsers. These results were compared with those obtained in the usual way, with an ultrafast p-i-n photodiodes connected to a sampling oscilloscope. We found results consistent with the SPAD resolution obtained. In fact, the obtained histograms had a lower FWHM value than the ones measured with the sampling oscilloscope.

In order to determine the width of the laser pulse, the 28 ps FWHM of the SPAD, the 16 ps FWHM contributed by the electronic circuitry and the 10 ps jitter of the electrical signal from the laser pulser had to be quadratically subtracted from the measured overall FWHM. A 27 ps laser pulse width is obtained for a 785 nm laser diode (Opto Electronics PPL30K). This is remarkably better than the 38 ps estimated by the manufacturer on the basis of the sampling oscilloscope measurement.

The performance of the SPAD cooled to -65 C was also investigated. In this experiment, the SPAD device was placed in a temperature controlled chamber, the active quenching circuit and the 785 nm laser diode were kept outside the chamber and coupled to the SPAD via coaxial cable and optical fiber respectively. The experimental result is shown in Fig.5. As before, the SPAD FWHM resolution under these conditions is determined by quadratically subtracting the 27 ps FWHM of the laser pulse (see above), the 16 ps FWHM contributed by the electronic circuits and the 10 ps synchronization jitter of the laser pulser electrical trigger output from the measured 38 ps

![Figure 5. Optical pulse of a laser diode.](image-url)
Single-Photon Solid-State Detector

overall FWHM. The FWHM resolution of the epitaxial SPAD at -65 C is thus found to improve down to 20 ps or less. To the best of our knowledge, this is the highest resolution ever reported for solid-state single-photon detectors.

From Fig.5, we note a small secondary peak in the histogram, occurring 150 ps after the principal one. While secondary peaks and ringings are normally present in the current pulse from fast p-i-n diodes, instrumental features of this type are not easily justified in the present type of measurement. The secondary pulse is therefore attributed to the laser and not to the detector. This is further demonstrated by the measurement reported in Fig.6, obtained with a different laser pulser (Hamamatsu CI308). No secondary peak is here noted, thus demonstrating that the secondary peak in Fig.5 is due to light (probably a second relaxation oscillation in the laser, or Fresnel reflection at the optical fiber entrance).

CONCLUSIONS

We have demonstrated that solid-state single-photon detectors are capable of 20 ps FWHM timing resolution. This result places these detectors at the same resolution level of the fastest available MCP-PMT. Physical quantities that allow to obtain such resolution have been investigated. From these studies, it may be expected that the physical limit to the SPAD resolution is at the level of a few picoseconds.

ACKNOWLEDGMENTS

The authors wish to thank the research and development laboratory of SGS-Thomson Microelectronics, Castelletto-Milano, for the fabrication of the epitaxial devices. This work was supported in part by CNR under programs "PF-MADESS" and "PS Fotorivelazione" and by Ministero della Pubblica Istruzione.

REFERENCES AND NOTES

9. Italian patent 22367A/88 pending.
15. We also made experiments with commercially available APDs (RCA C30921S) which have completely different structure (reach through avalanche photodiodes), and much higher breakdown voltage, \( \simeq 200 \) V. We found a similar dependence on the electric field, but much worse resolution, 460 ps FWHM at room temperature. See Ref. [19].


18. R. J. McIntyre, RCA, private communication.
