Flexible ultrathin-body single-photon avalanche diode sensors and CMOS integration

Pengfei Sun,[*] Ryoichi Ishihara, and Edoardo Charbon

Faculty of Electrical Engineering, Delft University of Technology, Feldmannweg 17, 2628 CT Delft, The Netherlands
[*]*p.sun@tudelft.nl*

Abstract: We proposed the world's first flexible ultrathin-body single-photon avalanche diode (SPAD) as photon counting device providing a suitable solution to advanced implantable bio-compatible chronic medical monitoring, diagnostics and other applications. In this paper, we investigate the Geiger-mode performance of this flexible ultrathin-body SPAD comprehensively and we extend this work to the first flexible SPAD image sensor with in-pixel and off-pixel electronics integrated in CMOS. Experimental results show that dark count rate (DCR) by band-to-band tunneling can be reduced by optimizing multiplication doping. DCR by trap-assisted avalanche, which is believed to be originated from the trench etching process, could be further reduced, resulting in a DCR density of tens to hundreds of Hertz per micrometer square at cryogenic temperature. The influence of the trench etching process onto DCR is also proved by comparison with planar ultrathin-body SPAD structures without trench. Photon detection probability (PDP) can be achieved by wider depletion and drift regions and by carefully optimizing body thickness. PDP in frontside-(FSI) and backside-illumination (BSI) are comparable, thus making this technology suitable for both modes of illumination. Afterpulsing and crosstalk are negligible at 2µs dead time, while it has been proved, for the first time, that a CMOS SPAD pixel of this kind could work in a cryogenic environment. By appropriate choice of substrate, this technology is amenable to implantation for biocompatible photon-counting applications and wherever bended imaging sensors are essential.

OCIS codes: (040.1345) Avalanche photodiodes (APDs); (040.0040) Detectors; (130.3120) Integrated optics devices; (280.4788) Optical sensing and sensors; (040.5160) Photodetectors; (230.5170) Photodiodes; (040.5250) Photomultipliers; (030.5260) Photon counting; (130.5990) Semiconductors; (130.6010) Sensors; (040.6040) Silicon; (160.5470) Polymers.

References and links

1. E. Charbon, "Single-photon imaging in complementary metal oxide semiconductor processes," Philos Trans A Math Phys Eng Sci **372**(2012), 20130100 (2014).
2. R. J. Walker, J. A. Richardson, and R. K. Henderson, "A 128×96 pixel event-driven phase-domain ΔΣ-based fully digital 3D camera in 0.13µm CMOS imaging technology," in *IEEE ISSCC Dig. Tech.* (IEEE, 2011), pp.410–412.
3. A. Gulinatti, I. Rech, P. Maccagnani, M. Ghioni, and S. Cova, "Improving the performance of silicon single-photon avalanche diodes," Proc. SPIE **8033**, 803302 (2011).
4. C. Niclass, M. Soga, H. Matsubara, S. Kato, and M. Kagami, "A 100-m Range 10-Frame/s 340x96-pixel Time-of-Flight Depth Sensor in 0.18-µm CMOS," IEEE J. Solid-State Circuits **48**(2), 559–572 (2013).
5. E. Charbon, H.-J. Yoon, and Y. Maruyama, "A Geiger mode APD fabricated in standard 65nm CMOS technology," in *IEEE International Electron Devices Meeting* (IEEE, 2013), paper 27.5.1.
6. M. S. Humayum, J. D. Weiland, G. Chader, and E. Greenbaum, *Artificial Sight* (Springer, 2011).
7. T. Tokuda, M. Takahashi, K. Uejima, K. Masuda, T. Kawamura, Y. Ohta, M. Motoyama, T. Noda, K. Sasagawa, T. Okitsu, S. Takeuchi, and J. Ohta, "CMOS image sensor-based implantable glucose sensor using glucose-responsive fluorescent hydrogel," Biomed. Opt. Express **5**(11), 3859–3870 (2014).
8. T. Noda, K. Sasagawa, T. Tokuda, H. Kanda, Y. Terasawa, H. Tashiro, T. Fujikado, and J. Ohta, "Fabrication of fork-shaped retinal stimulator integrated with CMOS Microchips for extension of viewing angle," Sensors Mater. **26**(8), 637–648 (2014).

9. G. Park, H.-J. Chung, K. Kim, S. A. Lim, J. Kim, Y.-S. Kim, Y. Liu, W.-H. Yeo, R.-H. Kim, S. S. Kim, J.-S. Kim, Y. H. Jung, T.-I. Kim, C. Yee, J. A. Rogers, and K.-M. Lee, "Immunologic and tissue biocompatibility of flexible/stretchable electronics and optoelectronics," Adv. Healthc. Mater. **3**(4), 515–525 (2014).
10. J. Yoon, S.-M. Lee, D. Kang, M. Meitl, C. Bower, and J. A. Rogers, "Heterogeneously integrated optoelectronic devices enabled by micro-transfer printing," Adv. Opt. Mater. **3**(10), 1313–1335 (2015).
11. D. R. Schuette, R. C. Westhoff, A. H. Loomis, D. J. Young, J. S. Ciampi, B. F. Aull, and R. K. Reich, "Hybridization process for backilluminated silicon Geiger-mode avalanche photodiode arrays," Proc. SPIE **7681**, 76810P (2010).
12. P. Sun, B. Mimoun, E. Charbon, and R. Ishihara, "A flexible ultra-thin-body SOI single-photon avalanche diode," in *IEEE International Electron Devices Meeting* (IEEE, 2013), pp. 1-4.
13. P. Sun, E. Charbon, and R. Ishihara, "A flexible ultra-thin-body single-photon avalanche diode with dual side illumination," IEEE J. Sel. Top. Quantum Electron. **20**(6), 3804708 (2014).
14. J. M. Hornibrook, J. I. Colless, I. D. Conway Lamb, S. J. Pauka, H. Lu, A. C. Gossard, J. D. Watson, G. C. Gardner, S. Fallahi, M. J. Manfra, and D. J. Reilly, "Cryogenic control architecture for large-scale quantum computing," Phys. Rev. Appl. **3**(2), 024010 (2015).
15. M. T. Rakher, L. Ma, O. Slattery, X. Tang, and K. Srinivasan, "Quantum transduction of telecommunications-band single photons from a quantum dot by frequency upconversion," Nat. Photonics **4**(11), 786–791 (2010).
16. P. Sun, E. Charbon, and R. Ishihara, "A flexible 32x32 SPAD image sensor with integrated microlenses," in *International Image Sensor Workshop* (IISW, 2015), pp. 336–339.
17. J. M. Pavia, M. Wolf, and E. Charbon, "Measurement and modeling of microlenses fabricated on single-photon avalanche diode arrays for fill factor recovery," Opt. Express **22**(4), 4202–4213 (2014).
18. S. Cova, A. Longoni, and A. Andreoni, "Towards picosecond resolution with single-photon avalanche diodes," Rev. Sci. Instrum. **52**(3), 408–412 (1981).
19. A. Spinelli, M. Ghioni, S. Cova, and L. M. Davis, "Avalanche detector with ultraclean response for time-resolved photon counting," IEEE J. Quantum Electron. **34**(5), 817–821 (1998).
20. C. Y. Chang, S. S. Chiu, and L. P. Hsu, "Temperature dependence of breakdown voltage in silicon abrupt P-N junctions," IEEE Trans. Electron. Dev. **18**(6), 391–393 (1971).
21. G. Collazuol, M. G. Bisogni, S. Marcatili, C. Piemonte, and A. Del Guerra, "Studies of silicon photomultipliers at cryogenic temperatures," Nucl. Instrum. Methods **628**(1), 389–392 (2011).
22. N. Serra, G. Giacomini, A. Piazza, C. Piemonte, A. Tarolli, and N. Zorzi, "Experimental and TCAD Study of breakdown voltage temperature behavior in n+p SiPMs," IEEE Trans. Nucl. Sci. **58**(3), 1233–1240 (2011).
23. M.-J. Lee, P. Sun, and E. Charbon, "A first single-photon avalanche diode fabricated in standard SOI CMOS technology with a full characterization of the device," Opt. Express **23**(10), 13200–13209 (2015).
24. I. Rech, I. Labanca, G. Armellini, A. Gulinatti, M. Ghioni, and S. Cova, "Operation of silicon single photon avalanche diodes at cryogenic temperature," Rev. Sci. Instrum. **78**(6), 063105 (2007).
25. S. Sze and G. Gibbons, "Effect of junction curvature on breakdown voltage in semiconductors," Solid-State Electron. **9**(9), 831–845 (1966).
26. A. Spinelli and A. L. Lacaita, "Physics and Numerical simulation of single-photon avalanche diodes," IEEE Trans. Electron. Dev. **44**(11), 1931–1943 (1997).
27. S. Mandai, M. W. Fishburn, Y. Maruyama, and E. Charbon, "A wide spectral range single-photon avalanche diode fabricated in an advanced 180 nm CMOS technology," Opt. Express **20**(6), 5849–5857 (2012).
28. C. Veerappan and E. Charbon, "A substrate isolated CMOS SPAD enabling wide spectral response and low electrical crosstalk," IEEE J. Sel. Top. Quantum Electron. **20**(6), 299–305 (2014).
29. S. Nikzad, M. Hoenk, J. Hennessy, A. Jewell, A. Carver, T. Jones, S. Cheng, T. Goodsall, and C. Shapiro, "High performance silicon imaging arrays for cosmology, planetary science, and other applications," in *IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2014), paper 10.7.1.
30. M. W. Fishburn, "Fundamentals of CMOS single-photon avalanche diodes," Ph.D. dissertation (Delft Univ. of Technology, 2012).

1. Introduction

Solid-state single-photon avalanche diode (SPAD) technology has existed for decades and is receiving wide attention for applications such as time-of-flight vision, time-correlated single-photon counting, fluorescence lifetime sensing and biomedical imaging. A SPAD is an avalanche photodiode (APD) operating above breakdown voltage, V_{BD}, in so-called Geiger mode and equipped with avalanche quenching and recharge mechanisms. Thanks to improved CMOS fabrication and Moore's Law, SPAD technology implemented in planar processes for imaging systems has significantly progressed in recent years [1–5].

Especially in the last decade, the push towards new applications, such as disposable assays, edible probes and implantable sensors has continued. A typical application in the field of biomedical sensing is the retinal prosthesis [6–8], where a single-photon imaging device, integrated with CMOS circuitry on flexible substrate, could be implanted into the eye and bent to match the curvature of the eyeball. Another application is chronic biomedical monitoring [9], where a wearable or implantable miniaturized single-photon sensor could be

left *in situ* to continuously monitor a person's health status, providing more accurate information about the progression of diseases such as cancer and other inflammatory or chronic ailments. However, current SPAD technology is generally implemented on bulk silicon and can't meet the requirements of implantable biomedical applications where backside-illumination and new substrate post processing are core technologies, while inherent CMOS compatibility is a requirement [10, 11]. In our previous work, we demonstrated the world's first flexible SPAD fabricated in an ultrathin-body silicon-on-insulator (SOI) process followed by transfer post-processing to flexible substrate [12], and we achieved a flexible SPAD with dual-side illumination for the first time [13].

In this work, we investigate the Geiger-mode performance of this flexible ultrathin-body SPAD comprehensively: dark count rate (DCR), V_{BD} and photon detection probability (PDP) are studied based on different junction parameters, operation temperature and device structures. Afterpulsing and crosstalk are negligible at the dead time, thanks to reduced parasitic capacitance, obtained through the integration of the output buffer. Thanks to the flexible substrate and dual-side illumination, the flexible SPAD chip could be implanted easily to human organ, such as eyeball, and integrated with stimulation components at any side of the retina. Also, thanks to the integration of CMOS process, it could pave the way to realize larger scale sensor array integrated with more electronics on flexible substrate. The DCR could be drastically reduced by operating the SPADs at temperatures as low as 80K. The measurements are significant, since they prove for the first time that cryogenic SPADs in CMOS operation is possible. It could be very useful for the telecommunication electronics [14] for quantum computer [15] in the future.

In section 2, the cross-section of device structure and pixel unit are presented, including a SPAD, a quenching resistor, and CMOS transistors. The flow chart of the flexible CMOS SPAD sensor technology is also presented. Section 3 presents measurements including DCR's dependence on enrichment doping, temperature and device structures, PDP, afterpulsing, and time jitter performance. The improvements by integrating CMOS buffer circuitries are analyzed. Conclusions are given in Section 4.

2. Device structure and fabrication flowchart

Figure 1(a) shows a cross-section of the trench-isolated SOI SPAD. It was fabricated by our previously proposed flexible ultrathin-body SPAD technology enabling dual-side illumination [12, 13]. The SPAD was realized in circular shape based on a N^+/P junction with P-type enrichment doping in the multiplication region. In order to prevent premature breakdown caused by the higher electric field at the edge of the junction, the enrichment doping is carefully designed to fit in the ultrathin-body SOI structure. Figure 1(b) shows a cross-section of the flexible trench-isolated SPAD integrated with the quenching resistor. Epitaxy silicon island isolated by trench is used as resistor to quench the avalanche. V_{OP} is the operation voltage of SPAD, which is V_{eb} above V_{BD} and biases the SPAD in Geiger mode.

As shown in Fig. 1(c), each of the two neighboring pixels contains a SPAD, a quenching resistor, and CMOS buffer circuits powered by supply voltage V_{DD}. All the components are based on a trench-isolated silicon island structure to achieve high levels of flexibility. Polyimide or polymer are used as flexible substrates and also act as a microlens to increase fill factor [16].
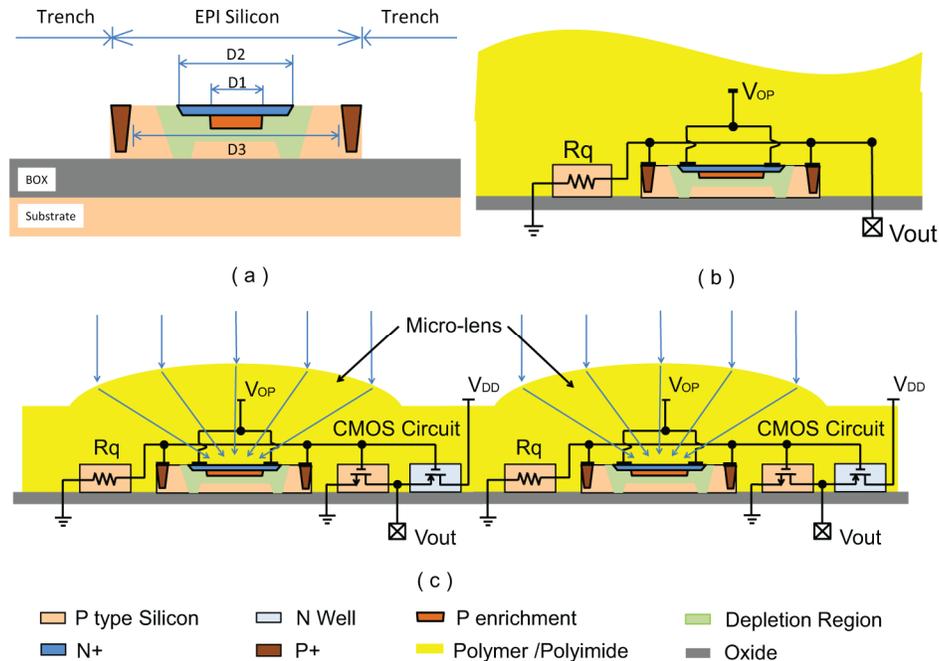
Fig. 1. (a) Cross-section of SOI ultrathin-body trench-isolated structure SPAD device; (b) Cross-section of flexible ultrathin-body trench-isolated structure SPAD integrated with quenching resistor; (c) Flexible CMOS SPAD sensor pixel schematic and SPAD cross-section.

For the flexible CMOS SPAD sensor, fabrication begins with a p-type SOI wafer prepared by epitaxy technology. The N-well is formed by implantation and followed by a thermal drive-in process. After CMOS transistors are built, a $N + P$ junction is formed by implantation. The manufacture flow is further optimized by variable process parameters such as enrichment doping dose and epitaxy layer thickness, which are analyzed comprehensively in section 3. After junction implantation, the device islands are formed and isolated with trenches by dry etching. Then, a passivation layer is deposited by two-time etch-back to form a spacer at the trench step. Contact holes are opened by dry etching. Then, a metallization layer is sputtered and patterned to realize the first metal interconnection. After the second metallization, which is similar to the first one, the wafer is sent to alloy [16]. The device fabrication flow is summarized in Fig. 2(a).

The flexible substrate transfer and microlens imprint process are summarized in Fig. 2(b). Followed by the backside oxide deposition in Fig. 2(a), the oxide is patterned as mask, as shown in step 1. The sol-gel polymer is coated on top of the device after the backside oxide mask is patterned. Then, the polymer is patterned and the quartz mold is brought into contact with the polymer. Pressure must be applied to form the microlens array on top of the SPAD sensor [17] with lateral alignment accuracy of less than 1µm. The imprinting process was being completed at the time of the writing of the paper. The silicon substrate under the buried oxide layer is then etched away. The etching stops at the buried oxide layer and the SPAD image sensor layer on new flexible substrate can easily be released. Thanks to the function of the polymer layer as both flexible substrate and microlenses, the SPAD sensors with CMOS circuit systems can act as flexible imager with higher fill factor. According to our current design, the fill factor is expected to be higher than 10% with microlenses.

The whole process (except microlens imprinting) was implemented in class-100 cleanroom and MEMS lab at Else Kooi Lab (formerly Delft Institute of Microsystems and Nanoelectronics). Tens of thousands of SPAD pixels from several batches have been fabricated with yield higher than 95% and results are quite consistent and reproducible.
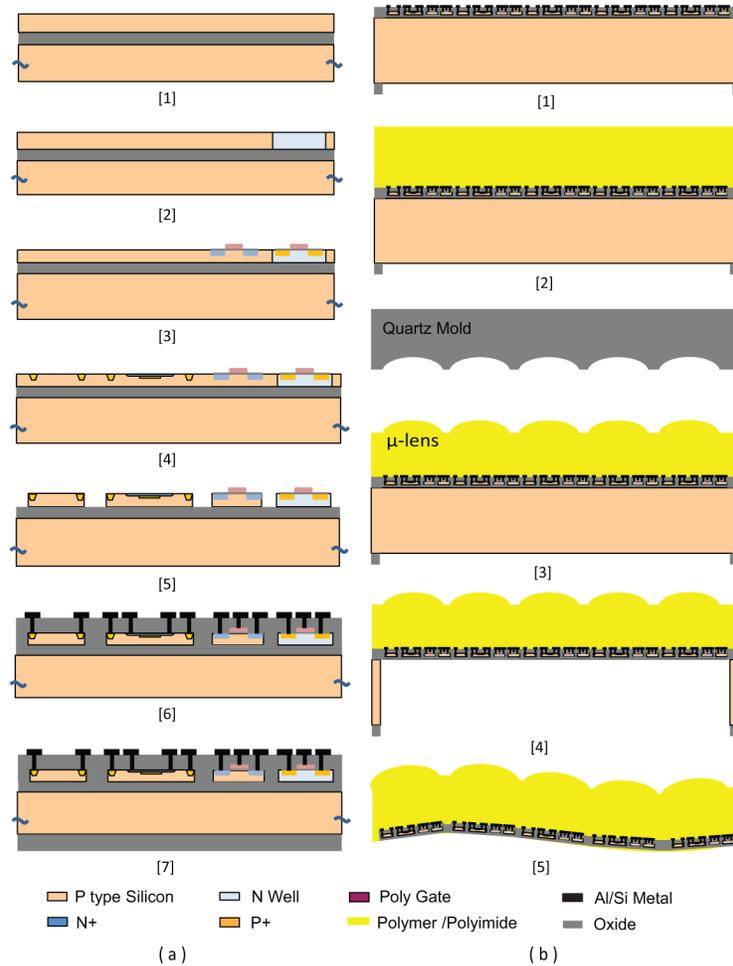
Fig. 2. (a). Pixel device fabrication flow chart. (1) Silicon epitaxy process on SOI wafer; (2) N-Well implantation and driving in; (3) CMOS transistor fabrication; (4) SPAD junction implantation; (5) trench etching process; (6) metallization; Note: Only 1st metallization is shown in this figure; (7) backside oxide deposition. (b). Flexible substrate transfer and microlens fabrication. (1) Oxide mask fabrication on backside; (2) sol-gel polymer coating and curing; (3) microlens imprinting by quartz mold; (4) substrate etching; (5) layer releasing.

3. Measurement results

For the flexible device characterization, the chip was packaged through wire bonding and then measured. Because the SPADs have been already equipped with quenching resistors, which were also integrated on flexible substrate, only the operation voltage V_{OP} and output need to be connected to pads. For the SPAD pixel integrated with CMOS buffering circuit, the power for circuit V_{DD} also needs to be provided through the pad. The SPADs were operated in Geiger mode and the output was monitored by a high speed oscilloscope (LeCroy Wavemaster 8600A). Measurements were done at room temperature except when noted.

3.1 characteristics on flexible trench-isolated SPAD integrated with resistor

The analysis of process parameters is based on flexible trench-isolated SPAD integrated with quenching resistor configuration. The SEM microphotograph of ultrathin-body trench-isolated SPAD integrated with quenching resistor is shown in Fig. 3. The DCR is measured by counting the number of avalanches while the SPAD is operating above V_{BD} in the dark. DCR

is caused by a combination of band-to-band tunneling and trap-assisted avalanching and it increased by excess bias voltage V_{eb}.

As shown in Fig. 1(a), a P + enhancement region is made to form the multiplication region, where the electric field is concentrated in this planar region and higher than that around the curvature region of the junction, thus preventing premature edge breakdown [18, 19]. It has been confirmed by DCR versus temperature measurements [13] that a large portion of dark noise comes from band-to-band tunneling due to high doping in the multiplication region. DCR as a function of excess bias based on devices with different doping levels in the multiplication region is reported in Fig. 4. To make sure other process parameters are consistent, the devices for comparison are fabricated on the same wafer as Fig. 3(b). When the P enrichment dose was reduced from $3.00 \times 10^{13} \mathrm{cm}^{-2}$ to $2.95 \times 10^{13} \mathrm{cm}^{-2}$, the DCR could be reduced by 50% while the SPAD was operating at the same excess bias. However, if the P enrichment dose was further reduced to $2.90 \times 10^{13} \mathrm{cm}^{-2}$, the DCR would increase by around 6%. If the P enrichment dose is reduced further to $2.85 \times 10^{13} \mathrm{cm}^{-2}$, the SPAD suffers from premature edge breakdown (PEB).

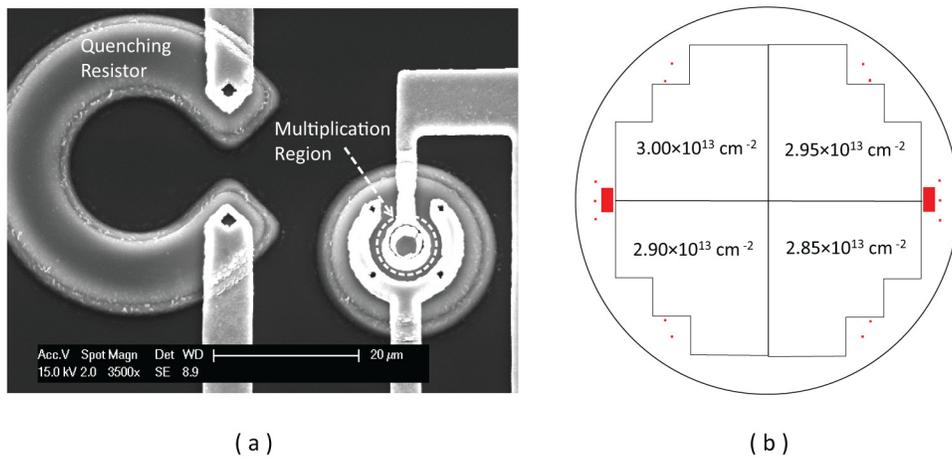(a)   (b)

Fig. 3. (a) SEM microphotograph of ultrathin-body trench-isolated SPAD integrated with quenching resistor. (b) Wafer map of devices with different enrichment implantation dose in the multiplication region.
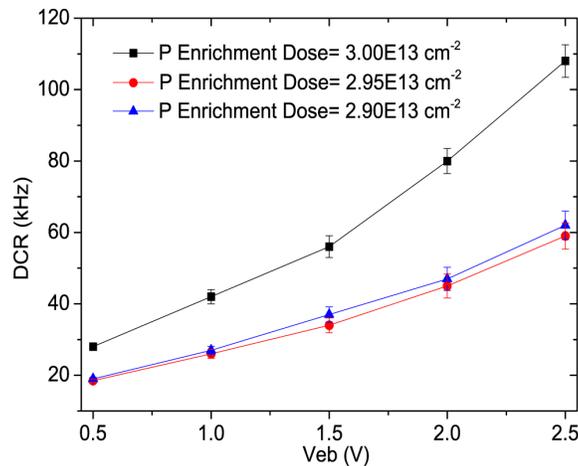

Fig. 4. DCR of flexible ultrathin-body SPAD (4μm-diameter multiplication region) with different doping levels in the multiplication region.

As dose decreased, V_{BD} would increase as shown in Fig. 5. The V_{BD} and DCR are further analyzed at cryogenic temperature. Compared with the breakdown voltages at different temperatures shown in Fig. 5, V_{BD} decreases with temperature due to the increase of ionization rates at low temperature. Based on the modified Baraff theory. As T decreased, the mean free path increases. The possibility of scattering process when a carrier transport across the junction decreased. Lower energy from the field is needed to ionize the carriers, thus decreasing V_{BD} [20].
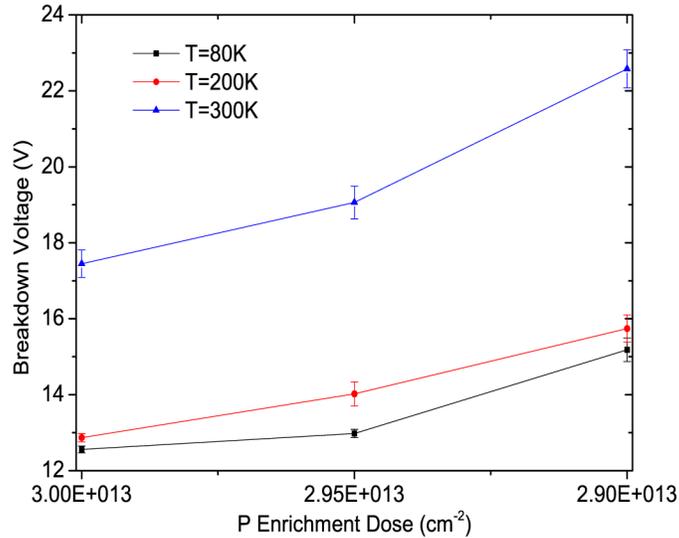
Fig. 5. Breakdown voltage of flexible ultrathin-body SPAD as a function of P enrichment dose at different temperatures.
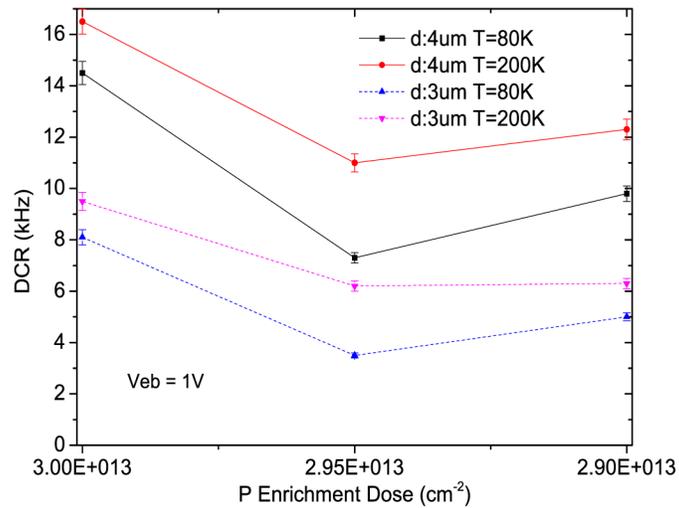
Fig. 6. DCR of flexible ultrathin-body SPAD at cryogenic temperatures. (d: diameter of multiplication region).

At the temperature below ~200K, the dark noise caused by Shockley-Read-Hall (SRH) generation and trap-assisted tunneling in the high electric field region is suppressed and DCR is dominated by band-to-band tunneling [21, 22]. The DCR measurements at cryogenic temperatures of 80K and 200K in Fig. 6 show consistency. The DCR decreases as the

enrichment doping decreases when the dose is reduced from 3.00×10^{13}cm^{-2} to 2.95×10^{13}cm^{-2}. However, it stops decreasing and even starts to increase when the dose is reduced to 2.90×10^{13}cm^{-2}, which is consistent with the results of Fig. 4. The DCR density could be reduced to hundreds of Hertz per micrometer square at 80K and it is believed to reduce to tens of Hertz per micrometer square [23] if trap-assisted noise from trench etching could be filtered out at extremely low temperatures. However, it needs future work to confirm possible increased afterpulsing below 70K [21, 24].
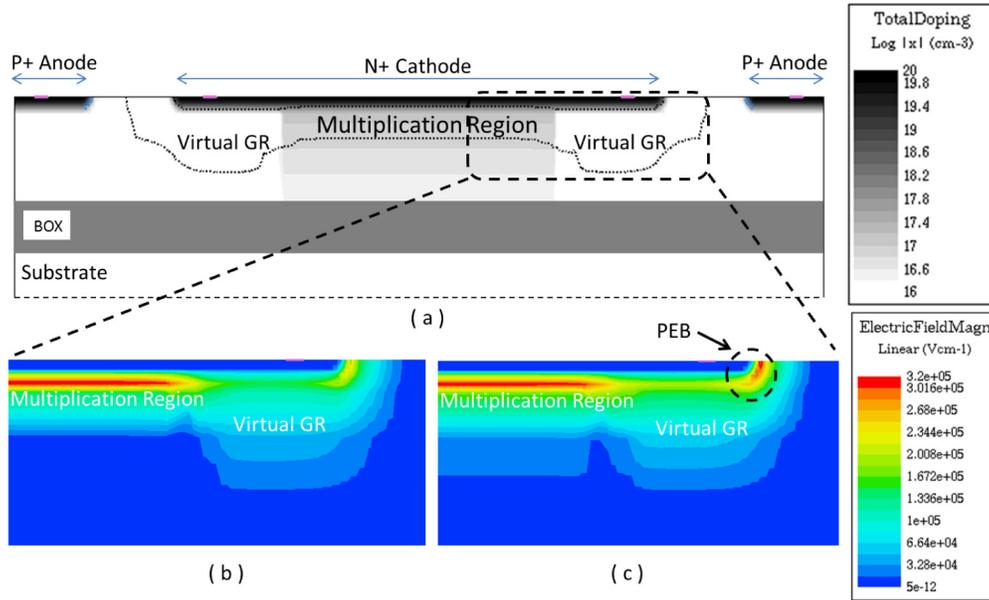

Fig. 7. (a) SPAD structure overlaid on a total doping plot; (b) zoom-in plot of electric field with enrichment doping concentrating (dose = 3.00×10^{13}cm^{-2})the high electric field at the multiplication region; (c) zoom-in plot of electric field with reduced enrichment doping (dose = 2.85×10^{13}cm^{-2}) causing PEB.

Electric field simulations (Fig. 7) were performed using MEDICI at device breakdown. If the breakdown voltage increased as enrichment doping decreased, the electric field around the virtual guard ring would be increased when the electric field at the multiplication region reached critical values (2.5×10^5 V/cm in silicon) [25, 26]. Shown by Figs. 7(b) and 7(c), when the enrichment doping was reduced, the electric field in the multiplication region still needed to reach the critical values so that the impact ionization avalanche could be triggered and consequently PEB is present due to stronger electric field at the curvature of the virtual guard ring. It shows consistency with the measurement results in Figs. 4 and 6.

To characterize the sensitivity of the SPAD, a monochromator (Oriel/Newport part 77250) was used to project into an integration sphere with a reference diode. By adjusting the intensity of the light source, the SPAD was operating in single-photon regime and PDP was characterized for incident light wavelength from 400nm to 950nm. In this work, the PDP was measured from flexible SPADs with different body thicknesses in both FSI and BSI, as shown in Fig. 8(a).

The junction cross-section is shown in Fig. 8(b). The photon detection probability will be the integral of electron-hole pairs generated at z multiplied by the probability that the injected electron-hole pair triggers an avalanche, signed by $p(z)$.

$$PDP = \int_{-\infty}^{\infty} T(\lambda) f_\lambda(z) p(z) dz, \qquad (1)$$

#251205
© 2016 OSA

Received 30 Sep 2015; revised 7 Nov 2015; accepted 18 Nov 2015; published 16 Feb 2016
22 Feb 2016 | Vol. 24, No. 4 | DOI:10.1364/OE.24.003734 | OPTICS EXPRESS 3741

where $T(\lambda)$ is the net transmission of light with wavelength λ and $f_\lambda(z)$ is the absorption probability at depth z:

$$f_\lambda(z) = \mu(\lambda)\exp(-\mu(\lambda)z), \qquad (2)$$

with $\mu(\lambda)$ being the mean penetration depth of the light into the silicon.

Equation (1) and (2) could be applied according to FSI and BSI respectively:
In FSI, the PDP can be modeled as:

$$PDP = \int_0^{Z_w} T(\lambda)f_\lambda(z)p(z)dz = \int_0^{Z_w} T(\lambda)\mu(\lambda)\exp(-\mu(\lambda)z)p(z)dz, \qquad (3)$$

While in BSI, the PDP can be modeled as:

$$PDP = \int_0^{Z_w} T(\lambda)f_\lambda(z_w - z)p(z)dz = \int_0^{Z_w} T(\lambda)\mu(\lambda)\exp(-\mu(\lambda)(z_w - z))p(z)dz, \quad (4)$$
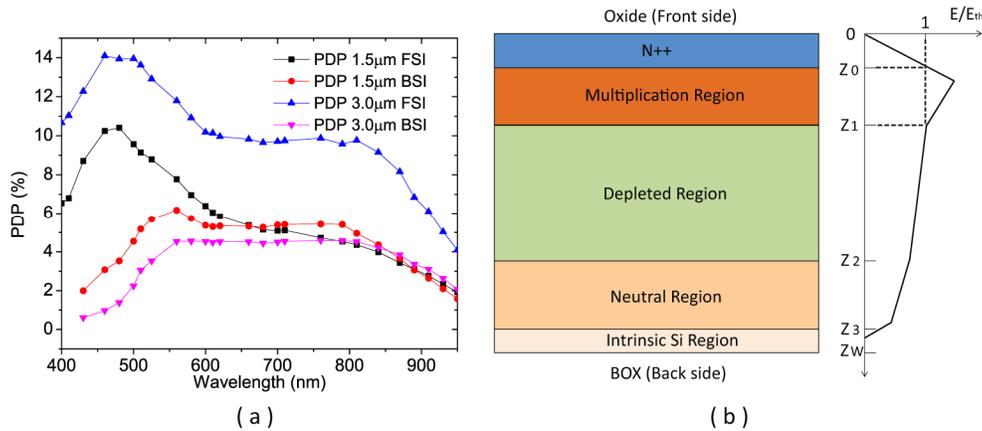
Fig. 8. (a) PDP of flexible SPAD with different body thickness in both FSI and BSI; (b) junction cross-section in ultrathin-body ($E_{th} = 2.5 \times 10^5$ V/cm).

By increasing the body thickness from 1.5µm to 3.0µm, the peak value of PDP in FSI could be increased from 10.7% to 14.3%. At long wavelengths above 650nm, it was enhanced by a factor of more than two. Thanks to the wider depletion and neutral regions below the multiplication region from Z1 to Z3 in the thicker body, the larger probability for carrier generated in the higher depth drifting to the multiplication region where avalanches were triggered, consequently PDP got enhanced especially at long wavelength, with which photons have longer penetration depth.

The PDP of 3.0µm-thick body was lower than that of 1.5 µm-thick body in BSI, which was different from the results in FSI. In BSI, most of carriers are generated at the bottom of silicon body due to absorption probability of photons in silicon. Carriers at the bottom have to drift across the neutral region to multiplication region and trigger the avalanches. The neutral region is wider for thicker body since the cathode and enrichment region profile are made from the front side, $\int_{Z0}^{Z1} T(\lambda)f_\lambda(z_w - z)p(z)dz$ decreases if z_w is increased and consequently the probability is reduced for carriers generated by photons with particular penetration depth to trigger avalanches.

3.2 Comparison between flexible trench-isolated SPAD and planar SOI SPAD

Trap assisted avalanche is another major source of dark noise in SPADs and it is caused by trap or defect states in silicon. To investigate the source of the defect states in the device structure, a planar SPAD structure is developed with LOCOS as isolation instead of trenches,

as shown in Fig. 9(b) [27, 28]. The SEM microphotograph of the planar SPAD in circular shape, which is isolated by 840nm-thick LOCOS by wet oxidation, is shown in Fig. 9(a).
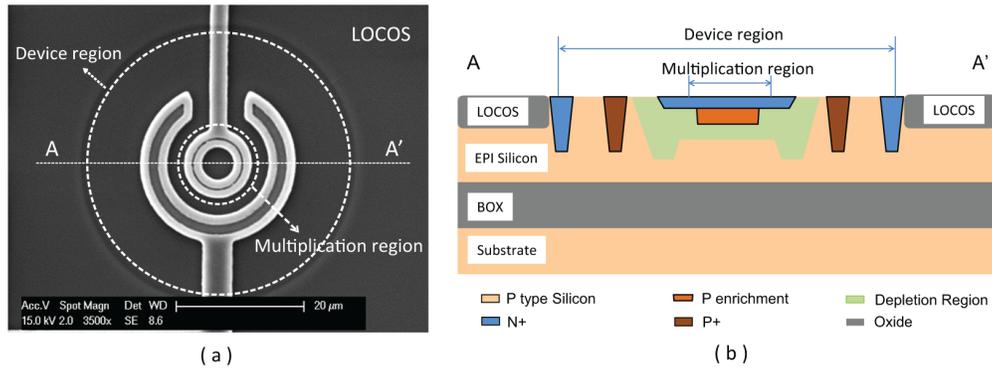

Fig. 9. (a) SEM top-view microphotograph of planar SPAD; (b) Cross-section of planar SPAD

The DCR measurement shows a much lower dark noise in the planar SPAD. While the diameter of the multiplication region was designed from 3µm to 6µm, the DCR density varied from 100Hz/µm^2 to 300Hz/µm^2, which is comparable with DCR density of standard SOI CMOS SPAD [23]. PDP was measured with peak value of 13% at 450nm. The comparison of DCR density and PDP based on trench-isolated SPAD and planar SPAD at different temperatures are summarized in Fig. 10.

Defects at the sidewall of trenches etched in plasma act as traps causing DCR by SRH generation and trap-assisted tunneling. It is proven by comparison of DCR density, while trench-isolated SPADs show 10 times higher DCR density than planar SPADs at room temperature. At cryogenic temperatures, as low as 80K, trap-assisted avalanche could be suppressed, showing trench-isolated SPADs with relatively similar DCR density if compared with planar SPADs. However, planar SPADs have worse flexibility because of the lack of silicon islands. Passivation technology [29] to filter out the defects around the trenches needs to be investigated in future work.
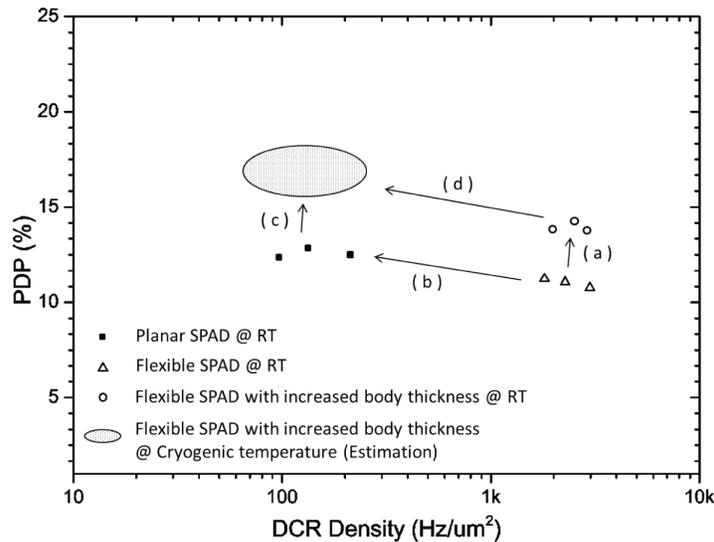

Fig. 10. Plot of DCR density vs. PDP in recent developments, each introducing a technological innovation or optimization with a consequent performance improvement. Note: PDP refers to FSI. (a) Increase of body thickness; (b) isolation by LOCOS; (c) isolation by trench and operation at cryogenic temperature; (d) operation at cryogenic temperature.

3.3 Characteristics on flexible CMOS SPAD sensor pixel

Following the fabrication in the proposed technology of Fig. 2, the first SPAD image sensor integrated with CMOS buffering circuit on flexible substrate is demonstrated as shown in the SEM microphotograph of Fig. 11(a). The photo of bent flexible SPAD pixel farm sample, which has been released and mounted to PDMS piece is shown in Fig. 11(b).
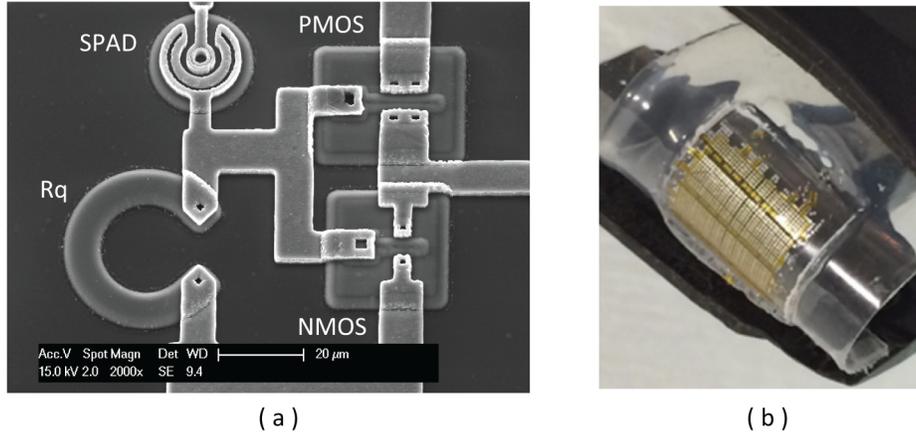

(a)   (b)

Fig. 11. (a) SEM microphotograph of a CMOS SPAD sensor pixel; (b) photo of bent flexible SPAD pixel farm sample.
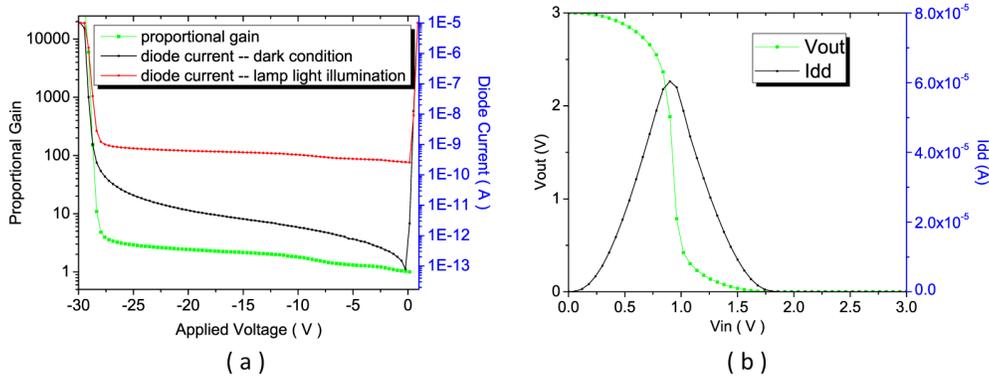

(a)   (b)

Fig. 12. (a) I-V characteristics of SPAD, note: Mitutoyo lamp link fiber source (Power: 150W)
(b) DC transfer curve of CMOS Inverter

The current-voltage (I-V) characteristics of the SPAD is shown in Fig. 12(a) for different light conditions. The breakdown voltage is around 26.5V at room temperature. The input-output response of the CMOS inverter, which consists of three-terminal SOI PMOS and NMOS transistors, is shown in Fig. 12(b).

Compared with our previous work [13], the excess bias could be increased to 4V, by integrating CMOS buffering. The characterization of DCR as a function of excess bias is shown in Fig. 13. The pixel was operated at cryogenic temperatures, as low as 80K, which is reported for the first time to our knowledge; DCR could be reduced dramatically, as shown in Fig. 14.

When the temperature decreased from 200K to 80K, the DCR was reduced by a factor of 4. DCR is dominated by band-to-band tunneling below 200K and has weaker dependence on T. At temperature of 80K, DCR density reaches around 150Hz/μm^2. The cryogenic dark noise is consistent with planar SPAD at room temperature in which the trap-assisted avalanches are suppressed.
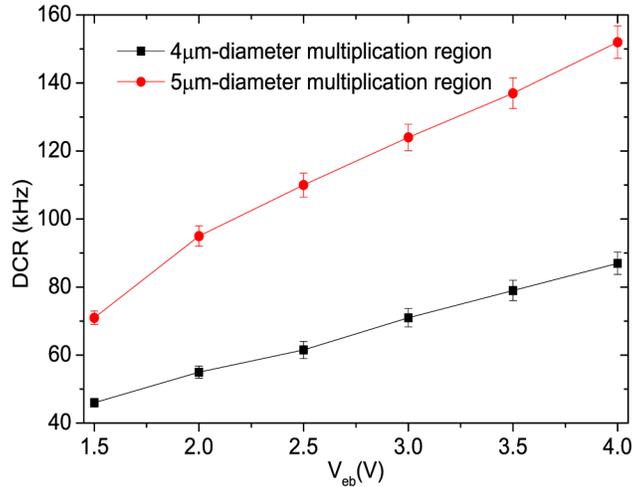
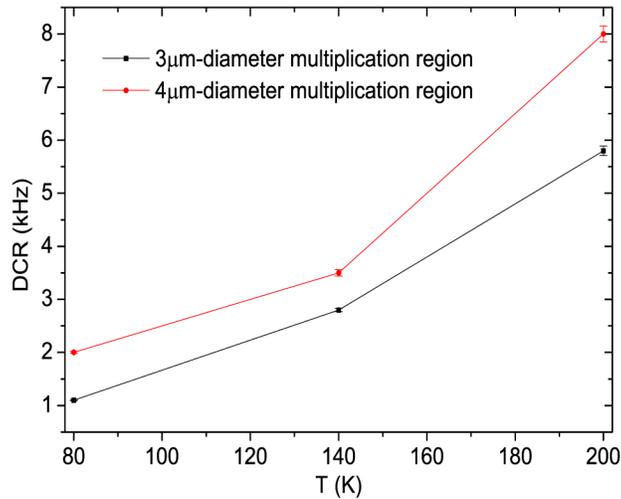Fig. 13. DCR as a function of excess bias in a fully integrated CMOS buffer SPAD pixel.

Fig. 14. DCR in a fully integrated CMOS buffer SPAD pixle as a function of cryogenic temperature. (V_{eb} = 1.5V; d: diameter of multiplication region).

To characterize the sensitivity of the pixel, the PDP was measured in both FSI and BSI, as shown in Fig. 15. The peak PDP was measured to 13% in FSI and 12.5% in BSI, showing similar PDP performance. Thanks to the buffering circuits, the dead time, which depends on RC recharge time in a passive quenching scheme, was reduced to around 160ns and consequently the PDP was enhanced because of the increased saturation frequency, which determines the upper limit of photon flux detectability. Furthermore, compared with our previous work [13], the PDP performance of BSI SPADs was also improved. The main reason is the epitaxy body doping diffusion into the intrinsic silicon layer on top of the buried oxide layer (Fig. 8(b)) with a large amount of thermal budget in CMOS process. Hence, the electric field across the neutral region near the buried oxide could be enhanced, resulting in increased $p(z)$ between z_3 and z_w and larger probability for the photon-generated carriers in BSI to drift to the depletion region where avalanches are triggered.
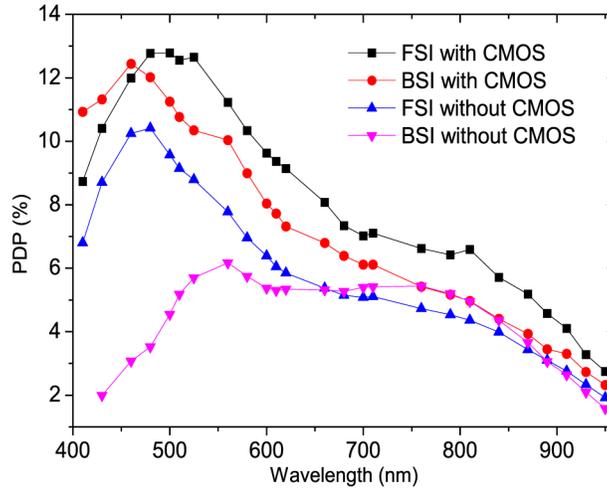
Fig. 15. PDP of frontside- and backside-illumination comparison between devices with/without co-integrated CMOS process.
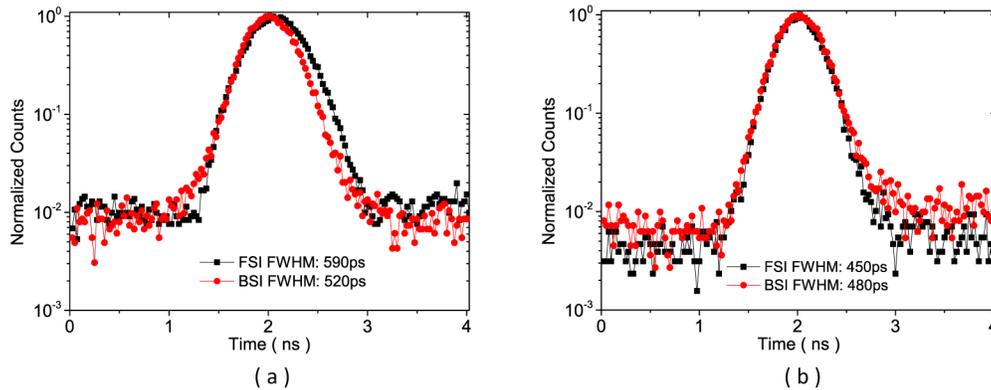

Fig. 16. Timing jitter measurements of flexible CMOS SPAD sensor pixel in FSI and BSI (a) using 405nm laser (b) using 637nm laser.

This observation is further confirmed by jitter performance measurements, which show the uncertainty of the time interval between the arrival time of the photon at the SPAD and the time when the pulse is generated.

To evaluate the timing jitter of the flexible CMOS SPAD sensor, the SPAD sensor was illuminated in both FSI and BSI modes by a laser source operated at 40MHz emitting light pulses with a few picoseconds of timing jitter. Neutral density filters were used to attenuate the light to single photon regimes. The time difference between the laser output trigger and the falling edge from the inverter buffer, which was synchronized with the avalanche rising pulse from a quenching resistor, was measured using a high speed oscilloscope (Lecroy Wavemaster 8600A, trigger jitter: less than 2.5ps). The statistical distribution of the time difference is used in estimating the device jitter in terms of full width half maximum (FWHM).

In this work, the jitter measurement performed using two different lasers with wavelengths of 405nm and 637nm (Advanced Laser Diode Systems GmBH) and timing jitter of less than 3ps. The results are summarized in Fig. 16. A 590ps and 520ps FWHM jitter was measured in FSI and BSI respectively, when the pixel was exposed to the 405nm laser. Compared with single SPAD without CMOS integration [13], the jitter mismatch between FSI and BSI

operation has been reduced significantly, indicating similar charge collection time uncertainty when the light reaches the front side and the back side of the sensor.
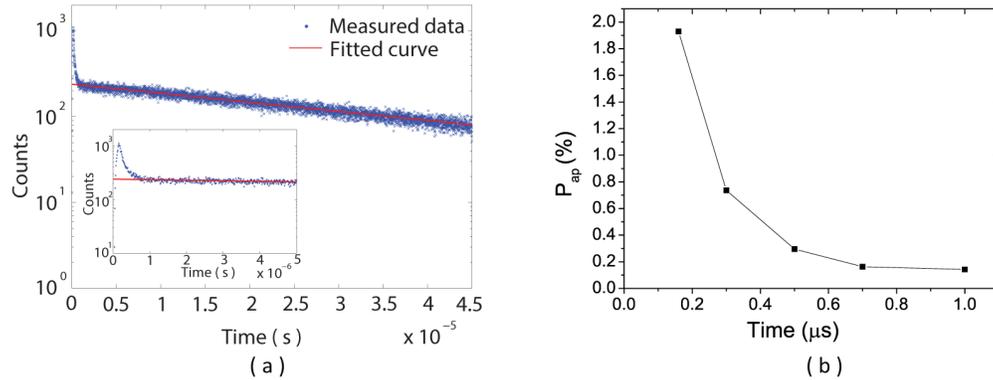

Fig. 17. (a) Exponentially fitted inter-arrival time histogram and detail of the histogram in the inset; (b) afterpulsing probability plot extracted from the distribution.

When the pixel was exposed to the 637nm laser, a FWHM jitter of 450ps was measured in FSI and 480ps was measured in BSI. The measured jitter had a smaller FWHM value when using a 637nm laser than a 405nm, indicating that the photons with higher penetration depth generate carriers closer to the multiplication region. The symmetric jitter performance in FSI and BSI is consistent with PDP measurements.

Afterpulsing is a type of correlated noise caused mostly by traps. Trapped carriers have activation energy levels close to the energy bands that can capture and hold carriers during the avalanche process within a release lifetime in the order of nanoseconds. The minimum hold off time is limited by afterpulsing [30]. A histogram of avalanche inter-arrival times is shown is Fig. 17(a) with an exponential fit. When a dead time of 160ns is implemented, the afterpulsing probability is around 1.9%, as shown in Fig. 17(b). At 1.5μs, afterpulsing probability is less than 0.1% or about 5 times better than in [13].

The comparison between flexible trench-isolated SPAD, planar SPAD, and flexible CMOS-buffering SPAD pixel is summarized in Table 1.

4. Conclusions

The Geiger-mode performance of flexible ultrathin-body SPADs has been comprehensively analyzed. Dark counts due to band-to-band tunneling and trap-assisted avalanche in this structure can be reduced to 0.15kHz/μm^2 by optimizing the multiplication doping and by operating at cryogenic temperatures, as low as 80K. By integrating pixel-level CMOS buffering circuits, the excess bias could reach 4V, thereby enabling a similar high peak PDP of 13% and better jitter performance with FWHM value of 450ps and 480ps in frontside- and backside-illumination respectively. Afterpulsing and crosstalk are negligible (less than 0.1% at nominal dead time larger than 1μs). This technology provides a suitable solution to advanced implantable photon counting devices for retinal prosthesis and other localized therapeutic/diagnostic solutions. Other applications include flexible multi-aperture imaging, anti-vignetting focal plane optimization, (implantable) bio-compatible chronic medical monitoring and wherever bended imaging sensors are essential.

Table 1. Performance comparison between flexible SPAD, planar SPAD and pixel

Performance	Flexible Trench-Isolated SPAD *			Planar SPAD			Flexible CMOS SPAD Pixel			Unit
	Min.	Typ.	Max.	Min.	Typ.	Max.	Min.	Typ.	Max.	
Avalanche breakdown voltage		17.2			25.5			26.5		V
DCR density @ RT (V_E = 1.5 V)	3.5		6.3	0.1		0.3		4.3		kHz/μm$_2$
DCR density @80K (V_E = 1.5 V)		0.4						0.15		kHz/μm$_2$
PDP @ FSI			14.3			13.0			13.0	%
PDP @BSI			6						12.5	%
Afterpulsing probability	0.5						0.15		1.9	%
Timing jitter @ FSI (637nm-wavelength laser)		900						450		ps
Timing jitter @ BSI (637nm-wavelength laser)		680						480		ps

* Configuration includes only SPAD and R$_q$. No CMOS transistor fabrication process.