Single-photon avalanche diodes in 0.18-μm high-voltage CMOS technology

L. D. HUANG,1 J. Y. WU,2 J. P. WANG,1 C. M. TSAI,1 Y. H. HUANG,1 D. R. WU,1 AND S. D. LIN1,*

1Department of Electronics Engineering, National Chiao Tung University, 1001 University Road, Hsinchu 300, Taiwan
2Information and Communications Research Laboratories, Industrial Technology Research Institute, Hsinchu 310, Taiwan
*sdlin@mail.nctu.edu.tw

Abstract: We have designed and fabricated high-performance single-photon avalanche diodes (SPADs) by using 0.18-μm high-voltage CMOS technology. Without any technology customization, the SPADs have low dark-count rate, high photon-detection probability, low afterpulsing probability, and acceptable timing jitter and breakdown voltage. Our design provides a low-cost and high-performance SPAD for various applications.

©2017 Optical Society of America

OCIS codes: (040.0040) Detectors; (040.1345) Avalanche photodiodes (APDs); (040.6040) Silicon.

References and links
1. Introduction

Since the birth in 1960s, single-photon avalanche diodes (SPADs) advanced with Si-based semiconductor technology [1,2]. The state-of-the-art CMOS single-photon avalanche diodes (SPADs) operated in Geiger mode exhibit low dark-count rate (DCR), high photon-detection probability (PDP), excellent timing resolution, and weak afterpulsing effect [3–6]. Benefited by the powerful integrated circuits (ICs) provided by mature silicon technology, CMOS SPADs outperform other photon-counting devices, such as photo-multiplier tube and charge-coupled devices. Previous works on 3-D imaging [7], light-detection and ranging (LiDAR) [8], fluorescence lifetime image microscopy [9, 10], time-resolved Raman spectroscopy [11], and radiometric temperature sensing [12], clearly demonstrated SPADs’ advantages. Very recently, commercial LiDAR products on driverless vehicles or safety assistant driving attract increasing attention [13] due to the foreseeable industrial and social impacts. For this particular application, a critical factor is the cost-down without sacrificing the system performance. At device level, CMOS SPAD’s performance could be optimized with adjustable layer arrangement and doping profile, which is in general not free, even not possible, for the existing CMOS technology. Therefore, a high-performance SPAD in low-cost CMOS technology without any customization is highly desirable. Fabricating SPADs with CMOS technology could face the other obstacle, the exact doping profiles and layer structures are not readily available because they are commercial secrets for the IC foundry. Each IC foundry has its own developed layers for each technology node, making the design of SPAD device structure much more difficult and highly constrained. In this aspect, it would be valuable to understand exact layer and doping information and its impact on device performance.

Up to date, SPADs have been fabricated with or without customization in many CMOS technology nodes from 0.8 µm to 65 nm in standard, high-voltage (HV), and imaging processes [14–25]. To boost the operation speed of the chips, advanced technology nodes are preferred. However, the advanced technology not only costs more but also could be problematic for the device itself. First, as the doping concentrations go higher, the band-to-band tunneling causes high DCRs. Second, the thick passivation layers decrease the light penetration so the PDP is reduced. So, in the present work, we focus on 0.18-µm technology that compromises the needs of the SPAD device and the accompanied circuits. In addition, our chips go through so-called multi-project wafer (MPW) service that is affordable for most academic researchers and start-up companies. The high-voltage CMOS process is chosen as it provides more layers for our use. Two high-performance SPADs fabricated side-by-side in 0.18-µm high-voltage (HV) CMOS technology have been extensively studied. Their I-V characteristics, DCR, bias-dependent PDP spectra, afterpulsing probability (APP), and timing jitters are measured, compared, and discussed. In addition, a detailed simulation for these two
devices is also presented. Our analysis reveals how the device structure determines the performance of a SPAD and provides a good reference for device design.

2. Measurement methods and device characteristics

2.1. Device structures and dark characteristics

Two SPADs, denoted as SPADs A and B, are fabricated by 0.18-μm HV CMOS process in TSMC and their schematic structures are shown in Figs. 1(a) and 1(b), respectively. The active region of SPAD A consists of DPW (deep p-well) and NBL (n-typed buried layer) with HVPW (high-voltage p-well) guard ring to prevent the corner breakdown [26]. SPAD B with the HVPW/NBL junction as its active region has a virtual guard-ring structure instead [5]. Note that the only difference between SPADs A and B is an additional DPW layer. The active regions of the devices are circular with a diameter of 20 μm.

![Fig. 1. Schematic structures of SPADs A (a), and B (b).](image)

To measure the I-V curves of the SPADs, we used a semiconductor parameter analyzer (Agilent B1500). The DCR measurement was carried out with a passive quenching circuit of 430-kΩ quenching resistor in series with the SPAD anode. The breakdown current was sensed through an ac coupled capacitor (10 nF) and the signal was amplified by a photon counting unit (C9744) and then counted by a dual channel counter (SR 400). The measured I-V curves and bias-dependent DCRs are plotted in Fig. 2. The breakdown voltages $V_{\text{BD}}$ of SPADs A and B are 49.9 V and 82.1 V as shown in the insets of Figs. 2(a) and 2(b), respectively. The inserted DPW junction reduces its $V_{\text{BD}}$ because the doping concentration of DPW layer is higher than that of HVPW. The similar DCRs of the two SPADs are 60 – 320 Hz in 50.5 – 60.0 V for SPAD A and 100 – 740 Hz in 82.3 – 92.0 V for SPAD B. The corresponding DCRs per unit area are about 0.68 and 1.06 Hz/μm² at the excess bias of 5 V for SPADs A and B, respectively.
Fig. 2. Bias-dependent DCRs of SPADs A (a) and B (b) and their I-V curves in the insets.

2.2. Photon detection probability (PDP)

The PDP measurement using the same passive quenching circuit were performed with the setup detailed in [5, 23]. In short, white light from a 1000 W halogen source was dispersed by a monochromator, coupled into a fiber, and then transmitted into a microscope in a dark box. A beam-splitter separates the incident photons into two beams, one for the real-time monitoring and the other one for the SPADs under test. With an iris and 100X near-infrared objective lens, the spot size is adjusted to about the size of the active area. In Figs. 3(a) and 3(b), the respective bias-dependent PDP spectra in the range of 450 – 900 nm for SPADs A and B are plotted. Note that, in order to compare the two devices with different breakdown voltages \( V_{BD} \), we define the normalized excess voltage \( V_{ex} \) as the excess voltage divided by the respective \( V_{BD} \) [5]. For SPAD A in Fig. 3(a), the PDP increases with the increasing \( V_{ex} \) and the peak value is about 22% at 570 nm at \( V_{ex} = 15\% \). In Fig. 3(b), the PDP spectra of SPAD B show slightly lower values and its peak PDP is about 19% also at 570 nm at \( V_{ex} = 15\% \). There are two features in the PDP spectra worthy noting. First, compared with SPAD B, SPAD A shows a slower decrease of PDP in the near infrared regime. At \( V_{ex} = 10\% \) (excess voltage of \(-5\) V), PDPs of SPAD A are 7.7% and 2.8% at 800 and 900 nm, respectively. The corresponding PDPs of SPAD B are 4.9% and 1.7%. Second, with the increasing \( V_{ex} \), the saturation of PDP of SPAD A comes slower than that of SPAD B. The PDP saturation is due to the saturation of breakdown trigger probability [22] so it indicates the different electric field distributions in two SPADs, which is to be discussed later herein. The better PDP of SPAD A in the near infrared regime is advantageous for LiDAR application on vehicles.

Fig. 3. Measured bias-dependent PDP spectra for SPADs A (a), and B (b).
In addition, we have performed the measurement of APP on SPAD A. The experiment was carried out in dark condition using the TCSPC technique with self-triggering. The breakdown signal from the SPAD was processed by a field-programmable gate-array (FPGA) to separate the adjacent events into two groups. One is used for timing start for TCSPC and the other one served as the stop signal. Note, for the APP measurement, the other device with the integrated circuit (not shown here) was used to shorten the dead time to about 20 ns because a long dead time will effectively eliminate all afterpulsing events. The measured APPs are 1.0% and 1.6% at the excess bias of 2.0 V and 5.0 V, respectively.

2.3. Timing jitter

Timing jitter is a key factor for timing application using SPADs. The light source for jitter measurement was a Ti-sapphire femto-second laser (Chameleon Ultra, Coherent) with a pulse picker (Model 9200). The repetition rate and pulse width is 100 kHz and <150 fs, respectively, in the wavelength range of 700 – 1000 nm. The light was focused into the active region of the SPADs by using a microscope. A time-correlated single-photon counting module (Pico-Harp 300) with bin resolution of 4 ps was used to receive the trigger signals from the pulse picker and from the SPADs to obtain the time-difference histograms [21]. Weak light condition was used to avoid the photon events in the dead time of SPADs.

The measured timing jitters using 720-nm laser for two devices biased at various voltages are plotted in semi-log scale in Fig. 4. Note that the time traces are normalized with the peak values and aligned with the peak time for clarity. For SPAD A in Fig. 4(a), the full-width at half maximum (FWHM) of the jitter is about 360 ps at 50.8 V (just above the breakdown voltage), and it decreases sharply to about 200 ps at 55.0 V. For SPAD B in Fig. 4(b), the FWHMs are 368 ps at 82.4 V and 184 ps at 92.4 V. Although the FWHMs for two devices are similar, a clear difference is spotted in their time traces. The diffusion tails of SPAD A are much more significant than those of SPAD B. This can be confirmed by the full-width at hundredth maximum (FW1/100M). For SPAD A, the FW1/100M is about 2.66 ns at 50.8 V and about 1.91 ns at 55.0 V. For SPAD B, they are 1.08 ns at 82.4 V and 0.59 ns at 92.4 V. Significant diffusion tails of SPAD A could be problematic for some timing applications and its physical reason will be discussed later in this paper. By using a two-exponential function, we fitted the timing jitter of SPAD A. With increasing biases, the fast time constants decreases from 197 ps to 91 ps and the slow ones from 890 ps to 567 ps.

![Fig. 4. Measured 720-nm timing jitters for SPADs A (a), and B (b) at various bias voltages.](image)

3. Simulations and discussions

3.1. Doping profiles and electric field distribution

Our two-dimensional device simulation was performed with Synopsys Sentaurus TCAD. The simulated structures were based on those in Fig. 1 with Gaussian doping profiles created by the Structure Editor tool. Figure 5 shows the in-depth doping distribution at the device center of two SPADs and their electric field distributions at the respective breakdown voltages. The
simulated breakdown voltages of SPADs A and B are 49.5 V and 81.8 V, respectively. Note that, the doping profile of SPAD A differs from that of SPAD B with the additional DPW layer. The DPW layer has a peak p-typed concentration of about 1E17 cm\(^{-3}\) and relatively narrow distribution so the breakdown voltage of SPAD A is lowered to around 50 V. As a result, 1) a neutral region spans from the device surface to about 2-μm deep in SPAD A, which does not exist in SPAD B, and 2) the electric field distribution of SPAD A is sharper than that of SPAD B. The former explains the significant diffusion tails observed in the timing jitter of SPAD A because the photon-carriers generated in the neutral region have to diffuse into the high-electric field before triggering a breakdown [4, 19]. The later could tell why the PDP saturation for SPAD A is slower than that for SPAD B. The triggering probability is not only highly dependent on the electric field strength but also on the size of high electric field region [22]. That is, to trigger a breakdown event, carriers have to travel in high field region for a long enough time to ensure the triggering. For SPAD B, the high-field region is more uniform and wider compared with that of SPAD A so its triggering probability approaches to one faster with the increasing excess bias. Thereby, the PDP saturation with applied bias is earlier for SPAD B.

Fig. 5. Simulated doping concentration profiles and electric field distributions in depth for SPAD A in (a) and SPAD B in (b).

3.2. Temperature-dependent breakdown voltages

We have also investigated the temperature-dependent breakdown voltages \(V_{BD}\). Here, the breakdown voltage is defined with the biased voltage for a breakdown current of 10 μA for simplicity. Figure 6 illustrates the measured and simulated \(V_{BD}\) of two SPADs in the temperature range of 290 – 320 K. For SPAD A in Fig. 6(a), the fitted \(V_{BD}\) shift rates are 40 mV/K in experiment and 34 mV/K in simulation. For SPAD B in Fig. 6(b), those are 65 mV/K and 53 mV/K, respectively. The shift rate difference between two SPADs comes from the doping concentration profiles [25]. However, the inconsistency (~20%) between measurement and simulation needs further investigations as it could be a good check point for the correctness of proposed doping profiles.

4. Conclusion

We have presented a detailed study on two high-performance SPADs in 0.18-mm high-voltage CMOS technology without any customization. The single difference of an additional layer between two SPADs makes the comparison simpler and more convincing. The device parameters including DCR, PDP, APP, and timing jitters have been discussed with the help of simulation tool. This work provides a design for low-cost and high-performance SPADs for various applications.
Fig. 6. Measured and simulated temperature-dependent breakdown voltages for SPAD A in (a) and SPAD B in (b).

Acknowledgments

We are grateful to the National Center for High-performance Computing for providing the simulation tool and the ITRI for their help on chip fabrication. This work was financially supported by the Ministry of Science and Technology and ATU program of MOE in Taiwan.