NEW SILICON EPITAXIAL AVALANCHE DIODE FOR SINGLE-PHOTON TIMING AT ROOM TEMPERATURE

Indexing terms: Optoelectronics, Avalanche diodes, Photo-diodes

A new silicon single-photon avalanche diode (SPAD) with epitaxial structure is presented. The carrier diffusion effect, which has plagued the time response of previous SPADs, is strongly reduced. The resolution obtained, less than 30 ps full width at half-maximum, is the highest so far reported in single-photon timing.

Avalanche photodiodes, designed to operate at bias higher than the breakdown voltage, can be employed for ultrafast detection of single optical photons. Such devices, called single-photon avalanche diodes (SPADs), are p–n junctions with uniform breakdown voltage and low thermal generation rate. In the operative bias condition a single carrier can trigger a self-sustaining avalanche current. If the carrier is photogenerated, the leading edge of the avalanche current marks the arrival time of the detected photon. The avalanche current continues to flow until an external circuit quenches the diode, lowering the bias below the breakdown voltage. After a finite dead time, the bias is rapidly restored to make possible the detection and timing of another photon. By using an active quenching circuit, SPADs can work in accurately controlled bias conditions, and the dead time following each pulse can be made very short. Remarkable performances have been achieved in various applications, such as optical fibre characterisation, laser ranging, fluorescent decay analysis etc.

We had previously implemented and tested prototype SPAD devices having the geometry described by Haitz. The active junction was formed by a shallow (0.3–2 μm) n− layer in a p−bulk substrate of 0.6–0.9 Ω cm resistivity. A deep diffused (5–8 μm) n− guard ring surrounded it, avoiding edge breakdown. These devices had about 28 V breakdown voltage and could be biased up to 33 V, i.e. up to the guard ring breakdown level.

The time resolution curve of a single-photon detector is given by the statistical distribution of the delays from the true arrival time of the photon at the device to the actual detection time, marked by the electrical output signal of the detector. It can be measured in a time-correlated single-photon counting set-up, by employing picosecond optical pulses. In the resolution curve of the prototype devices, photons absorbed in the depletion layer of the active junction produced a fast peak; we repeated full width at half-maximum (FWHM) values down to 60 ps. The experiments also showed that (i) the peak width is mainly related to the statistical fluctuations in the avalanche build-up times, and (ii) the peak width is improved as the maximum electric field is increased. As shown in Fig. 1, however, the resolution curve of the previous SPAD devices was plagued by a slow tail after the peak (which in the time scale of Fig. 1 is compressed to a single point). The effect is due to minority carriers, photogenerated in the deep neutral region beneath the junction, that succeed in reaching the depletion layer by diffusion. Since the optical absorption coefficient decreases at longer wavelengths, the intensity and the duration of the diffusion tail correspondingly increase. The diffusion tail and its wavelength dependence represent a serious drawback to high-resolution measurements of fast optical signals. In all cases where the optical signal is non-monochromatic, the data should be deconvolved by a time response of the detector which depends on the actual detected optical spectrum and must be very accurately known. This is at least impractical, if not impossible.

A primary objective was therefore a strong reduction of the diffusion tail. Our approach was to design an epitaxial device structure, in which the substrate acts as a sink for photogenerated carriers and the active junction is located in a suitable epistripe. A Monte Carlo program, previously developed by two of us, was used to compute the time-dependent diffusion effect in the devised structures and, consequently, to define the actual device design. Furthermore, to reduce the width of the resolution peak, the new devices were designed to attain higher values of the maximum electric field, up to 550 kV/cm.

In Fig. 2 the schematic cross-section of the new device is shown. The active junction is built in the p-epi-layer and the substrate/epitaxial np junction is zero or reverse-biased. The electrons photogenerated in the substrate cannot reach the epilayer. Furthermore, the substrate/epitaxial junction competes with the active junction in collecting the electrons generated in the neutral p-epilayer. Note that the avalanche current flows now laterally to the side ohmic contact. Therefore, the series resistance of the new device depends on the thickness of the neutral epilayer beneath the guard ring. To avoid impairing the output signal, a value of a few kΩ should not be exceeded. This requirement sets a limit to the reduction of the epilayer thickness and, therefore, to the reduction of the diffusion effect. A 2 μm-thick guard ring on a 12 μm-thick epitaxial was chosen. The actual thickness of the neutral epilayer in operating conditions can be reduced, to some extent, by increasing the reverse bias on the substrate/epitaxial junction. Anyhow, the series resistance never exceeds 10 kΩ. Boron ions were implanted in the active junction region, to lower its breakdown voltage to 13 V and correspondingly increase the breakdown electric field. A field plate was introduced to increase the breakdown voltage of the guard ring up to 70 V. The active junction can thus operate at excess bias voltage up to 55 V above its breakdown voltage, without suffering edge breakdown.

The experiments fully confirmed the reduction of the diffusion tail. As shown in Fig. 1 (curve b), the improvement is particularly striking in the long timescale. A remarkable improvement was also found in shorter timescales, where a narrower peak was observed. Accurate measurements of the peak FWHM against the excess bias were carried out. Fig. 3 shows the pulse of a gain-switched diode laser at 833 nm measured with the new SPAD device biased at 6 V above the breakdown and at room temperature; the FWHM is 43 ps. At such a resolution level, the contributions of the electronic circuit jitter and of the laser pulse width become significant and must be quadratically subtracted, to evaluate the true resolution of the SPAD. The time jitter due to electronic noise was measured using a current pulse generator for emulating the SPAD current pulses. A value of 15 ps FWHM was measured. The optical pulse width of the 833 nm diode laser was known to be about 28 ps FWHM. It is thus found

![Fig. 2 Schematic cross-section of new devices in p-epitaxial layer over n-substrate](image-url)
that the device resolution at room temperature is better than 30 ps. To our knowledge, this is the highest time resolution so far reported in single-photon timing measurements.

Fig. 3 Pulse of gain-switched diode laser, measured with new epitaxial SPADs at room temperature

In conclusion, the carrier diffusion tail in the resolution curve is strongly reduced in *pn* epitaxial SPADs. Further reduction can be obtained by designing more sophisticated device structures. The time resolution has improved from 60 to 30 ps FWHM using a high electric field in the avalanche junction. A more detailed knowledge of the physics underlying the statistics of the avalanche build-up time may lead to further progress.

Acknowledgement: We thank the R&D laboratory of SGS-Thomson Microelectronics, Castelletto-Milano, for the fabrication of the epitaxial devices. This work was supported in part by the Consiglio Nazionale delle Ricerche under the programmes ‘PF-MADESS Materiali e Dispositivi per Elettronica allo Stato Solido’ and ‘PS Fotorivelazione’ and by the Ministero Pubblica Istruzione.

M. GHIONI
S. COVA
A. LACAITA
G. RIPAMONTI
Dipartimento di Elettronica & Centro di Elettronica Politecnico di Milano
Quantistica e Strumentazione Elettronica CNR
Piazza L. da Vinci 32, 20133 Milano, Italy

References

**PERFORMANCE OF TIME DIVERSITY/CONVOLUTIONAL CODING SCHEME**

**Indexing terms**: Codes and coding, Convolutional codes, Burst trapping, Time diversity, Information theory

A rate $R = 1/3$ coding scheme, comprising a time diversity delay line and an $R = 2/3$ convolutional code, is investigated. The bit error rate performance of the scheme on a binary symmetric channel and a burst error channel is presented.

The encoder for the scheme investigated here is depicted in Fig. 1. The two $m$-bit shift registers of the $R = 2/3$ convolutional code of constraint length $m$ sub-blocks are separated by a delay line of length $l$ bits. The scheme is capable of both random error and burst error correction. The parameter $l$ is determined by the desired burst error correction capability. Any $R = 2/3$ convolutional code, or equivalently $R = 2/3$ block code and a time diversity delay line can be used in the scheme. As can be seen from Fig. 1, each information bit appears in two disjoint sets of parity equations, each set producing a separate syndrome.

The decoding strategy is as follows. The received code bits are decoded with an $R = 2/3$ decoder. When each two-bit sub-block of information bits is about to leave the output of the decoder, the assumption is made that the only channel errors capable of affecting its decoding, occurred in the $3m$ bit constraint length of received bits including and immediately following it. An $l$-bit delay line is also included in the decoder, and the final decoding decision is delayed until both three-bit code sub-blocks containing a certain information bit have been received and decoded initially. The information bit from the sub-block with corresponding syndrome indicating a lower weight error pattern, is selected as the final decoding decision.

On a bursty channel, the decoder operates essentially as a burst trapper, since it attempts to detect burst errors affecting code sequences of length $l$ or fewer $3$-bit sub-blocks, and to contain the information bits received for the first time during the burst within the $l$-bit delay line, while the time diversity is later used for substituting these information bits.

On a binary symmetric channel, the scheme is a more efficient random error corrector than the $R = 2/3$ code. If the $R = 2/3$ code can correct $2$ errors within a constraint length of $3m$ code bits, at least $i + 1$ errors must occur in each of the two $3m$-bit constraint lengths affecting the decoding of an information bit, before a decoding error can occur. Consequently, the scheme is capable of correcting up to $2 + 1$ errors within $6m$ code bits.

On a compound channel with both burst and random errors, the random errors in the guard spaces separating bursts, are corrected with the efficacy of the $t$-error-correcting $R = 2/3$ code. The scheme can thus be considered as an adaptive convolutional coding scheme.

The scheme was stimulated with $l = 83$ and the $d_{\text{min}} = 4$, single error-correcting CSOC with generators

\[ G_1^D(D) = 1 + D + D^4 \]

\[ G_2^D(D) = 1 + D + D^3 \]