TU Delft

Delft University of Technology

Flexible CMOS Single-Photon Avalanche Diode Image Sensor Technology

Sun, Pengfei

DOI
10.4233/uuid:11717f7d-51c9-471b-8f9e-ee1a70e7f032

Publication date
2016

Document Version
Final published version

Citation (APA)
Sun, P. (2016). *Flexible CMOS Single-Photon Avalanche Diode Image Sensor Technology*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:11717f7d-51c9-471b-8f9e-ee1a70e7f032

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Flexible CMOS Single-Photon Avalanche Diode Image Sensor Technology

Pengfei Sun

孙鹏飞

Flexible CMOS Single-Photon Avalanche Diode Image Sensor Technology

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rectro Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College van Promoties,
in het openbaar te verdedigen op 16 september 2016 om 10:00 uur

door

Pengfei SUN
Elektrotechnisch ingenieur

Geboren te Jingdezhen, China

This dissertation has been approved by the
promotor: Prof. dr. P. M. Sarro
promotor: Prof. dr. E. Charbon
copromotor: Dr. R. Ishihara

Composition of the doctoral committee:

Rector Magnificus	Chairman
Prof. dr. P. M. Sarro	Delft University of Technology, promotor
Prof. dr. E. Charbon	Delft University of Technology, promotor
Dr. R. Ishihara	Delft University of Technology, copromotor

Independent members:

Prof. dr. M. Wolf	University Hospital Zurich, Switzerland
Prof. dr. P. Magnan	ISAE, Toulouse, France
Dr. S. Nikzad	Jet Propulsion Laboratory, USA
Prof. dr. W. Uhring	University of Strasbourg, France
Prof. dr. P. French	Delft University of Technology, reserve member

Keywords: single-photon avalanche diode, SOI, flexible substrate, CMOS integration, photon counting and image sensor

Copyright © 2016 by P. Sun

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission of the author.

To my family

Contents

Nomenclature

3D	Three dimension
APD	Avalanche photodiode
APS	Active-pixel sensor
BOX	Buried oxide
BSI	Back-side illumination
CCD	Charge-coupled device
CPG	Ceramic pin grid
CMP	Chemical-mechanical polishing
DCR	Dark count rate
DRIE	Deep reactive ion etching
ECR	Excess count rate
EMCCD	Electron multiplying charge-coupled device
FPGA	Field programmable gate array
FSI	Front-side illumination
FWHM	Full-width-at-half-maximum
IGZO	Indium gallium zinc oxide
IRF	Instrument response function

LOCOS	Local oxidation of silicon
LPCVD	Low-pressure chemical vapor deposition
MRI	Magnetic resonance imaging
NIRS	Near-infrared spectroscopy
NDF	Neutral density filter
OPD	Organic printed photodetectors
OTFT	Organic thin film transistor
PCB	Printed circuit board
PDE	Photon detection efficiency
PDMS	Polydimethylsiloxane
PDP	Photo detection probability
PEB	Premature edge breakdown
PECVD	Plasma-enhanced chemical vapor deposition
PEN	Polyethylene naphthalate
PET	Positron Emission Tomography
PI	Polyimide
PMT	Photomultiplier tube
PVD	Physical vapor deposition
QE	Quantum efficiency

SEM	Scanning electron microscopy
SOI	Silicon on insulator
SPAD	Single-photon avalanche diode
SRH	Shockley-Read-Hall
TCSPC	Time-correlated single photon counting
TEOS	Tetraethyl orthosilicate
TFT	Thin film transistor
TOF	Time of flight

Chapter 1

Introduction

This thesis proposed and demonstrated the world's first flexible CMOS single-photon avalanche diode image sensor, providing a suitable solution for implantable biomedical imaging or monitoring applications and wherever a curved imaging plane is essential. This chapter serves as introduction to this thesis. First, the photon-counting technology in state-of-art biomedical applications is explained in Section 1.1. This is followed by a discussion in Section 1.2 on novel biomedical devices such as implantable retinal prosthesis and wearable near-infrared spectroscopy (NIRS) monitor, where flexible single-photon sensors are critical. Overviews of different photon sensor devices and typical flexible electronics solutions are presented and compared in Section 1.3 and 1.4 respectively. Finally, the goal of this research work is presented in Section 1.5 and this thesis is outlined in Section 1.6.

1.1 Photon-counting for biomedical imaging applications

Biomedical imaging [1] is receiving more and more attention as a collection of technologies to create visual representations of the interior of a body for clinical analysis and medical intervention. It could be classified into several branches, including radiography, magnetic resonance imaging (MRI), nuclear medicine, ultrasound, elastography, photoacoustic imaging , tomography and so on [2]–[5].

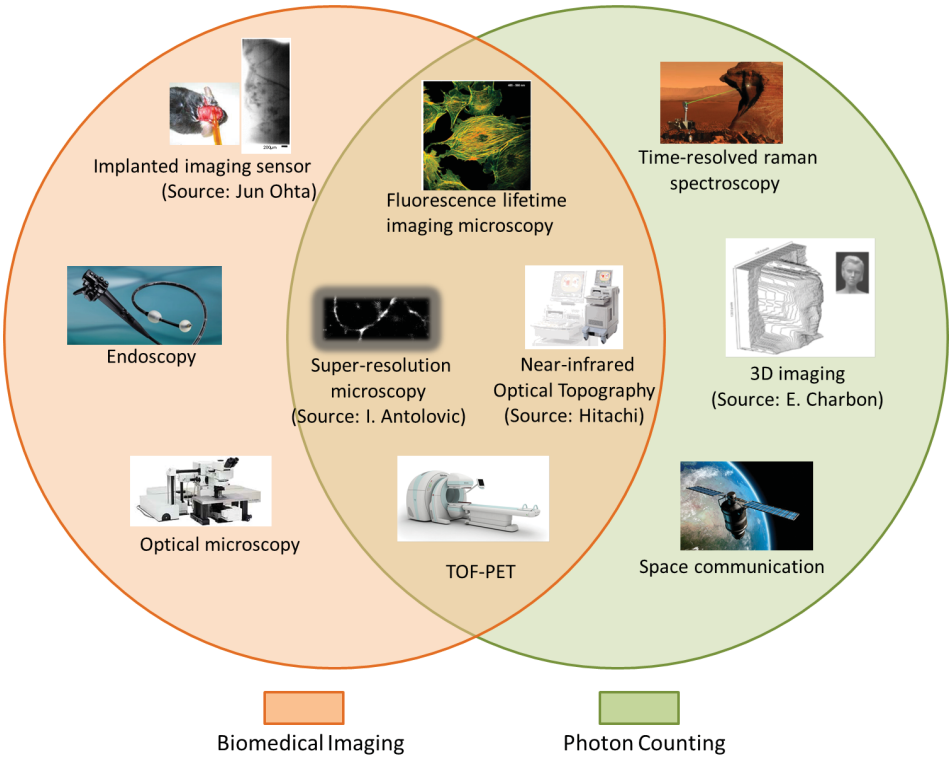


Fig. 1.1: Image sensor for biomedical applications.

Advances in imaging technologies are enabling us to create innovatively the way to treat disease, which are exciting, often non-invasive methods for probing and imaging the internal aspect of body or cell. Thanks to imaging technology, more and more biomedical instruments and facilities for clinical analysis and scientific research have been developed, as shown in Fig. 1.1 [6][7].

As a promising imaging technology, photon counting is being developed in biomedical imaging applications and many other fields such as time-resolved Raman spectroscopy, 3D imaging and even space communications [8]–[10].

Advanced biomedical imaging, such as microenvironment study of fluorescent molecules, require precise measurement of subnanosecond lifetimes based on very weak optical signals [8]. The requirement to detect single photon, which is the ultimate level of sensitivity in optical radiation sensor, makes it very challenging for conventional imaging technique, which mainly relies on sufficient light intensity. So photon-counting technique becomes more and more important in biomedical imaging fields in recent decades, such as time-correlated single photon counting (TCSPC) offers both single photon sensitivity and picoseconds temporal resolution. Whether it is for accurately locating the position of a tumour, providing an early diagnosis of a disease, achieving unique insight into the fluorescence lifetime imaging or even getting a super-resolution image of a biomolecule, the development and application of photon-counting technologies offers an exciting opportunity for micro-electrical researchers to collaborate with life scientists and clinicians.

1.2 Novel development of biomedical imaging technology

Tab. 1.1: Trend of biomedical imaging technology development. (source: Jun Ohta)

Hospital (Conventional)	Personal Use (New trend)
X-ray camera	Pill camera
MRI, PET	Healthcare chip
NIRS	Mobile NIRS
Endoscopy	Capsule endoscopy
DNA sequencer	DNA chip
Stimulation needle	Retinal prosthesis

Especially in the last decade, the push towards new applications, such as pill cameras, healthcare chips, capsule endoscopy, retinal prosthesis, edible probes and implantable sensors has continued [6][11][12].

Different from conventional applications, as Tab. 1.1, these new configurations contain novel CMOS image sensor chips, which allow bio-materials or living tissues to be in direct contact with the surface, enabling more compact biomedical imaging systems.

Such a compact system can also be implanted into a living body. The implantation of an imaging system also enables new applications of clinical devices. A typical application in the field of biomedical sensing is the retinal prosthesis [13]–[15], which is a non-living, electronic substitute for the retina. It aims to restore vision to someone blinded by retinal eye disease. Another application is chronic biomedical monitoring [16], where a wearable or implantable miniaturized image 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.

To develop a imaging device such as retina prosthesis, current imaging sensor technology has many challenges and will be explained in detail in the following sections.

1.2.1 Retinal prosthesis

As shown in Fig. 1.2 (a), the human retina is a thin layer of tissue with a thickness of 0.1–0.4 mm lining the inner surface of the eyeball. Incident light is imaged onto the retina by the lens. Photon flux with light pattern information is absorbed by the layer of photoreceptor cells, shown in Fig. 1.2 (b), and translated into electrical signals. The optic nerve sends photo-signals to the brain to form the image. So the retina plays an important role in visual information collection and processing, and so dysfunction can result in blindness [17].

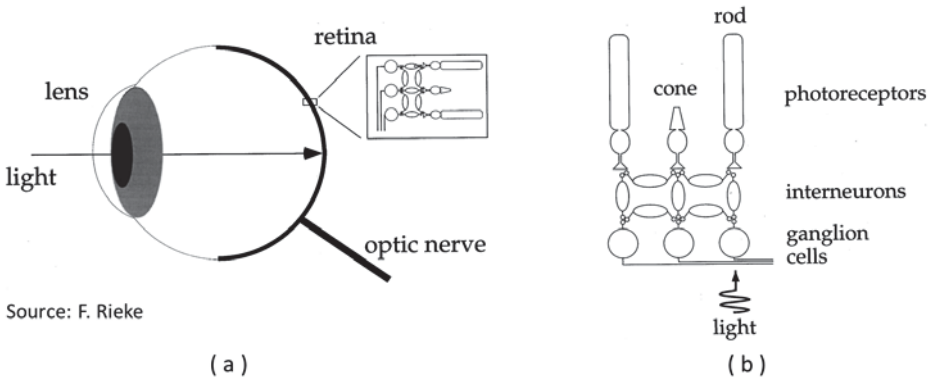


Fig. 1.2: (a) Structure of eye and retina; (b) Cell structure of photoreceptor. (Source: F. Rieke)

As early as the 1990s, the first retinal prosthesis prototype was proposed by researchers from John's Hopkins University and North Carolina State

University. The retinal device is composed of several parts. A camera is mounted on a pair of glasses to receive images. The image is then compressed by a video chip into a data stream and wirelessly sent to a receiving chip, implanted in the eye cavity. The receiving chip decodes the data and transforms it into electrical impulses that are then relayed to a silicon chip resting on the retina. The chip stimulates the ganglion cells, which sends the signal to the visual cortex for interpretation. This prototype achieved preliminary success to help the blind to see [18].

Researchers from Nara Institute of Science and Technology (NAIST) proposed an intraocular retinal prosthesis, which has been receiving wide attention; the device is shown in Fig. 1.3 [12]. By using CMOS technology, the stimulation chip with large number of electrodes and scanning electronics could be implanted into retinal tissues. An array of CMOS stimulation chips could further be packaged to flexible substrate by flip chip. According to the stimulator implantation site, this method could be classified into three types: epi-retinal, sub-retinal and STS stimulations as shown in Fig. 1.3(b).

In these methods, the power supply and stimulus pattern data are generated from input image data and transmitted by wireless technology such as electromagnetic coupling coils.

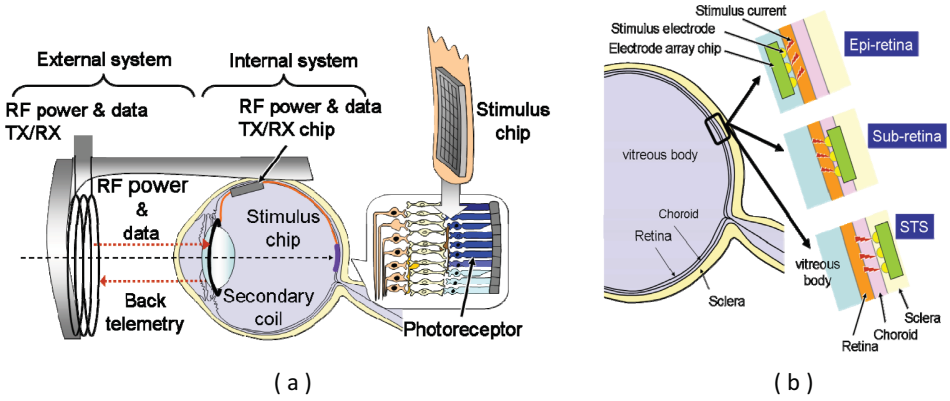


Fig. 1.3: (a) Typical configuration of an intraocular retinal prosthesis system proposed by NAIST; (b) Retinal stimulation types.

A better solution could be to implant a stimulation microelectronic chip integrated with photodiode array, which could take the function of degenerated photo-receptor cells would strongly enhance the capability of retinal prosthesis. As proposed by a medical start-up company, Retina Implant AG, the system shown in Fig. 1.4 comprises a 3x3mm^2 microelectronic chip (0.1-millimeter thick) containing ~1,500 light-sensitive photodiodes, amplifiers and electrodes.

It was reported that it could help generate at least partial vision by stimulating intact nerve cells in the retina [19].

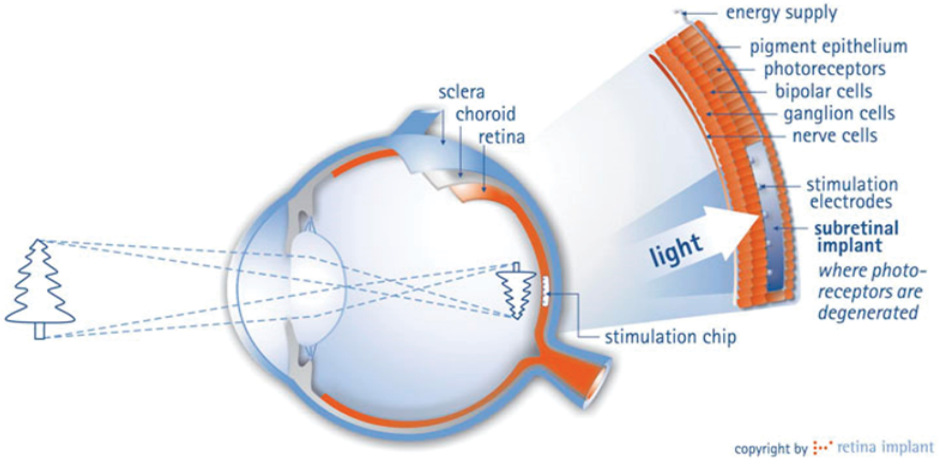


Fig. 1.4: Retinal prosthesis prototype proposed by Retina Implant AG.
(Source: Retina Implant AG)

Tab. 1.2: Comparison of different retina prosthesis technologies.

	Photoreceptor	***Stimulus***	***Transmission***	***Package***
JHU/NCU	External	Pads	Wireless	Chip
NAIST	External	*Electrode Array*	RF	*Flexible PI*
R. I. AG	*Photodiodes*	Electrodes	*Interconnection*	Chip

By comparing these methods listed in Tab. 1.2, an ideal retina prosthesis includes a photoreceptor and a high density stimulus array assembled in a compact and flexible substrate, which can be packaged with flexible materials. As far as we know, the main photoreceptor-function device in current retina prostheses is a conventional photodiode, which can't fully take the replacement of photoreceptor due to the limited sensitivity in a dark environment.

As shown in Fig.1.2(b), the photoreceptor cell is defined as a special neuron in retina, which could be classified into rod, cone and photosensitive retinal ganglion cells. Different from the other two cells, rod cell is extremely sensitive and capable of detecting the absorbing single photons. At very low light levels, visual experience is based on the rod cells.

Photon-counting devices in retina prostheses could be an ideal solution to take the function of rod cells in photoreceptor, with several prerequisites: (1) high fill factor, (2) low dark noise, (3) CMOS buffering and addressing circuits integration, and (4) flexible substrate fitting within the eyeball.

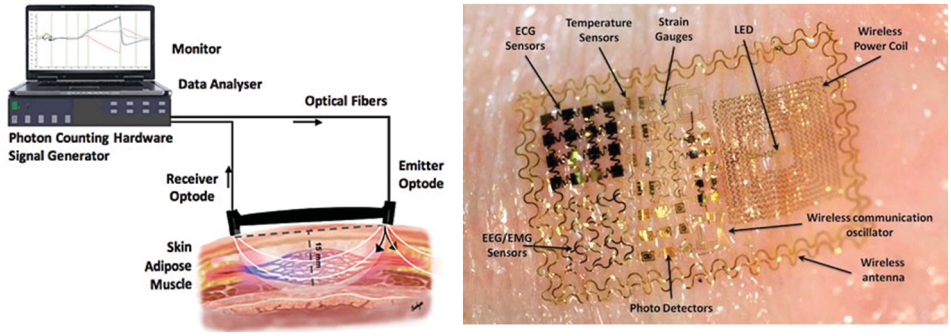
1.2.2 Implantable biomedical monitor

In the last decades, photon-counting technology has been used widely in both industry and academia for biomedical monitoring applications, such as molecule confocal microscopy, PET, TCSPC fluorescence lifetime imaging, and time resolved near-infrared spectroscopy (NIRS) [9][20]. Thanks to the development of advanced photon-counting devices, high photon detection efficiencies, sub-nanosecond timing jitter and low dark noise performance of devices, such as APD or SPAD, have been enhancing photon-counting biomedical instruments [21].

As shown in Fig. 1.5(a), NIRS was developed based on photon-counting mechanism to monitor changes in local tissue oxygen and hemodynamics in real time [22]. In this instrument, lasers or diodes transmit pulses of light into tissue and optical sensors detect the returning photons. Raw optical data could be extracted from the changes in absorption of light at discrete wavelengths and then converted by mathematical software algorithms, which is based on a modification of the Beer-Lambert law, into real-time concentration changes for each chromophore [23].

Nowadays, thanks to advanced CMOS and ultra-fine fabrication technology, implantable and wearable biomedical monitoring devices are receiving wide attention as a promising alternative for the traditional biomedical monitoring instruments. As shown in Fig. 1.5(b), an "electronic skin" proposed by UIUC and NIBIB [24], which adheres non-invasively to human skin, provides a detailed temperature map of any surface of the body.

Based on this idea, it is conceivable to create photon-counting biomedical monitor instruments, which are wearable or implantable, such as mobile and *in-situ* NIRS.



(a) (b)

Fig. 1.5: (a) NIRS instrumentation setup is configured for transcutaneous monitoring of muscle; (b) "electronic skin" to monitor body temperature.

1.2.3 Flexible single-photon sensor solution

To meet the requirements of implantable or wearable biomedical applications, novel photo detector solution needs to be developed, in which backside-illumination and new substrate post processing are core technologies, while inherent CMOS compatibility is a prerequisite [25][26] as shown in Fig. 1.6.

The ideal solution should address two challenges, one of which is that the photon detector needs to be highly sensitive to detect single photons with low noise while at the same time being CMOS compatible. The other challenge is that the photon detector should have high fill factor to let a high enough photon flux impinging from both front- and back-sides through. Furthermore, the substrate material should be biocompatible and bendable to fit in the curved imaging plane in the human eyeball or other body surface.

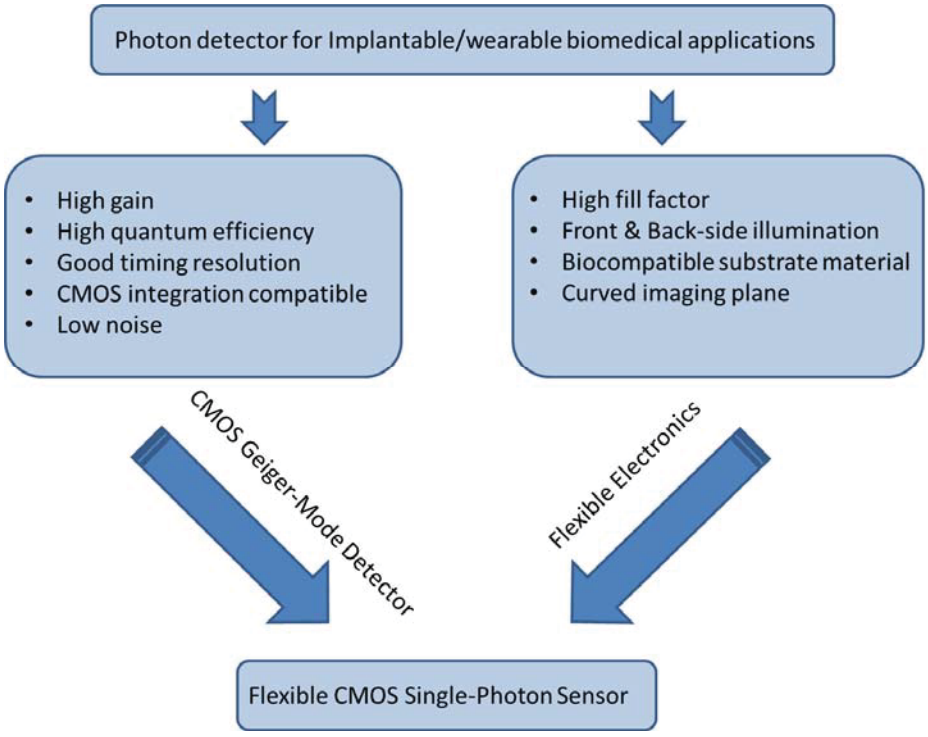


Fig. 1.6: Flexible single-photon sensor solution.

1.3 Photon-counting devices

1.3.1 Photomultiplier tube (PMT)

A PMT is a vacuum tube consisting of a photo cathode, a serial of electron multiplier dynodes and an anode sealed in an evacuated glass enclosure. A high voltage, typically in the range of hundreds to thousands of volts, is used to bias the dynodes (electrodes used for electron multiplication). When a photon of sufficient energy strikes the cathode, a photoelectron is ejected by photoelectric effect and accelerated towards to a serial of multiplier dynodes. At each dynode, a photoelectron would generate 10^5 to 10^7 electrons. There, electrons are collected at the anode and generate an electrical signal. As a photon-counting device, PMTs have a high gain and good timing resolution but often suffer from significant noise. Furthermore, they are bulky and sensitive to B fields.

1.3.2 CMOS active-pixel sensor (APS)

APS is a mainstream imager widely used in consuming electronics in recent years. Each APS pixel has its own in-pixel individual amplifier, which is configured to boost the photo-generated charge signal towards the column bus. Originally APS pixels suffer from fix-pattern noise, reset noise and other thermal noise sources [27]. By introducing the pinned photodiode and correlated double sampling, the noise level could be significantly reduced. However, due to the low gain, the APS needs external components to amplify photon-generated carriers in order to act as photon counter [28]. In addition, timing resolution is still not good enough for most time-resolved applications.

1.3.3 Charge-coupled device (CCD)

Charge-coupled devices (CCDs) are widely used as solid-state imagers. In CCD, photon-generated charges are transferred between MOS capacitors. Thanks to an improved technology, over these years high quantum efficiency (QE) could be achieved. A gain stage could be added between device and readout to make it sensitive to single photons, known as Electron Multiplying CCD (EMCCD) [29]. The state-of-art EMCCD is capable of single-photon sensitivity. A thresholding scheme is usually used to distinguish single-photon events from noisy events. However, excess noise, due to the stochastic nature of the electron multiplication through cascade amplifications, is still a significant drawback for many biomedical applications. Furthermore, The use of gating and limited frame rate don’t allow free-running and high-speed imaging applications.

1.3.4 Avalanche photodiode (APD)

APDs are essentially a pn junction, reverse-biased below breakdown. The electric field is high enough to cause impact ionization, allowing active amplification of photo-generated carriers. As the bias is close to the breakdown voltage, APDs can achieve an optical gain exponentially related to the bias, reaching typically a factor as high as ten thousand. APDs have been used as photon counters in many applications [30], however, poor timing resolution and sizable non-uniformity are a limiting factor.

1.3.5 Single-photon avalanche diode (SPAD)

A SPAD is Geiger-mode APD with special structure design allowing for operation above breakdown. In this mode, the electric field is high enough to cause a self-sustaining avalanche. To avoid the destruction of the device, quenching circuitry (either external or integrated) is required. Thanks to the maturity of CMOS technology, SPADs could be integrated with CMOS circuitry on chip, thus largely reducing the performance, compactness, and cost of SPAD devices, thus widening the range of applications, such as time-of-fight imaging and LIDAR, positron emission tomography (PET), time-resolved Raman spectroscopy, and near-infrared optical spectroscopy, to name a few [31].

1.4 Flexible device fabrication solutions

1.4.1 Silicon substrate thinning

Silicon wafer thinning, down to less than 50μm, causes silicon to be flexible up to some certain extent; this is a straightforward way to realize flexible devices with conventional IC technology.

There are several techniques to thin down silicon wafers such as wet or dry etching, grinding, and chemical-mechanical polishing (CMP) [32]. However, all of them still have difficulties to reach a very accurate thickness and always generate defects at the bottom surface of the silicon substrate. It is also very challenging to handle a substrate with such a small thickness. Many solutions have been developed to make this process more reliable for production [33]. At this moment, it is still not a proper solution for high performance flexible CMOS sensors.

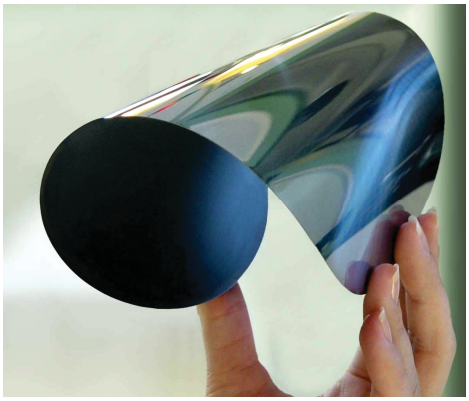
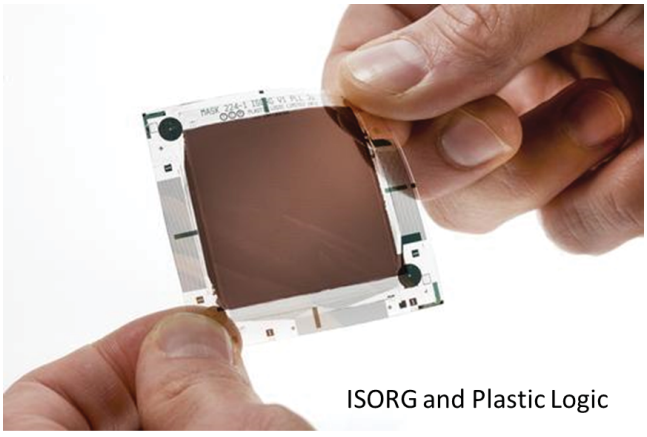
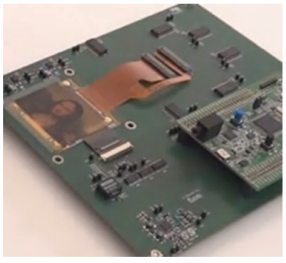


Fig. 1.7: Wafer thinning down to 30µm by remote plasma etching. (Source: PVA TePla)

1.4.2 Low-temperature thin film processing



(a)



(b)

(c)

Fig. 1.8: (a) Flexible plastic 96x96 image sensor; (b) demonstration board; (c) transparent image of Leonardo da Vinci's Mona Lisa.

In the large-area electronic field such as flat display panel, laser-crystallized thin film transistor (TFT) technology has been widely used because of the low thermal budget. The semiconductor material, which could include silicon, germanium, indium gallium zinc oxide(IGZO), cyclopentasilane, and so on, could be directly coated, deposited, or sputtered to form thin film onto flexible substrate and then crystallized to a polymorph state from amorphous state [34][35]. The drawbacks are non-uniformity and a relatively poor device performance due to the defects during (re-)crystallization.

Recently, organic printed photodetectors (OPD), single-grain TFT and organic TFT (OTFT) are receiving attention in the imaging community [36]–[38]. As shown in Fig. 1.8, ISORG and Plastic Logic proposed the first flexible plastic image sensor and the transparent image of Mona Lisa. The organic thin film was deposited on plastic substrate and then form the TFT and photodiode. Due to the very low mobility of organic materials, only static and transparent image was demonstrated in this technology.

Overall, this solution could be suitable in large-area electronics rather than applications where operating speed or integration density are critical.

1.4.3 High performance device and circuitry chip transfer.

In applications where high-performance CMOS transistors or high-frequency circuitry are critical, substrate transfer solutions were developed to avoid the limitations of low temperature thin film processing. Thanks to the ultra-fine transfer technologies [39], several approaches have been demonstrated to realize high-performance transistor and circuitry modules on flexible substrate. Researchers are currently developing new materials, such as carbon nanotubes, graphene platelets, nanowires and so on for a semiconductor layer [40][41], which could be easily lifted off and transferred to a flexible substrate. By smart package techniques, photodiode arrays and other optics and electronic modules could be assembled to stretchable substrates to form arbitrary surfaces, such as a hemispherical compound eyes [42], such as the one shown in fig. 1.9.

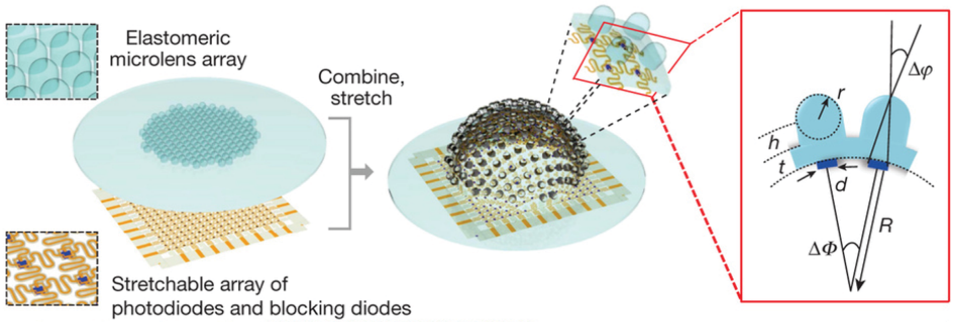


Fig. 1.9: Digital camera in form of a hemispherical, apposition compound eye.

Recently, by introducing thin, low modulus elastomers to isolate the active circuit materials from applied strains, it is a promising way to integrate high performance CMOS circuits on diverse classes of substrates such as fabric, leather or even paper [43]. However, it is necessary that ultrathin CMOS circuits be planar; serpentine mesh geometries have been fabricated before they could be released through lifting up by PDMS stamp. However, the dimension of individual CMOS circuitry modules is hundreds of micrometers, as shown in the figure. Furthermore, the whole process is quite complex and high cost.

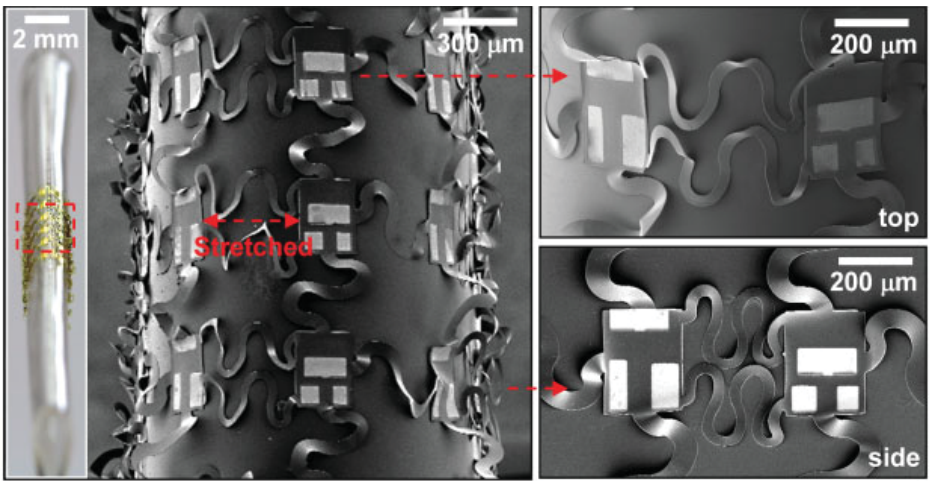


Fig. 1.10: Optical image of a folded circuit (left) consisting of an array of CMOS inverters and SEM images at different places.

1.4.4 SOI substrate transfer

The Circonflex technology, proposed by Philips Research, provides a suitable solution to realize high-performance large-scale ICs on flexible substrates. After high-performance devices are first processed on silicon-on-insulator (SOI) wafers, a thin polyimide layer is coated and cured on top of the circuitry layer and a glass carrier wafer is glued in place. The whole stack is then flipped upside-down and the whole Si substrate is removed by grinding or etching, using the BOX layer as stop layer. The polyimide foil can be easily de-laminated from the glass substrate thanks to poor adhesion. This method has been applied successfully in making flexible RFID taps [44].

Similarly, the FleX™ Silicon-on Polymer™ process was proposed by American Semiconductor, Inc. to transform standard full thickness SOI wafers into flexible wafers, as shown in Fig. 1.11, enabling new attachment and connection methods that are compatible with flexible circuit manufacturing. Furthermore, based on flexible IC, Flexible Hybrid Systems have been developed to stack printed electronics according to the report in [45].

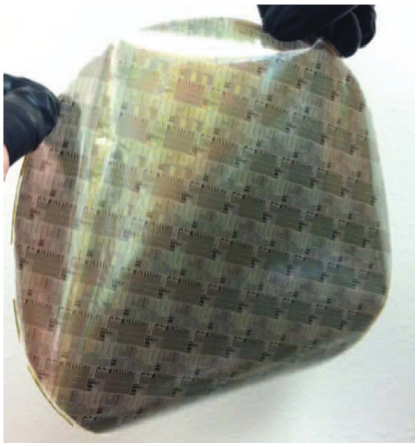


Fig. 1.11: FleX wafer – 130nm CMOS. (Source: American Sem. Inc.)

Table 1.3 shows the comparison of different flexible electronic solutions. Device transfer and SOI substrate transfer are preferred for high-performance CMOS imaging devices. Considering the large-scale integration, micrometer device dimensions and process cost, SOI substrate transfer is a more suitable solution.

Many demonstrations such as FleX-MCU, FleX-ADC and FleX-RFIC have been successfully achieved. In this thesis, we extend this solution to the imaging sensor field, to enable flexible high-performance image sensors with several

innovations, such as dual-side illumination, for implantable biomedical applications.

Tab. 1.3: Comparison of flexible electronic solutions.

	Substrate thinning	***Thin film processing***	***Device transfer***	***SOI substrate transfer***
Device performance	Good	Poor	Good	Good
Integration scale	High	Low	Middle	High
Mechanical flexibility	Middle	High	High	Good
Process complexity - cost	Low	Low	High	Middle

1.5 Goals of this thesis

The **main goal** of this work is to achieve the world's first flexible CMOS single-photon sensor for applications targeted towards implantable biomedical devices, such as retinal prostheses, and wearable medical monitor devices. This thesis also provides an innovative platform for many other applications, such as flexible multi-aperture imaging, anti-vignetting focal plane cameras and wherever bended imaging sensors are essential.

In order to realize a flexible CMOS single-photon sensor, SPAD would be the core photodiode in this work because of high gain, good timing resolution and CMOS integration compatibility, compared with other photon-counting devices. By comparing typical flexible electronics solutions, SOI substrate transfer technology would be a preferred solution to realize CMOS large-scale SPAD imaging sensor system.

The **first contribution** is to design and fabricate an ultrathin-body SPAD device structure in SOI process. Special isolations of devices should be designed to release mechanical stress when the chip was bended. Thorough measurement was done to investigate Geiger-mode performance.

The **second contribution** is to propose a flowchart and to implement SOI substrate transfer process. Thorough measurements would be performed to compare the SPAD performances before and after substrate transfer. The world's first flexible ultrathin-body SPAD would be demonstrated based on this work.

The **third contribution** is to comprehensively analyze the Geiger-mode performance of flexible SPAD, which is configured with passive quenching

resistor on the same flexible substrate. Dark noise is thoroughly analyzed by different multiplication region doping, device isolation structures and cryogenic experiments. Optic parameters such as photo detection probability (PDP), and timing jitter are compared based on different illumination sides – front-side illumination (FSI) and back-side illumination (BSI).

The **forth contribution** is to integrate CMOS transistors with SPAD on the same substrate. The purpose is to pave the way to realize larger scale sensor array integrated with more electronics on flexible substrate. Based on single pixel of CMOS SPAD sensor, thorough optical and electrical measurements would be performed. Based on this foundation work, the ultimate objective is to realize the imaging system by designing and fabricating flexible 32x32 CMOS SPAD image sensor.

1.6 Thesis organization

In chapter 2, the structure of SOI SPAD is shown with device simulation. Fabrication flowchart and measurement results before and after substrate transfer are presented. It demonstrates the world's first flexible ultrathin-body SPAD.

In chapter 3, comprehensive analysis of dark noise, PDP, timing jitter in dual-side illumination and correlated noise of flexible SPAD are presented.

In chapter 4, results of SPAD integration with CMOS transistors on flexible substrate are presented. Based on that, the design and fabrication of flexible 32x32 CMOS SPAD image sensor are presented in chapter 5. Finally, chapter 6 gives the conclusions of this thesis.

Chapter 2

A novel photon counter: Flexible ultrathin-body SOI single-photon avalanche diode

This chapter focuses on the world's first flexible SPAD fabricated in an ultra-thin-body silicon-on-insulator (SOI) process followed by transfer post-processing to flexible substrate. Section 2.1 provides some background of SPAD device and the requirements in this work. The device theory of SPAD is presented in Section 2.2, followed by the introduction of terminologies of SPAD in Section 2.3. An overview of conventional SPAD fabrication based on CMOS technology is given in Section 2.4. Then the proposed SOI SPAD design and fabrication are presented in Section 2.5, followed by thorough characteristics in Section 2.6. Substrate transfer process is presented in Section 2.7 to achieve the flexible SPAD. Characteristics before and after substrate transfer are compared in Section 2.8 and conclusions are given in Section 2.9. This chapter is based on results published in [46][47].

2.1 Introduction

Single-photon avalanche diode (SPAD) technology has been 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. Due to improved CMOS fabrication, novel device structures, and new readout techniques, SPAD technology has significantly progressed in recent years [48]–[52].

Moore's Law has enabled the design of larger and larger imaging systems based on SPADs and, thanks to pixel scaling, higher fill factors [53][54]. When operating in front-side illumination (FSI) mode, the sensitive area of the pixel is partially blocked by metal layers; furthermore, the optical stack's layers generally create interference, due to mismatched refractive indices, which cause further light attenuation. When operating in backside-illumination (BSI) mode, SPADs enjoy potentially thinner optical stacks and no metal blockage, thus minimizing optical interference and achieving a virtually perfect fill factor [55]. Additionally, thanks to a different geometry, it is possible to implement deeper junctions, so as to enhance sensitivity at longer wavelengths.

Current SPAD technology is generally implemented on bulk silicon and it is still difficult to realize backside-illuminated devices [56]. Furthermore, to our knowledge, no SPAD has ever been fabricated on flexible substrate, as demanded by advanced biological and life sciences research and by future medical (implantable) applications.

In this chapter, we present the world's first flexible APD fabricated in an ultra-thin-body silicon-on-insulator (SOI) process followed by transfer post-processing to flexible substrate [46][47]. The flexible APD can operate both in proportional and in Geiger mode (i.e. SPAD), both in FSI and BSI modes. The SPAD multiplication region is located in a silicon layer island with a thickness of 1.5μm, thereby enabling a photon detection probability (PDP) of 11% at 460nm in FSI mode and 6% at 550nm and wider sensitive spectrum in BSI mode. The dark noise, characterized in SPADs with 3-μm diameter active area in terms of dark count rate (DCR) is less than 20kHz with an excess bias of 0.5V at room temperature and it is dominated by band-to-band tunneling, which tends to be less temperature dependent. Afterpulsing is negligible at a dead time longer than 1μs; crosstalk is negligible at a distance of 10μm. Applications include flexible multi-aperture imaging, anti-vignetting focal plane optimization, and (implantable) bio-compatible chronic medical monitoring.

2.2 Theory of SPAD

2.2.1 Steady-state current-voltage characteristics

Typically, the steady-state behavior of a p-n junction diode contains three regions: a forward region, a reverse region and a breakdown region.

As shown in Fig. 2.1, in forward region, when the applied voltage is larger than the inherent potential, there is a flow of current increasing exponentially with applied voltage. While in the reverse region, very little current flows but, as the electrical field increases, current gain also increases.

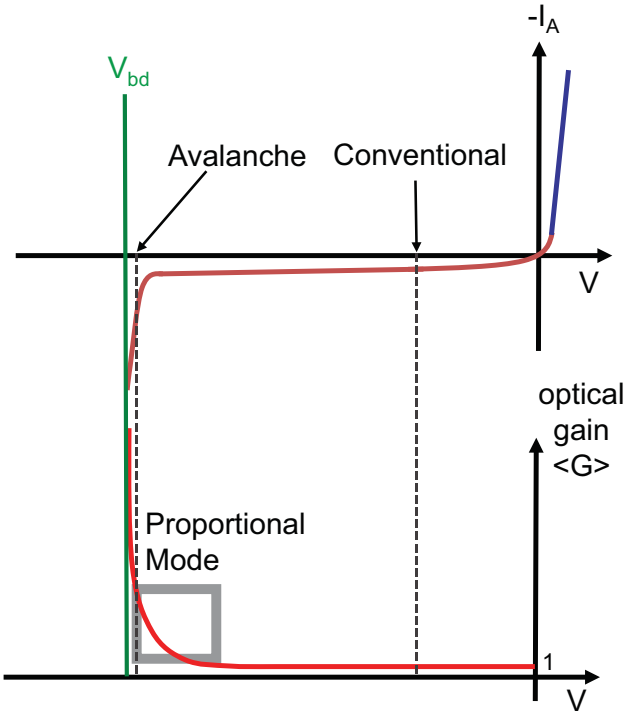


Fig. 2.1: Current-Voltage characterization and optical gain of APD below $V_{bd.}$

Before breakdown voltage, the ideal current density characteristics can be described as below:

$$J = J_p + J_n = J_s(e^{qV/kT} - 1), \quad (2.1)$$

$$J_s = \frac{qD_p p_{no}}{L_p} + \frac{qD_n n_{po}}{L_n}. \quad (2.2)$$

with D_n and D_p being diffusion coefficient; L_n and L_p being diffusion length for electrons and holes respectively.

Avalanche Photodiode is a pn junction operating below breakdown. We can achieve a current gain as much as ten thousand but the gain varies exponentially with bias.

2.2.2 Breakdown voltage and excess bias

Due to the impact ionization mechanism [57], the breakdown voltage is defined as the voltage where the multiplication factor approaches infinity M_n (Assume the avalanche process is initiated by electrons).
In this condition, the ionization integral is:

$$1-\frac{1}{M_n}=\int_0^W \alpha_n \exp\left[-\int_x^W (\alpha_n-\alpha_p)dx'\right]dx=1. \qquad (2.3)$$

Term α_n is the average ionization rate of electrons, and α_p the average ionization rate of holes. The ionization rates will increase as the applied voltage increases. Quantitatively, the breakdown voltage is defined at the applied voltage at which the breakdown integral reaches unity. The difference between the applied voltage, given by V_{op} and V_{bd} is termed the excess bias, V_{eb} .

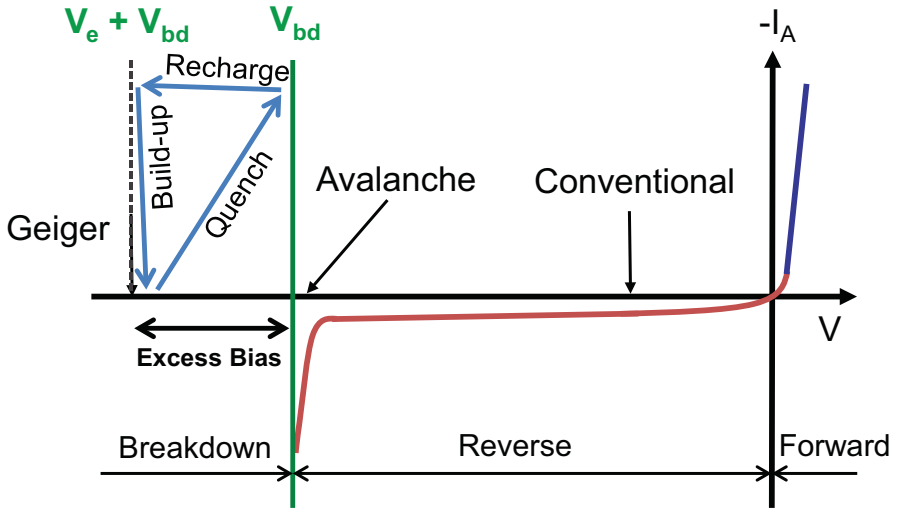


Fig. 2.2: SPAD's states in Geiger mode.

In breakdown region, which is above breakdown voltage, the electric field is so high that impact ionization would happen. An injection of hot carrier or photon-generated carrier would cause a self-sustaining avalanche of carriers as shown in Fig. 2.2 [58]. However, before the injection of the first carrier, there is

a transient state when the diode operates above breakdown voltage, only with leakage flowing through the junction. In order to use it as photon counter, quenching scheme is introduced and explained in Section 2.2.4.

2.2.3 Multiplication region and guard ring

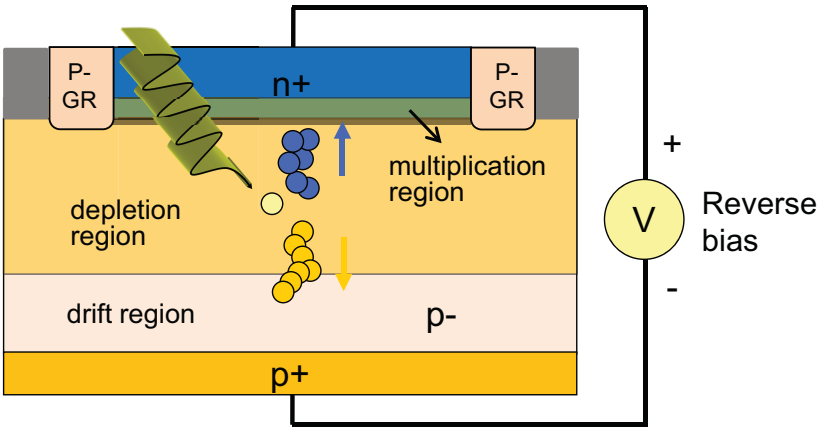


Fig. 2.3: Cross-section of SPAD junction regions.

As shown in Fig. 2.3, the reverse biased junction contains three regions, which are multiplication region, depletion region and drift region. The multiplication region is defined as the region where more than 95% impact ionizations occur. The electric field in the multiplication region is above the critical threshold for silicon and it should be higher than the other depletion regions and drift regions. The photo carrier generated in the multiplication region would cause self-sustaining avalanche immediately, while the photo generated carrier in other depletion region and lower-field drift region would drift into the multiplication region before causing an avalanche.

In order to form a highly uniform electric field in multiplication region, it is necessary to separate the multiplication region from the surrounding area. The structure to implement the separation is called “guard ring”. The purpose of the guard ring is to prevent premature edge breakdown (PEB), which is usually caused by higher electric field at the edges of doping regions. PEB is undesirable in a SPAD because it creates micro-active regions at the corner, while suppressing avalanching in the multiplication region [59].

In Fig. 2.3, the guard ring is implemented by an extra P^- doping. "Virtual" guard rings can also be possible by introducing higher implantation in the active region. The details will be reviewed in section 2.4.

2.2.4 Quenching scheme

One simple but effective way to stop the avalanche is the use of a ballast resistor between the anode of the diode and the ground or the cathode and positive voltage. By doing this, upon photon arrival, the avalanche current grows producing a voltage rise across the resistor. This brings the voltage across the diode below the breakdown voltage, thus, quenching the avalanche process. Then, the SPAD is recharged again above its breakdown voltage so it is ready to detect another photon. The time needed to perform a full detection cycle defines the dead time of the SPAD.

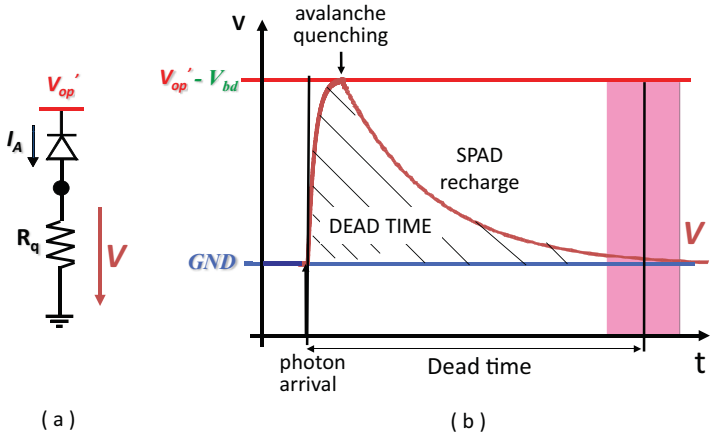
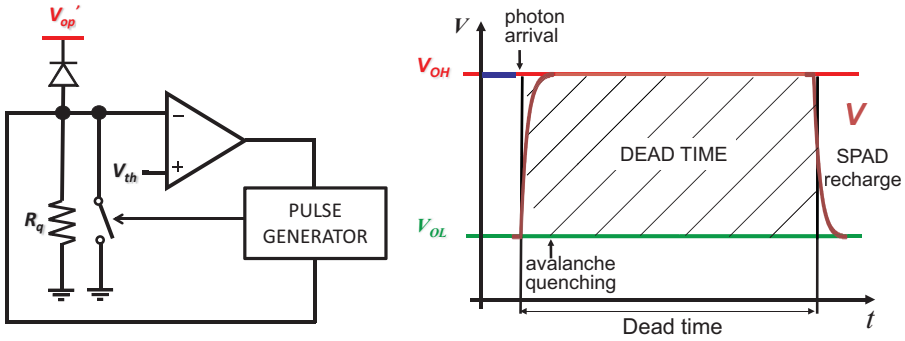


Fig. 2.4: (a) Passive quenching configuration; (b) Output waveform in passive quenching configuration.

The drawbacks of passive quenching circuits are mainly long recharge time and not well-controlled dead time. To overcome these drawbacks, active quenching circuits have been introduced [60]. As shown in Fig. 2.5, the operation consists of detecting the avalanche and control the recharge of the device via a reduced recharge resistance or a switch to ground (or Vdd). Both quenching and recharge may be controlled by active pulse generator circuits to control the dead time precisely.



(a) (b)

Fig. 2.5: (a) Active quenching configuration; (b) Output waveform in active quenching configuration.

2.3 Performance parameters for single SPAD device

2.3.1 Dark count rate

Dark count rate (DCR) is defined as the frequency of spurious pulses generated by non-photon-generated carriers, which are mainly thermally generated or tunneling-assisted carriers within the junction. DCR is an important parameter, which defines the noise level in the dark and limits the detection dynamic range from the low end. It is usually measured when Geiger-mode SPAD is shielded in completely darkness.

In principle, both trap-assisted avalanche and band-to-band tunneling contribute to DCR. It will be analyzed thoroughly in the Chapter 3.

2.3.2 Photon detection probability

Photon detection probability (PDP) is defined as the probability that a photon imping on the active region triggers an avalanche. Similar with so-called quantum efficiency in conventional photodiodes, PDP expresses the sensitivity of SPAD, which is always less than 100% due to factors such as parasitic reflection, non-ideal impact ionization and light's penetration depth.

Detailed and analytical insights of PDP will be reported in Chapter 3.

2.3.3 Dead time

Dead time is termed as the period when SPAD is under lower biasing conditions following an avalanche and unable to detect any further carriers or photons.

Dead time t_{dead} is an important parameter since it is related to the saturation count rate f_{sat}.

$f_{sat} \cong \frac{1/e}{t_{dead}}$ for passive quenching (Fig 2.4 (a)),

$f_{sat} \cong \frac{1}{t_{dead}}$ for active quenching (Fig. 2.5 (a)).

2.3.4 Correlated noise – Afterpulsing

Afterpulsing is a type of correlated noise caused by traps allowing energy levels close to energy bands to capture and hold carriers during avalanche process. Carriers are captured and hold by traps with a release lifetime on the order of nanoseconds.

A histogram of inter-avalanche arrival times is made to extract the dead time and afterpulsing probability as shown in Fig. 2.6.

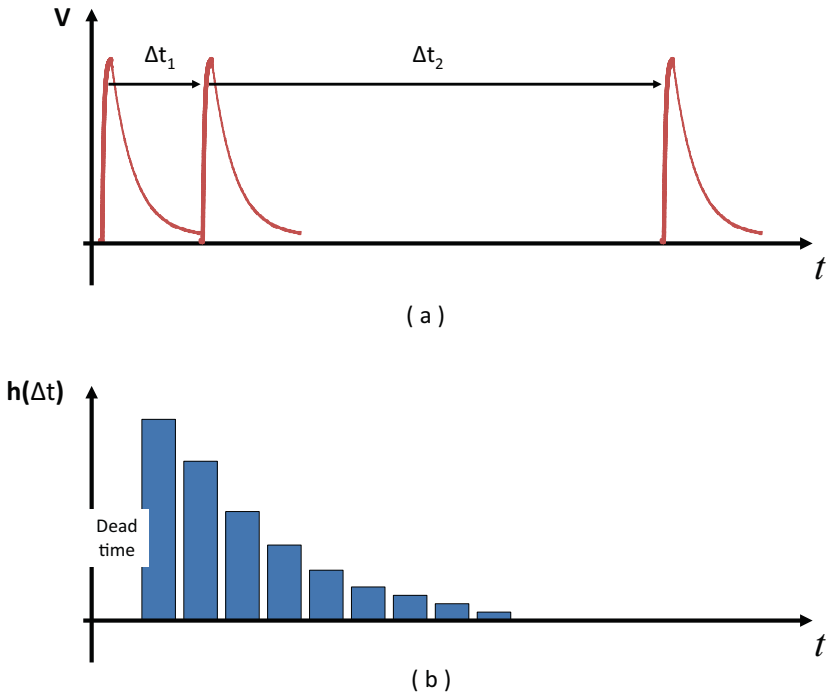


Fig. 2.6: (a) Afterpulsing pulses shown with inter-avalanche arrival times;
(b) Histogram of inter-avalanche arrival times.

Afterpulsing limits the minimum hold-off time, which is corresponding to dead time of SPAD. When dead time is too short, afterpulsing may increase and introduce correlated noise. By using advanced and cleaner device processes, trap density could be reduced, thus reducing afterpulsing. Afterpulsing can also be reduced by increasing dead time to allow for trap relaxation.

Afterpulsing probability P_{ap} is defined as the probability of an afterpulse per avalanche. P_{ap} at a specific dead time can be extracted by the inter-avalanche time histogram, fitting an exponential to the uncorrelated noise source. P_{ap} is then calculated due to the fraction of events above the fit curve but below the experimental curve. The details are described in [61].

2.3.5 Timing resolution

Timing jitter of SPADs is the uncertainty of the time interval between the arrival time of the photon and the time when the pulse is triggered. It is due to the statistical nature of the avalanche build-up. Instead of an identical build-up for every injected carrier, there will be uncertainty in the avalanche process.

To characterize the timing jitter, the SPAD device is usually illuminated by a picosecond laser emitting pulses with certain wavelength in time-correlated single-photon counting (TCSPC), whereby a histogram of the response of the detector is constructed and the standard deviation or FWHM of the instrument response function (IRF) is extracted. Neutral density filter (NDF) is inserted between the laser and the SPAD non-perpendicularly to attenuate the photon flux so that the SPAD under illumination could be operated in single-photon scheme.

The jitter is often modeled as a Gaussian curve convolved with the sum of a delta function and one or more exponential functions, mainly depending on the carrier diffusion and junction profile.

2.4 Conventional SPAD fabrication

2.4.1 Reach-through SPAD

The first avalanche photodiode with reach-through structure was proposed in [62]. It contains a thick portion of intrinsic silicon and can be scaled up to large area, thus resulting in high sensitivity. The diode configuration and impurity profiles are illuminated in Fig. 2.7. Thanks to the wide depletion region, which spans tens or hundreds of μm, such structure can detect imping photons of a large range of wavelength. Especially, silicon reach-through structure SPAD is very sensitive to near-infrared (NIR) light.

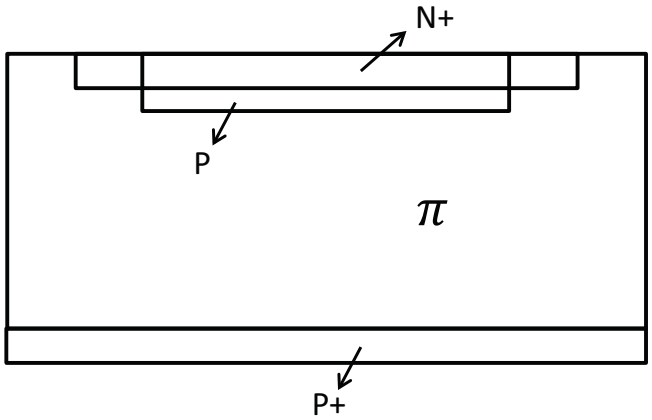


Fig. 2.7: Reach-through avalanche diode structure.

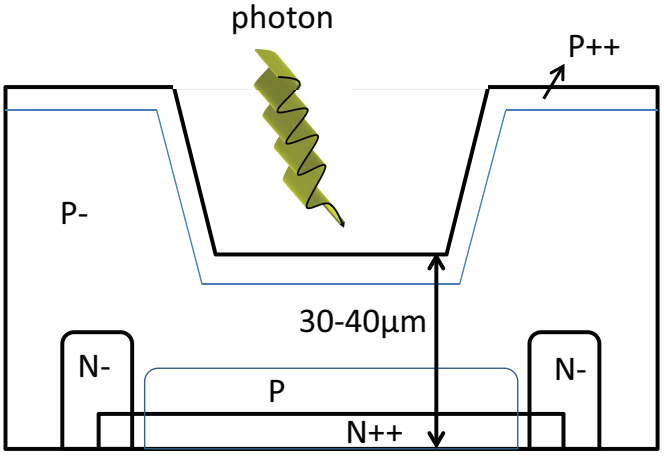


Fig. 2.8: Cross-section of a reach-through SPAD.

The cross-section of another typical reach-through SPAD is shown in Fig. 2.8 [63]. The active area is defined by an n++ phosphorous diffusion together with a p diffusion to increasing the p− quasi-intrinsic substrate concentration. Low-doped n− phosphorous guard rings are fabricated at the junction edges. The wafer is then flipped and etched back down to 30–40 µm from the active area. A p++ boron diffusion is then created to provide a low resistance path for the avalanche current and a good ohmic contact with the anode metal deposition. The backside contact is the cathode. This structure is designed to deplete all the 30–40 µm thickness, thus providing both a thick depletion region and a high electric field all over the space-charge zone.

2.4.2 Planar SPAD

Planar SPADs are usually fabricated with junctions near the surface of the semiconductor layer. A conventional pn junction is implemented in a planar process through implantation and annealing. Similar with thick reach-through SPADs, the structures are carefully designed to avoid PEB, which is due in this case to the fact that the electric field at the junction is maximized at the corners of the junction when applying a high voltage in reverse bias.

Several PEB prevention mechanisms with different structures are shown and compared in [48]. In figure 2.9(a), the guard ring is formed by the lightly doped p− implant, reducing the electric field at the edge of the p+ implant [64]. In figure 2.9(b), the PEB prevention mechanism is implemented by a floating p implant together with a polysilicon gate, which can further extend the depletion region as implicit guard ring [65].

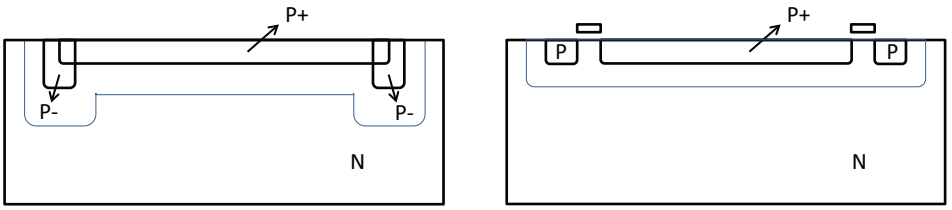


Fig. 2.9: Premature breakdown mechanisms by (a) extra light doped p^- guard ring; (b) floating p implant locally with polysilicon gate.

As shown in Fig. 2.10(a), it is possible to decrease the electric field using the geometry of shallow trench isolation in CMOS process. However, this suffers from high noise owing to trapping centers induced by trench fabrication [66]; to overcome this disadvantage, techniques are developed to prevent traps accumulated during the trench fabrication from triggering avalanches. As shown in Fig. 2.10(b), an effective technique was proposed in [67], by using several layers of doped semiconductor material with decreasing doping levels from the trench to the multiplication region. By shrinking mean free paths of carriers close to the trench, it is achieved to suppress noise by forcing carriers generated there to recombine before reaching the multiplication region.

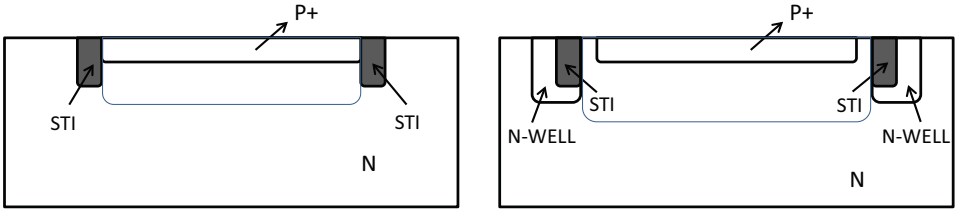


Fig. 2.10: Premature edge breakdown prevention mechanisms by (a) STI guard ring, (b) STI guard ring encapsulated in doping layer.

In figure 2.11, a virtual guard ring mechanism is proposed. By implanting an n- layer, the electric field in the middle of the diode, which is also the active region, would be maximized, thus preventing PEB at corners [68]. This structure could also be implemented efficiently in thin planar body layer, where body thickness is a design parameter. The details would be explained in section 2.5.

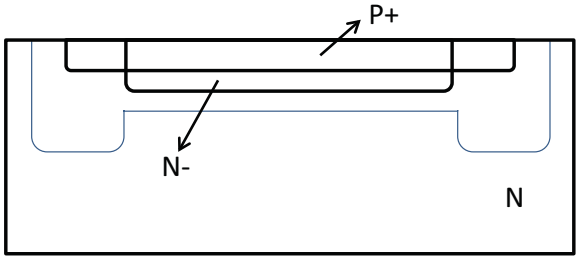


Fig. 2.11: Premature edge breakdown prevention mechanisms achieved by extra doping in active region (virtual guard ring).

2.5 SOI SPAD fabrication

Different from traditional CMOS-compatible SPADs, which are generally implemented in bulk silicon wafer, SPADs using silicon on insulator (SOI) technology are proposed in this work, providing a promising solution to realize BSI and flexible image sensor further involving substrate transfer technology.

The proposed SOI SPAD cross-section and model are shown in Fig.2.12 (a) and (b), respectively. This SPAD structure is similar to that of Fig. 2.11, where a virtual guard ring is implemented. Nonetheless, the introduction of epitaxy on SOI occupy the whole SPAD device. Since the thickness of the epitaxy layer is a design parameter, the body doping needs to be carefully optimized to avoid PEB while the silicon body under virtual guard ring is fully depleted. Enrichment doping also needs to be optimized, otherwise either too much band-

to-band tunneling noise or PEB causing by too high breakdown voltage will occur; this will be investigated in Chapter 3.

Fabrication begins with a p-type SOI wafer prepared by epitaxy technology. To ensure good mechanical compatibility and the application of dual side illumination, the thickness of the top silicon layer was optimized to be 1.5µm. To the best of our knowledge, the SPAD proposed in this work has the thinnest body to date, if compared with conventional SPAD structures, including reach-through SPADs and planar SPADs [62][67][69][70].

The N+P junction, with depth of approximately 100nm, is formed by implantation. A P+ enhancement region is made to form the multiplication region and an all-around virtual guard ring is defined implicitly by the existing doping difference. Guard rings are used to isolate active regions, so as to prevent electrical crosstalk, and to prevent premature edge breakdown, so as to minimize DCR [59].

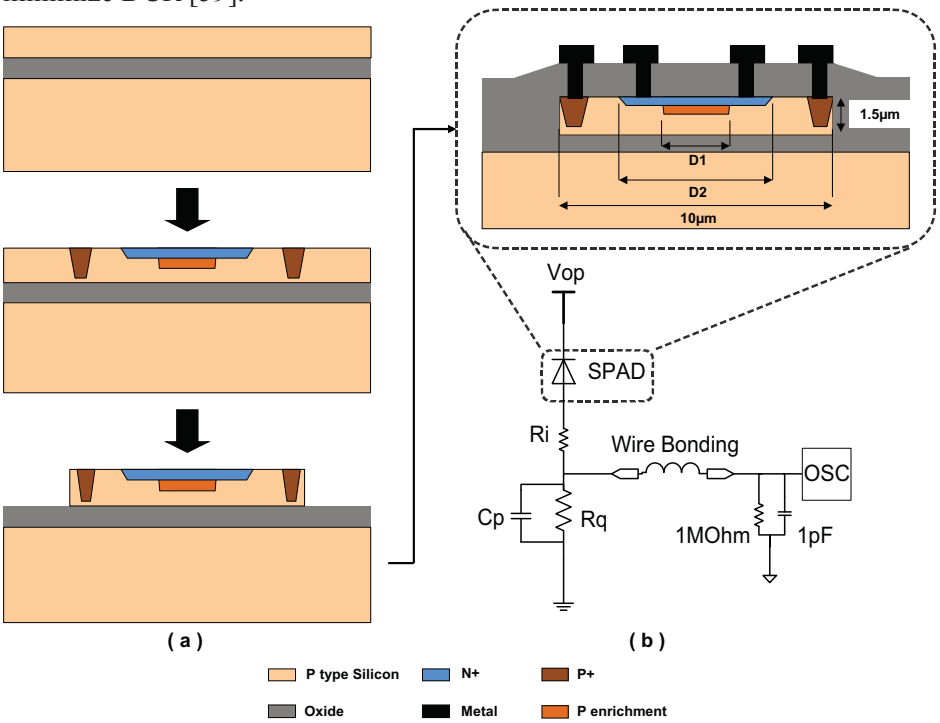
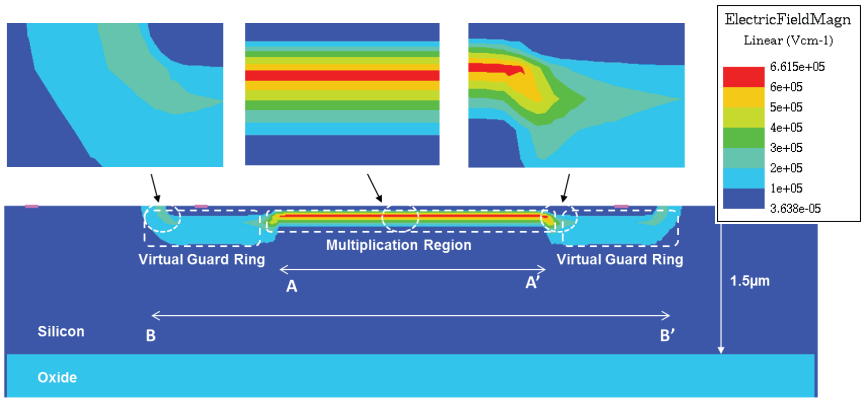


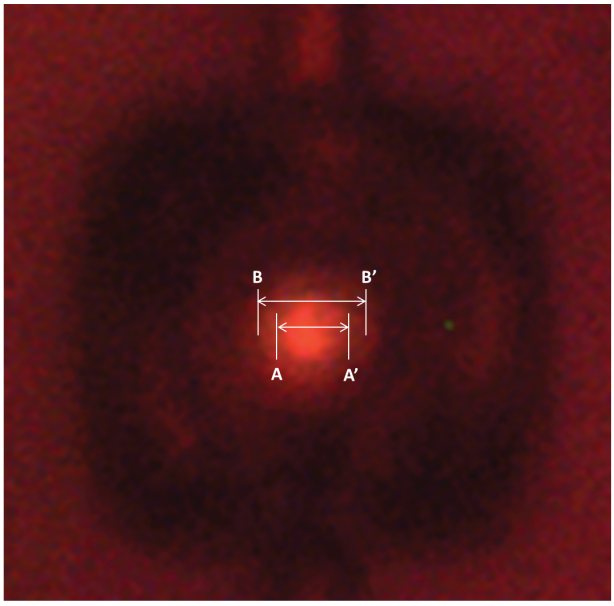
Fig.2.12: (a) Device Fabrication and cross-section of device structure: D1 is the diameter of the enhancement layer, D2 is the diameter of the implicit guard ring; (b) Schematic of passive quenching circuit and simplified SPAD model.

The Medici-simulated electric field contours are shown in Fig. 2.13. By optimizing the implantation and junction profiles, the electric field can be

constrained in the multiplication region and PEB can be suppressed effectively. It is also proved by the ionization light emission image in Fig. 2.13(b).



(a)



(b)

Fig. 2.13: (a) 10μm-diameter SPAD device overlaid on a Medici-simulated electric field plot. In silicon, impact ionization occurs with electric fields higher than 2.5×10^5V/cm. The multiplication region is in correspondence with the enhancement layer, where the absolute value of breakdown voltage, $|V_{BD}|$, is lower but uniform. In the virtual guard ring areas, the electric field is reduced, thus preventing premature edge breakdown. (b) Light emission by impact ionization test indicating the high electric field across the central multiplication region.

After junction implantation, device islands are formed by dry etching. A highly doped surrounding ring along the periphery of the island is made to separate the device from defect and trap centers at the island edges. 1.5μm TEOS PECVD oxide is deposited as insulator by two-time etch-back to form a spacer at the silicon island step. Contact holes are opened by dry etching. Then, a 3μm physical-vapor-deposited (PVD) Al/Si(1%) is used to form anode and cathode contacts. The typical device diameter is 10μm, with different pitch available (see D1 and D2 in Fig. 2.12).

A passive quenching circuit, which is simple but effective, was used in this work as shown in Fig. 2.12 (b). The anode of the SPAD with internal series resistance R_i is connected to a quenching resistance R_q in parallel with parasitic capacitance C_p. Due to the presence of R_q between the anode of the diode and ground, the avalanche current on photon arrival produces a voltage rise across the quenching resistor and brings the voltage across the diode below breakdown voltage, thus quenching the avalanche. A R_q of 50kOhm is chosen by considering the trade-off between the excess bias and avalanche recharge time. R_i was estimated to be several kOhms and it is due to a fully depleted area under the virtual guard ring in the 1.5 μm-thin silicon body layer. The total capacitance during avalanche recharge, including diode parasitic capacitance and external capacitance from package, was estimated to be tens of picofarad, which could be further reduced by integrating the quenching resistor. The output signal in Geiger mode is probed by an oscilloscope. A SEM micrograph of a single device is shown in Fig. 2.14.

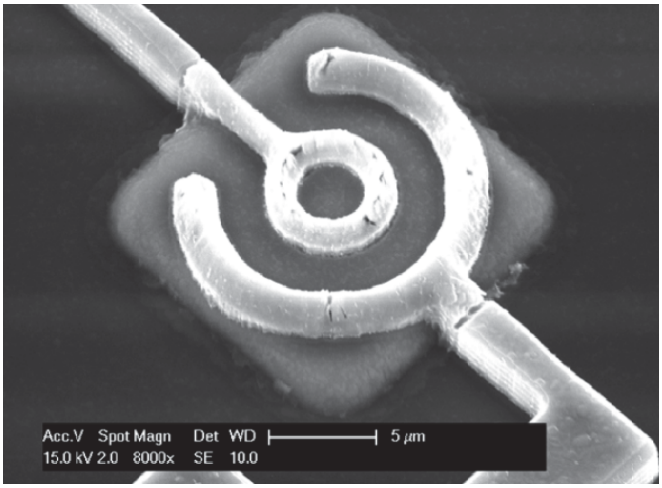


Fig.2.14: SEM image of an individual SOI SPAD.

2.6 SOI SPAD characteristics

The SPAD was implemented in several versions varying parameters such as guard ring dimensions, doping profiles, etc. The current-voltage (I-V) characteristics of the SOI SPAD are shown in Fig. 2.15 for different light conditions. The breakdown voltage is 12.5V.

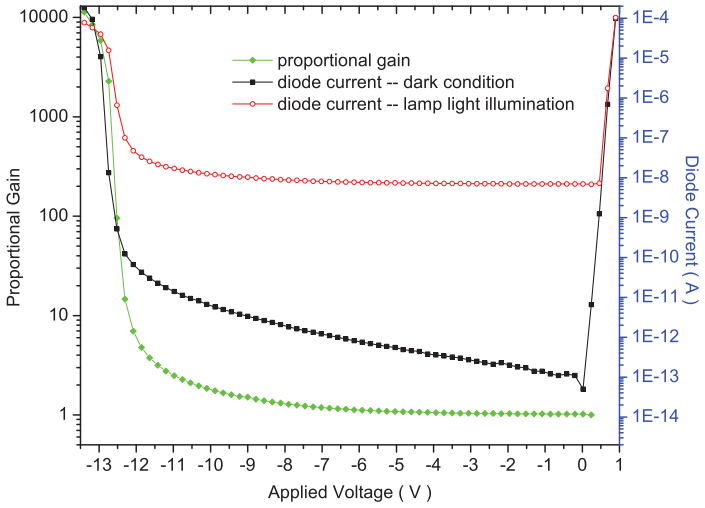


Fig. 2.15: I-V characteristics of SOI SPAD before transfer.

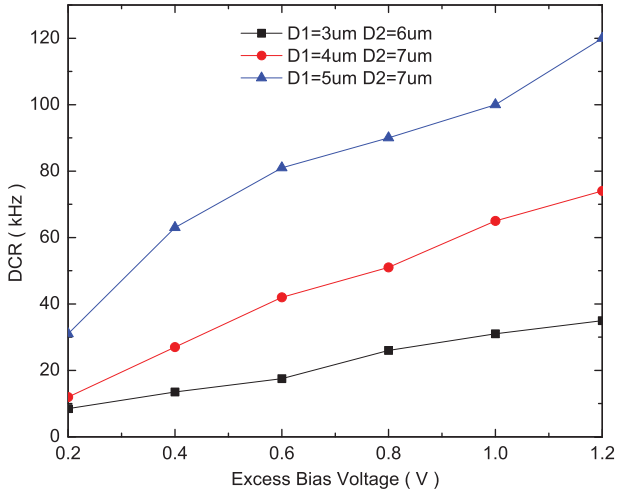


Fig. 2.16: Dark count rate (DCR) as a function of excess bias for different device pitch values.

By optimizing the dimensions of multiplication region and virtual guard ring, a DCR less than 20kHz was achieved, as shown in Fig. 5 that plots DCR as a function of excess bias voltage. The characterization of DCR as a function of temperature is shown in Fig. 6; the DCR is dominated by band-to-band tunneling instead of trap-assisted avalanching, proving the high quality of the top silicon layer.

In order to characterize sensitivity, photocurrent I_d of the SOI SPAD is measured at different wavelengths and compared with photocurrent I_{ref} that is measured on a reference photodiode with known quantum efficiency. The normalized ratio I_d/I_{ref}, shown in Fig. 2.17, has some resonant-like peaks above 700nm, due to a cavity-like structure in the multiplication region with an interface between silicon and buried oxide similar to [71][72]; a similar behavior was also seen in PDP (Fig. 2.17), where PDP at long wavelengths (700nm ~ 900nm) is enhanced.

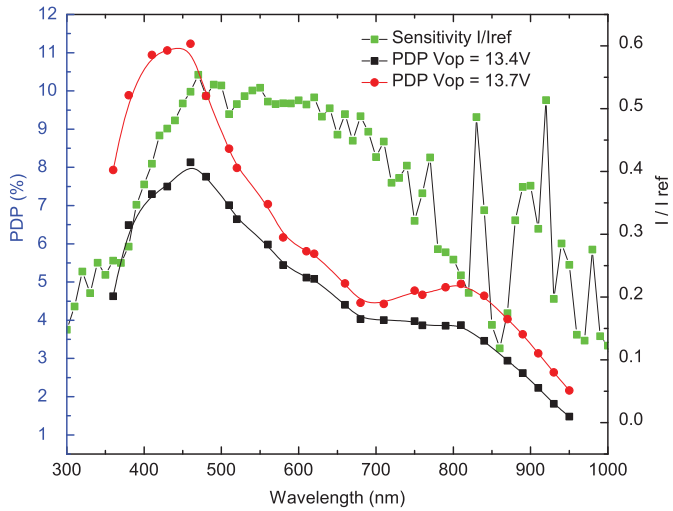


Fig. 2.17: Photon detection probability of the SOI APD when operating in Geiger mode; sensitivity of normalized I_d/I_{ref} when operating in proportional mode as a function of wavelength (D1=5µm, D2=7µm). I_{ref} is from reference diode with quantum efficiency accounted for all wavelengths.

Temporal performance is shown in Fig. 2.18 for a 10µm-diameter SPAD (D1=5µm, D2=7µm, V_{eb}=1V) with a full-width-at-half-maximum (FWHM) jitter of 500ps.

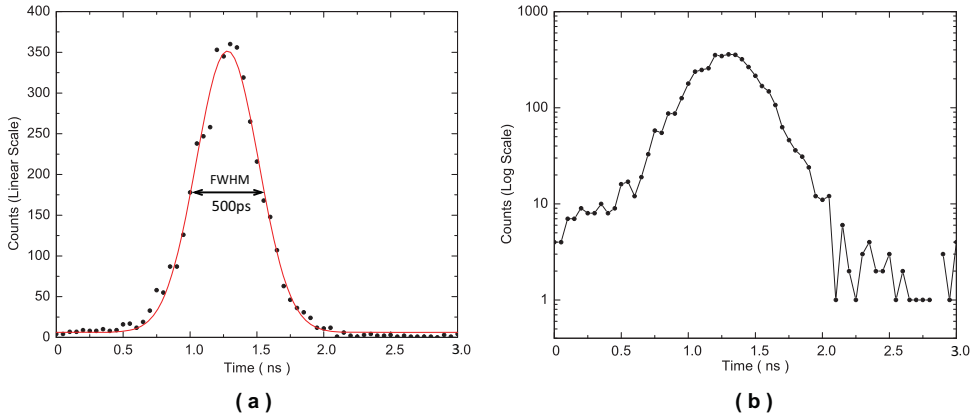


Fig. 2.18: Jitter performance of the APD when operating in Geiger mode (a) in linear scale (b) in logarithmic scale.

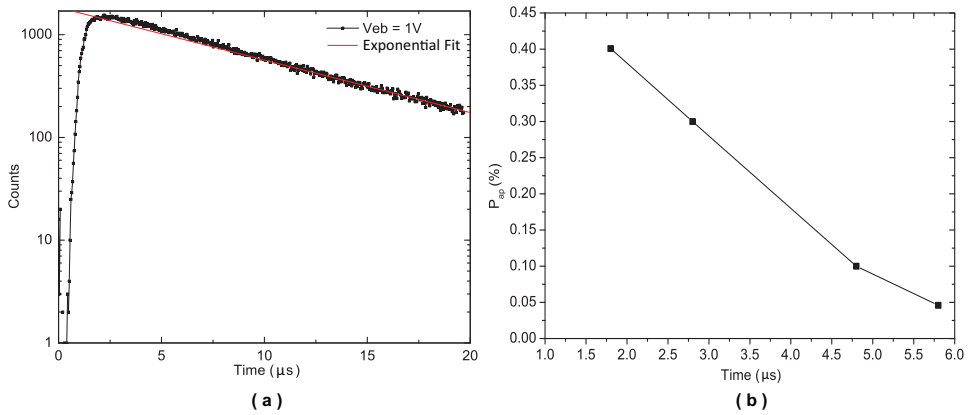


Fig. 2.19: (a) Exponential fit of inter-arrival response histogram at V_{eb} = 1V of the SOI SPAD. (b) Afterpulsing probability plot extracted from the histogram.

Afterpulsing is a type of correlated noise caused by the fact that, during an avalanche, traps could capture and hold carriers within a certain lifetime. A histogram of the inter-avalanche arrival times for the SOI SPAD and flexible SPAD respectively is shown in Fig. 2.19, along with the extracted afterpulsing probability as a function of dead time, as proposed in [73].

With a dead time longer than 1µs, as shown in the figure, the afterpulsing probability of the SOI SPAD is less than 1%. Because of the full depletion in the ultra-thin body under virtual guard rings, the internal series resistance is around several kilo Ohms. In this case, a relatively long dead time would be

generated in the passive quenching scheme. Additionally, the low amount of trap states in the silicon layer is also a positive reason for negligible correlated noise.

2.7 Substrate transfer process

The layer transfer process is summarized in Fig. 2.20. After SOI SPAD devices are processed, PECVD oxide is deposited at the backside of the SOI wafer and patterned as mask as shown in Fig. 2.20 (a) for the following deep reactive ion etching (DRIE). The polyimide is coated and cured on the top of the device at 400°C. Then, the polyimide is patterned to expose the metal contact and light absorption area on the frontside of SPADs in Fig. 2.20 (b). The silicon substrate under the buried oxide layer is etched away by DRIE. It is however well known that the etch rate of silicon strongly depends on both the area of silicon exposed and the aspect ratio [74][75]. In order to minimize the backside silicon substrate etching variation, the process is designed to include end-point detection and optimized over-etching time. The over-etching stops at 1µm-thick buried oxide layer as shown in Fig. 2.20 (c). Generally, it is quite challenging to handle a polyimide-based substrate with device layer. In this work, it is achieved by leaving some small polyimide anchors to the bulk substrate frame. The anchors are designed strong enough to keep the chips in place during DRIE processing but fragile sufficiently to allow breaking off. Hence, the SPAD device layer on polyimide can easily be released by means of reasonable mechanical stress as shown in Fig. 2.20 (d). As a result, the SPAD's layer has been successfully transferred to flexible polyimide layer and was further mounted to PEN (Polyethylene naphthalate) or other flexible substrates for package.

As shown in Fig. 2.20 (d), SPADs can operate in dual side illumination (DSI) since the light absorption area on both the frontside and backside are exposed.

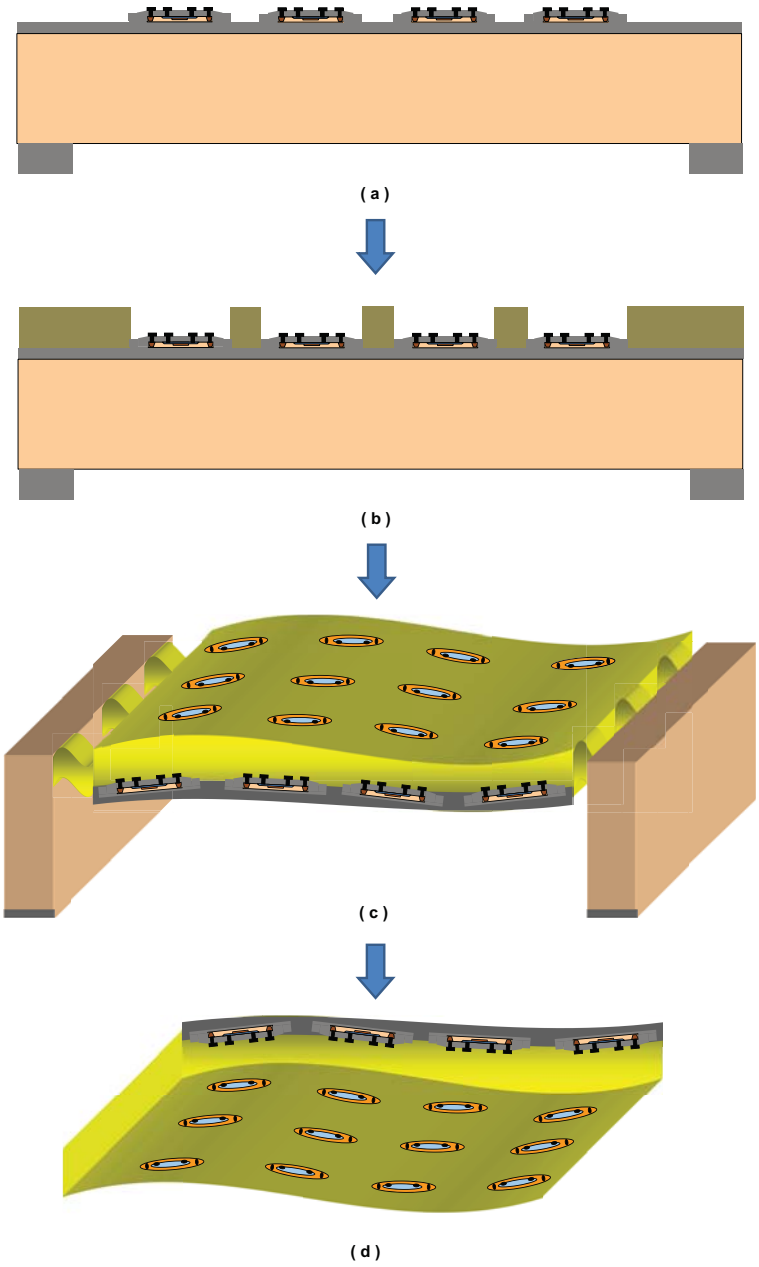


Fig. 2.20: SPAD layer transfer process. (a) Backside mask patterning; (b) Polyimide coating and patterning; (c) Substrate etching by DRIE; (d) Flexible device layer releasing.

The maximum temperature of the transfer process is 400°C (polyimide curing process in oven with temperature monitor), thus there is little thermal impact on the fabricated devices during transfer process.

The whole processing work is done in the class-100 cleanroom and MEMS lab at Else Kooi Lab. We have fabricated several batches of 4-inch wafers. The process is quite reliable and repeatable with a yield higher than 95%.

The persistent image of DCR quenching pulses of flexible SPAD on polyimide foil, which is shown in Fig. 2.21 (a) and (b), is given in (c), indicating flexible SPAD working in Geiger mode.

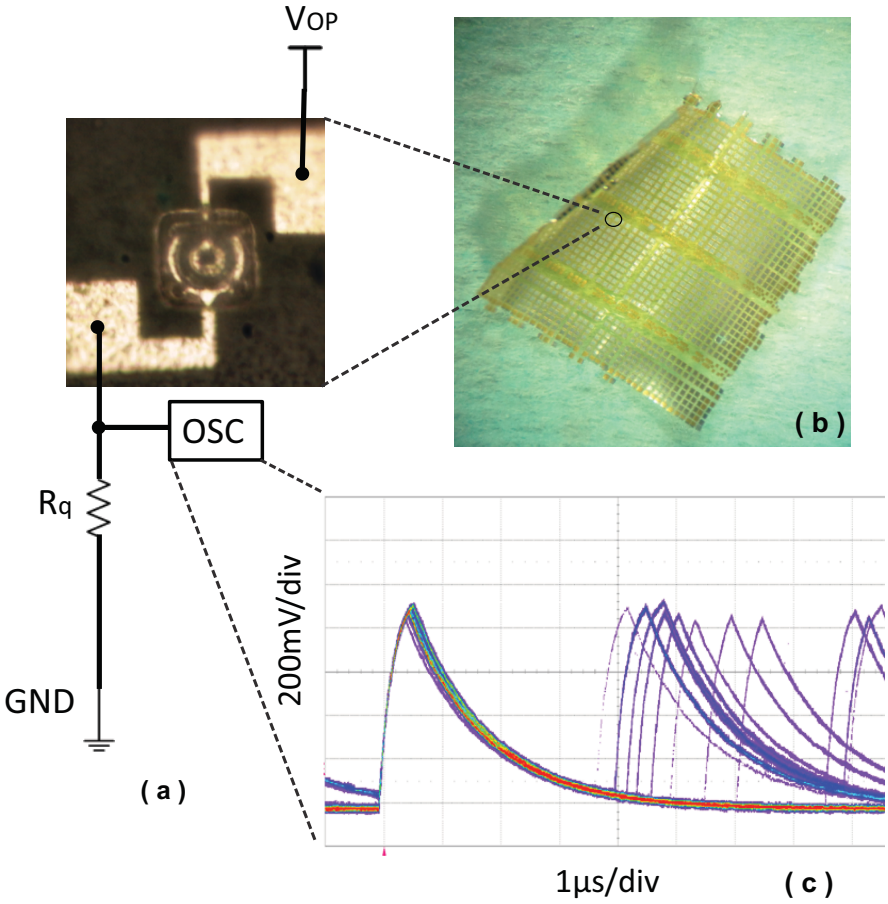


Fig. 2.21: (a) SPAD (microscopic image) configuration of passive quenching with external resistor; (b) Flexible polyimide foil containing SPADs released from the SOI wafer; (c) Persistent image of DCR quenching pulses of flexible SPAD.

2.8 Device performance comparison

The current-voltage (I-V) characteristics of the SPAD before (SOI) and after transfer (flexible) are shown in Fig.2.22 for different light conditions. The current gain varies exponentially on the bias voltage in proportional mode and a gain of as much as 10000 has been achieved below breakdown. The breakdown voltage before transfer is 12.7V. After transfer, it typically increases to 13.4V.

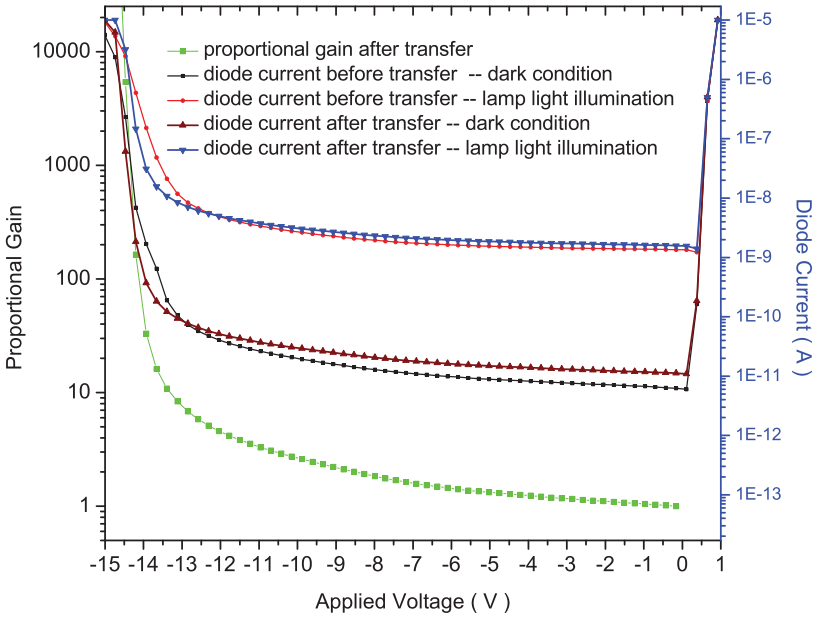


Fig. 2.22: I-V characteristics of SPAD before and after transfer.

By optimizing the dimensions of the multiplication region and of the virtual guard ring, a DCR of less than 20kHz was achieved, as shown in Fig. 2.23. The DCR has a similar value with that before transfer. However, with a DCR value of tens of kilohertz, it is difficult to evaluate how the transfer process affects DCR of the device. The figure plots DCR as a function of excess bias voltage. The characterization of DCR as a function of temperature is shown in Fig. 2.24. In principle, both band-to-band tunneling and trap-assisted avalanching, which is the combination of trap-assisted tunneling and trap-assisted thermal generation (SRH generation), contribute to DCR [76][77]. By analyzing the temperature sweep, one recognizes that DCR varies by less than 20% per decade. So it is concluded that DCR is dominated by band-to-band tunneling. Furthermore, superficial carrier generation from the trap states of box oxide

surface, especially in the fully depleted area below the junction, has an impact on DCR.

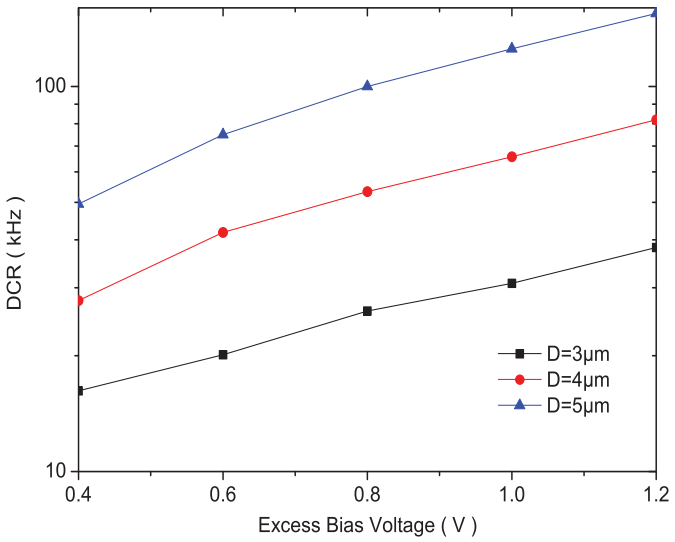


Fig. 2.23: Dark count rate (DCR) of flexible SPADs as a function of excess bias for different device pitch values.

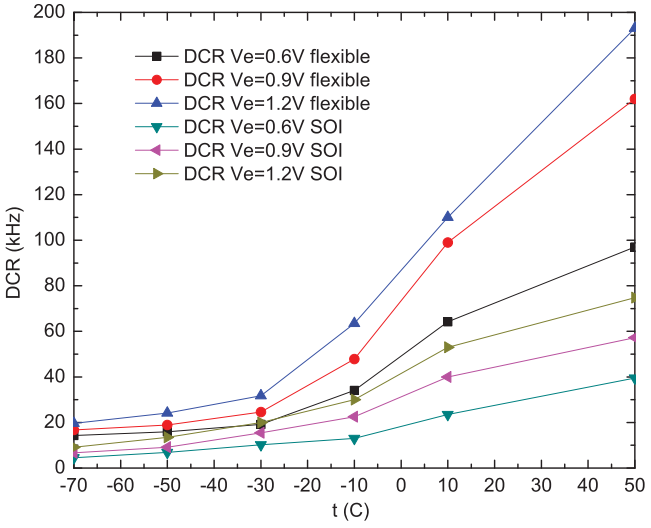


Fig. 2.24: DCR versus temperature in flexible SPADs at different excess bias voltages (D1=3µm, D2=6µm).

By comparison to SOI and flexible devices in Fig. 2.21, the breakdown voltage increases less than 6% after transfer. We believe that self-heating is the cause of the increase when the device is encapsulated by polyimide. It is also confirmed by the plot of DCR as a function of temperature in Fig. 2.24 because the maximum DCR of SPADs on polyimide is higher than that of SOI SPADs, especially at high temperatures, as self-heating becomes less at room temperature.

By comparing it with the afterpulsing distribution of a flexible SPAD, shown in Fig. 2.25, it is concluded that there is no intrinsic change of afterpulsing from a SOI SPAD to a flexible SPAD.

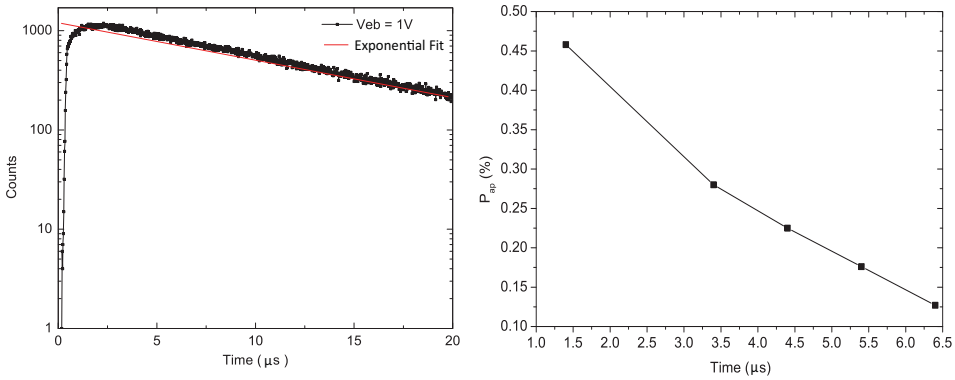


Fig. 2.25: (a) Afterpulsing distribution at V_{eb} = 1V of flexible SPAD; (b) Afterpulsing probability plot extracted from distribution.

Fig. 2.26 shows photographs of SPADs after being transferred onto a flexible polyimide layer and images of the front and back of flexible SPADs bonded on PCB.

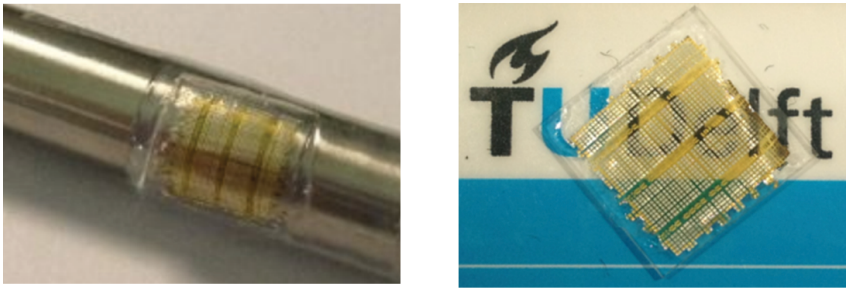


Fig. 2.26: (a) Flexible SPADs on PEN substrate bent onto 10mm-diameter cylinder; (b) Flexible SPADs mounted onto quartz.

2.9 Conclusion

To the best of our knowledge, the reported device is the first flexible SPAD and also the first ultra-thin-body SPAD with dual side illumination at the time of its publication [46][47][78]. DCR, PDP, afterpulsing, and time jitter performance of the device on flexible substrate is consistent with that of a device before transfer. The performance of the proposed device compares favorably with that of CMOS SPADs, while it can operate in dual-side illumination. The peak PDP can reach 11% in FSI mode and 6% in BSI mode and the minimum DCR was less than 20kHz with negligible afterpulsing probability. In BSI, the sensitive spectrum may be wider and PDP, especially at long wavelengths, could be enhanced. Furthermore, fill factor could be improved significantly, due to a wider carrier collection region and the absence of blocking metal layers in BSI mode. It could be improved further by designing larger active-area SPADs and by optimizing the doping profiles, so as to widen the ratio of active to guard ring areas. This technology is CMOS compatible, enabling CMOS circuit monolithic integration and large-scale SPAD array with readout circuit. A further improvement can be introduced with a buried layer at the backside interface or some other material to solve the superficial carrier generation and charge collection issue [79]. Such a dual side illumination SPAD provides a novel methodology to overcome the limits of BSI applications in CMOS technology, simultaneously enhancing the fill factor while pixel pitch keeps scaling down. Flexible SPADs can be used in anti-vignetting image sensors, bio-inspired composite-eye or flexible multi-aperture cameras, and wherever the sensor plane must follow a certain curvature in advanced implantable biological and future life sciences research.

Chapter 3

Comprehensive study on flexible ultrathin-body single-photon avalanche diode

Based on the flexible ultrathin-body SPAD proposed in Chapter 2, the optical characterizations including PDP and timing jitter in both frontside and backside illuminations are performed and compared. Furthermore, SPADs integrated with quenching resistors on flexible substrate were fabricated. The Geiger-mode performances of this flexible SPAD configuration are investigated comprehensively: dark count rate (DCR), V_{BD} and photon detection probability (PDP) are studied based on different junction parameters, operation temperature and device structures.

This chapter is based on results published in [47][80][81]

3.1 Dual-side illumination characterizations

For the flexible SPAD dual-side illumination characterizations, the chip with SPAD devices on polyimide foil was packaged onto PCB through wire bonding as shown in Fig. 3.1 and then measured in both sides. The cathode was wire bonded to the pad of operation voltage V_{OP} and anode was wire bonded to output pad with an external quenching resistor connected between anode and GND. 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.

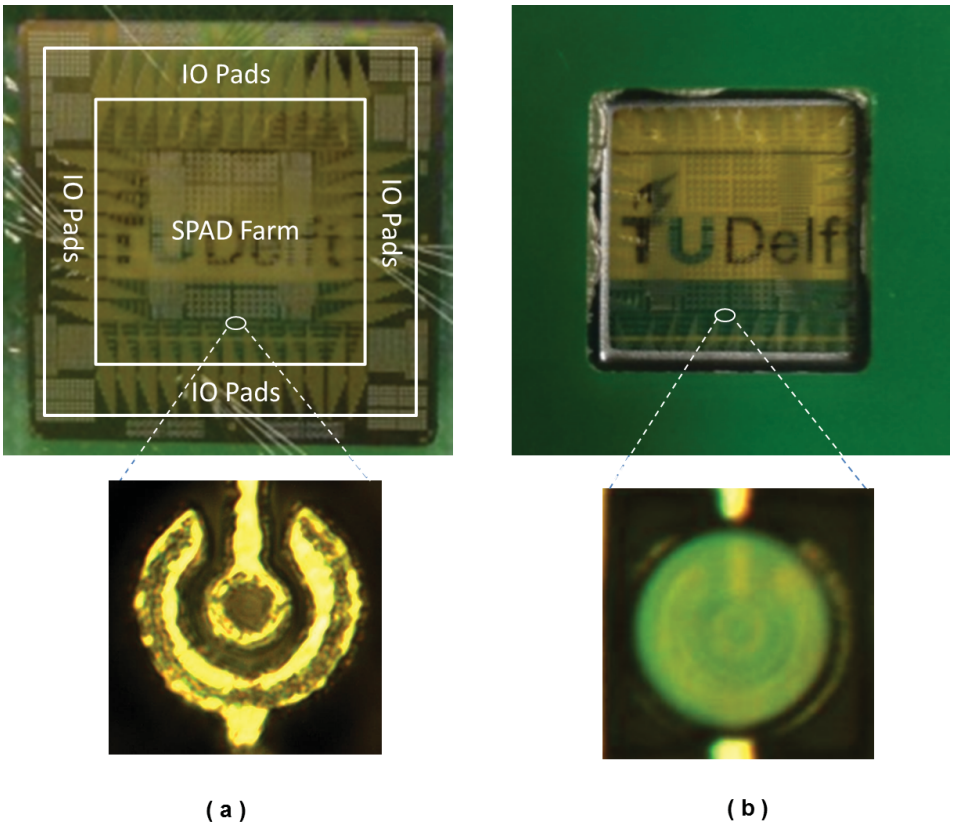


Fig. 3.1: (a) Front of flexible SPADs bonded on PCB; (b) Back of flexible SPADs bonded on PCB.

3.1.1 PDP comparison of FSI and BSI

A SPAD's sensitivity is usually described by photon detection probability (PDP), which quantifies the probabilities that a photon impinging on the detection triggers an avalanche pulse. It is defined as:

$$PDP = QE \times P_{avalanche}. \quad (3.1)$$

QE is the quantum efficiency and $P_{avalanche}$ the avalanche probability. The PDP of particular wavelength depends on the photon penetration depth and the chance that photons generate carriers at particular depth multiplied by avalanche probability. PDP in FSI and BSI has different conditions due to the fact that a shallow junction in FSI becomes a deep junction in BSI.

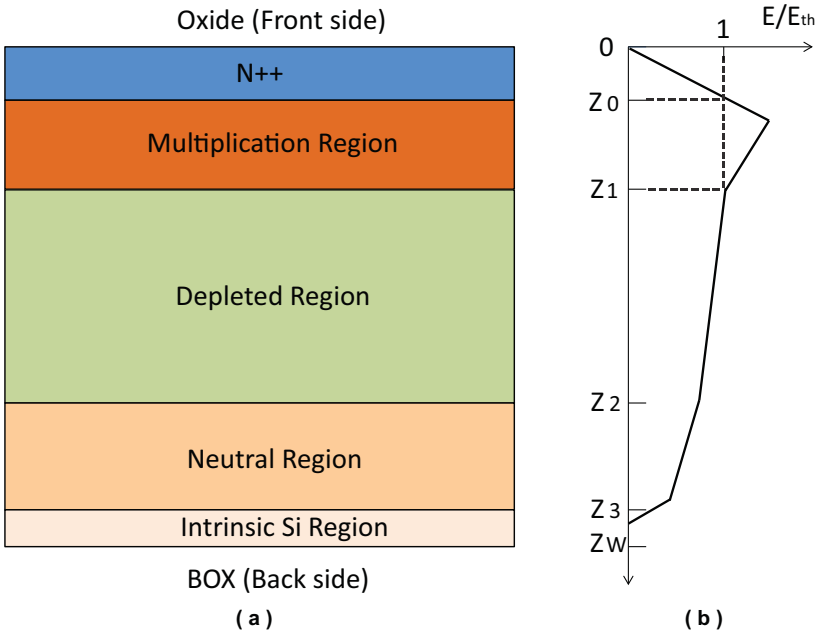


Fig. 3.2: (a) junction cross-section of active region in ultrathin-body; (b) Electric field across the body in arbitrary scale ($E_{th} = 2.5\times10^5$ V/cm).

The junction cross-section is shown in Fig. 3.2 (a). 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 $p(z)$.

$$PDP = \int_{-\infty}^{\infty} T(\lambda) f_{\lambda}(z) p(z) dz, \quad (3.2)$$

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

$$f_\lambda(z) = \mu(\lambda)\exp(-\mu(\lambda)z), \qquad (3.3)$$

with $\mu(\lambda)$ being the mean penetration depth of the light into the silicon [82].

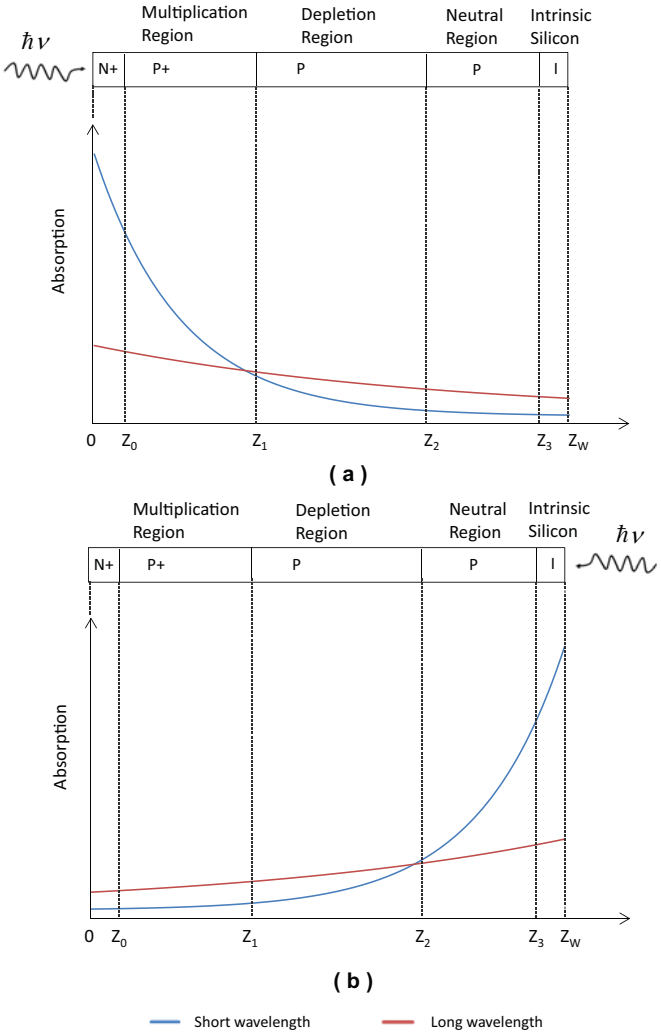


Fig. 3.3: (a) Absorption of photons overlaid with junction structure in FSI;
(b) Absorption of photons overlaid with junction structure in BSI (arbitrary scale).

Equation (3.2) and (3.3) could be applied according to FSI and BSI respectively. In FSI, the PDP can be modeled as:

$$PDP = \int_0^{Z_w} T(\lambda) f_\lambda(z) p(z) dz = \int_0^{Z_w} T(\lambda)\mu(\lambda)\exp(-\mu(\lambda)z) p(z) dz, \quad (3.4)$$

while in BSI, the PDP can be modeled as:

$$PDP = \int_0^{Z_w} T(\lambda) f_\lambda(z_w - z) p(z) dz = \int_0^{Z_w} T(\lambda)\mu(\lambda)\exp(-\mu(\lambda)(z_w - z)) p(z) dz, \quad (3.5)$$

where the parameters of the equation are seen in Fig. 3.3. As shown in the figure, the absorption is described by $f_\lambda(z) = \mu(\lambda)\exp(-\mu(\lambda)z)$ for incident light overlaid with junction cross-section in both FSI and BSI. By comparing photon absorptions with different wavelengths, the absorption probability $f_\lambda(z)$ across the junction for photons with short wavelength drops dramatically, but less impact on the absorption probability of photons with higher penetration depths, such as red and infrared photons.

Since the avalanche can only occur in the multiplication region, the PDP in FSI, especially for photons of short wavelength, should be higher than in BSI due to the fact the junction stays shallow in frontside. Also, photons of longer wavelengths with higher penetration depth generate carriers across diffusion, depletion region and multiplication regions with relatively high absorption probability in BSI. However, the PDP difference for long-wavelength photons between FSI and BSI should be smaller than short-wavelength photons principally. Furthermore, the intrinsic silicon region, which is used as seed layer for the epitaxy process should suppress the electric field, as shown in Fig. 3.2, resulting in a negative impact for PDP in BSI. It will be explained in Chapter 4 that dopant diffusion into this region by a high thermal budget would enhance the avalanche probability, thus improving the PDP in BSI.

On the other hand, carriers by lateral collection need to be considered for BSI. As shown in Fig. 3.4 (a), the photon absorption area in BSI is larger than that in FSI due to the photon-generated carrier lateral collection region around the multiplication area and the lack of metal layers blocking light, resulting that the fill factor in BSI could be improved significantly.

The photons of high penetration depth from BSI absorbed in the depletion region around the active area such as the fully depleted area below virtual guard ring also contribute to the carriers generated and ionized because of carrier's lateral drifting into multiplication region, which is shown in Fig. 3.4 (b) and (c).

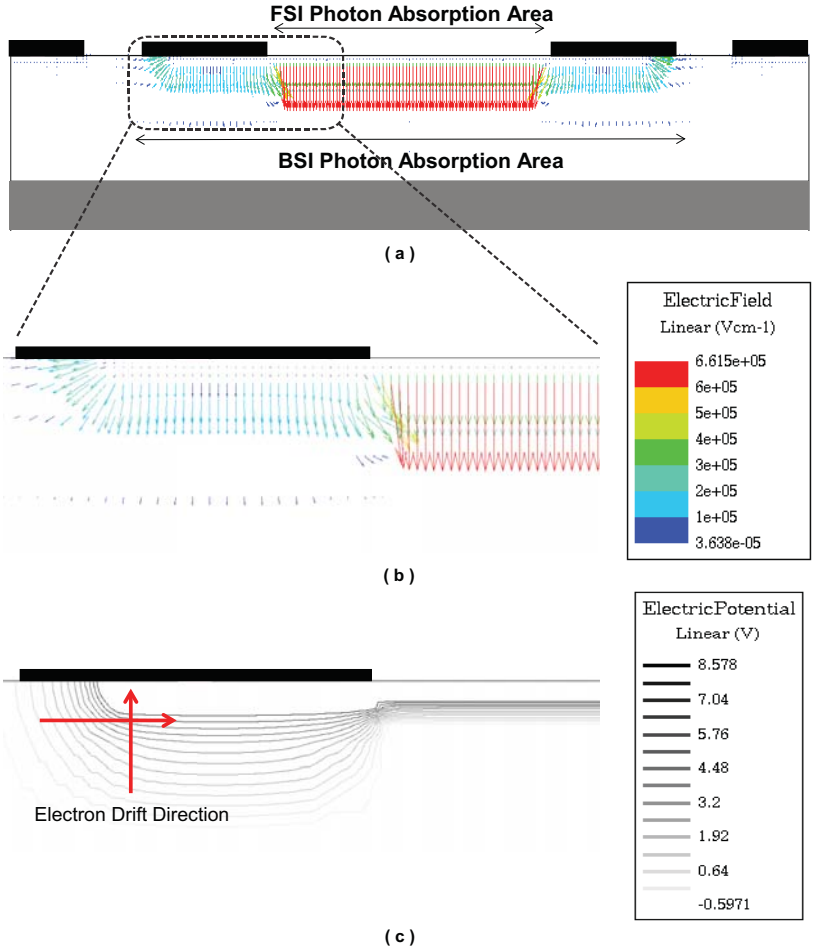


Fig. 3.4: (a) Comparison of photon absorption in FSI and BSI overlaid in electric field plot; (b) Zoom-in plot of electric field lines at the carrier lateral collection area; (c) Zoom-in plot of equipotential contours at the carrier lateral collection area.

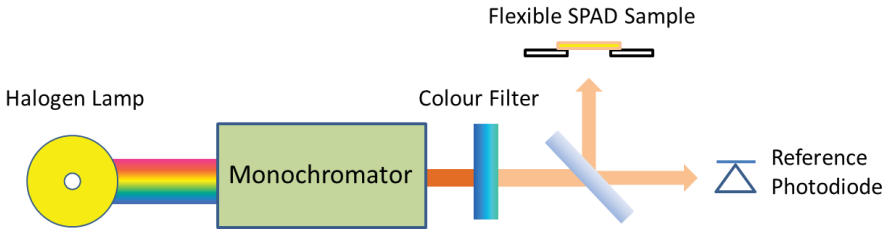


Fig. 3.5: PDP measurement setup.

Fig. 3.5 shows the PDP measurement setup. The light from a monochromator (Oriel/Newport part 77250) was projected into an integration sphere (Oriel/Newport part 819D-SL-2), shown in the figure as a beamsplitter, with a reference diode (Hamamatsu part S1226-BQ) at one port of the sphere and the SPAD at the other. The reflectance of the monochromator is around 99% for the wavelength from 400nm to 900nm. By the integration sphere, the amount of photons is same for reference diode and the SPAD.

Photon flux (intensity) is different at each wavelength and it was calculated by measuring the reference current of reference diode, whose quantum efficiency is already known. The formula can be described as below:

$$Flux(\lambda) = \frac{I(\lambda) - I_{dark}}{r(\lambda) \cdot A_{ref} \cdot (\hbar c / \lambda)}, \tag{3.6}$$

where $I(\lambda)$ is the reference photodiode's current, I_{dark} is the reference diode's dark current, $r(\lambda)$ is the reference diode's responsivity (already known from manual), A_{ref} is the reference diode' active area, $\hbar$ is Planck's constant and c is the speed of light.

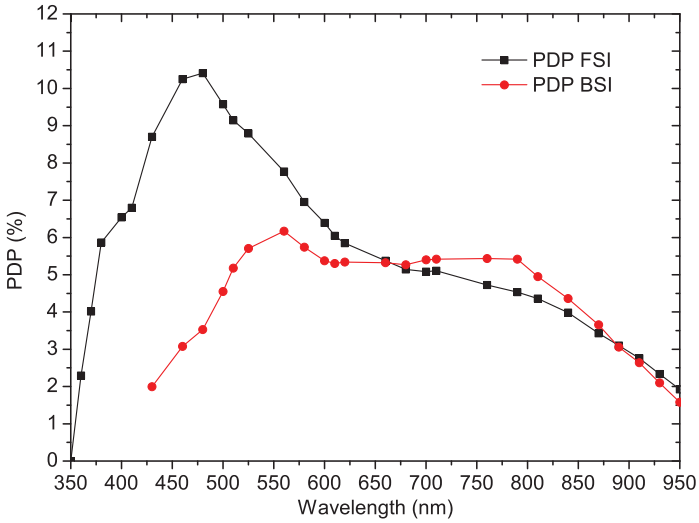


Fig. 3.6: Photon detection probability of both flexible SPADs in FSI and BSI mode. (V_{eb}=1.0V).

The PDP measurements of the flexible SPAD in FSI and BSI are shown in Fig. 3.6. In FSI mode, the flexible SPAD has similar PDP as the SOI SPAD. However, in BSI mode the flexible SPAD has lower PDP at waveform from 350nm to 600nm due to the lower absorption probability of that waveform range at relatively deeper depletion region and multiplication region.

At long wavelengths (600–900 nm), the PDP in BSI mode is similar of that in FSI mode and especially for the spectrum higher than 700nm, the PDP is enhanced in BSI mode, possibly because of the carrier's lateral drifting as discussed with Fig. 3.4 (b) and (c).

3.1.2 Timing jitter of FSI and BSI

In principle, the timing jitter of SPADs is due to the statistical nature of the avalanche build-up. There will be uncertainty in the waveform following carrier injection, with some of the jitter coming from static effects such as trigger position. If the carriers need to diffuse or drift to the depletion region, the uncertainty increases. In this section, the temporal performance of timing jitters of flexible SPAD in both FSI and BSI are analyzed comprehensively with two different waveform picosecond lasers. As shown in Fig. 3.7, the flexible SPAD was illuminated by a picosecond laser (Advanced Laser Diode Systems GmBH) emitting pulses with a repetition rate of 40MHz. A neutral density filter (NDF) was inserted between pulse laser source and the flexible SPAD perpendicularly to attenuate the light density. The purpose is to illuminate the SPAD in photon-starved regime. The light absorption module was used to absorb the reflected photon flux to avoid external interference. Because the laser light can partially go through polyimide and trenches between SPADs, light absorption module was also set up behind sample [83].

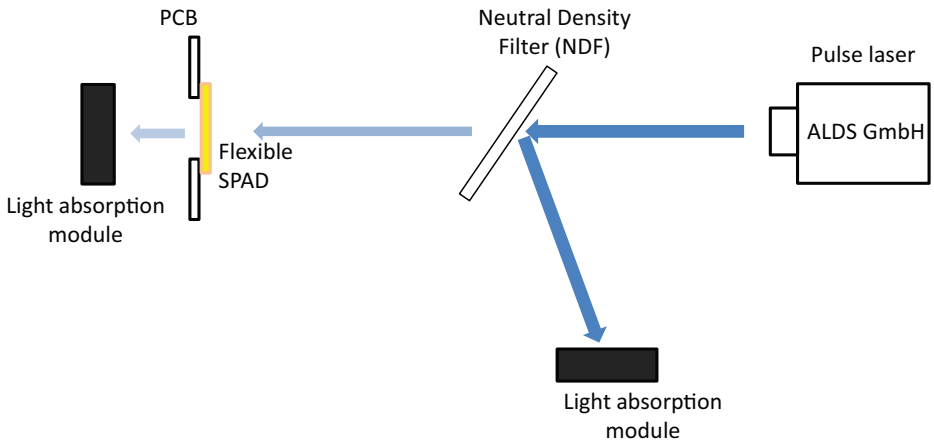


Fig. 3.7: Timing jitter measurement setup.

The flexible sample was assembled onto PCB with a hole opening to enable dual-side illumination. The laser with a wavelength of 405nm and a laser with a wavelength of 637nm were used in this experiment. A jitter of 380ps FWHM

was measured in a 4μm-diameter active area FSI SPAD exposed to the blue laser with a pulse width of 40ps FWHM, as shown in Fig. 3.8. The jitter of the response of the same FSI SPAD to the red laser was measured to be 900ps FWHM (Fig. 3.8).

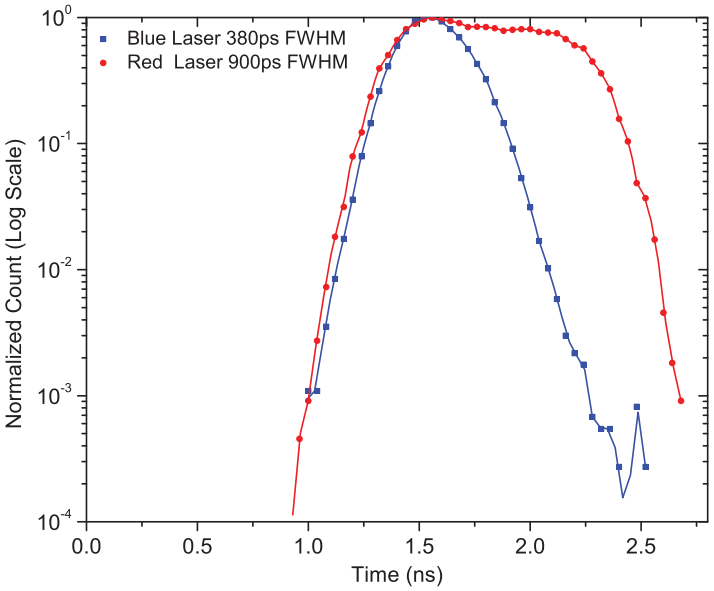


Fig. 3.8: Jitter performance of the FSI flexible APD when operating in Geiger mode.

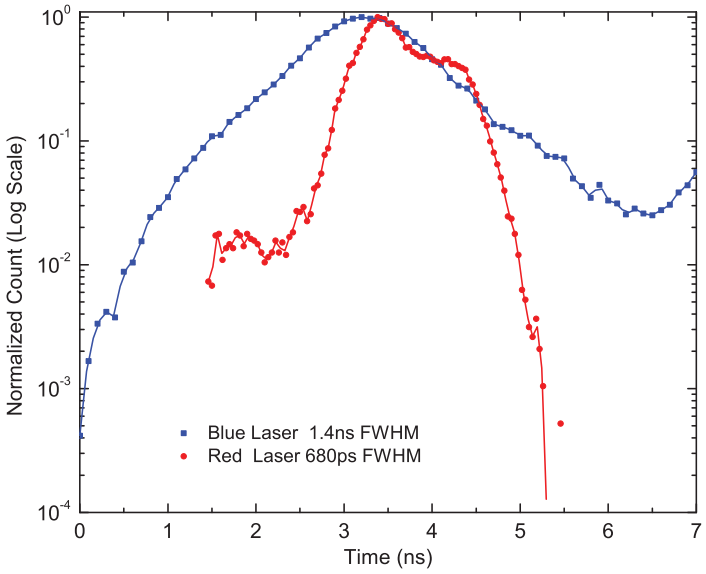


Fig. 3.9: Jitter performance of the BSI flexible APD when operating in Geiger mode.

Fig. 3.9 shows the response of the flexible SPAD operating in BSI mode. The response to the 637nm-wavelength laser (red) laser is characterized by a FWHM jitter of 680ps, while the response to the 405nm-wavelength laser (blue) by a FWHM jitter of 1.4ns.

This behavior is illustrated in Fig. 3.10 [58][57]. In FSI mode, see Fig. 3.10 (a), photons from a red laser have higher penetration depth and could be detected at larger depths. Thus, the photo carriers generated at larger depth are collected to the multiplication region from depletion and neutral regions. The charge collection time from depletion and neutral region, which is deeper, is distributed, thus causing the jitter to be a convolution of a Gaussian distribution with a charge collection time distribution between 1.5ns and 2.5ns. The jitter performance is consistent with measurements. Alternatively, blue photons are detected very near the shallow junction and need not transport before a multiplication is triggered, thus the Gaussian, blue-laser time distribution of the response is achieved and the jitter is improved.

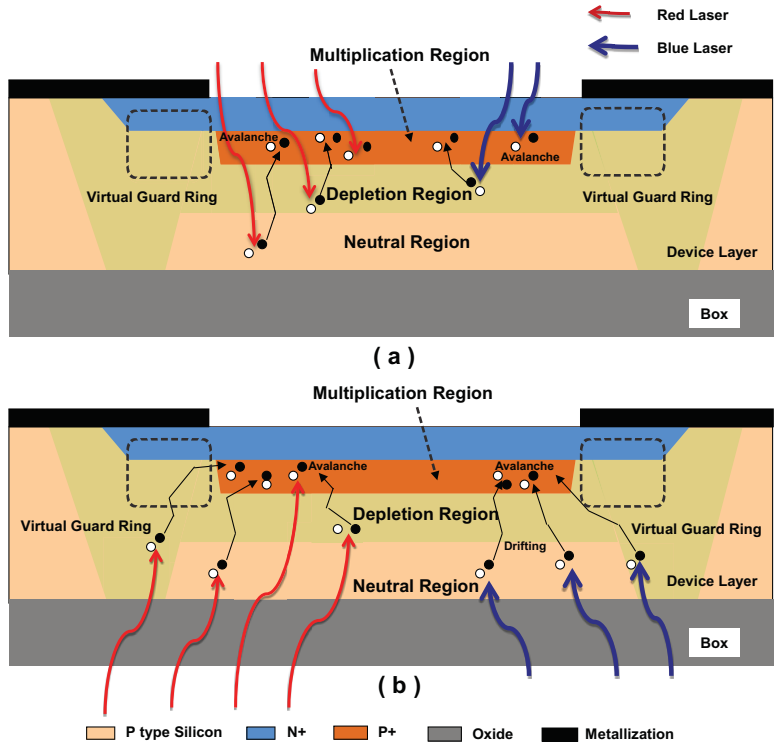


Fig. 3.10: (a) Avalanche triggered by red laser and blue laser from the frontside. (b) Avalanche triggered by red laser and blue laser from the backside.

Fig. 3.10 (b) shows the case of a SPAD operating in BSI mode. In this case, carriers generated by red photons result in a response with a suppressed convolution of the Gaussian distribution and a charge collection time distribution in the temporal response, while carriers generated by blue photons require higher average collection time and thus exhibit a worse response.

Overall, BSI SPADs have higher jitter; this behavior could be due not only to the combination of collection times from the depletion region and neutral region, but also to the lateral collection of carriers since the fill factor is quite large in BSI mode. Superficial charge generation also contributes to the variation of the collection time due to the fully depleted region at the box oxide surface.

3.1.3 Summary

The characterization of APDs operating in proportional and Geiger modes before and after substrate transfer, in FSI and BSI respectively, are summarized in Tab.3.1.

Table. 3.1: Summary of APD performance before and after transfer.

	SOI SPAD			**Flexible SPAD (FSI)**			**Flexible SPAD (BSI)**			
Performance	**Min.**	**Typ.**	**Max.**	**Min.**	**Typ.**	**Max.**	**Min.**	**Typ.**	**Max.**	**Unit**
Avalanche breakdown voltage		12.7			13.4			13.4		V
Excess bias	0		1.2	0		1.2	0		1.2	V
DCR	20		120	20		177	20		177	kHz
PDP			11.3			11.0			6.1	%
Timing jitter (405nm-wavelength laser)		500			380			1400		ps
Timing jitter (637nm-wavelength laser)					900			680		ps

3.2 Configuration of flexible SPAD integrated with quenching resistor

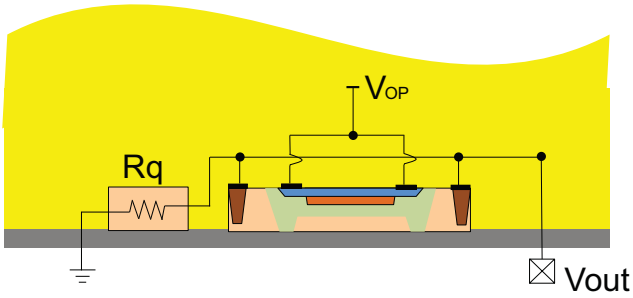
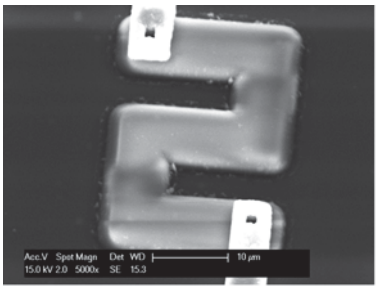
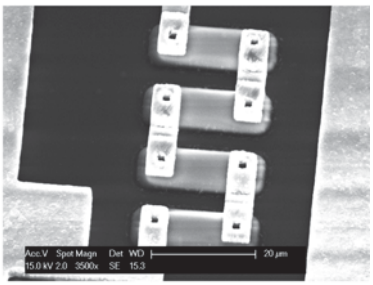


Fig. 3.11: Cross-section of flexible ultrathin-body trench-isolated structure SPAD integrated with quenching resistor.

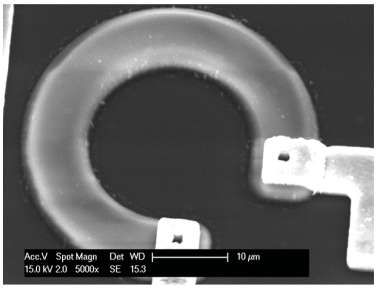
By making use of trench isolations, epitaxy silicon body can be patterned to different body islands acting as resistors. Fig. 3.11 shows a cross-section of the flexible trench-isolated SPAD integrated with the quenching resistor. The resistor pattern can be executed at the same lithography and etching procedure with SPAD trench isolation, so no additional fabrication needed.



(a)



(b)



(c)

	Resistance (kOhm)	Area (µm²)
a	157	20 x 30
b	100	25 x 45
c	195	30 x 30

Fig. 3.12: SEM images of trenched isolated silicon body island resistors (EPI Si: thickness: 1.5µm; Boron doped concentration: 1E16 cm^{-3}).

Different designs of silicon island resistors are shown in Fig. 3.12. By comparing different designs, pattern (c) was selected to be the optimized option by trading off between the resistive value and area.

In this design, ultrathin-body SPADs with different silicon body thicknesses were fabricated. The effect of body thickness on the SPAD performance is investigated in later sections. Two typical configurations of ultrathin-body SPAD integrated with resistors with body thicknesses of 1.5μm and 3.0μm are shown in Fig. 3.13 (a) and (b) respectively.

For 3.0μm trench etching, it is divided into two 1.5 μm trenches stacked to reduce the topology in half, which is helpful for the following passivation and metallization layer processes. It is also shown in Fig. 3.13 (b).

As presented in Chapter 2, body resistivity is important for SOI since it is directly related to the channel resistivity between anode and multiplication region. By increasing the body thickness from 1.5μm to 3.0μm, the boron doped concentration needs to be reduced from 1E16 cm^{-3} to 5E15 cm^{-3} to make the channel resistance consistent.

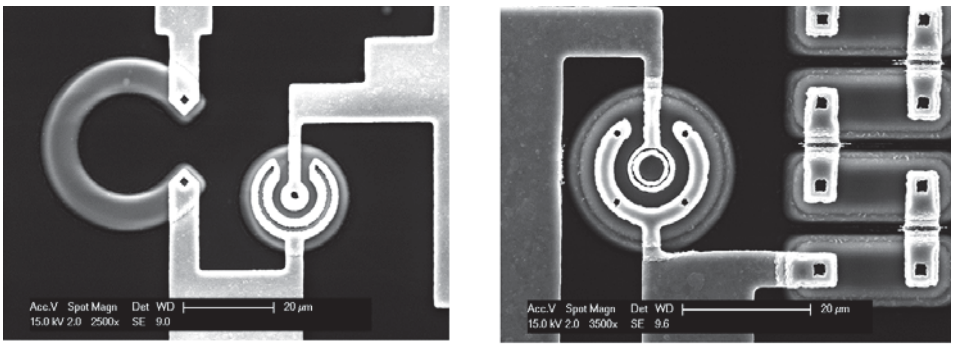


Fig. 3.13: SEM images of configurations of ultrathin-body SPAD integrated with resistor (a) EPI Si: thickness: 1.5μm; Boron doped concentration: 1E16 cm^{-3}; (b) EPI Si: thickness: 3.0μm; Boron doped concentration: 5E15 cm^{-3}.

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.

3.3 DCR and V_{BD} analysis on process parameters and cryogenic measurements

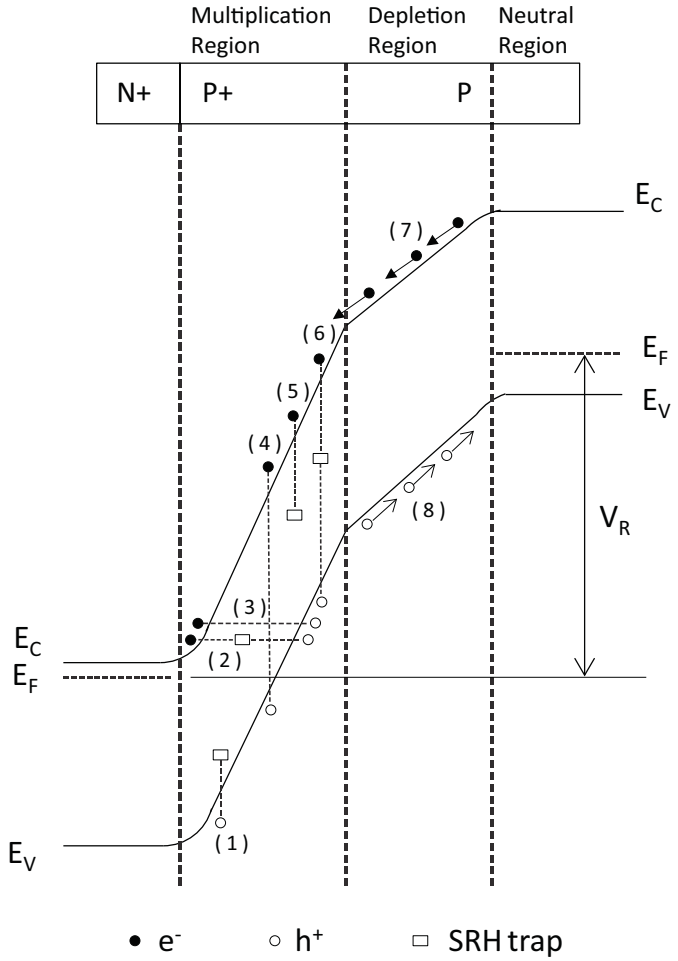


Fig. 3.14: Dark noise source overlaid in energy-band diagram in N+P SPAD structure under reverse bias: (1) trap-released holes; (2) trap-assisted tunneling; (3) band-to-band tunneling; (4) directly thermal injection; (5) trap-released electrons; (6) trap-assisted thermal injection; (7) electrons drift; (8) holes drift. Note: Dark noise (1)-(6) could also happen at depletion region.

As mentioned in Chapter 2, the breakdown voltage V_{BD} is defined as the reverse bias voltage, which could make the ionization integral approach 1. When the SPAD is biased above V_{BD}, it is operated in Geiger mode with energy-band diagram shown in Fig. 3.14. Due to the strong electric field in the multiplication region, avalanches could be triggered by non-photon-generated carriers, which

is measured as dark count rate (DCR). The noise source could be due to different mechanisms such as carrier diffusion, Schottky-Read-Hall (SRH) directly thermal injection or trap-assisted thermal injection, directly or trap-assisted band-to-band tunneling, and trap released carriers which is the cause of afterpulsing.

In principle, the combination of trap-assisted noise and tunneling noise dominates the dark noise of SPADs. Especially for the enrichment region, the electric field is so high (above ~1MV/cm) that tunneling increases dramatically.

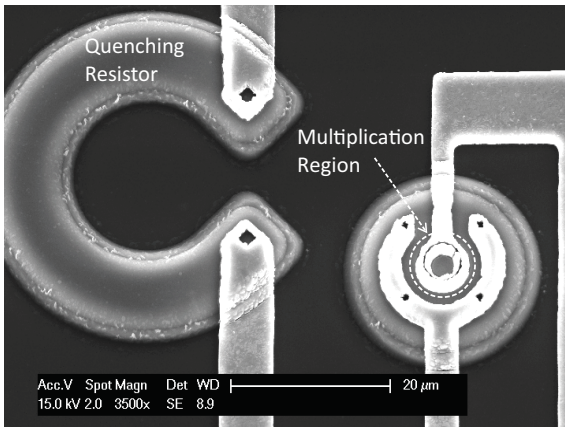
Trap-assisted noise, based on SRH theory, is due to carrier capture and release with silicon lattice defects. Compared with tunneling rates, it is more temperature dependent. The presence of lattice defects, which acts as traps, increases the probability of thermal carrier injection [84].

Table. 3.2: Summary of different noise mechanisms

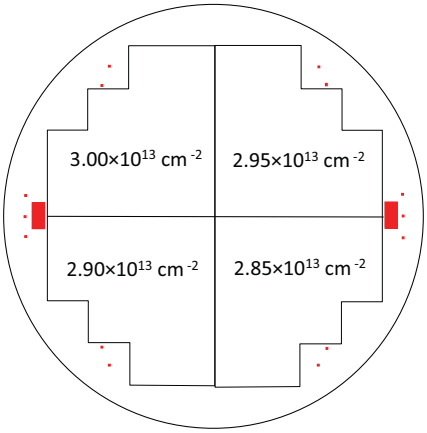
	Thermal generation		**Tunneling**		**Afterpulsing**
	Direct injection	SRH	Trap-assisted tunneling	Band-to-band tunneling	
Type	Uncorrelated (4)	Uncorrelated (6)	Uncorrelated (2)	Uncorrelated (3)	Correlated (1), (5)
Distribution	Poisson	Poisson	Poisson	Poisson	Exponential
Doping	Weak dependence	Strong dependence	Very strong dependence	Very strong dependence	Weak dependence
Excess bias	Weak dependence	Strong dependence	Very strong dependence	Very strong dependence	Strong dependence
Temperature	Strong dependence	Very strong dependence	Strong dependence	Weak dependence	Inverse dependence

During fabrication, the traps are mostly due to implantation and etching. The analysis will be given in later sections. Different noise mechanisms are summarized in Table. 3.2.

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. 3.15 (a). 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}.



(a)



(b)

Fig. 3.15: (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.

As shown in Fig. 2.12 (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 [64][85]. It has been confirmed by DCR versus temperature measurements in Fig. 2.23 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. 3.16. To make sure other

process parameters are consistent, the devices for comparison are fabricated on the same wafer as shown in Fig. 3.15 (b). When the P enrichment dose was reduced from $3.00\times10^{13}cm^{-2}$ to $2.95\times10^{13}cm^{-2}$, the DCR was reduced by 50% while the SPAD was operating at the same excess bias. However, if the P enrichment dose was further reduced to $2.90\times10^{13}cm^{-2}$, the DCR would increase by around 6%. If the P enrichment dose is reduced further to $2.85\times10^{13}cm^{-2}$, the SPAD suffers from premature edge breakdown (PEB).

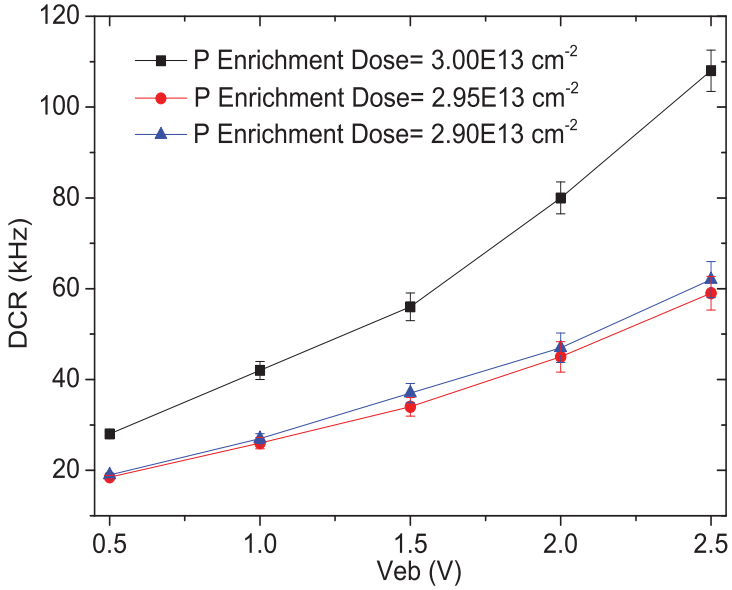


Fig. 3.16: DCR of flexible ultrathin-body SPAD (4μm-diameter multiplication region) with different doping levels in the multiplication region.

To investigate the effect of temperature, a cryogenic chamber – PMC150, as shown in Fig. 3.17, is used in this work to perform DCR and V_{BD} measurement at cryogenic temperature.

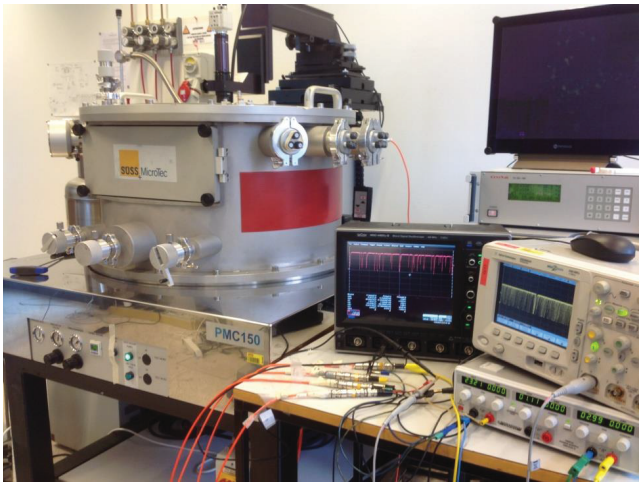


Fig. 3.17: Cryogenic measurement chamber PMC150.

The V_{BD} and DCR are further analyzed at cryogenic temperature. Compared with the breakdown voltages at different temperatures shown in Fig. 3.18, 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 chance of a scattering process decreases when a carrier transport across the junction is also decreased. Lower energy from the field is needed to ionize the carriers, thus decreasing V_{BD} [86].

At the temperatures below ~200K, the dark noise caused by SRH generation and trap-assisted tunneling in the high electric field region is suppressed and the DCR is dominated by band-to-band tunneling [87][88]. The DCR measurements at cryogenic temperatures of 80K and 200K in Fig. 3.19 show consistency. The DCR decreases as the enrichment doping decreases when the dose is reduced from $3.00\times10^{13}cm^{-2}$ to $2.95\times10^{13}cm^{-2}$. However, it stops decreasing and starts increasing when the dose is reduced to $2.90\times10^{13}cm^{-2}$, which is consistent with the results of Fig. 3.19. The DCR density was reduced to hundreds of Hertz per square micrometer at 80K and it is believed to reduce to tens of Hertz per micrometer square [80] if trap-assisted noise from trench etching could be filtered out at extremely low temperatures. However, it needs further research to confirm possible increased afterpulsing below 70K [87][89].

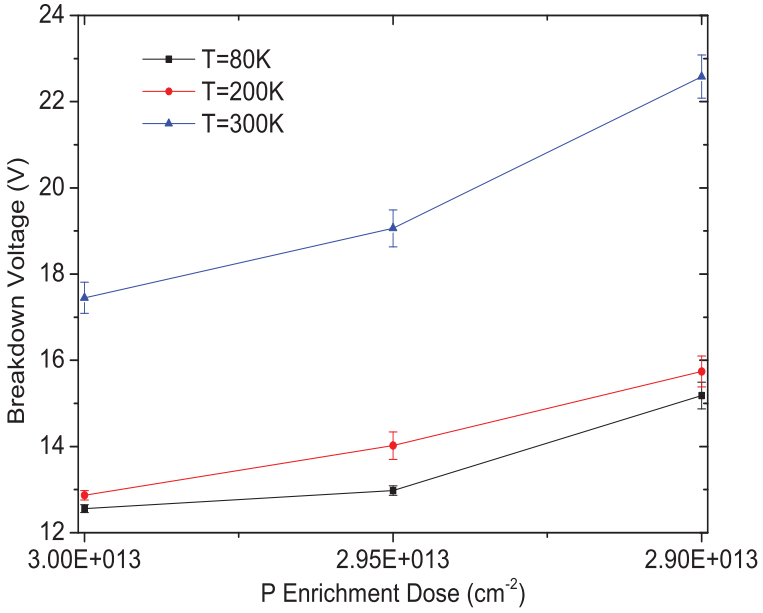


Fig. 3.18: Breakdown voltage of flexible ultrathin-body SPAD as a function of P enrichment dose at different temperatures.

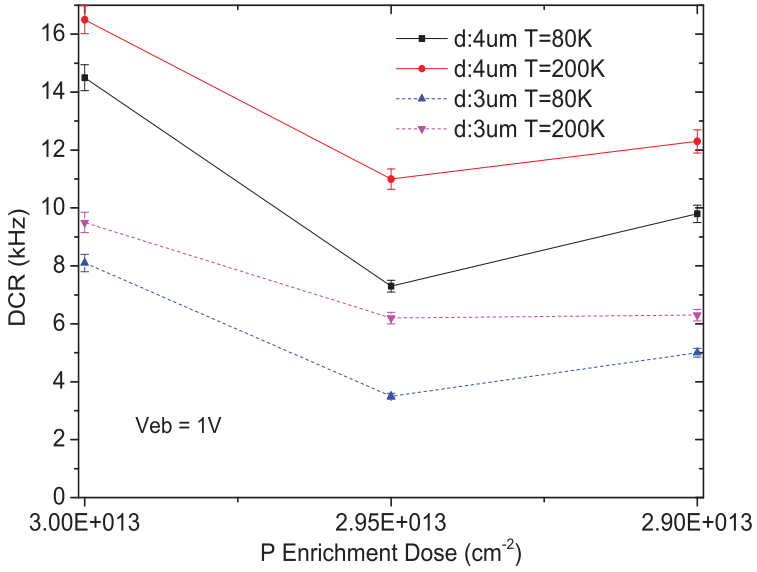


Fig. 3.19: DCR of flexible ultrathin-body SPAD at cryogenic temperatures. (d: diameter of multiplication region).

Electric field simulations (Fig. 3.20) 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) [90][91]. Shown by Fig. 3.20 (b) and Fig. 3.20 (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 shown in Fig. 3.16 and Fig. 3.19.

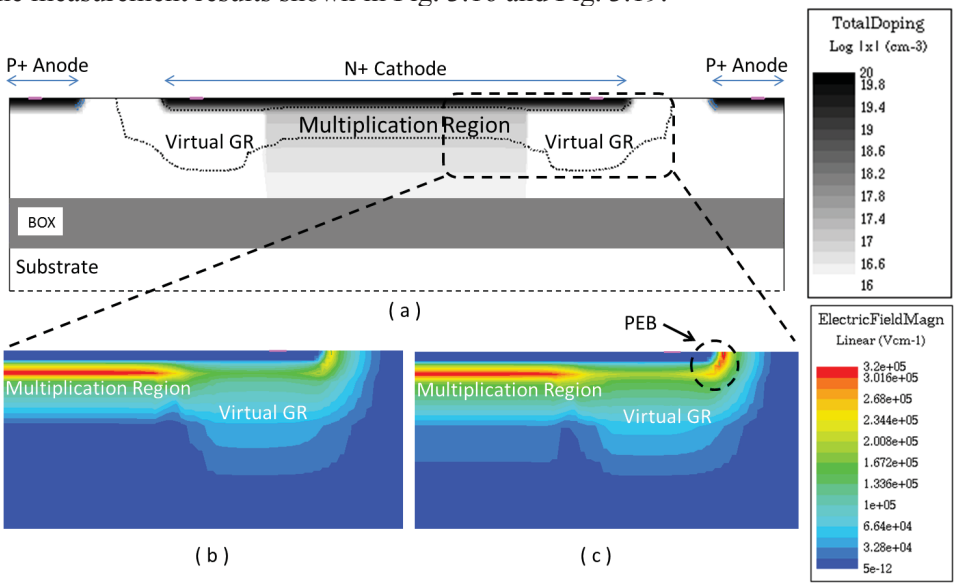


Fig. 3.20: (a) SPAD structure overlaid on a total doping plot; (b) zoom-in plot of electric field with enrichment doping concentrating (dose = 3.00×10^{13}cm^{-2}) with the high electric field at the multiplication region; (c) zoom-in plot of electric field with reduced enrichment doping (dose = 2.85×10^{13}cm^{-2}) causing PEB.

3.4 PDP analysis based on body thickness

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. 3.21.

Based on Equations 3.3 and 3.4, Z_w of different values could be applied according to FSI and BSI, respectively: In FSI, the PDP for device thickness of 1.5μm and 3.0μm can be modeled as equation 3.7 and 3.8, respectively;

$$PDP = \int_0^{1.5\mu m} T(\lambda)\mu(\lambda)\exp(-\mu(\lambda)z)\,p(z)dz\,; \qquad (3.7)$$

$$PDP = \int_0^{3.0\mu m} T(\lambda)\mu(\lambda)\exp(-\mu(\lambda)z)\,p(z)dz\,; \qquad (3.8)$$

while in BSI, the PDP can be modeled as equation 3.9 for thickness of 1.5μm and equation 3.10 for thickness of 3.0μm, respectively;

$$PDP = \int_0^{1.5\mu m} T(\lambda)\mu(\lambda)\exp(-\mu(\lambda)(1.5\mu m - z))\,p(z)dz\,; \qquad (3.9)$$

$$PDP = \int_0^{3.0\mu m} T(\lambda)\mu(\lambda)\exp(-\mu(\lambda)(3.0\mu m - z))\,p(z)dz\,. \qquad (3.10)$$

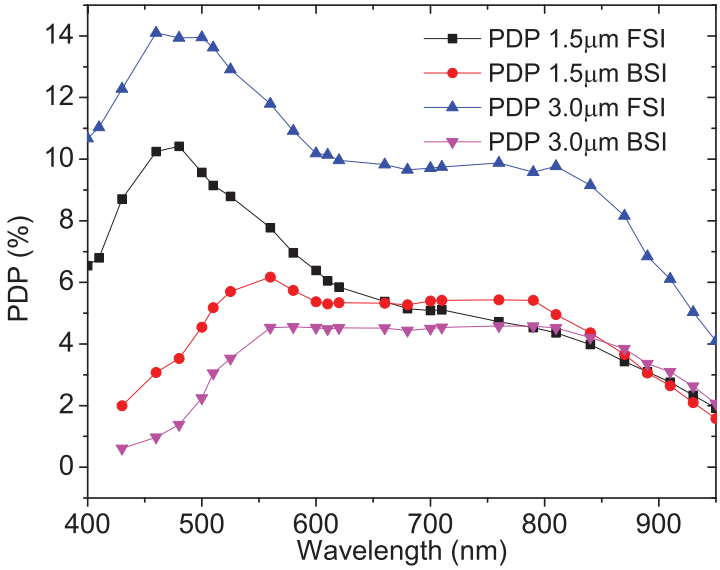


Fig. 3.21: PDP of flexible SPAD with different body thickness in both FSI and BSI.

By increasing the body thickness from 1.5μm to 3.0μm, the peak value of PDP in FSI was 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 the multiplication region and trigger the avalanches. The neutral region is wider for thicker body since the cathode and enrichment region profiles 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.

3.5 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 local oxidation of silicon (LOCOS) as isolation, instead of trench, as shown in Fig. 3.22 (a) and (b) [80]. The SEM microphotograph of the planar SPAD in circular shape, which is isolated by 840nm-thick LOCOS by wet oxidation, is shown in Fig. 3.22 (c).

As summarized in Fig. 3.23, the fabrication starts with a silicon SOI wafer prepared by epitaxy technology. 100nm silicon nitride layer was deposited by LPCVD and patterned to define the LOCOS area. Then, LOCOS is formed by thermal oxidation followed by silicon nitride stripping using wet H3PO4 at 157 ºC. Enrichment doping, cathode, anode and surrounding ring were formed by implantation. After annealing, metallization layer was formed and wafer was sent to alloy.

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 [80]. 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. 3.24.

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 [92] to filter out the defects around the trenches needs to be investigated in future work.

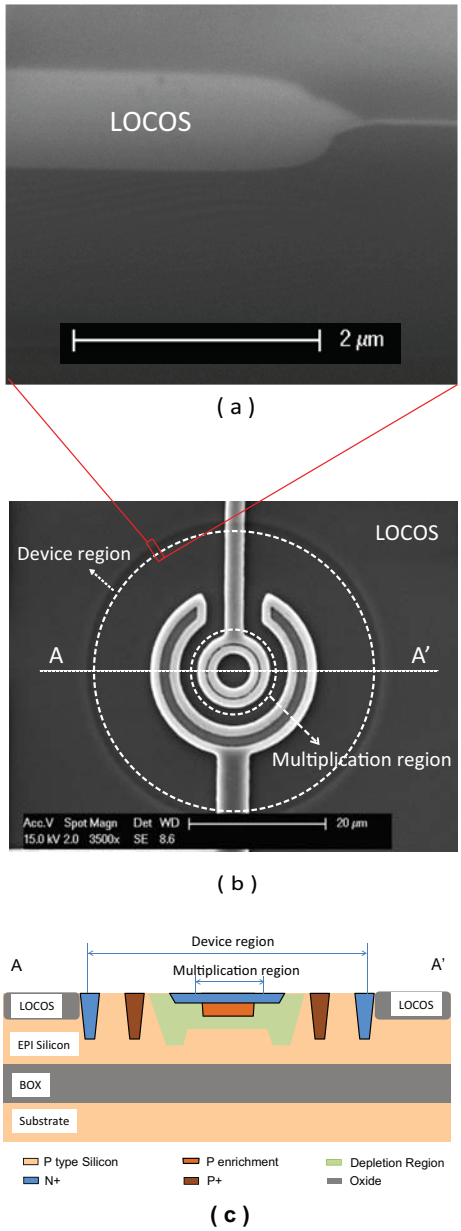


Fig. 3.22: (a) SEM top-view microphotograph of planar SPAD. (b) Cross-section of planar SPAD.

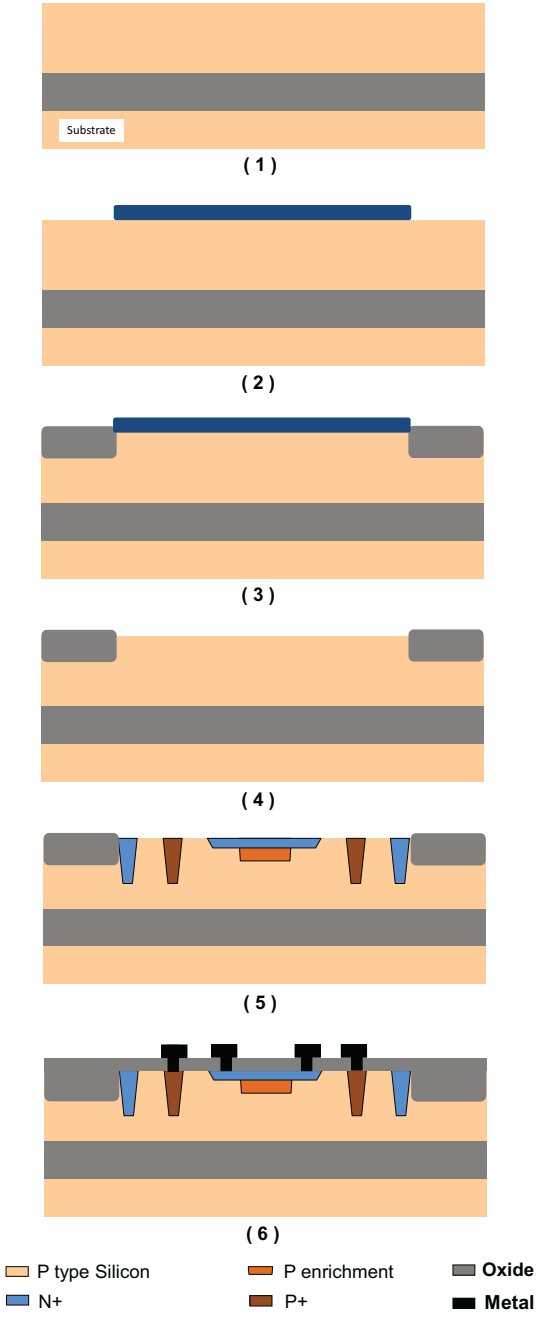


Fig. 3.23: (1) SOI epitaxy; (2) Si_3N_4 layer deposition by LPCVD and patterning; (3) LOCOS oxidation; (4) Si_3N_4 stripping; (5) Junction implantations; (6) Metallization.

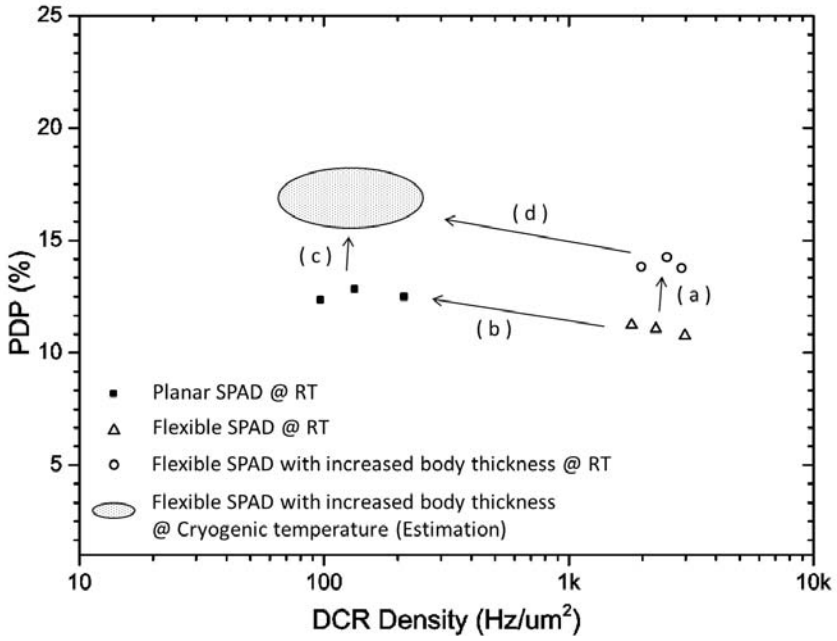


Fig. 3.24: Plot of DCR density vs. PDP in recent developments, each introducing a technological innovation or optimization with a consequent performance improvement. Note: PDP refers to FSI. (a) Increase of body thickness; (b) isolation by LOCOS; (c) isolation by trench and operation at cryogenic temperature; (d) operation at cryogenic temperature.

3.6 Conclusion

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, was further reduced, resulting in a DCR density of tens to hundreds of Hertz per square micrometer 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.

Chapter 4

Flexible single-photon avalanche diode sensor integration with CMOS technology

In this chapter, we extend this work to the first flexible SPAD image sensor with in-pixel and off-pixel electronics integrated in CMOS. In section 4.1, an overview is given to explain why CMOS technology is needed for SPAD. Some key parameters about CMOS SPAD and process specifications are given in section 4.2. Details of pixel design and fabrication are explained in section 4.3. Thorough characterizations including static characteristics and Geiger-mode performances are given and compared in section 4.4, followed by conclusions in section 4.5.

This chapter is based on results published in [81][93].

4.1 Overview

In order to perform single-photon imaging, many efforts have been devoted to imaging device integration. Thanks to improved CMOS fabrication and Moore's Law, SPAD technology implemented in planar processes for imaging systems has significantly progressed in recent years [48]. In the foreseeable future, CMOS will no doubt remain the dominant technology defining the mainstream integration technology [94]. Therefore, it is very important to make any new devices CMOS compatible.

With the goal of enabling the flexible SPAD implemented as large-scale flexible single-photon imaging sensor, this thesis work includes the first flexible SPAD image sensor with in-pixel and off-pixel electronics integrated in CMOS. It could pave the way to larger scale sensor arrays integrated with more electronics on flexible substrate. Based on this architecture and dual-side illumination, the flexible SPAD chip could be implanted into a human organ, such as an eyeball, and integrated with stimulation components at any side of the retina.

Thanks to reduced parasitic capacitance, obtained through the integration of the output buffer, the dynamic range of SPADs could be enhanced. Afterpulsing and crosstalk are negligible at the dead time. Intrinsic silicon layer could be doped by diffusion from epitaxy layer during a large amount of thermal budget during CMOS process, resulting comparable PDP in FSI and BSI.

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 [95] and for electronics to drive a quantum computer [96].

4.2 Key parameters

4.2.1 Dynamic range and dead time

Dynamic range is defined as the ratio between saturation count rate and DCR.

$$DR = 20log\left[f_{sat} / DCR \right], \tag{4.1}$$

where f_{sat} is the saturation or maximum count rate.

f_{sat} can be maximized by reducing the hold-off time; this is achieved with integrated quenching resistors and buffer circuitry to reduce parasitic capacitance is an effective way, while keeping afterpulsing in check.

4.2.2 CMOS technology specifications

The whole flow of fabrication implemented at the Else Kooi Lab (EKL) is shown in Fig. 4.1. In design phase, variable processes including oxidation, deposition, implantation, and annealing are simulated and optimized based on TSUPREM-4. SPAD and CMOS transistors based on the SOI structure are simulated using Medici. After combining different simulation results, the final layout was designed. The processing phase includes the CMOS SPAD process in class 100 cleanroom and substrate transfer process in the MEMS lab. The CMOS process features are summarized in Tab. 4.1. After that, the wafer was diced and packaged for characterization.

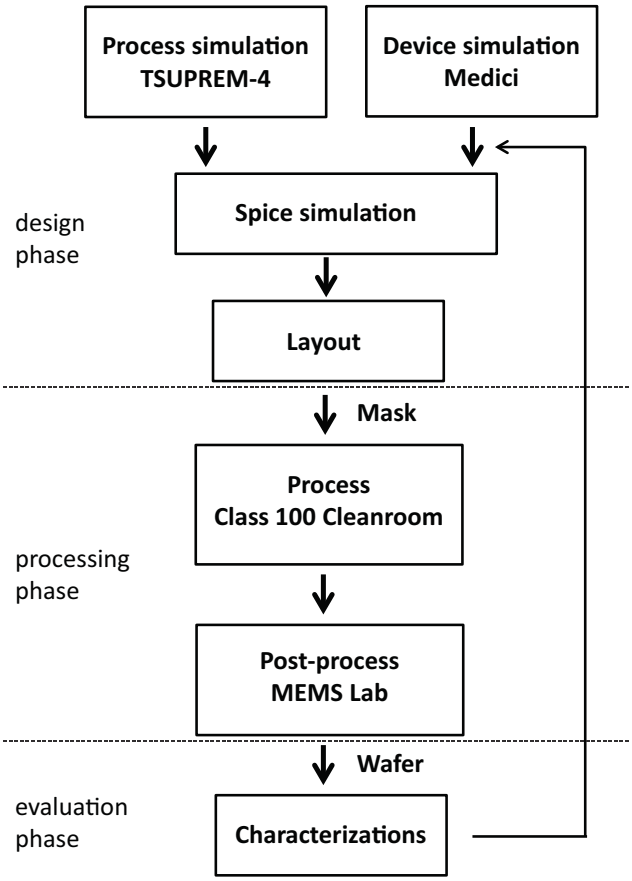


Fig. 4.1: Overview EKL CMOS fabrication flow.

Table. 4.1: CMOS process features.

Item	*Parameter*
Wafer	4 inch SOI
Gate material	Poly silicon
Channel isolation	LOCOS
Epitaxy thickness	1.65 ± 0.05 µm
Minimum lithography size	1.0 µm
Channel width	2.0 ± 0.1 µm
Gate oxide thickness	25 ± 1nm

4.2.3 Fill factor

Fill factor is defined as the ratio of active area and pixel area. CMOS transistors and other components require a considerable area in pixel, thus reducing fill factor and photon detection efficiency (PDE). Improving the active area of the SPAD would increase the fill factor but introduce more dark noise at the same time. An effective way to reclaim some of the lost light is the use of microlenses as shown in Fig. 4.2. The proposed microlenses imprinting on flexible CMOS SPAD sensor will be explained in Section 4.3.

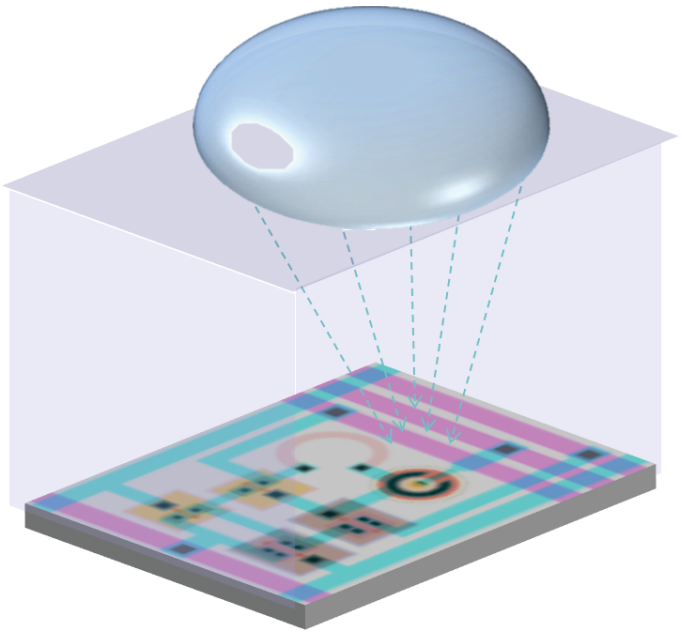


Fig. 4.2: Detail view of micro-lens imprinted on SPAD pixel.

4.3 Pixel structure and fabrication flowchart

As shown in Fig. 4.3, 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 [93].

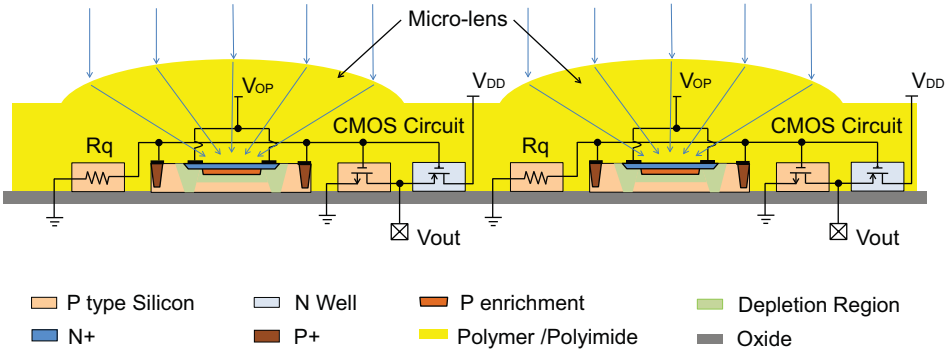


Fig.4.3: Flexible CMOS SPAD sensor pixel schematic and SPAD cross-section.

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. LOCOS, formed similarly as flowchart in Fig. 3.23, is used to isolate transistor channels. After gate oxide growth, poly silicon is deposited and doped by phosphorous diffusion as gate. Source and drain are formed by implantations.

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. The device fabrication flow is summarized in Fig. 4.4.

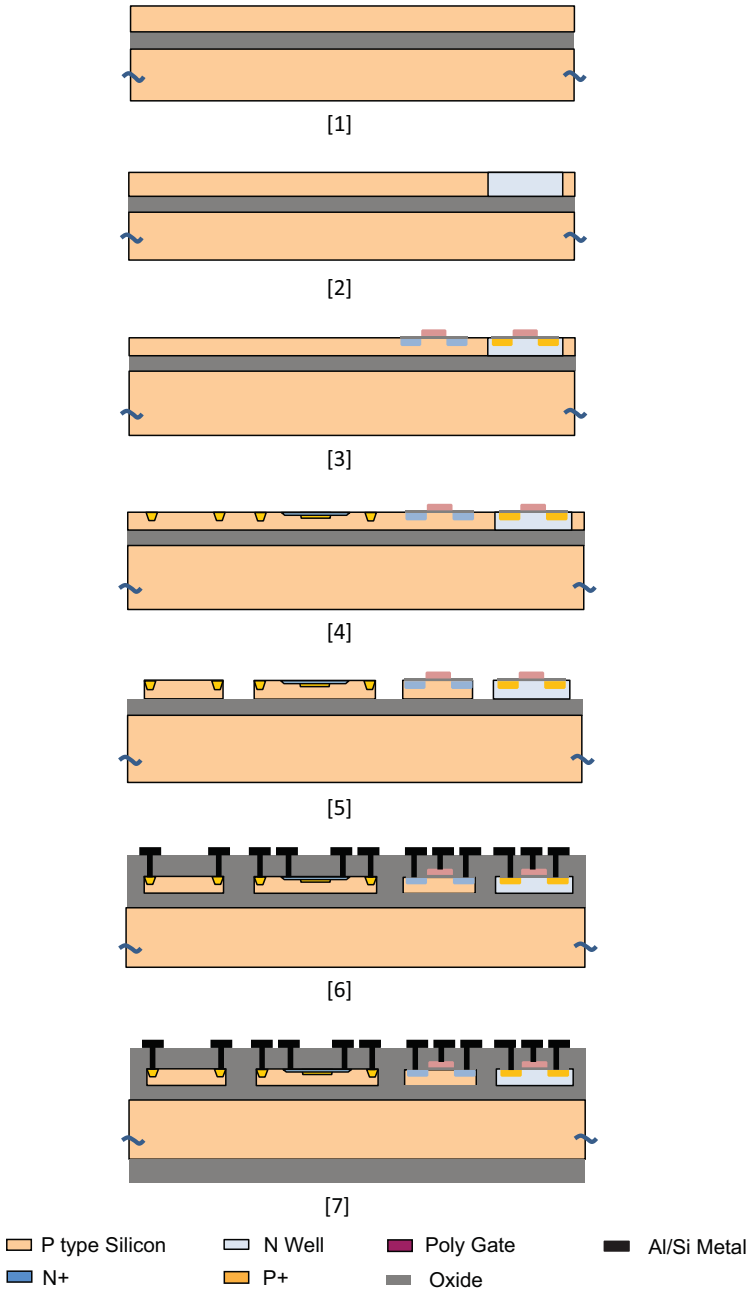


Fig. 4.4: (a). Pixel device fabrication flow chart. (1) Silicon epitaxy process on SOI wafer; (2) N-Well implantation and driving in; (3) CMOS transistor fabrication; (4) SPAD junction implantation; (5) trench etching process; (6) metallization; Note: Only 1st metallization is shown in this figure; (7) backside oxide deposition.

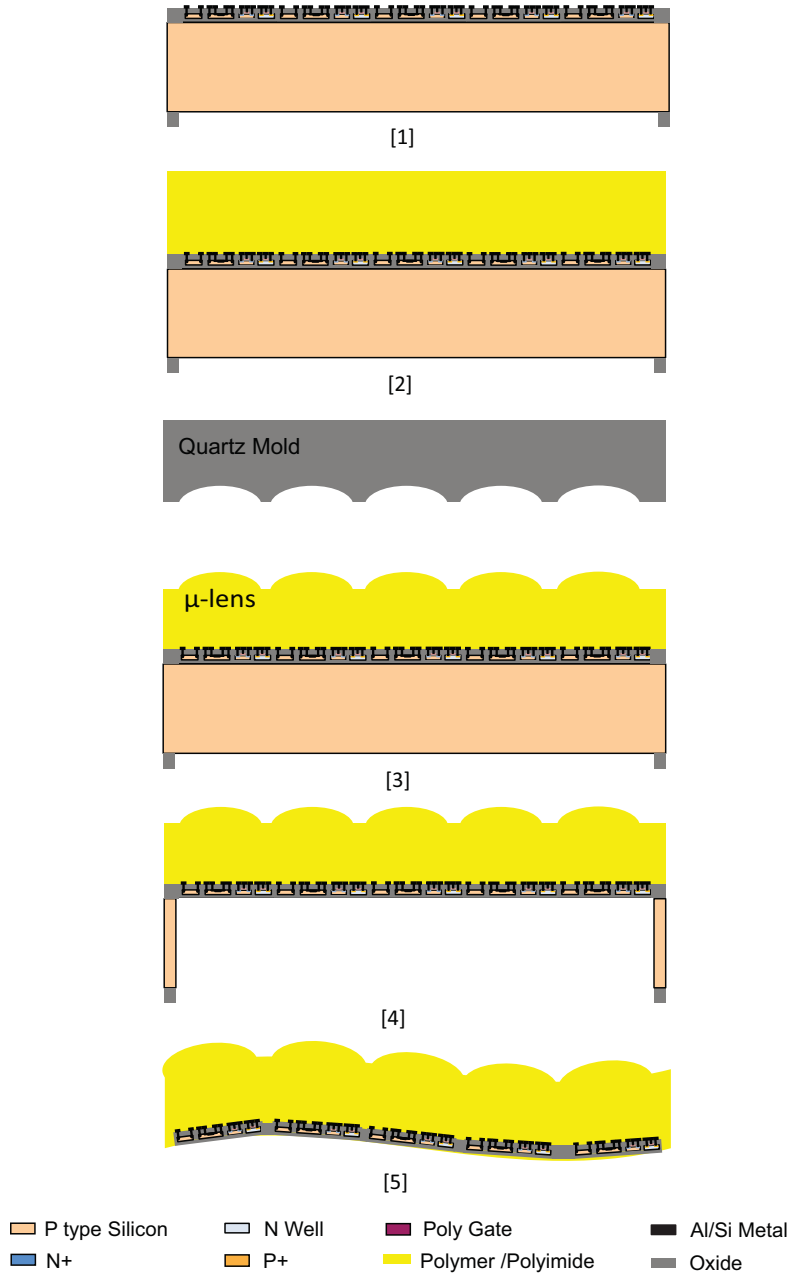


Fig. 4.5: Flexible substrate transfer and microlens fabrication. (1) Oxide mask fabrication on backside; (2) sol-gel polymer coating and curing; (3) microlens imprinting by quartz mold; (4) substrate etching; (5) layer releasing.

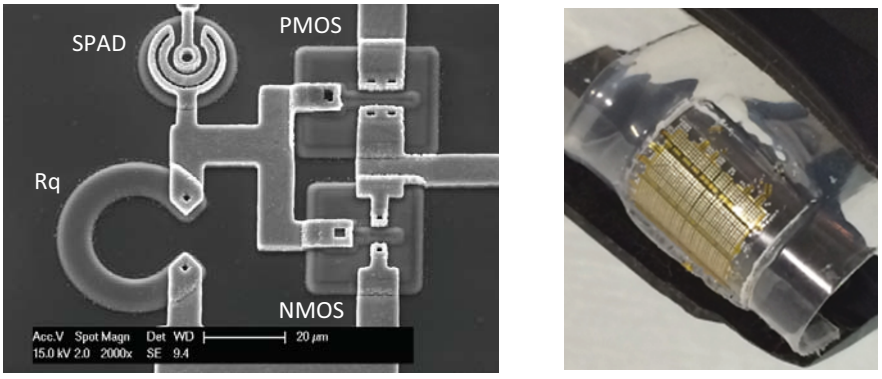
The flexible substrate transfer and microlens imprint process are summarized in Fig. 4.5. Followed by the backside oxide deposition in Fig. 4.4, the oxide is patterned as mask, as shown in step 1. The sol-gel polymer is coated on top of the device after the backside oxide mask is patterned. Then, the 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 [97] with lateral alignment accuracy of less than 1μm. The imprinting process was being completed at the time of the writing of the thesis. 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. Tens of thousands of SPAD pixels from several batches have been fabricated with yield higher than 95% and results are quite consistent and reproducible.

4.4 Measurement results

Following the fabrication in the proposed technology of Fig. 4.4-4.5, the first SPAD image sensor integrated with CMOS buffering circuit on flexible substrate is demonstrated as shown in the SEM microphotograph of Fig. 4.6(a). The photo of bent flexible SPAD pixel farm sample, which has been released and mounted to PDMS piece is shown in Fig. 4.6(b).



(a) (b)

Fig. 4.6: (a) SEM microphotograph of a CMOS SPAD sensor pixel; (b) photo of bent flexible SPAD pixel farm sample.

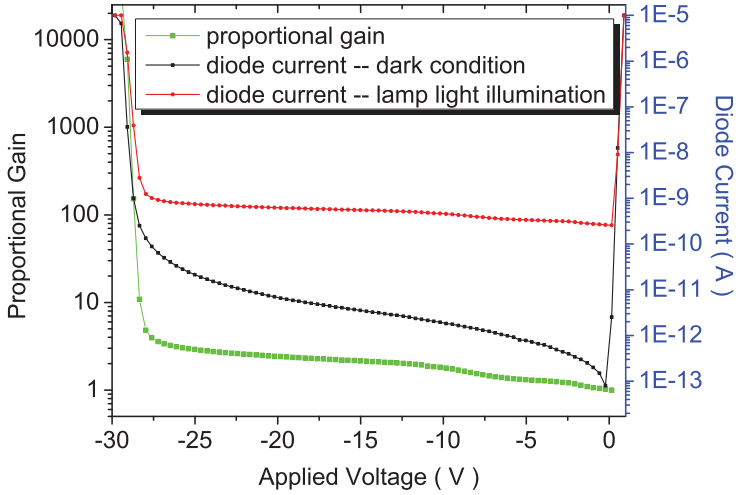


Fig. 4.7: I-V characteristics of SPAD, note: Mitutoyo lamp link fiber source. (Power: 150W)

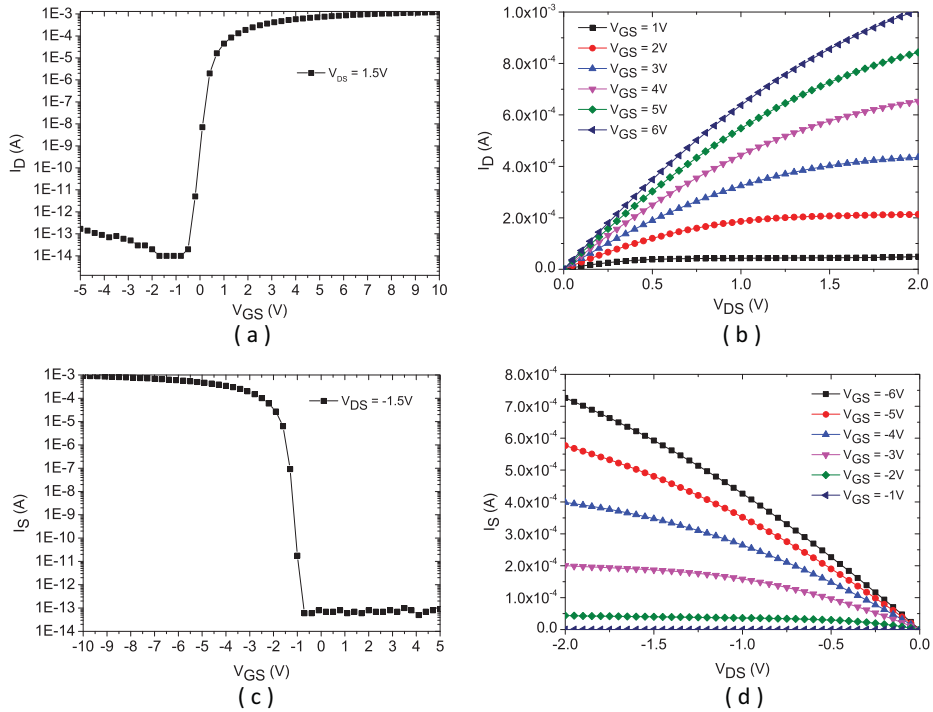


Fig. 4.8: (a) NMOS transfer I-V characteristic. (b) NMOS output I-V characteristic. (c) PMOS transfer I-V characteristic. (d) PMOS output I-V characteristic.

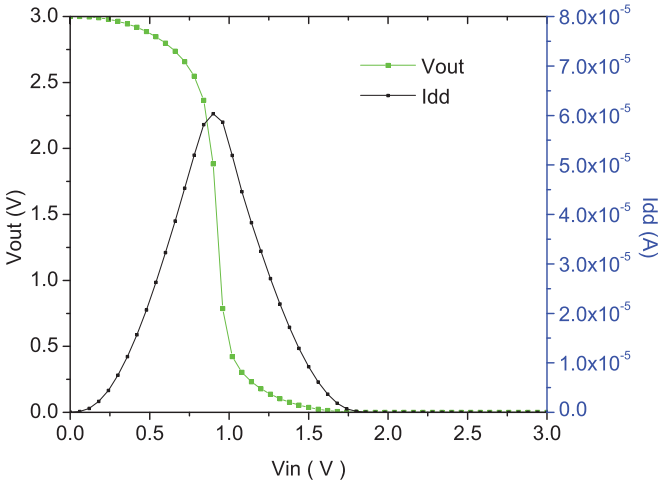


Fig. 4.9: DC transfer curve of CMOS Inverter.

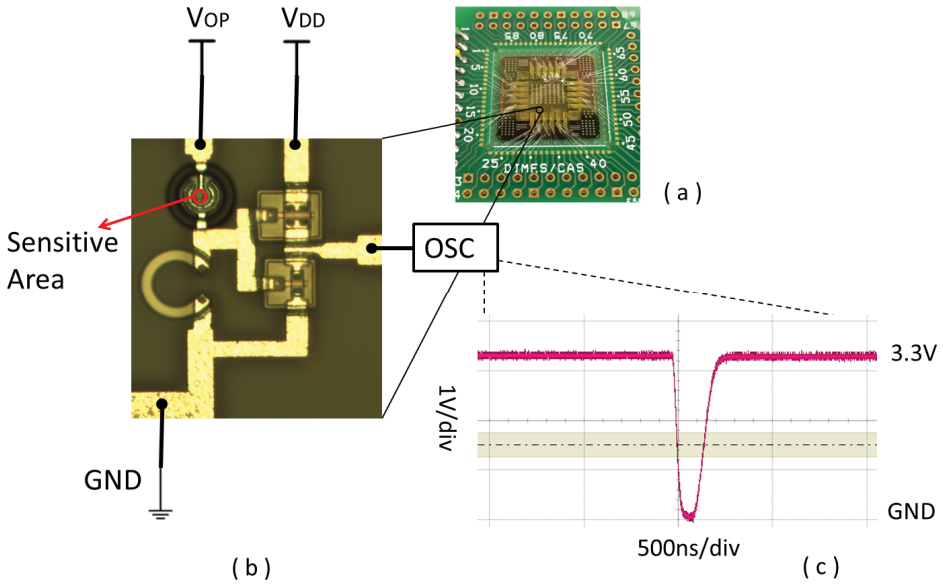


Fig. 4.10: (a) Flexible CMOS SPAD pixel farm sample mounted onto PCB; (b) Microscopic image of flexible CMOS SPAD pixel; (c) Oscilloscope image of dark count quenching pulse of pixel output.

The current-voltage (I-V) characteristics of the SPAD is shown in Fig. 4.7 for different light conditions. The breakdown voltage is around 26.5V at room temperature. The transfer and output characteristics of NMOS and PMOS are shown in Fig. 4.8 respectively. The input-output response of the CMOS inverter, which consists of three-terminal SOI PMOS and NMOS transistors, is shown in Fig. 4.9.

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 and CMOS buffering circuits, which were also integrated on flexible substrate, the operation voltage V_{OP} and the power for circuit V_{DD} need 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). Fig. 4.10 shows the flexible SPAD package on PCB and the whole measurement set-up.

Compared with our previous work [46], the excess bias was increased to 4V, by integrating CMOS buffering. The characterization of DCR as a function of excess bias is shown in Fig. 4.11. The pixel was operated at cryogenic temperatures, as low as 80K, which is reported for the first time to our knowledge; DCR was reduced dramatically, as shown in Fig. 4.12.

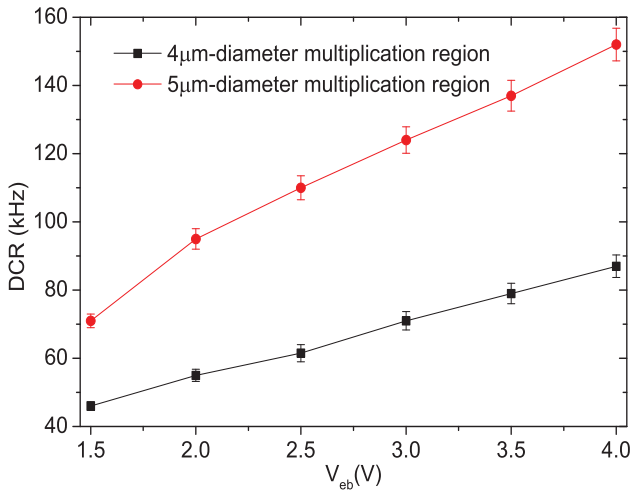


Fig. 4.11: DCR as a function of excess bias in a fully integrated CMOS buffer SPAD pixel.

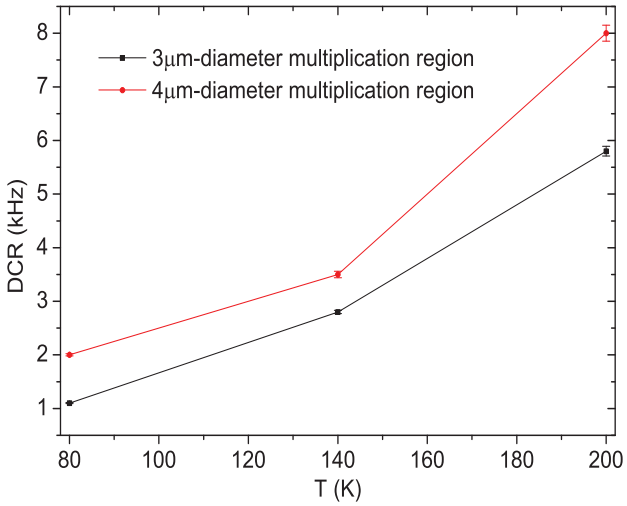


Fig. 4.12: DCR in a fully integrated CMOS buffer SPAD pixel as a function of cryogenic temperature. (V_{eb}=1.5V; d: diameter of multiplication region).

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.

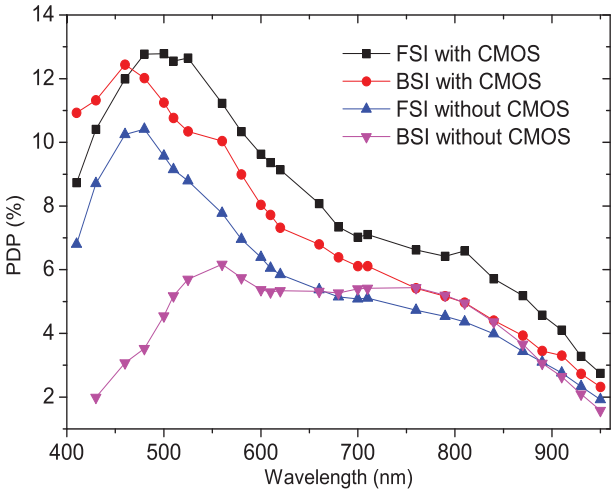


Fig. 4.13: PDP of frontside- and backside-illumination comparison between devices with/without co-integrated CMOS process.

To characterize the sensitivity of the pixel, the PDP was measured in both FSI and BSI, as shown in Fig. 4.13. 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.

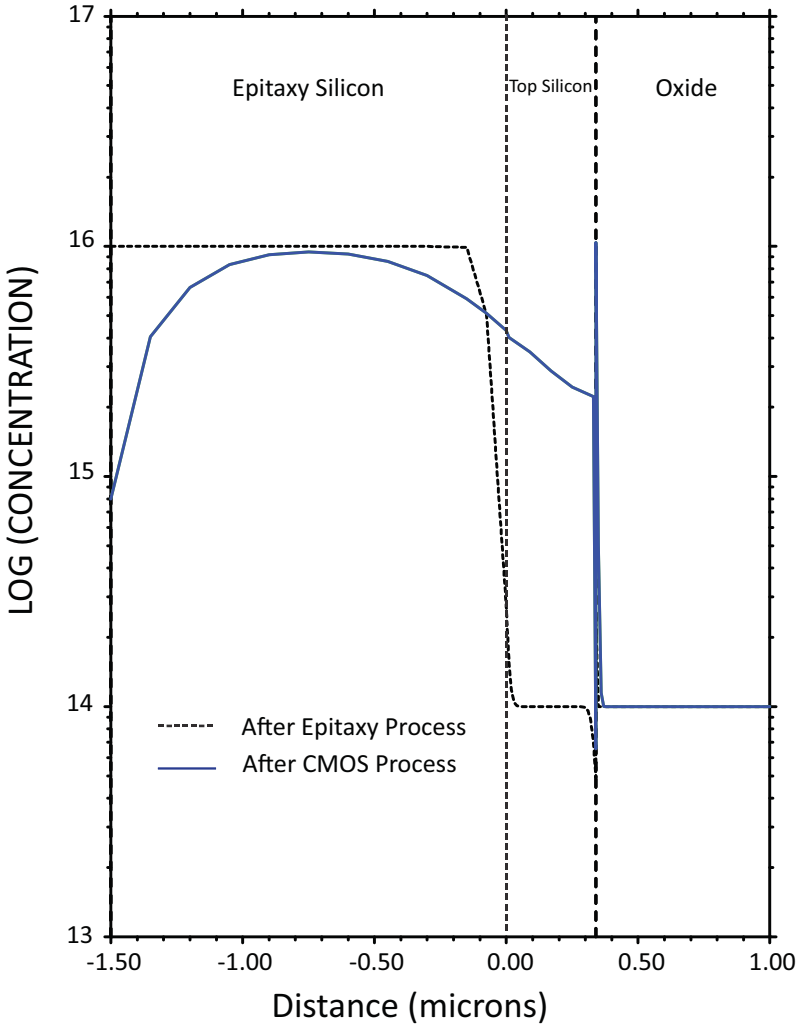


Fig 4.14: SOI Body doping profile comparison before and after high thermal budget CMOS process simulated by TSUPRM4.

Furthermore, compared with our previous work [47], 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 with a large amount of thermal budget in CMOS process. As the doping profile simulation in Fig. 4.14, P-type doping (boron) concentration in top silicon area (originally intrinsic) got increased after high-thermal-budget CMOS process, which includes long-time high-temperature driving, oxidation, LPCVD deposition. The SPAD junction is implanted after 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 (Fig. 3.2) and larger probability for the photon-generated carriers in BSI to drift to the depletion region where avalanches are triggered. On the other hand, the P doping near the epitaxy silicon surface decreased due to the surface oxidation and out-diffusion. So, the N+P junction after CMOS process becomes deeper than [47], which is also positive to get comparable PDP from both front-side and back-side illumination.

This observation is further supported 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.

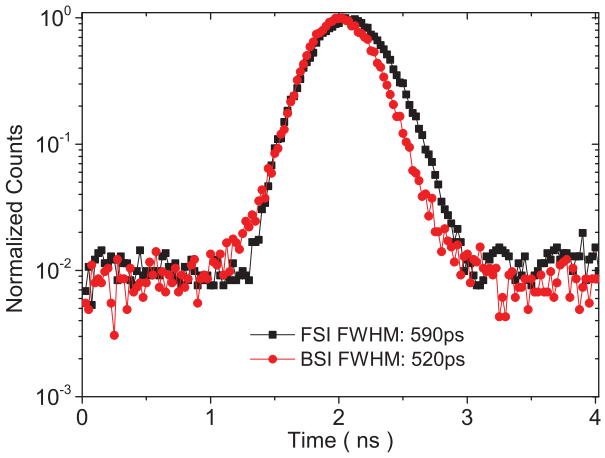


Fig. 4.15: Timing jitter measurements of flexible CMOS SPAD sensor pixel in FSI and BSI using 405nm laser.

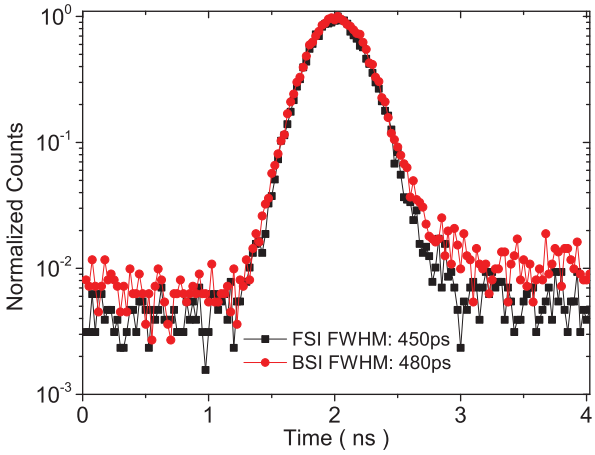


Fig. 4.16: Timing jitter measurements of flexible CMOS SPAD sensor pixel in FSI and BSI using 637nm laser.

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. 4.15 and Fig. 4.16. 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 [47], 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.

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.

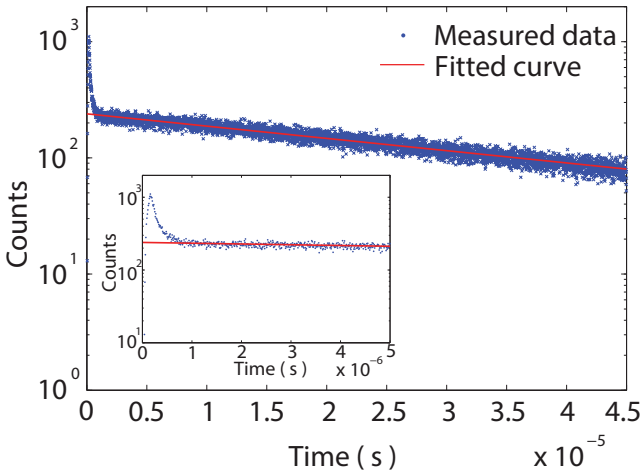


Fig. 4.17: Exponentially fitted inter-arrival time histogram and detail of the histogram in the inset.

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. A histogram of avalanche inter-arrival times is shown is Fig. 4.17 with an exponential fit. When a dead time of 160ns is implemented, the afterpulsing probability is around 1.9%, as shown in Fig. 4.18. At 1.5µs, afterpulsing probability is less than 0.1% or about 5 times better than in [47].

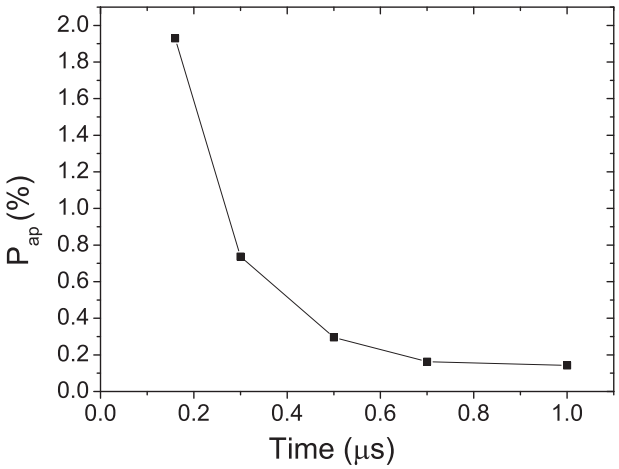


Fig. 4.18: Afterpulsing probability plot extracted from the distribution.

The comparison between flexible trench-isolated SPAD, planar SPAD, and flexible CMOS-buffering SPAD pixel is summarized in Tab. 4.2.

4.5 Conclusion

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). 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.

Tab. 4.2: Performance comparison between flexible SPAD, planar SPAD and pixel.

	Flexible Trench-Isolated SPAD *			**Planar SPAD**			**Flexible CMOS SPAD Pixel**			
Performance	**Min.**	**Typ.**	**Max.**	**Min.**	**Typ.**	**Max.**	**Min.**	**Typ.**	**Max.**	**Unit**
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.

Chapter 5

Flexible 32x32 CMOS single-photon avalanche diode image sensor

In this chapter*, we extend the flexible CMOS SPAD sensor to array format, in which the first flexible 32x32 SPAD image sensor with in-pixel and off-pixel electronics integrated in CMOS is presented. The concept was first proposed in [93]. For the first time, an array of SPADs with integrated CMOS circuitry was implemented on flexible polyimide substrate. After an overview in section 5.1, the architecture design of the array is given in section 5.2. Fabrication details are presented in section 5.3. Imaging readout and signal processing based on FPGA are presented in section 5.4. Characterizations on both SOI chip and flexible chip are given in section 5.5. Dual-side imaging results are shown in section 5.6. Resolution enhancement results are shown in section 5.7. Conclusions are given in section 5.8.

* Part of this chapter's work is shared with Mr. J. Weng's master thesis.

5.1 Overview

Thanks to the development of CMOS technology, CMOS SPAD performance has steadily advanced with improving correlated and non-correlated noise [98][99][100], sensitivity [101] and timing resolution [80]. While the SPAD performance has met the requirements in many applications, SPADs are receiving more and more attention associated with arrays and imager. The first fully integrated SPAD array in standard CMOS technology was demonstrated in [102]. In the last decade, the large scale SPAD array in deep submicron CMOS technology has advanced significantly in different applications such as fluorescence spectroscopy [103], remote sensing [104] and super-resolution imaging [105].

Based on our work in previous chapters, the world's first flexible CMOS SPAD and the corresponding pixel have been characterized extensively [47][81]. Here, we extend this technology to the array format, with focus on statistical performance.

The architecture of the image sensor with in-pixel and off-pixel CMOS electronics was designed; the SPAD device and quenching scheme are based on the work presented in previous chapters, which has been investigated thoroughly. In particular, when designing an array, it becomes important to distribute V_{OP}, which is a relatively high voltage, globally to all the 32x32 pixels, with little variability and negligible IR drop. It is also required to make the pixel and interconnections as compact as possible to have better fill factor, thus sufficient PDE.

Regarding the complexity of the array structure and substrate transfer post-processing, a special layout and fabrication flow has been developed. Imaging readout and signal processing is implemented on an FPGA, owing to the digital nature of SPADs. Performance characterizations are reported proving the fundamental functions.

The work presented in this chapter is significant, since it demonstrates the flexible CMOS SPAD image sensor at large scale for the first time, thus paving the way to applications of implantable biomedical photon sensors or wherever a flexible imaging sensor plane is essential.

5.2 Flexible 32x32 CMOS SPAD sensor architecture

The functional diagram of the flexible CMOS SPAD sensor comprising 32x32 pixels is shown in Fig. 5.1. The sensor uses two power supplies: operational voltage V_{OP} and digital CMOS circuit power V_{DD} of 3.3V. The readout circuitry consists of a 5-to-32 channel decoder for row addressing. Each column has a output buffer. The main sensor array is implemented on flexible substrate while the input and output pads are designed on the peripheral silicon substrate frame.

As shown in Fig. 5.2, the pixel consists of a SPAD, a quenching resistor, 4-transistor circuit which functions as buffer and switch. The N^+ cathode is biased to a high operation voltage V_{OP}, around 25V, which is common to all the pixels in the array. The size of single pixel is 162µm x 125µm and the microscopic image is shown in Fig. 5.3. The inverter stage converts the Geiger-mode voltage pulse to a digital pulse, with the input threshold voltage of around 1V as shown in Fig. 4.9. The switch stage, which is a transmission gate, feeds the digital pulse to output bus when the row is addressed. All the functional circuitry are implemented on flexible substrate.

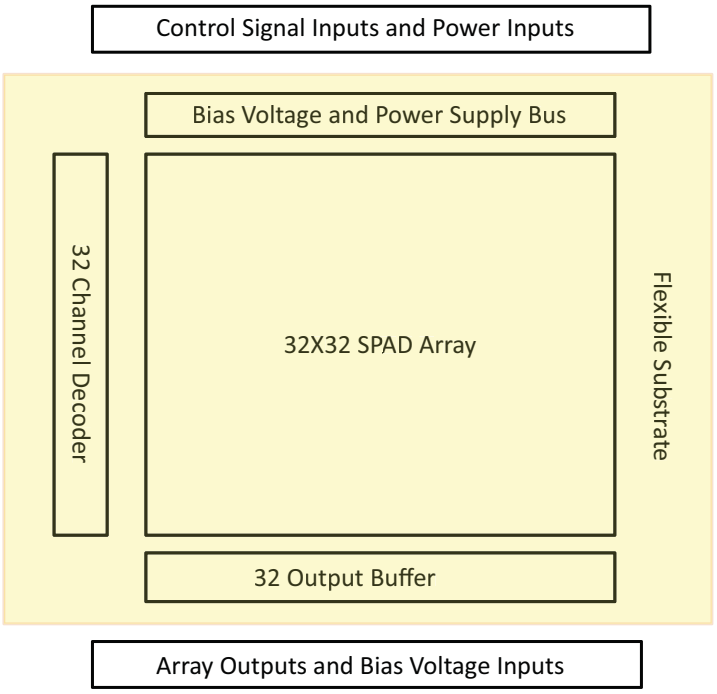


Fig. 5.1: Functional architecture of the flexible CMOS SPAD sensor.

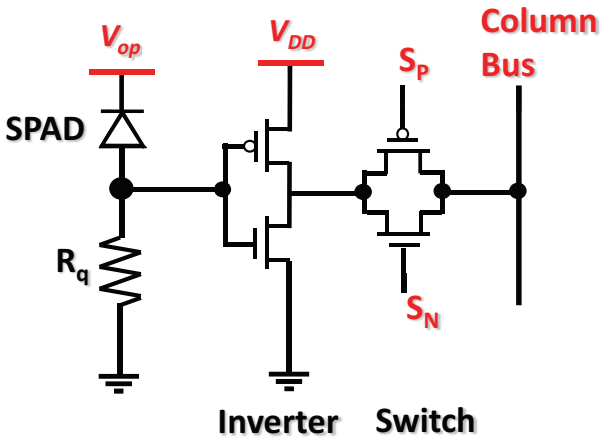


Fig. 5.2: Schematic of single pixel.

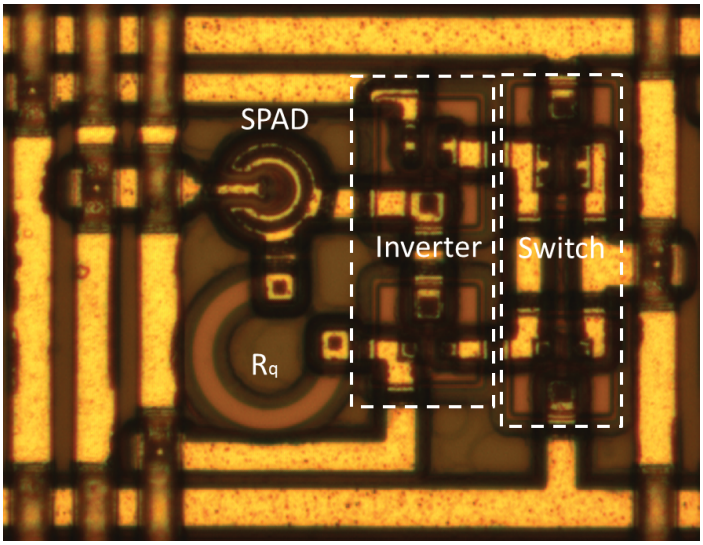


Fig. 5.3: Micrograph of a pixel.

A 5-to-32-bit decoder is implemented as 32-row scanning circuitry to address the array row by row as shown in Fig. 5.4. Each unit is a 5-input NAND gate consisted of two 3-input NAND gates, two 2-input NAND gates and two inverters, all with V_{DD} of 3.3V. The schematic and microscopic layout are shown in Fig. 5.5 and Fig. 5.6, respectively.

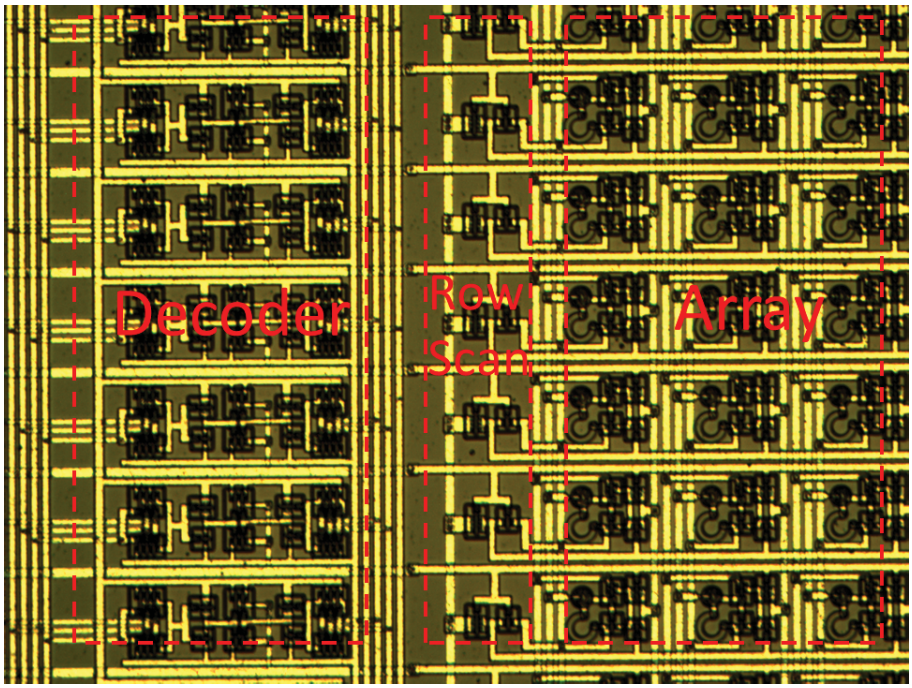


Fig. 5.4: Micrograph of the layout of pixels in the array.

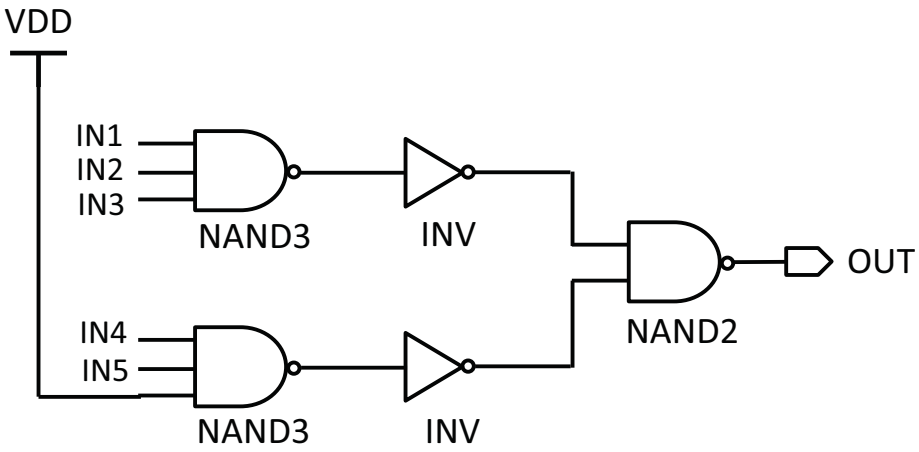


Fig. 5.5: Schematics of 5-input NAND decoder.

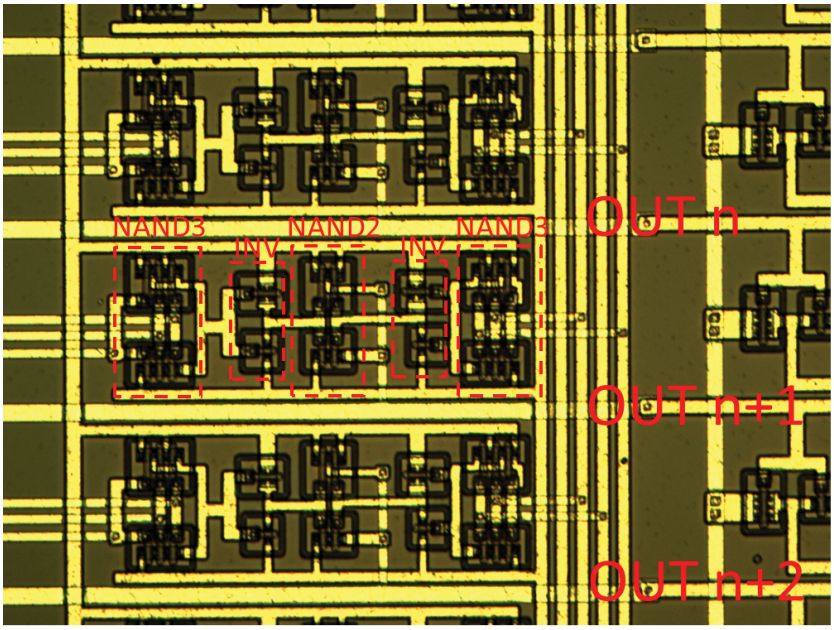


Fig. 5.6: Micrograph of the 5-input NAND decoder.

Tab. 5.1: 32-row addressing circuitry digital value

IN5	IN4	IN3	IN2	IN1	OUT = NAND(IN1, IN2, IN3, IN4, IN5)
$\overline{S5}$	$\overline{S4}$	$\overline{S3}$	$\overline{S2}$	$\overline{S1}$	OUT (Row1)
$\overline{S5}$	$\overline{S4}$	$\overline{S3}$	$\overline{S2}$	$S1$	OUT (Row2)
$\overline{S5}$	$\overline{S4}$	$\overline{S3}$	$S2$	$\overline{S1}$	OUT (Row3)
$\overline{S5}$	$\overline{S4}$	$\overline{S3}$	$S2$	$S1$	OUT (Row4)
$\overline{S5}$	$\overline{S4}$	$S3$	$\overline{S2}$	$\overline{S1}$	OUT (Row5)
					• • •
$S5$	$S4$	$S3$	$S2$	$\overline{S1}$	OUT (Row31)
$S5$	$S4$	$S3$	$S2$	$S1$	OUT (Row32)

The digital value of the 32-row addressing circuitry is shown in Table 5.1. It needs to be noted that each output has also been connected to one particular bond pad outside of the array with the purpose to test during experiments in lab.

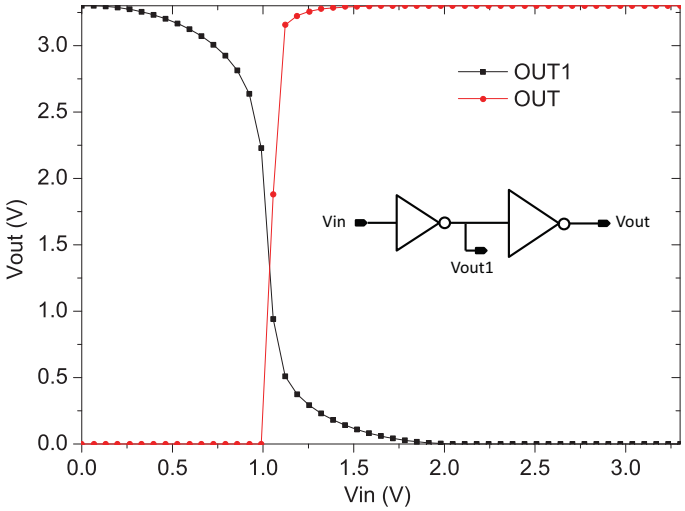


Fig. 5.7: DC transfer curves of two-cascade-inverter buffer.

The output buffer is composed of two cascaded inverters to boost the digital signal on the column bus. The DC transfer curves of two cascade-inverter buffer are shown in Fig. 5.7 and the microscopic layout of the buffer block in the array is shown in Fig. 5.8.

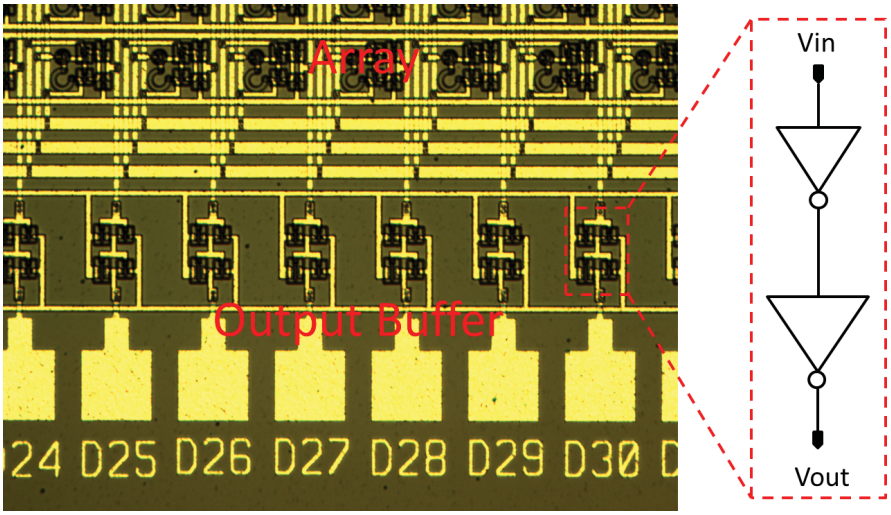


Fig. 5.8: Micrograph of the output buffer.

5.3 Fabrication flow

Based on the fabrication in Chapter 4 (Fig. 4.4), the flowchart is developed as shown in Fig. 5.9. It starts from silicon epitaxy on SOI wafer lightly doped with Boron. The front-end process includes CMOS transistor fabrication, which consists of N-Well implantation and thermal drive-in, LOCOS isolation, gate oxide growth, gate patterning, source and drain implantations. SPAD junction is formed by three implantations. Trenches are etched for device isolation and wafer are sent to anneal. It takes 10 lithography masks in total to implement the front-end process.

Different from fabrication in Chapter 4, the back-end process is much more complicated because two-layer metallization needs to be implemented on top of the device array with a topography step of 1.85 μm. Etch-back and re-deposition of passivation layers is used to form spacers during the topography steps. According to the layout, back-end process parameters, such as spacer dimension, thicknesses of different passivation layers and metal layers, are optimized. Overall, there are 4 lithography masks in back-end process.

Fig. 5.9 shows the SEM images of devices after 1st metallization. In Fig. 5.10, the cross-section shows the spacer at the step could cover the topology well. The 2nd metallization is performed on top on 1st metallization layer so another optimized-thickness oxide is deposited on 1st metal, followed by VIA opening and 2nd metal sputtering deposition. The images of 2nd metallization and cross-section are shown in Fig. 5.12 and Fig. 5.13 respectively.

After the back-end process, the chip is functional as SOI and followed by substrate transfer post-process, which would take two lithography masks. After oxide deposition on backside, the first lithography, which is also implemented on backside, forms the patterns for substrate etching. Polyimide is then patterned to expose SPAD sensitive area and pad area. Substrate etching would be implemented by using DRIE as shown in Fig. 5.14 (a) and die was released by reasonable mechanical force, which is shown in Fig. 5.14 (b).

The whole flowchart is summarized in Fig. 5.9. The whole process is implemented at class-100 cleanroom and MEMS lab at Else Kooi Lab (TUDelft).

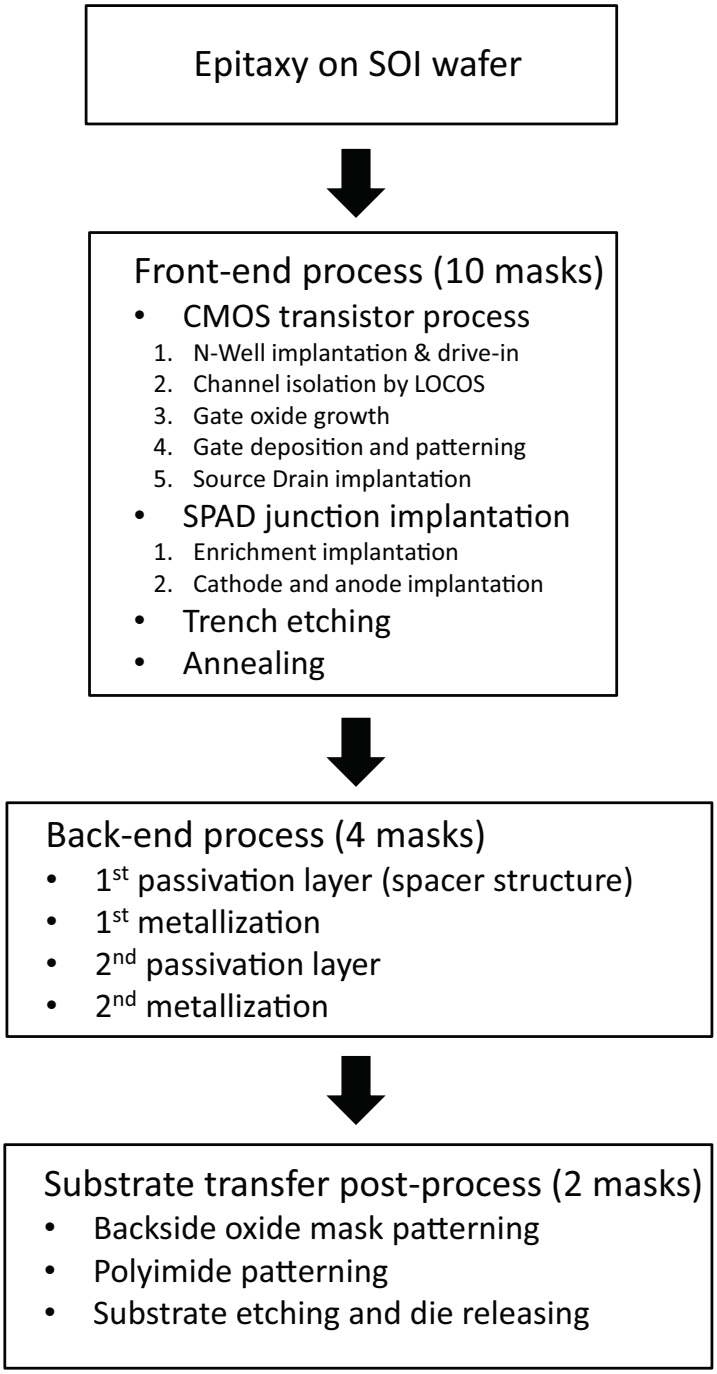


Fig. 5.9: Flowchart diagram of flexible CMOS SPAD sensor fabrication.

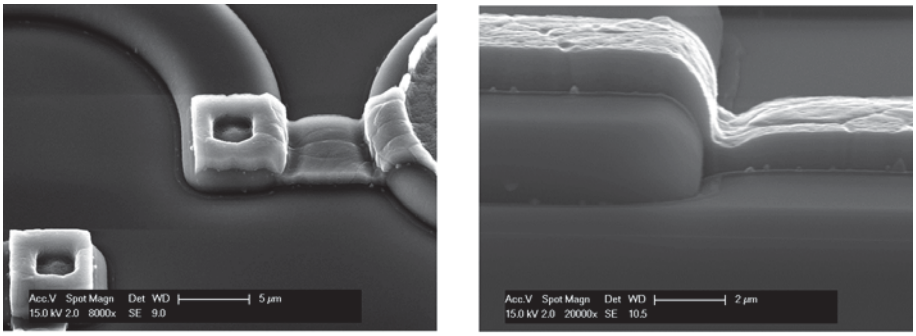


Fig. 5.10: (a) SEM image of 1st metallization. (b) SEM image of the 1st metal layer at topography step.

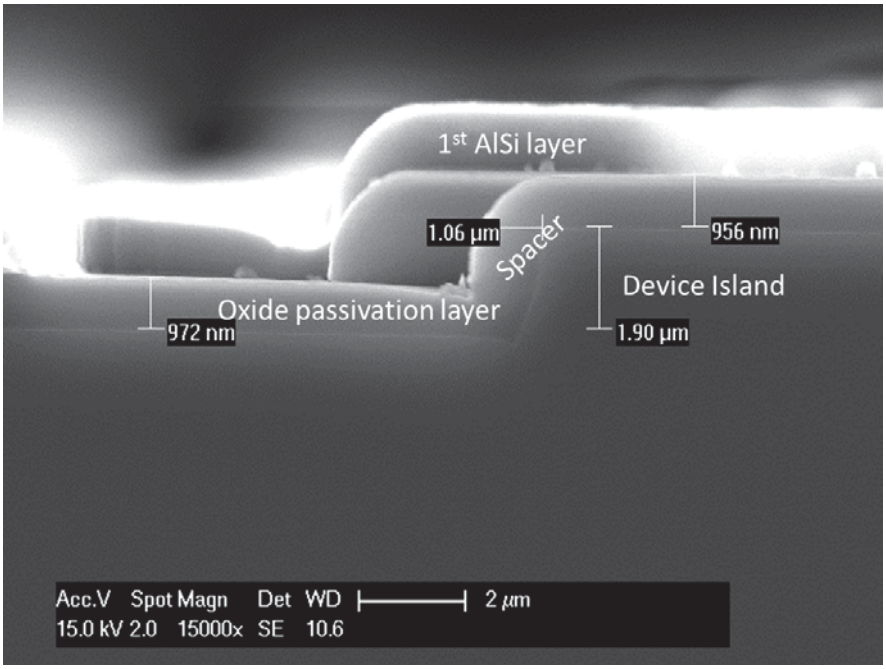


Fig. 5.11: Cross-section image of 1st metallization.

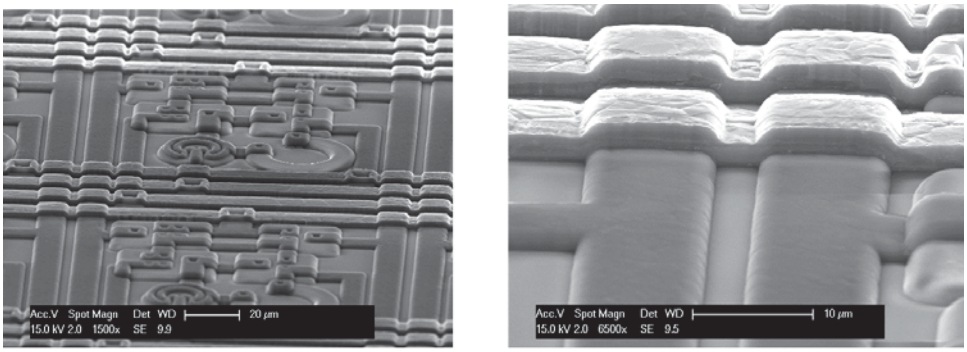


Fig. 5.12: (a) SEM image of 2nd metallization. (b) SEM image of the 2nd metal layer at topography step.

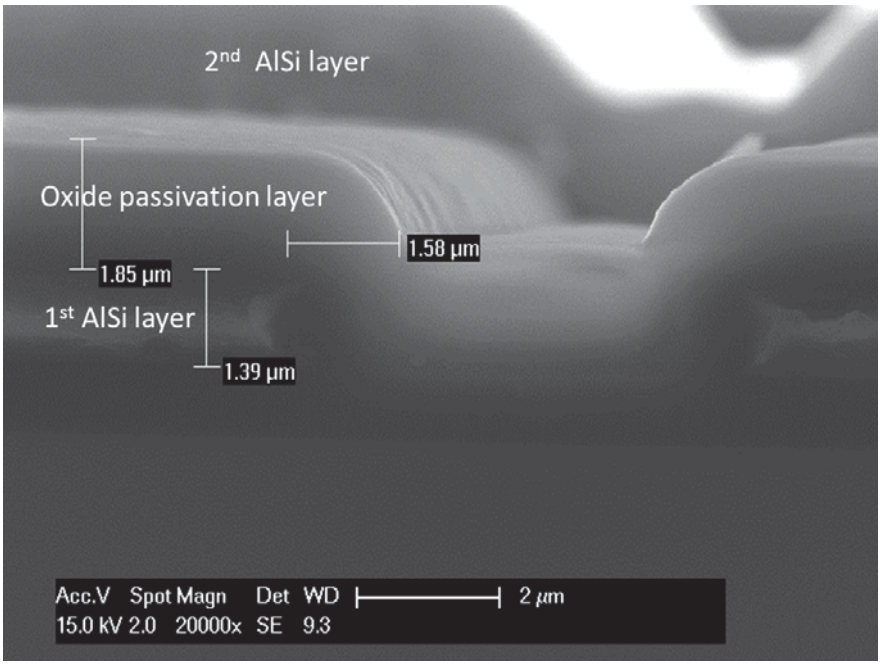


Fig. 5.13: Cross-section image of 2nd metallization.

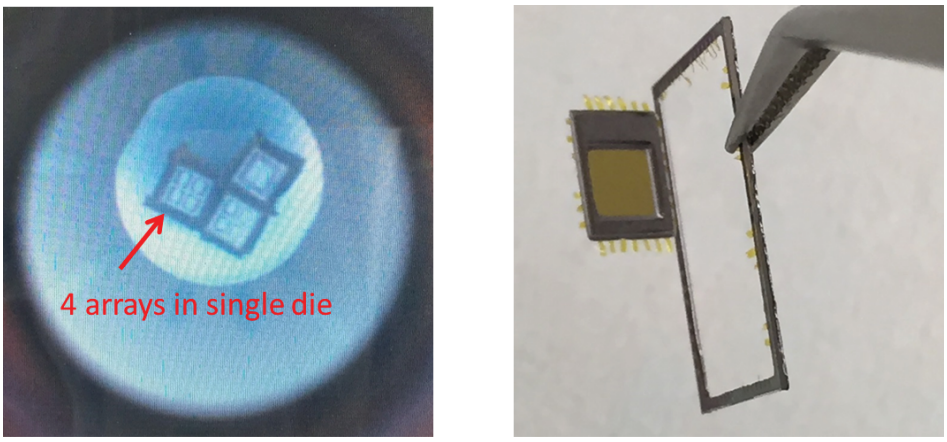


Fig. 5.14: (a) Image during substrate etching in plasma. (b) Last one of the four arrays releasing from a die after substrate etching.

5.4 Imaging readout system and signal processing

The chip was mounted on a CPG 144 package or PCB for wire bonding. V_{OP} and V_{DD} are supplied through the external power supply. An FPGA is implemented to generate row scanning signal and receive output signal on column bus. Signal data are processed by logic units and FIFOs in FPGA and transmitted by the USB interface module to the computer. The diagram of the readout system is shown in Fig. 5.15.

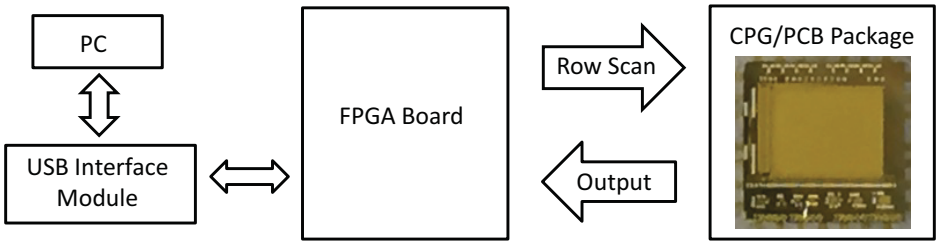


Fig. 5.15: Diagram of the image readout and signal processing system.

The hardware system, which is composed of a chip-on-board, mother board, USB communication module and FPGA board, is shown in Fig. 5.16.

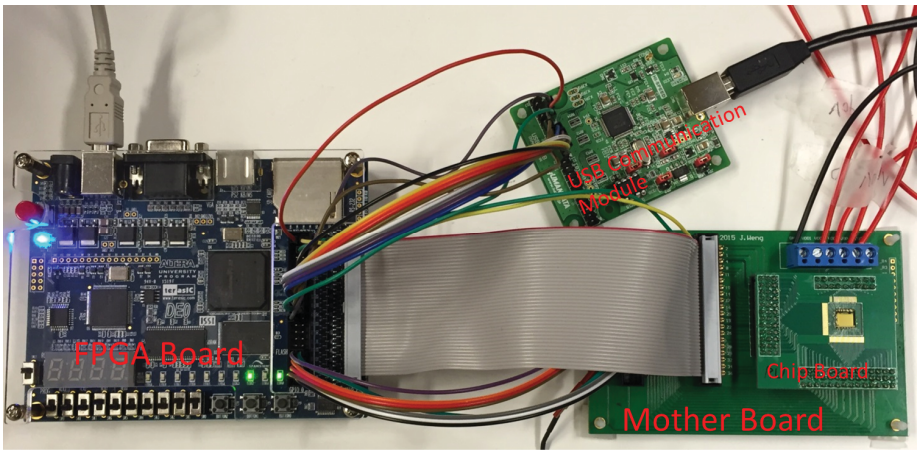


Fig. 5.16: Photo of image readout and signal processing hardware.

5.5 Array characterizations

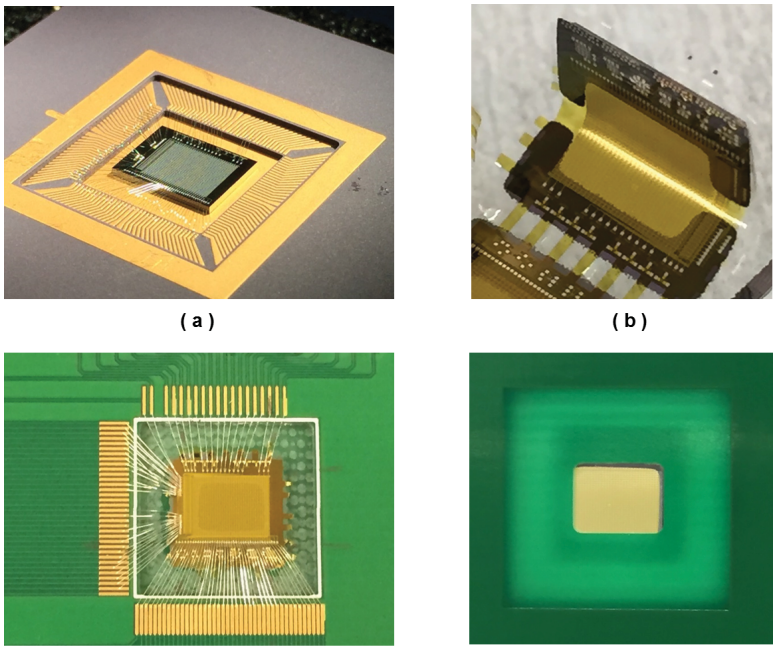


Fig. 5.17: (a) SOI chip mounted in CPG-144; (b) bendable chip released after substrate etching; (c) front side of flexible chip mounted onto PCB; (d) backside of flexible chip mounted onto PCB.

The characterizations were performed on both SOI chip and flexible chip. As shown in Fig. 5.17 (a), the SOI chip was mounted onto CPG-144 package by wire bonding. After substrate transfer post-processing, presented in Chapter 4, the flexible chip was released as shown in Fig. 5.17 (b). To characterize the flexible chip, it also needed to be mounted onto special designed chip board and wire bonded to pads. The front side and back side images are shown in Fig. 5.17 (c) and (d) respectively.

5.5.1 SOI chip characterizations

Figure 5.18 plots DCR as a function of pixel location for the 32x32 SOI chip. It was measured under nominal V_E at room temperature. Row addressing signals are generated from FPGA to scan the whole array. Besides several hot spots observed, a mean value of around 35kHz under V_{OP} of 21.5V was measured, which is consistent with the results in Chapter 3. A DCR cumulative distribution over 32x32 SOI pixels was plotted in Fig. 5.19 with inset of histogram distribution.

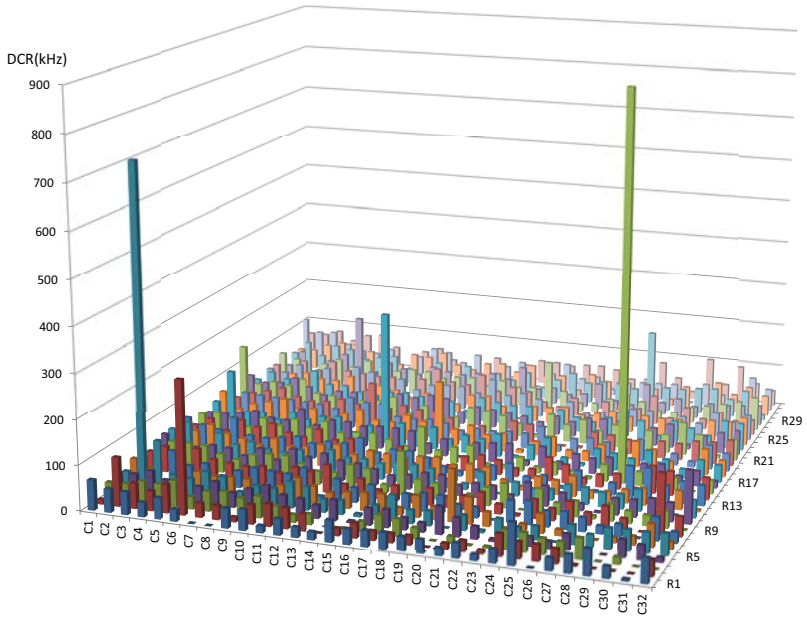


Fig. 5.18: DCR as a function of pixel location of the 32x32 SOI chip under V_{OP}=21.5V.

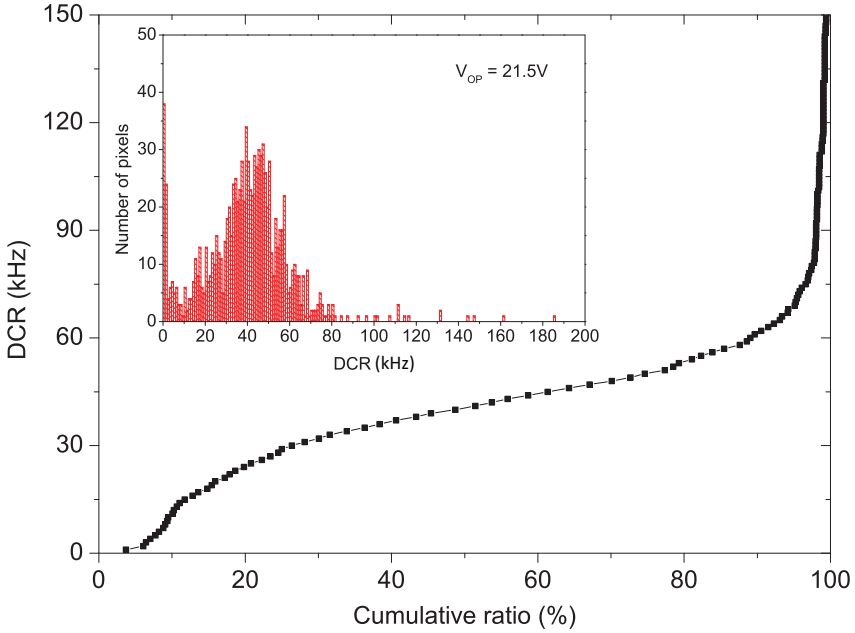


Fig. 5.19: DCR cumulative distribution over 32x32 SOI pixels under V_{OP}=21.5V (Inset: Histogram distribution).

5.5.2 Flexible chip characterizations

The flexible chip was first investigated by DCR measurements. The DCR distribution as a function of pixel location is plotted in Fig. 5.20.

The DCR measurements were performed under different V_{OP} , which ranges from 22V to 24V. The DCR value of flexible chip is higher than SOI chip at the same V_{OP}, which is consistent with results in section 2.8. A mean value of around 120kHz DCR under V_{OP} of 22V was measured and a DCR cumulative distribution over 32x32 pixels was plotted in Fig. 5.21 with inset of histogram distribution.

DCR distribution histograms over an array under different operation voltages are measured and compared as shown in Fig. 5.22. In general, the DCR distribution was modeled and fitted with a Gaussian distribution, excluding the hot spots. The mean value of DCR increases as the operation voltage increases and the FWHM value of distribution becomes larger.

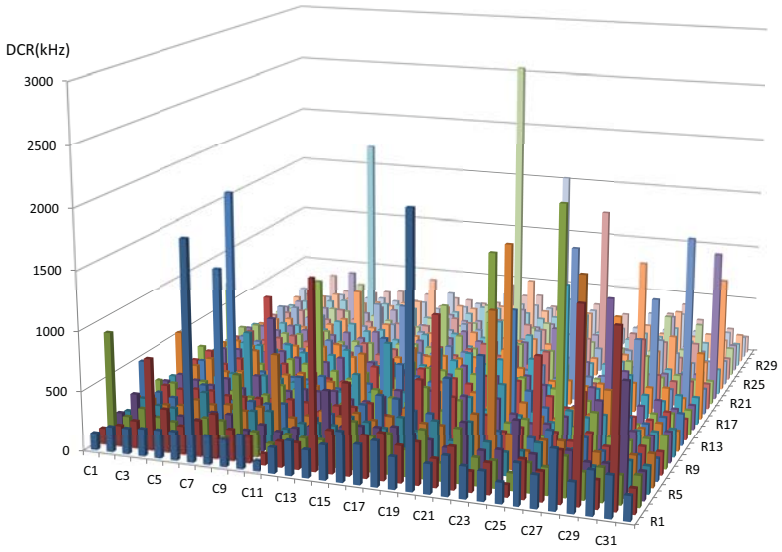


Fig. 5.20: DCR as a function of pixel location of the flexible chip under V_{OP} of 23.3V.

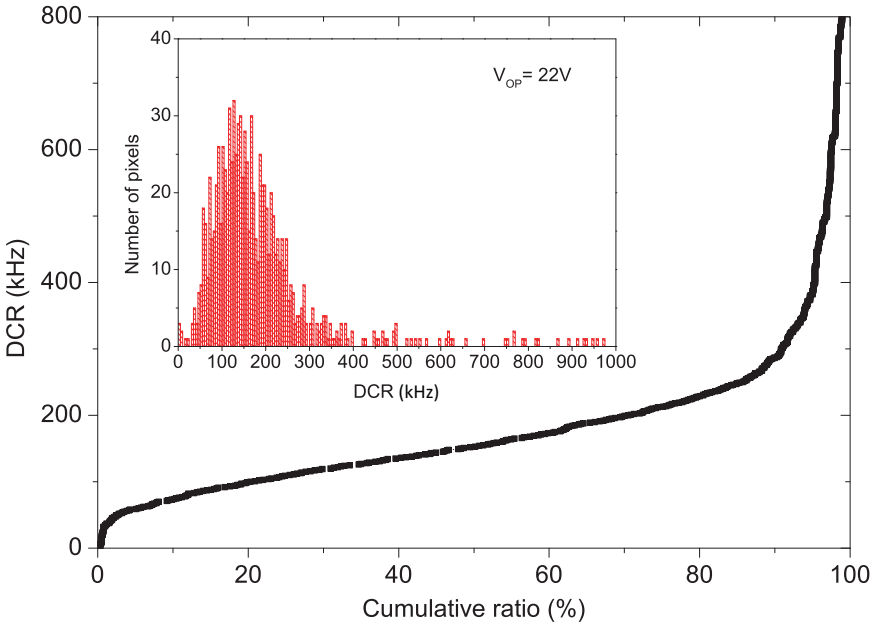


Fig. 5.21: DCR contour plot of the flexible chip under V_{OP} of 22V.

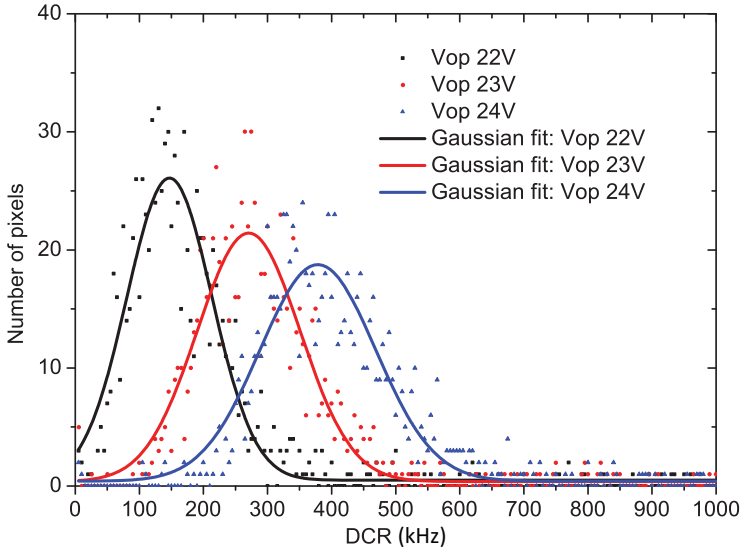


Fig. 5.22: DCR distribution histograms at different V_{OP}.

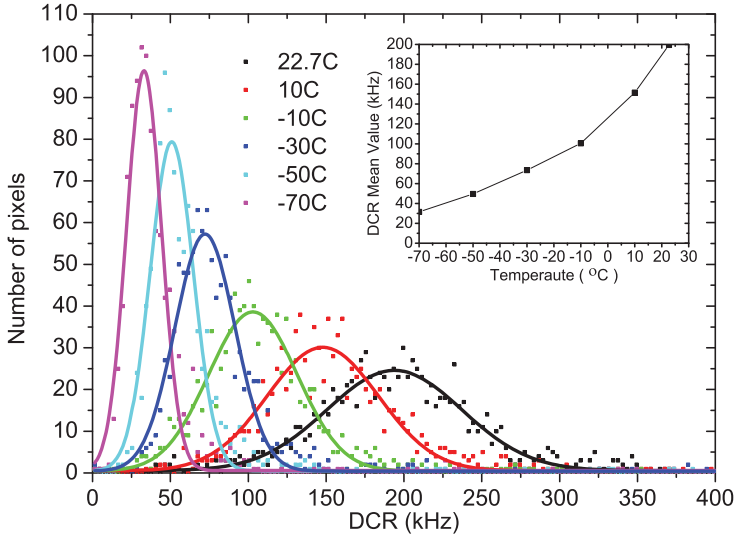


Fig. 5.23: DCR distribution histograms at different temperatures (Inset: DCR mean value at different temperatures)

By cooling down the flexible chip from 22.7ºC (room temperature) to -70ºC, mean value of DCR with V_{eb} around 4V was reduced from 200kHz to 30kHz exponentially with the temperature. FWHM of distribution histogram became smaller as the temperature went down, as shown in Fig. 5.23.

To characterize the optical sensitivity, the sample was illuminated in FSI mode with a uniform light from the integration sphere and monochromator. The light source set up is the same with Fig. 3.5. Power density of light P_{Light} could be represented as below:

$$P_{Light} = \frac{I_{ref}}{R_{ref} \cdot A_{ref}}, \tag{5.1}$$

The wavelength of the light is fixed at 500nm and the response sensitivity of the reference diode at 500nm is 260mA/W. The sensitive area of the reference diode is 33mm^2. The 32x32 flexible SPAD array was operated at V_{OP} of 22V, 23V and 24V respectively.

The mean ECR value of the flexible chip was measured under different illumination power is plotted as shown in Fig. 5.24. It scales linearly with P_{Light}. In principle, if P_{Light} keeps increasing, it would reach saturation as shown in the inset, in which f_{sat} is around 2.2 MHz according to equation $f_{sat} \cong \frac{1/e}{t_{dead}}$. However, due to the limit of our light source, the measurement is performed lower than saturation, as shown in the inset. At the same illumination power density, ECR mean value increases as V_{OP} increases from 22V to 23V, however it would approach to saturation, where V_{OP} reaches 24V in this case.

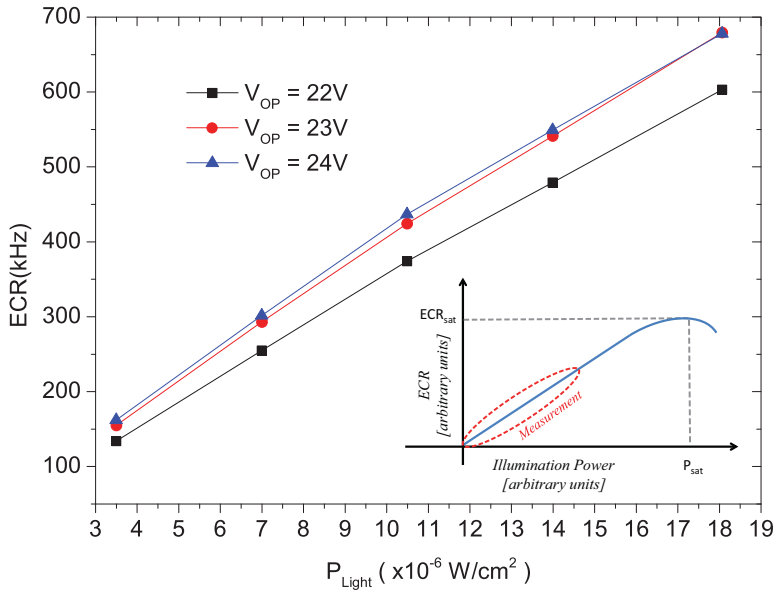


Fig. 5.24: ECR mean value under different P_{Light}.

The relative sensitivity of the flexible chip at different V_{OP} is characterized in FSI under different wavelengths, as shown in Fig. 5.25. It is described by the normalized mean ECR over photon flux. The measurement is consistent with previous PDP results in Fig. 4.13 (Chapter 4).

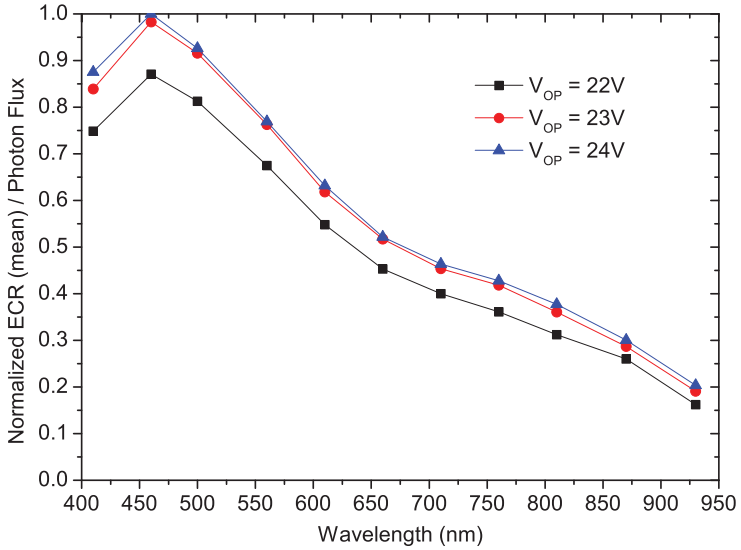


Fig. 5.25: Normalized sensitivity under different wavelength.

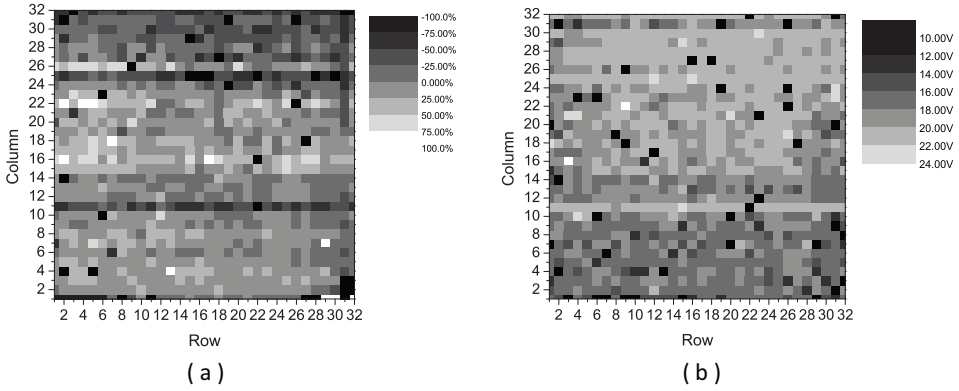


Fig. 5.26: (a) PDP non-uniformity across the sensor; (b) V_{BD} across the sensor.

A contour image of PDP non-uniformity across the 32x32 array is shown in Fig. 5.26 (a), when the sensor was illuminated with a uniform light. DCR has been measured prior to the illumination and subtracted from every pixel across

the whole chip. A contour image of V_{BD} across the chip is shown in Fig. 5.26 (b) and V_{BD} of SPAD in each pixel is extracted by "DCR fit" method [61].

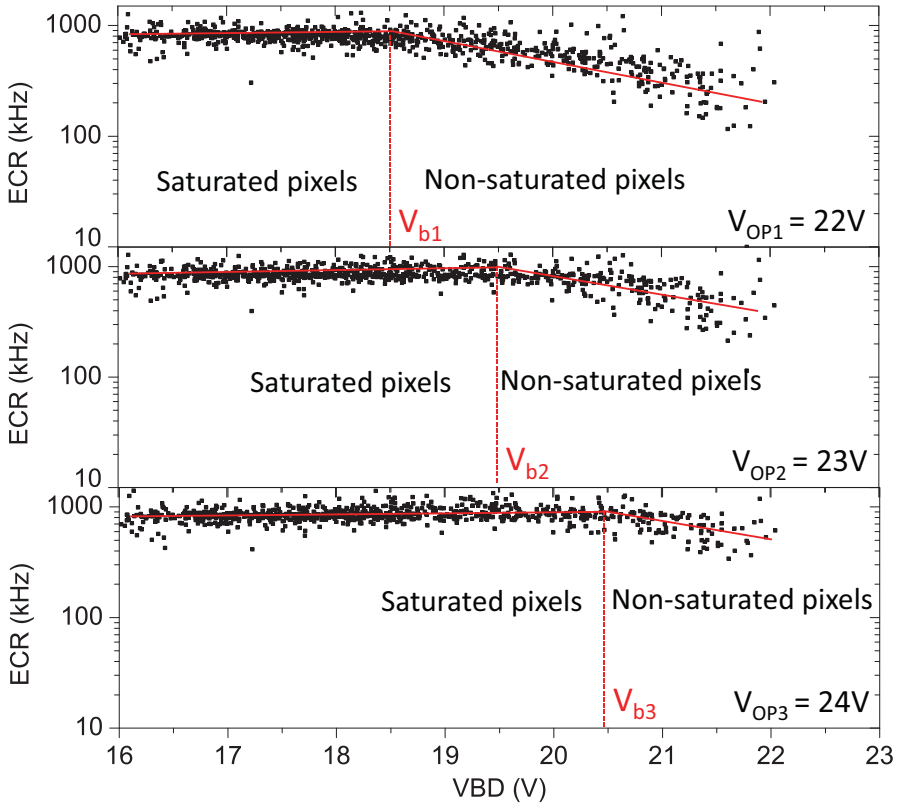


Fig. 5.27: ECR-V_{BD} distributions across the sensor under different V_{OP}.

The flexible chip was illuminated by uniform 660nm-wavelength monochromatic light flux with power of 29μW/cm^2. ECR-V_{BD} distributions across the 32x32 pixels under different V_{OP} are compared in Fig. 5.27. V_b, extracted from the cross point of fitting lines of saturated and non-saturated pixel distributions, is defined as threshold value of breakdown voltages V_{BD} between saturated pixels and non-saturated pixels. V_b decreases linearly as V_{OP} reduced. For the non-saturated pixels, ECR decreases as V_{BD} increases, which is consistent with the theory [61].

Crosstalk is negligible due to the fact that both electrical and optical isolations are suppressed in the flexible chip. Electrical isolation is ensured by the fact that SPADs and ancillary circuitry are SOI trench-isolated structures. Optical isolation is ensured by large pixel pitch, when compared with the multiplication region.

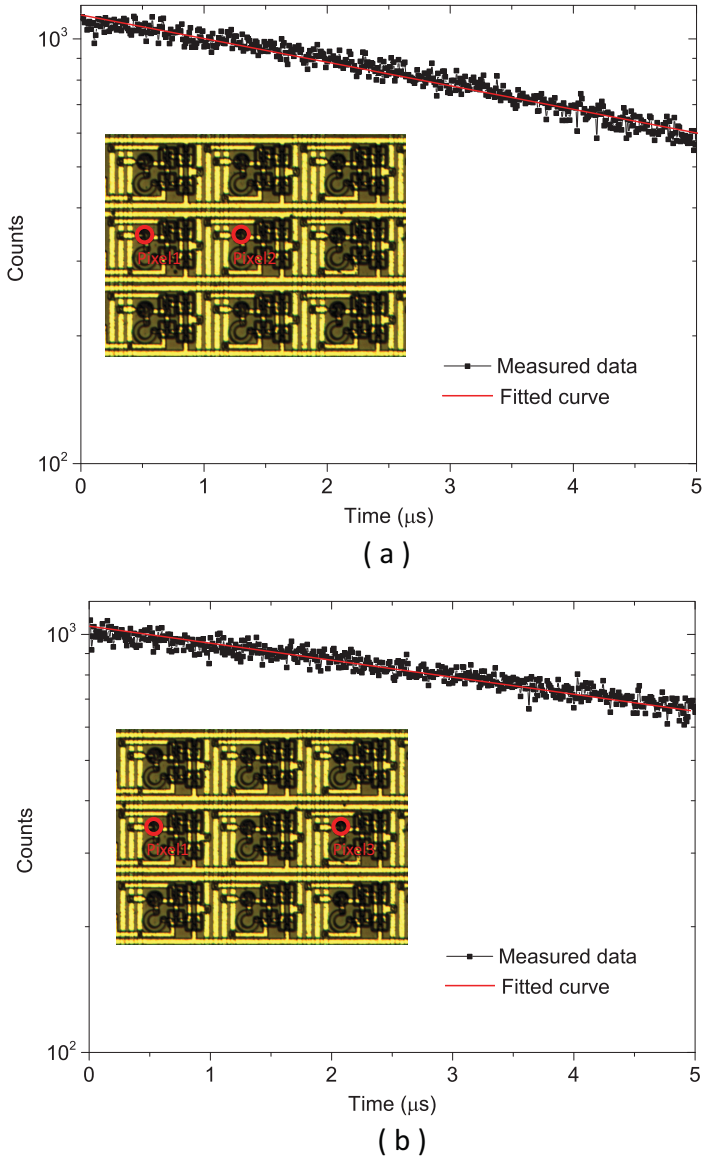


Fig. 5.28: Cross-inter-arrival time histogram measured from (a) two adjacent pixels; (b) two non-adjacent pixels.

It is further proved by gathering histogram information on inter-arrival times between the counts of two adjacent pixels, as shown in Fig. 5.28(a). The inter-arrival time histogram shows a distribution fitted well with an exponential,

proving negligible crosstalk probability. The inter-arrival times between the counts of two non-adjacent pixels also shows negligible crosstalk probability, as shown in Fig. 5.28(b).

5.6 Dual-side imaging

A compact signal processing module was developed containing chip board, USB communication board and FPGA board. Camera modules are mounted on both front side and back side. The detail setups of dual-side imaging camera are shown in Fig. 5.29. Light source is composed of surface light together with mask (TUD logo). It is mounted on a moving stage during imaging and resolution enhancement experiments.

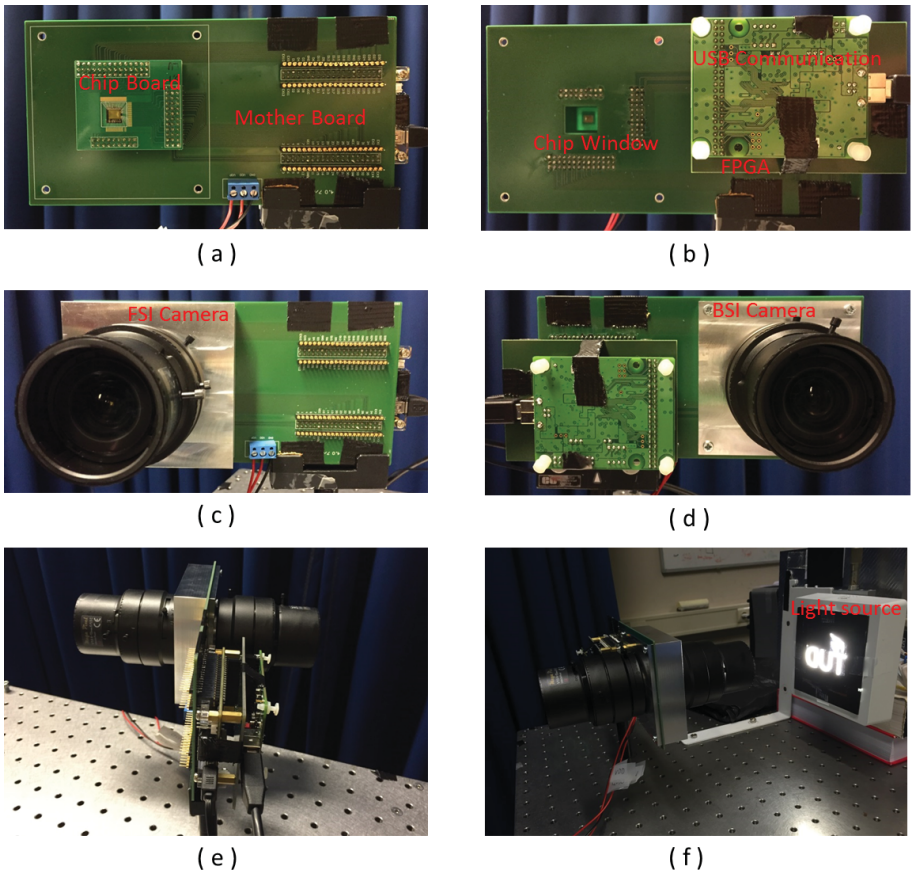
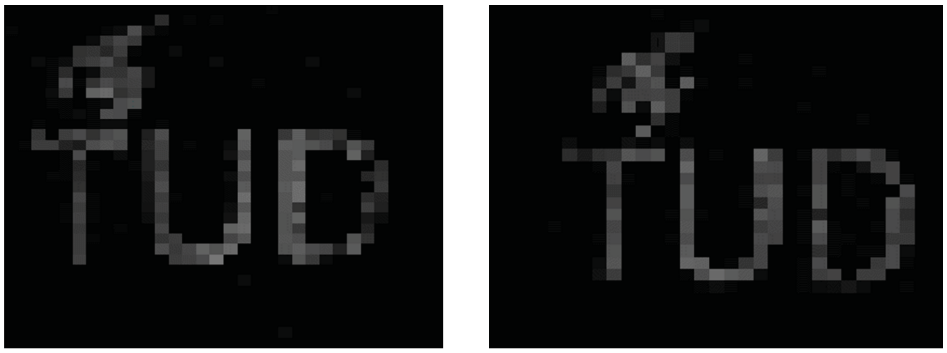


Fig. 5.29: (a) Front-side imaging signal processing module; (b) Back-side of signal processing module; (c) Front-side imaging camera module; (d) Back-side imaging camera module; (e) Dual-side imaging compact module; (f) Imaging setup.



(a) (b)

Fig. 5.30: “TUD” logo imaging on flexible CMOS SPAD image sensor (a) Front-side imaging; (b) Back-side imaging.

Fig. 5.30 shows the imaging results of FSI and BSI respectively. The 32x32 chip was scanned at frame rate of 32ms/frame and readout by 32 column buffers parallel. By optimizing the optic setup, the imaging from both sides are comparable.

5.7 Resolution enhancement

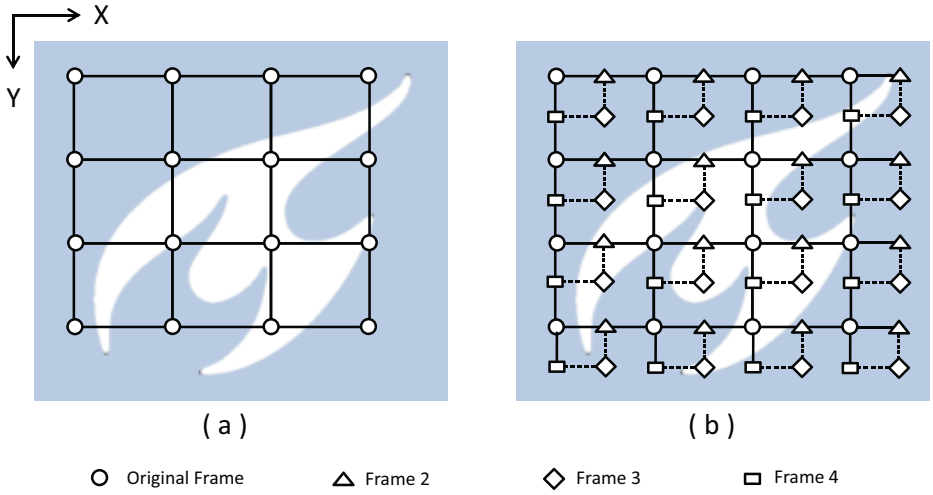


Fig. 5.31: Resolution enhancement method. Frame 2: Shifting half pixel dimension in X direction. Frame 3: Shifting half pixel dimension in both X and Y directions. Frame 4: Shifting half pixel dimension in Y direction.

As shown in Fig. 5.31, by shifting the imaging object by half pixel dimension in X, Y and both X and Y directions, multiple frames could be integrated together to get an image with resolution enhanced by two times in both X and Y directions [106].

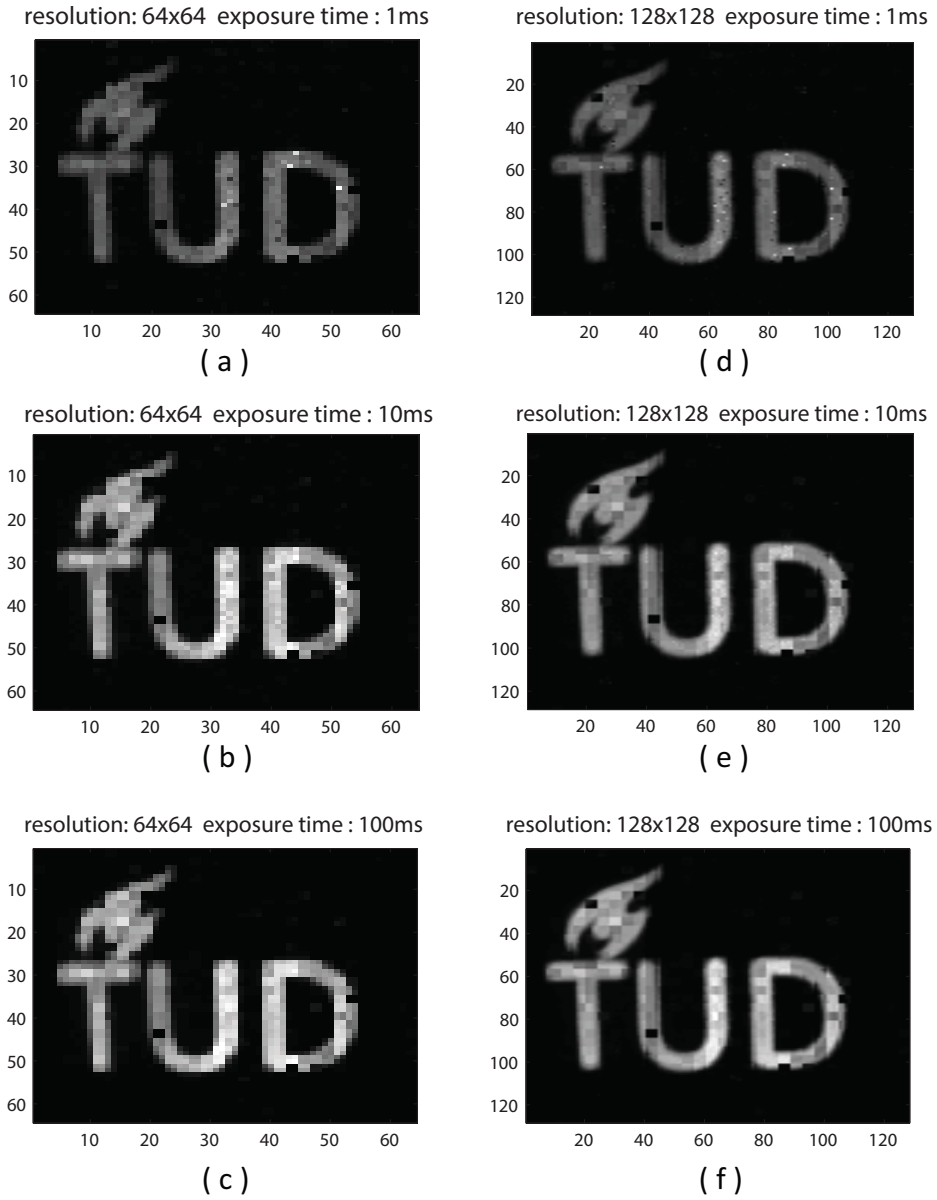


Fig. 5.32: Comparing of imaging based on different resolution and exposure time.

This method is applied in the imaging experiments with flexible CMOS SPAD sensor chip and different resolutions such as 64x64 and 128x128 imaging results are achieved, which is shown in Fig. 5.32. Imaging results based on different exposure time are also compared. Higher resolution and longer exposure time enable higher contrast and imaging quality.

The electrical and optical performance are summarized in Tab. 5.1.

Tab. 5.1: Performance summary of flexible CMOS SPAD image sensor chip.

Item	Min.	Typ.	Max.	Unit	Comments
Resolution	32x32		128x128		
Row rate		1k		fps	
Frame rate		30		fps	
Digital supply voltage		3.3		V	
SPAD V_{OP}	22		25	V	
Power dissipation			30	mW	Entire flexible chip
DCR		120		kHz	V_{OP} = 22V
DCR temperature dependence		1.5X/20°C			
V_{BD} non-uniformity		11%			
PDP non-uniformity		22.1%			@560nm; V_{OP} = 23V
PRNU		19.7%			
Signal-Noise Ratio		36		dB	
Dynamic range		25		dB	Estimated up to 52dB by active quenching at low temperature
Detectable light intensity	3.5			$\mu W/cm^2$	

5.8 Conclusions

In this chapter, the design and fabrication of flexible 32x32 CMOS SPAD image sensor was presented. Both of the SOI chip and flexible chip were demonstrated successfully for the first time. The measurements are consistent

with the previous research results in chapter 4. The statistical distribution of DCR and PDP shows reasonable uniformity, which mainly depends on the fabrication yield in the lab. SOI chip shows better uniformity than flexible chip. The flexible chip shows consistent optical sensitivity. It also shows negligible electrical and optical crosstalk thanks to good electrical and optical isolations.

It is achieved for the first time that dual-side photon counting imaging by flexible SPAD image sensor. The imaging quality is even improved by resolution enhancements.

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 bended imaging sensors.

Chapter 6

Conclusions and recommendations

6.1 Conclusions

The primary goal of this thesis was to explore and develop a novel single-photon avalanche diode technology in applications of flexible and implantable biomedical electronics for which we defined 4 contributions. In order to meet the requirements, focus was placed on the investigation of the potential features of the system, such as ultra-thin body, frontside and backside illumination modes, high fill factor and sensitivity, flexible substrate post-processing and CMOS integration compatibility. The challenges of this investigation have been successfully addressed. The world's first flexible ultra-thin-body SPAD was demonstrated and extended to flexible 32x32 CMOS SPAD image sensor level. The state of the art in SPAD technology, as well as in flexible biomedical sensors has been considerably enhanced. The main conclusions of this thesis are summarized as follow:

- The first ultra-thin body SOI SPAD was demonstrated by epitaxy and implantation technologies. The epitaxy silicon body thickness was designed to be around 1.5μm. The body doping and junction profile were carefully optimized to avoid PEB while virtual guard ring region was fully depleted. The performance of the proposed device compares favorably with that of conventional CMOS SPADs, while The minimum DCR was less than 20kHz with negligible afterpulsing. The sensitive spectrum was much wider and PDP, especially at long wavelengths, was enhanced, which is the feature having buried oxide in SOI. Thanks to the buried oxide, we paved the road for the flexible substrate transfer post-processing.

- By using flexible substrate transfer post-processing, the flexible SPAD with dual side illumination was successfully demonstrated and reported for the first time, to the best of our knowledge. DCR, PDP, afterpulsing, and time jitter performance of the device on flexible substrate is consistent with that of a device before transfer. It can operate both front- and backside illuminations while the peak PDP can reach 11% in FSI mode and 6% in BSI mode. Furthermore, fill factor could be improved significantly, due to a wider carrier collection region and the absence of blocking metal layers in BSI mode. It could be even improved by designing larger active-area SPADs and optimizing the doping profiles such as the ratio with guard ring area. Such a dual side illumination SPAD provides a novel methodology to overcome the limits of BSI applications in CMOS technology, simultaneously enhancing the fill factor while pixel pitch keeps scaling down. On the other hand, BSI SPADs have higher jitter; this behavior could be due not only to the combination of collection times from the depletion region and neutral region, but also to the lateral collection of carriers since the fill factor is quite large in BSI mode. Superficial charge generation also contributes to the variation of the collection time due to the fully depleted region at the box oxide surface.

- Geiger-mode performance of flexible SPAD was thoroughly characterized. Experimental results show that 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, was further reduced, resulting in a DCR density of tens to hundreds of Hertz per micrometer square at cryogenic temperatures. The influence of the trench etching process onto DCR is also proved by comparison with planar ultrathin-body SPAD structures without trench. By integrating quenching resistor, excess voltage and dynamic range was enhanced. Relatively high PDP can be achieved by wider depletion and drift regions and by carefully optimizing body thickness.

- The first flexible fully integrated CMOS SPAD sensor pixel was demonstrated successfully by integrating a CMOS process to flexible SPAD fabrication. 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 front- and backside illumination respectively. Afterpulsing and crosstalk are negligible (less than 0.1% at nominal dead time larger than 1μs). Thanks to the high-thermal-budget CMOS process before

SPAD junction implantation, P-type doping (boron) concentration in top silicon area (originally intrinsic) got increased, enhancing the electric field across the neutral region near the buried oxide, which is positive for PDP in BSI, especially at short wavelength. By enabling the CMOS integration compatibility of flexible SPAD technology, it could pave the way to realize larger scale sensor array integrated with more electronics on flexible substrate, enabling CMOS circuit monolithic integration and large-scale SPAD array with readout circuit.

- In order to implement large arrays of SPADs based on the demonstrated flexible CMOS SPAD technology, area-optimized front-end and readout circuits were designed. Each pixel contains a SPAD, a quenching resistor, inverter buffer and transmission-gate switch circuitry and all the devices are isolated by trenches. Complex flowchart was developed, including epitaxy, front-end process, back-end process, and flexible substrate transfer post-processing. The world's flexible 32x32 CMOS SPAD image sensor was successfully implemented. Regarding the question about the statistical distribution of performance, DCR distribution and PDP uniformity were characterized through the image readout and signal processing system based on FPGA. The Flexible CMOS SPAD sensor can be used in anti-vignetting image sensors, bio-inspired composite-eye or flexible multi-aperture cameras, and wherever the sensor plane must follow a certain curvature in advanced implantable biological and future life sciences research.

6.2 Recommendations

This thesis demonstrated several breakthroughs in flexible CMOS SPAD sensor technology. While some innovative progress achieved in our laboratory, more research from different fields are needed to push this technology into industry application level.

- SOI provides a good solution to transfer ultrathin-body SPAD to flexible substrate. This ultrathin-body SPAD suffers from relatively high DCR. The traps at the interface between silicon and BOX layer might be a source of dark noise. A further improvement can be introduced with a buried layer at the backside interface or some other material to solve the superficial carrier generation and charge collection issues [107].

- 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 between different structures (trench-isolated SPAD and planar SPAD) and different temperatures (RT and cryogenic temperature). Passivation technology [108] to filter out the defects around the trenches needs to be investigated in future work.

- The microlenses are currently being processed by CSEM SA. For the 32x32 array we are currently processing the chip, each pixel contains a SPAD, quenching resistor, CMOS inverter, and CMOS transmission gate as switch. The pixel size is 162µm x 125µm. The typical diameter of the SPAD active region is 4µm. Without microlens, the fill-factor is less than 1%. The current design of microlens has a 60 µm diameter. The fill factor will be enhanced to more than 10% using the microlens array. The same process of microlenses by CSEM SA has been been successfully been integrated to other SPAD chip [109], which would be a useful preliminary reference for future work.

- Substrate transfer process could be improved to double-polyimide process. In principle, there is a stress neutral plain in a certain depth of the film. By using the sandwiched structure of double-polyimide process, the mechanical stress neutral plain could be shifted to device layer [110]. However, for this thesis, the bottom polyimide layer could only be evaporated in a special way to the device layer without causing any stress.

- In the application of implantable retinal prosthesis and biomedical monitor, electrodes could be integrated with the flexible SPAD array to stimulate human nerves or for monitoring applications in the body. Thanks to the dual-side illumination of the flexible CMOS SPAD sensor array, the electrode could be placed at any side of the flexible chip. The material could be CMOS compatible metal or some novel multi-scale material for heterogeneous integration [111].

- Fill factor could be further improved by 3D integration. µ-Czochralski single grain crystallization technology [112] provides a good solution to form crystalline silicon or germanium locally with low thermal budget. CMOS transistor could be built up based on crystalline silicon area. For single pixel of SPAD array, the transistors could be further scaled down and stacked on top of ultrathin-body SPAD.

Bibliography

[1] A. Webb and G. C. Kagadis, "Introduction to Biomedical Imaging," *Med. Phys.*, vol. 30, no. 8, p. 2267, 2003.

[2] L. V. Wang and Hsin-I Wu, *Biomedical Optics: Principles and Imaging*. 2012.

[3] A. F. Laine, "In the Spotlight: Biomedical Imaging.," *IEEE reviews in biomedical engineering*, vol. 1. pp. 4–7, 2008.

[4] V. Acikel and E. Atalar, "Biomedical Imaging," *Biomed. Imaging*, pp. 186–213, 2014.

[5] S. Chua and A. Groves, *Biomedical Imaging*. 2014.

[6] Z. Gorocs and A. Ozcan, "On-chip biomedical imaging," *IEEE Rev. Biomed. Eng.*, vol. 6, pp. 29–46, 2013.

[7] C. M. Tempany and B. J. McNeil, "Advances in biomedical imaging.," *JAMA*, vol. 285, no. 5, pp. 562–567, 2001.

[8] W. Becker, A. Bergmann, G. Biscotti, and A. Rueck, "Advanced time-correlated single photon counting technique for spectroscopy and imaging of biological systems," in *Proc. SPIE, Commercial and Biomedical Applications of Ultrafast Lasers IV*, 2004, vol. 5340, pp. 1–9.

[9] N. Bertone, M. Wabuyele, A. Team, H. Team, and M. Davies, "Single Photon Counting with a focus On Biomedical Applications," in *Proc. of Pittcon*, 2003, pp. 1–18.

[10] S. Lebid, R. O'Neill, C. Markham, T. Ward, and S. Coyle, "Functional Brain Signals: A photon counting system for brain activity monitoring," in *IEEE Conference Proceedings of the Irish Signals and Systems Conference*, 2004, pp. 469–474.

[11] J. Ohta, "Implantable CMOS imaging devices for bio-medical applications," in *IEEE 54th International Midwest Symposium on Circuits and Systems*, 2011, pp. 1–4.

[12] J. Ohta, T. Tokuda, K. Sasagawa, and T. Noda, "Implantable CMOS Biomedical Devices," *Sensors*, vol. 9, no. 11, pp. 9073–9093, 2009.

[13] M. S. Humayun, J. D. Weiland, G. Chader, and E. Greenbaum, *Artificial Sight*. 2007.

[14] 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*, vol. 5, no. 11, pp. 3859–70, 2014.

[15] T. Noda, K. Sasagawa, and T. Tokuda, "Fabrication of Fork-Shaped Retinal Stimulator Integrated with CMOS Microchips for Extension of Viewing Angle," *Sensors Mater.*, vol. 26, no. 8, pp. 637–648, 2014.

[16] 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. il Kim, C. Yee, J. A. Rogers, and K. M. Lee, "Immunologic and tissue biocompatibility of flexible/stretchable electronics and optoelectronics," *Adv. Healthc. Mater.*, vol. 3, no. 4, pp. 515–525, 2014.

[17] F. Rieke and D. Baylor, "Single-photon detection by rod cells of the retina," *Rev. Mod. Phys.*, vol. 70, no. 3, pp. 1027–1036, 1998.

[18] M. S. Humayun, E. de Juan, G. Dagnelie, R. J. Greenberg, R. H. Propst, and D. H. Phillips, "Visual perception elicited by electrical stimulation of retina in blind humans.," *Arch. Ophthalmol.*, vol. 114, no. 1, pp. 40–46, 1996.

[19] Y. H.-L. Luo and L. da Cruz, "The Argus® II Retinal Prosthesis System," *Prog. Retin. Eye Res.*, vol. 50, pp. 89–107, 2015.

[20] M. N. Mohd Zain, A. Musa, M. H. Hisham, A. R. Laili, and Z. M. Yusof, "Photon counting polarimetry measurement towards non-invasive biomedical glucose monitoring," *2014 IEEE 5th Int. Conf. Photonics*, pp. 156–158, 2014.

[21] D. Palubiak, M. M. El-Desouki, O. Marinov, M. J. Deen, and Q. Fang, "High-speed, single-photon avalanche-photodiode imager for biomedical applications," *IEEE Sens. J.*, vol. 11, no. 10, pp. 2401–2412, 2011.

[22] M. Wolf, M. Ferrari, and V. Quaresima, "Progress of near-infrared spectroscopy and topography for brain and muscle clinical applications.," *J. Biomed. Opt.*, vol. 12, no. 6, p. 062104, 2007.

[23] B. Shadgan, M. Fareghi, L. Stothers, A. Macnab, and A. M. Kajbafzadeh, "Diagnosis of testicular torsion using near infrared spectroscopy : A novel diagnostic approach," vol. 8, no. 3–4, pp. 249–252, 2014.

[24] J. A. Rogers, "Wearable Electronic Sensors Can Now Be Printed Directly on the Skin," *MIT Technology Review*, 2013.

[25] J. Yoon, S. M. Lee, D. Kang, M. A. Meitl, C. A. Bower, and J. A. Rogers, "Heterogeneously Integrated Optoelectronic Devices Enabled by Micro-Transfer Printing," *Adv. Opt. Mater.*, vol. 3, no. 10, pp. 1313–1335, 2015.

[26] 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 back-illuminated silicon Geiger-mode avalanche photodiode arrays," in *SPIE Defense, Security, and Sensing*, 2010, p. 76810P–76810P–7.

[27] H. Tian, B. Fowler, and A. El Gamal, "Analysis of temporal noise in CMOS photodiode active pixel sensor," *IEEE J. Solid-State Circuits*, vol. 36, no. 1, pp. 92–101, 2001.

[28] G. Bonanno, M. Belluso, A. Calì, R. Cosentino, A. Modica, and S. Scuderi, "A new photon counting detector: intensified CMOS-APS," in *Scientific Detectors for Astronomy*, 2004, pp. 21–28.

[29] K. B. W. Harpsøe, M. I. Andersen, and P. Kjægaard, "Bayesian photon counting with electron-multiplying charge coupled devices (EMCCDs)," *Astron. Astrophys.*, vol. 537, p. A50, 2012.

[30] A. S. Huntington, M. A. Compton, and G. M. Williams, "Linear-mode single-photon APD detectors," *Proc. SPIE Adv. Phot. Count. Tech. II*, vol. 6771, p. 67710Q, 2007.

[31] Wikipedia, "Single-Photon Avalanche Diode," *Wikipedia*, 2012. [Online]. Available: http://en.wikipedia.org/wiki/Single-photon_avalanche_diode.

[32] C. Miyazaki, H. Shimamoto, T. Uematsu, and Y. Abe, "Development of wafer thinning and dicing technology for thin wafer," in *IEEE International Conference on 3D System Integration, 3DIC*, 2009.

[33] Y. S. Kim, N. Maeda, H. Kitada, K. Fujimoto, S. Kodama, A. Kawai, K. Arai, K. Suzuki, T. Nakamura, and T. Ohba, "Advanced wafer thinning technology and feasibility test for 3D integration," *Microelectron. Eng.*, vol. 107, pp. 65–71, 2013.

[34] H. Kwok, M. Wong, Z. Meng, and S. Zhao, "Polycrystalline silicon thin film transistors with bridged-grain structures," US 8,426,865 B2, 2013.

[35] N. Münzenrieder, C. Zysset, L. Petti, T. Kinkeldei, G. A. Salvatore, and G. Tröster, "Flexible double gate a-IGZO TFT fabricated on free standing polyimide foil," *Solid. State. Electron.*, vol. 84, pp. 198–204, 2013.

[36] G. Pace, A. Grimoldi, D. Natali, M. Sampietro, J. E. Coughlin, G. C. Bazan, and M. Caironi, "All-organic and fully-printed semitransparent photodetectors based on narrow bandgap conjugated molecules," *Adv. Mater.*, vol. 26, no. 39, pp. 6773–6777, 2014.

[37] A. Arslan, R. Ishiahara, and C. I. M. Beenakker, "Single grain TFTs and lateral photodiodes for large area X-ray detection," in *Proc. of IISW*, 2011, pp. 301–304.

[38] A. Pierre, I. Deckman, P. B. Lechene, and A. C. Arias, "High Detectivity All-Printed Organic Photodiodes," *Adv. Mater.*, vol. 27, no. 41, pp. 6411–6417, 2015.

[39] D.-H. Kim, J. Song, W. M. Choi, H.-S. Kim, R.-H. Kim, Z. Liu, Y. Y. Huang, K.-C. Hwang, Y. Zhang, and J. a Rogers, "Materials and noncoplanar mesh designs for integrated circuits with linear elastic responses to extreme mechanical deformations.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 48, pp. 18675–18680, 2008.

[40] S. W. Hong, T. Banks, and J. A. Rogers, "Improved density in aligned arrays of single-walled carbon nanotubes by sequential chemical vapor

deposition on quartz," *Adv. Mater.*, vol. 22, no. 16, pp. 1826–1830, 2010.

[41] A. K. Geim and K. S. Novoselov, "The rise of graphene.," *Nat. Mater.*, pp. 183–191, 2007.

[42] Y. M. Song, Y. Xie, V. Malyarchuk, J. Xiao, I. Jung, K. J. Choi, Z. Liu, H. Park, C. Lu, R. H. Kim, R. Li, K. B. Crozier, Y. Huang, and J. A. Rogers, "Digital cameras with designs inspired by the arthropod eye," *Nature*, vol. 497, no. 7447, pp. 95–99, 2013.

[43] D. H. Kim, Y. S. Kim, J. Wu, Z. Liu, J. Song, H. S. Kim, Y. Y. Huang, K. C. Hwang, and J. A. Rogers, "Ultrathin silicon circuits with strain-isolation layers and mesh layouts for high-performance electronics on fabric, vinyl, leather, and paper," *Adv. Mater.*, vol. 21, no. 36, 2009.

[44] B. Mimoun, V. Henneken, A. Van Der Horst, and R. Dekker, "Flex-to-rigid (F2R): A generic platform for the fabrication and assembly of flexible sensors for minimally invasive instruments," *IEEE Sens. J.*, vol. 13, no. 10, pp. 3873–3882, 2013.

[45] R. L. Chaney, D. R. Hackler, D. G. Wilson, and B. N. Meek, "Physically Flexible High Performance Single Crystal CMOS Integrated with Printed Electronics," 2014.

[46] P. Sun, B. Mimoun, E. Charbon, and R. Ishihara, "A Flexible Ultra-Thin-Body SOI Single-Photon Avalanche Diode," *Int. Electron Devices Meet.*, vol. 11, no. 1, pp. 284–287, 2013.

[47] P. Sun, E. Charbon, and R. Ishihara, "A Flexible Ultrathin-Body Single-Photon Avalanche Diode With Dual-Side Illumination," *IEEE J. Sel. Top. Quantum Electron.*, vol. 20, no. 6, pp. 276–283, 2014.

[48] E. Charbon, "Single-photon imaging in complementary metal oxide semiconductor processes.," *Philos. Trans. A. Math. Phys. Eng. Sci.*, vol. 372, no. 2012, p. 20130100, 2014.

[49] R. J. Walker, J. A. Richardson, and R. K. Henderson, "A 128x96 pixel event-driven phase-domain ΔΣ-based fully digital 3D camera in 0.13μm CMOS imaging technology," in *Digest of Technical Papers - IEEE International Solid-State Circuits Conference*, 2011, pp. 410–411.

[50] A. Gulinatti, I. Rech, P. Maccagnani, M. Ghioni, and S. Cova, "Improving the performance of silicon single-photon avalanche diodes," *Proc. SPIE 8033, Adv. Phot. Count. Tech. V*, vol. 8033, pp. 803302–803302–11, 2011.

[51] C. Niclass, M. Soga, H. Matsubara, and S. Kato, "A 100m-range 10-frame/s 340x96-pixel time-of-flight depth sensor in 0.18μm CMOS," *Eur. Solid-State Circuits Conf.*, pp. 107–110, 2011.

[52] E. Charbon, H. J. Yoon, and Y. Maruyama, "A Geiger mode APD fabricated in standard 65nm CMOS technology," in *Technical Digest - International Electron Devices Meeting, IEDM*, 2013.

[53] S. G. Wuu, C. C. Wang, B. C. Hseih, Y. L. Tu, C. H. Tseng, T. H. Hsu,

R. S. Hsiao, S. Takahashi, R. J. Lin, C. S. Tsai, Y. P. Chao, K. Y. Chou, P. S. Chou, H. Y. Tu, F. L. Hsueh, and L. Tran, "A leading-edge 0.9μm pixel CMOS image sensor technology with backside illumination: Future challenges for pixel scaling," in *Technical Digest - International Electron Devices Meeting, IEDM*, 2010.

[54] D. N. Yaung, B. C. Hsieh, C. C. Wang, J. C. Liu, T. J. Wang, W. D. Wang, C. C. Chuang, C. Chao, Y. L. Tu, C. S. Tsai, T. L. Hsu, F. Ramberg, W. P. Mo, H. Rhodes, D. Tai, V. C. Venezia, and S. G. Wuu, "High performance 300mm backside illumination technology for continuous pixel shrinkage," in *Technical Digest - International Electron Devices Meeting, IEDM*, 2011.

[55] M. S. Oh, H. K. Kong, H. S. Lee, K. Il Kim, K. H. Bae, S. B. Kim, S. K. Kim, M. S. Lim, J. C. Ahn, T. C. Kim, G. Hiroshige, S. H. Kim, D. K. Min, and Y. J. Lee, "Backside-illumination 14μm-pixel QVGA time-of-flight CMOS imager," in *2012 IEEE 10th International New Circuits and Systems Conference, NEWCAS 2012*, 2012, pp. 325–328.

[56] N. Teranishi, H. Watanabe, T. Ueda, and N. Sengoku, "Evolution of optical structure in image sensors," in *Technical Digest - International Electron Devices Meeting, IEDM*, 2012.

[57] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*. 2007.

[58] P. Seitz and A. J. P. Theuwissen, *Single-Photon Imaging*. 2011.

[59] F. Zappa, S. Tisa, A. Tosi, and S. Cova, "Principles and features of single-photon avalanche diode arrays," *Sensors Actuators, A Phys.*, vol. 140, no. 1, pp. 103–112, 2007.

[60] A. Gallivanoni, I. Rech, and M. Ghioni, "Progess in Quenching Circuits for Single Photon Avalanche diodes," *IEEE Transations Nucl. Sci.*, vol. 57, no. 6, pp. 3815–3826, 2010.

[61] M. W. Fishburn, "Fundamentals of CMOS Single-Photon Avalanche Diodes," Delft University of Technology, 2012.

[62] P. Webb and A. . Jones, "Large Area Reach-Through Avalanche Diodes for Radiation Monitoring," *IEEE Transations Nucl. Sci.*, vol. 21, no. 1, pp. 151–158, 1974.

[63] R. Mclntyre and P. Webb, "Low-noise, reach-through avalanche photodiodes," US5583352 A, 1996.

[64] S. Cova, A. Longoni, and A. Andreoni, "Towards picosecond resolution with single-photon avalanche diodes," *Rev. Sci. Instrum.*, vol. 52, no. 3, pp. 408–412, 1981.

[65] C. Niclass, M. Gersbach, R. Henderson, L. Grant, and E. Charbon, "A single photon avalanche diode implemented in 130-nm CMOS technology," *IEEE J. Sel. Top. Quantum Electron.*, vol. 13, no. 4, pp. 863–869, 2007.

[66] H. Finkelstein, M. J. Hsu, and S. C. Esener, "STI-bounded single-photon

avalanche diode in a deep-submicrometer CMOS technology," *IEEE Electron Device Lett.*, vol. 27, no. 11, pp. 887–889, 2006.

[67] M. Gersbach, J. Richardson, E. Mazaleyrat, S. Hardillier, C. Niclass, R. Henderson, L. Grant, and E. Charbon, "A low-noise single-photon detector implemented in a 130 nm CMOS imaging process," *Solid. State. Electron.*, vol. 53, no. 7, pp. 803–808, 2009.

[68] M. Ghioni, S. Cova, a. Lacaita, and G. Ripamonti, "New silicon epitaxial avalanche diode for single-photon timing at room temperature," *Electron. Lett.*, vol. 24, no. 24, p. 1476, 1988.

[69] M. A. Karami, M. Gersbach, H.-J. Yoon, and E. Charbon, "A new single-photon avalanche diode in 90nm standard CMOS technology.," *Opt. Express*, vol. 18, no. 21, pp. 22158–22166, 2010.

[70] R. K. Henderson, E. A. G. Webster, R. Walker, J. A. Richardson, and L. A. Grant, "A 3x3, 5µm pitch, 3-transistor single photon avalanche diode array with integrated 11V bias generation in 90nm CMOS technology," in *Technical Digest - International Electron Devices Meeting, IEDM*, 2010.

[71] M. Ghioni, G. Armellini, P. MacCagnani, I. Rech, M. K. Emsley, and M. Selim Ünlü, "Resonant-cavity-enhanced single-photon avalanche diodes on reflecting silicon substrates," *IEEE Photonics Technol. Lett.*, vol. 20, no. 6, pp. 413–415, 2008.

[72] M. Ghioni, G. Armellini, P. Maccagnani, I. Rech, M. K. Emsley, and M. S. Ünlü, "Resonant-cavity-enhanced single photon avalanche diodes on double silicon-on-insulator substrates," *J. Mod. Opt.*, vol. 56, no. 2–3, pp. 309–316, 2009.

[73] S. Cova, A. Lacaita, and G. Ripamonti, "Trapping Phenomena in Avalanche Photodiodes on Nanosecond Scale," *IEEE Electron Device Lett.*, vol. 12, no. 12, pp. 685–687, 1991.

[74] H. Jansen, H. Gardeniers, M. De Boer, M. Elwenspoek, and J. Fluitman, "A survey on the reactive ion etching of silicon in microtechnology," *J. Micromech. Microeng.*, vol. 6, pp. 14–28, 1996.

[75] K. R. Williams, K. Gupta, and M. Wasilik, "Etch rates for micromachining processing - Part II," *J. Microelectromechanical Syst.*, vol. 12, no. 6, pp. 761–778, 2003.

[76] G. A. M. Hurkx, D. B. M. Klaassen, and M. P. G. Knuvers, "A New Recombination Model for Device Simulation Including Tunneling," *IEEE Trans. Electron Devices*, vol. 39, no. 2, pp. 331–338, 1992.

[77] W. Shockley and W. T. Read, "Statistics of the Recombination of Holes and Electrons," *Phys. Rev.*, vol. 87, no. 46, pp. 835–842, 1952.

[78] S. Nikzad, M . Hoenk, A. Jewell, T. Jones, and R. Mcmillan, "Curved focal plane arrays," in *IEEE Workshop CCDs Adv. Imag. Sens.*, 2003.

[79] S. Nikzad, T. J. Cunningham, M. E. Hoenk, R. P. Ruiz, D. M. Soules,

and S. E. Holland, “Direct detection of 0.1-20 keV electrons with delta doped, fully depleted, high purity silicon p-i-n diode arrays,” *Appl. Phys. Lett.*, vol. 89, no. 18, 2006.

[80] M. 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*, vol. 23, no. 10, pp. 13200–13209, 2015.

[81] P. Sun, R. Ishihara, and E. Charbon, “Flexible Ultrathin-Body Single-Photon Avalanche Diode Sensors and CMOS Integration,” *Opt. Express*, vol. 24, no. 4, pp. 3734–3748, 2016.

[82] B. N. Brockhouse, “Lattice Vibrations in Silicon and Germanium,” *Phys. Rev. Lett.*, vol. 2, no. 6, pp. 256–258, 1959.

[83] S. Mandai, “Multichannel Digital Silicon Photomultipliers for Time-of-Flight PET,” Delft University of Technology, 2014.

[84] M. A. Karami, “Deep-submicron CMOS Single Photon Detectors and Quantum Effects,” Delft University of Technology, 2011.

[85] A. Spinelli, M. A. Ghioni, S. D. Cova, and L. M. Davis, “Avalanche detector with ultraclean response for time-resolved photon counting,” *IEEE J. Quantum Electron.*, vol. 34, no. 5, pp. 817–821, 1998.

[86] C. Y. Chang and S. S. Chiu, “Temperature Dependence of Breakdown Voltage in Silicon Abrupt P-N Junctions,” *IEEE Trans. Electron Devices*, vol. 18, no. 6, pp. 391–393, 1971.

[87] G. Collazuol, M. G. Bisogni, S. Marcatili, C. Piemonte, and A. Del Guerra, “Studies of silicon photomultipliers at cryogenic temperatures,” in *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2011, vol. 628, no. 1, pp. 389–392.

[88] N. Serra, G. Giacomini, A. Piazza, C. Piemonte, and A. Del Guerra, “Experimental and TCAD Study of Breakdown Voltage Temperature Behavior in n+p SiPMs,” *IEEE Trans. Nucl. Sci.*, vol. 58, no. 3, pp. 1233–1240, 2011.

[89] 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.*, vol. 78, no. 6, 2007.

[90] S. Sze and G. Gibbons, “Effect of junction curvature on breakdown voltage in semiconductors,” *Solid. State. Electron.*, vol. 9, no. 9, pp. 831–845, 1966.

[91] A. Spinelli and A. L. Lacaita, “Physics and numerical simulation of single photon avalanche diodes,” *IEEE Trans. Electron Devices*, vol. 44, no. 11, pp. 1931–1943, 1997.

[92] 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, & Other Applications,” *IEEE*

Int. Electron Devices Meet., vol. 10, no. 7, 2014.

[93] 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.

[94] Semiconductor Industry Association, *International Technology Roadmap for Semiconductors (ITRS)*. 2013.

[95] 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.*, vol. 3, no. 2, 2015.

[96] 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*, vol. 4, no. October, p. 7, 2010.

[97] J. Pavia, M. Wolf, and E. Charbon, "Measurement and modeling of microlenses fabricated on single-photon avalanche diode arrays for fill factor recovery," *Opt. Express*, vol. 22, no. 4, pp. 4203–4213, 2014.

[98] D. Bronzi, F. Villa, S. Bellisai, B. Markovic, S. Tisa, A. Tosi, F. Zappa, S. Weyers, D. Durini, W. Brockherde, and U. Paschen, "Low-noise and large-area CMOS SPADs with timing response free from slow tails," in *European Solid-State Device Research Conference*, 2012, pp. 230–233.

[99] A. Tosi, F. Villa, D. Bronzi, Y. Zou, R. Lussana, D. Tamborini, S. Tisa, D. Durini, S. Weyers, U. Pashen, W. Brockherde, and F. Zappa, "Low-noise CMOS SPAD arrays with in-pixel time-to-digital converters," in *SPIE*, 2014, vol. 9114, pp. 91140C:1–91140C:8.

[100] C. Veerappan and E. Charbon, "A substrate isolated CMOS SPAD enabling wide spectral response and low electrical crosstalk," *IEEE J. Sel. Top. Quantum Electron.*, vol. 20, no. 6, p. 3801507, 2014.

[101] E. A. G. Webster, J. A. Richardson, L. A. Grant, D. Renshaw, and R. K. Henderson, "A single-photon avalanche diode in 90-nm CMOS imaging technology with 44% photon detection efficiency at 690 nm," *IEEE Electron Device Lett.*, vol. 33, no. 5, pp. 694–696, 2012.

[102] A. Rochas, M. Gosch, A. Serov, P. A. Besse, R. S. Popovic, T. Lasser, and R. Rigler, "First fully integrated 2-D array of single-photon detectors in standard CMOS technology," *IEEE Photonics Technol. Lett.*, vol. 15, no. 7, pp. 963–965, 2003.

[103] A. P. Singh, J. W. Krieger, J. Buchholz, E. Charbon, J. Langowski, and T. Wohland, "The performance of 2D array detectors for light sheet based fluorescence correlation spectroscopy.," *Opt. Express*, vol. 21, no. 7, pp. 8652–8668, 2013.

[104] F. Villa, D. Bronzi, S. Bellisai, G. Boso, A. Bahgat Shehata, C. Scarcella, A. Tosi, F. Zappa, S. Tisa, D. Durini, S. Weyers, and W. Brockherde,

"SPAD imagers for remote sensing at the single-photon level," in *Proc. SPIE 8542, Electro-Optical Remote Sensing, Photonic Technologies, and Applications VI*, 2012, p. 85420G.

[105] I. M. Antolovic, S. Burri, C. Bruschini, R. Hoebe, and E. Charbon, "Nonuniformity Analysis of a 65-kpixel CMOS SPAD Imager," *IEEE Trans. Electron Devices*, vol. 63, no. 1, pp. 1–8, 2015.

[106] B. Sajadi, D. Qoc-Lai, A. T. Ihler, M. Gopi, and A. Majumder, "Image enhancement in projectors via optical pixel shift and overlay," in *IEEE International Conference on Computational Photography*, 2013, pp. 1–10.

[107] S. Nikzad, T. J. Cunningham, M. E. Hoenk, R. P. Ruiz, D. M. Soules, and S. E. Holland, "Direct detection of 0.1-20 keV electrons with delta doped, fully depleted, high purity silicon p-i-n diode arrays," *Appl. Phys. Lett.*, vol. 89, no. 182114, 2006.

[108] S. Nikzad, *High Performance Silicon Imaging*. 2014.

[109] J. Pavia, M. Wolf, and E. Charbon, "Measurement and modeling of microlenses fabricated on single-photon avalanche diode arrays for fill factor recovery," *Opt. Express*, vol. 22, no. 4, pp. 312–314, 2014.

[110] H. Irie, K. Kita, K. Kyuno, and A. Toriumi, "In-Plane Mobility Anisotropy and Universality Under Uni-axial Strains in n- and p-MOS Inversion Layers on (100), (110), and (111) Si," *IEDM Tech. Dig.*, vol. 9, no. 5, pp. 225–228, 2004.

[111] R. H. Poelma, "Multi-Scale Material and Technology for Heterogeneous Integration," Delft University of Technology, 2016.

[112] R. Ishihara, J. Derakhshandeh, M. R. Tajari Mofrad, T. Chen, N. Golshani, and C. I. M. Beenakker, "Monolithic 3D-ICs with single grain Si thin film transistors," *Solid. State. Electron.*, vol. 71, pp. 80–87, 2012.

Summary

Photon counting imaging technology is being developed widely in biomedical applications. This thesis proposes a novel photon-counting imager – flexible CMOS single-photon avalanche diode image sensor and presents detail research on the design, fabrication and thorough characterizations.

Motivated by novel biomedical applications emerging especially in the last decade, such as healthcare chip, retinal prosthesis and implantable sensors, Chapter 1 covers the features and requirements for these new configurations, using the nature of light and single photons. Special attention is focused on the photon counting device which needs to fit into human organ or body. Figures of merit, state-of-the-art detectors, and current flexible electronics process solutions are presented. After comparing and in-depth understanding different technology features, the flexible single-photon avalanche diode solution is figured out, providing a suitable way to retinal prosthesis, implantable biomedical monitor and wherever curved single-photon imaging plane is essential.

Different from typical PN junction, single-photon avalanche diode (SPAD) works above breakdown voltage (V_{BD}), in so-called Geiger mode. In this mode, the electric field is so strong that an injection of carrier would cause a self-sustaining avalanche. Physical principles of SPAD are discussed and fundamental metrologies are explained in Chapter 2. Reach-through structure and planar structure SPADs are discussed. Ultrathin-body SOI SPAD is reported for the first time. Thanks to the substrate transfer post-process, the world's first flexible ultrathin-body SOI SPAD is demonstrated successfully. The SPAD characteristics before and after substrate transfer are consistent, showing little impact of the post processing.

Chapter 3 focuses the optical characteristics of dual-side illumination, which is a new feature of this work compared with state-of-the-art SPAD technology. PDP and timing jitter characteristics are thoroughly measured from both front side and back side. A good match is shown between an experimentally measurement and theoretical analysis based on ultrathin-body junction profile and light absorption. This chapter also presents the configuration of SPAD integrated with trench-isolated quenching resistor on the same flexible substrate. Furthermore, this chapter thoroughly analyzes Geiger-mode characteristics such as DCR, VBD, PDP performances based on different process parameters including multiplication doping concentration, junction body thickness,isolation structures and temperature conditions.

The analysis are consistent with measurement results and the effect of each parameter is compared and concluded to the optimized structure.

Chapter 4 reports the first flexible SPAD imager pixel integrated with in-pixel CMOS electronics, by integrating CMOS process with the proposed flexible SPAD fabrication. This chapter gives an overview of the contributions of CMOS advances to SPAD technology, showing how important CMOS integration is for the flexible SPAD in implantable biomedical applications. An enhanced dynamic range is achieved, thanks to reduced parasitic capacitance, obtained through the integration of CMOS buffer. Optical performances such as PDP and timing jitter are comparable in both FSI and BSI. This chapter also demonstrates the significant CMOS SPAD Geiger-mode operation in the cryogenic environment as low as 80K. Furthermore, this chapter proposes the micro-lens imprinting on polyimide or polymer layer to increase the pixel fill factor.

Based on work in Chapter 4, which acts as the preliminary investigation and paves the road to larger scale sensor array, Chapter 5 presents the world's first flexible 32x32 CMOS SPAD image sensor, extending this thesis to array format in practice. This chapter not only addresses the design considerations such as the architecture circuitry function blocks and pixel structure, but also discusses the fabrication flowchart including front-end CMOS SPAD process, special back-end process on topology and flexible substrate transfer post-processing. External signal read out and process hardware system are implemented based on FPGA. The flexible SPAD array shows reasonable fundamental characteristics DCR distribution, PDP non-uniformity and negligible crosstalk. It is achieved for the first time that dual-side photon counting imaging by flexible SPAD image sensor with thorough characterizations. The imaging quality is even improved by resolution enhancement methods.

The proposed flexible CMOS SPAD technology is a breakthrough since it provides an innovative solution to apply photon counting device in the implantable biomedical sensor fields for the first time. This feature could also apply wherever curved single-photon imaging plane is essential.

Samenvatting

Imaging technologie door het tellen van fotonen, wordt op grote schaal ontwikkeld in biomedische toepassingen. Dit proefschrift stelt een nieuw foton-tel imager voor - flexibele CMOS enkel-foton lawinediodebeeldsensor en presenteert gedetailleerd onderzoek naar het ontwerp, de fabricage en grondige karakteriseringen.

Gemotiveerd door nieuwe biomedische toepassingen in opkomst, vooral in de afgelopen tien jaar zoals de gezondheidszorgchip, retinale prothese en implanteerbare sensoren, wordt in Hoofdstuk 1 de functies en eisen voor deze nieuwe configuraties onthuld, met behulp van de aard van het licht en enkele fotonen. Speciale aandacht gaat uit naar het foton-telapparaat welke moet passen in het menselijk orgaan of lichaamsdeel. Kwaliteitsfactor, de allernieuwste detectoren, en de huidige flexibele-elektronica procesoplossingen worden gepresenteerd. Na het vergelijken en diepgaand inzicht in de verschillende technologische eigenschappen, is de flexibele enkel-foton lawinediode-oplossing bedacht, welke voor een geschikte manier zorgt voor het netvlies prothese, implanteerbare biomedische monitor en overal waar gebogen enkel-foton imaging van essentieel belang is.

Anders dan de typische PN-verbinding, werken enkel-foton lawinedioden (SPAD) boven doorslagspanning (VBD), in de zogenaamde Geiger-modus. In deze modus, is het elektrisch veld zo sterk dat een injectie van een ladingsdrager een zelfdragende lawine zou veroorzaken. Fysische eigenschappen van de SPAD worden besproken en fundamentele metrologiëen worden toegelicht in Hoofdstuk 2. Doordringend structuur en vlakke structuur SPADS worden besproken. Ultradunne-lichaam SOI SPAD wordt voor de eerste keer gerapporteerd. Dankzij het substraatoverdracht in de nabewerking, is 's werelds eerste flexibele ultradunne-lichaam SOI SPAD met succes aangetoond. De SPAD karakteristieken vóór en na substraatoverdracht zijn consistent, met weinig invloed van de nabewerking.

Hoofdstuk 3 richt zich op de optische eigenschappen van tweekantenverlichting, een nieuw eigenschap van dit werk ten opzichte van de allernieuwste SPAD technologie. PDP en timing-jittereigenschappen zijn grondig gemeten aan zowel de voor- als achterzijde. Een goede match wordt getoond tussen een experimentele meting en theoretische analyse op basis van ultradunne-lichaamsverbindingsprofiel en lichtabsorptie. In dit hoofdstuk worden ook de configuratie van SPAD geïntegreerd met gleuf-geïsoleerde quench-weerstand op dezelfde flexibele substraat gepresenteerd. Bovendien

wordt in dit hoofdstuk een grondige analyse van Geiger-modekenmerken gemaakt zoals DCR, VBD, PDP prestaties op basis van verschillende parameters waaronder vermenigvuldiging-doteringsconcentratie, verbinding-lichaamsdikte, isolatiestructuren en temperatuuromstandigheden. De analysen komen overeen met meetresultaten en het effect van elke parameter wordt vergeleken met de geoptimaliseerde structuur.

Hoofdstuk 4 rapporteert de eerste flexibele SPAD imager-pixel geïntegreerd met in-pixel CMOS elektronica, door het CMOS-proces te integreren met de voorgestelde flexibele SPAD fabricage. Dit hoofdstuk geeft een overzicht van de bijdragen van de CMOS ontwikelingen aan SPAD technologie, waaruit blijkt hoe belangrijk CMOS-integratie is voor de flexibele SPAD in implanteerbare biomedische toepassingen. Een verbreed dynamische bereik is verkregen, dankzij gereduceerde parasitaire capaciteit, verkregen door de integratie van de CMOS buffer. Optische prestaties zoals PDP en timing jitter zijn vergelijkbaar in zowel FSI als BSI. Dit hoofdstuk toont ook de aanzienlijke CMOS SPAD Geiger-moduswerking in de cryogene omgeving met minimale temperaturen van 80K. Bovendien stelt dit hoofdstuk microlens imprinten voor op polyimide of een polymeerlaag om de pixel-vulfactor te verhogen.

Op basis van het werk in Hoofdstuk 4, welke fungeert als het vooronderzoek en de weg naar grotere schaal sensorarray opent, presenteert Hoofdstuk 5 's werelds eerste flexibele 32x32 CMOS SPAD-beeldsensor, de uitbreiding van dit proefschrift array-formaat in de praktijk. Dit hoofdstuk behandelt niet alleen de designoverwegingen, zoals de architectuur-circuit-functieblokken en pixelstructuur, maar bespreekt ook het fabricagestappenplan waaronder het front-end CMOS-SPAD-proces, speciale back-end-proces op topologie en flexibel-substraatoverdrachtsnabewerking. Extern-signaaluitlezing en proces-hardwaresysteem zijn geïmplementeerd op basis van FPGA. De flexibele SPAD array toont redelijke fundamentele kenmerken, DCR distributie, PDP non-uniformiteit en verwaarloosbare kruisspraak. Voor het eerst is een twee-zijdig foton-telbeeldvorming bereikt door een flexibel SPAD image sensor met grondige karakteriseringen. De beeldkwaliteit is zelfs verbeterd door resolutie versterkmethoden.

De voorgestelde flexibele SPAD CMOS technologie is een doorbraak omdat het een innovatieve oplossing biedt om voor het eerst de fotoon-telapparaat toe te passen in implanteerbare biomedische sensorenvelden. Deze functie kan ook worden toegepastn waar gebogen enkel-fotoon imaging vlak van essentieel belang is.

Acknowledgements

Around four years ago, I made a decision to pursue PhD degree at Delft University of Technology. Now, I am typing the last part of my thesis and realize so many people showed up, encouraged, guided and contributed to my work and life along this long road.

First, I would like to thank Prof. dr. Edoardo Charbon for the great supervision. Without his substantial guidance and encourage, it is impossible to finish this thesis. He always gives me unlimited research resource and support. I have benefited a lot from his wide knowledge, comprehensive analysis, brilliant ideas and sometimes sense of humor.

Second, I would like to thank Dr. Ryoichi Ishihara for offering this PhD opportunity and providing his great support all the time. He is a very wise and patient scientist, who always opens his door whenever I need to discuss. It is really nice to work together with him and thanks for bring us delicious Japanese food.

Third, I would like to thank Prof. dr. Lina Sarro for her great supports during my work. She is a very kind person who always gives me warm encouragements. I would like to express my appreciation for her guidance.

Next to them, I would like to thank Dr. Shouleh Nikzad, Prof. Pierre Mangan, Prof. Martin Wolf, Prof. Wilfried Uhring and Prof. Paddy French for accepting to be my defence committee. It is my great honor to have these top experts reviewing my thesis manuscript, travelling and serving in my defence at Delft.

Furthermore, I would like to thank Prof. Kees Beenakker, Prof. Kouchi Zhang, Prof. Lis K. Nanver and Prof. Ronald Dekker for their advices and collaborations during my PhD period. I really appreciate their excellent academic contributions making our institute a world-class research center.

It was an exciting experience to work at EKL BICMOS processing line and MEMS lab. I would like to express my appreciation to Dr. Henk van Zeijl, Dr. Gregory Pandraud and Silvana Milosavljevic for the valuable discussions and advices. Furthermore, I owe special thanks to Tom Scholtes for great support in epitaxy, ion implantation, lithography and valuable opinions when I got processing problems. I would like to express special gratitude to Willem van der Vlist, who come here to help me with flexible chip bonding during his free time. I would also like to thank Koos van Hartingsveldt and Loek Steenweg for great help with flexible chip packaging and bonding. Many thanks to Mario for the various deposition and etching supports and experience sharing and all other clean room members: Johannes van Wingerden, Johan van der Cingel, Cassan

Visser, Jan Cornelis Wolff, Jan Groeneweg, Ruud Klerks, Wim Wien, Alex van den Bogaard, Joost Berendse, Robert Verhoeven, Ron van Viersen and Jan Warmerdam. I would like to thank Marian Roozenburg, Lidwina Tromp, Bianca Knot for nice help and support.

It was such a great time to work with my talent group mates during last four years. Many thanks to Dr. Shingo Mandai and Dr. Chockalingam Veerappan for your great help in SPAD knowledge, Dr. Sten Vollebregt, Dr. Jin Zhang and Dr. Aslihan Arslan, Dr. Miki Trifunovic, Dr. Paolo Sberna for your kind support and experience sharing in clean room processing, Dr. M. J. Lee, Chao Zhang, Ivan Michel Antolovic, Tin Gong, Esteban Venialgo Augusto Carimatto for the SPAD measurement support and opinions; Andres Alfaro Barrantes, Nello Franzese, Yuanxing Xu, Gautham Rangasamy, Harald Homulle, Masoud Babaie, Jeroen van Dijk, Rosario Marco Incandela, Scott Lindner, Bishnu Patra, Preethi Padmanabhan, Siddharth Sinha, Lin Song, Arin Ulku, Bahador Valizadeh Pasha, Junjie Weng, Dali Zhang, Andres Alfaro Barrantes, Nello Franzese, Yuanxiing Xu, Dr. Fabio Sebastiano, and visiting professors Dr. Mashiro Akiyama and Dr. Yudong Li with whom I have worked with, in making my PhD life memorable.

Many thanks to my dear colleagues and friends around: Dr. Lin Qi, Dr. Jin Zhang, Xueming Li, we would continue looking for nice food everywhere, maybe with little Lucas in future; Dr. Agata Sakic, Dr. Caroline Mok, it was a great pleasure to share chocolate with you; Dr. Jiaqi Tang, I am very happy to see Jiaco Company growing up and wish you all the best; Dr. Jing Zhang and Pan Liu, very happy for your marriage and cute daughter; Dr. Hairen Tan, I would express my best wishes for you and your family, Dr. Yu Bi, a beautiful female scientist and stays young forever; Dr. Lihao Han, Fai tong Si, I would never forget our exciting trip to Aachen; Dr. Rene Poelma, a calm scientist giving others great support; Dr. Amir Sammak, Dr. Bruno Morana, Dr. Gianpaolo Lorito, Yelena Grachova, William Solano, Cinzia Silvestri, Violeta Prodanovic, Nikolas Gaio, Marta Kluba, Aleksandar Jovic, Bo Sun, Jianlin Huang, Mingzhi Dong, Thu Hang Bui, Hong Wah Chan. It was really a pleasure time with you, making life colorful.

How can I forget my supervisors in China who have taught, trained, motivated and helped me grow up. I would express my special appreciation to Prof. Hoi-Sing Kwok and Prof. Man Wong for excellent supervision when I was within Hong Kong University of Science and Technology.

I would express my deep memory to Prof. Zhiguo Meng at Nankai University for his great knowledge, strong will and nice personality, which would guide and motivate me forever. I would like to thank Prof. Shaozhen Xiong and Prof. Chunya Wu for nursing and helping me in many ways.

I would like to thank all my friends when I was studying and working at Tianjin, Hong Kong and any other places.

Finally, I am very grateful to my family. I would like to thank my parents who are supporting me at back all the time and providing endless love. As well, I would like to thank my in-laws for various support and dedications. Being a source of inspirations, I appreciate your love forever.

Pengfei Sun
August, 2016
Delft, The Netherlands

List of Publications

Technical Journals

1. **Pengfei Sun**, Ryoichi Ishihara and Edoardo Charbon, "Flexible Ultrathin-Body Single-Photon Avalanche Diode Sensors and CMOS Integration," ***Optics Express,*** Vol. 24 Issue 4, pp. 3734-3748, 2016.
(DOI: 10.1364/OE.24.003734)

2. Myung-Jae Lee, **Pengfei Sun**, and Edoardo Charbon, "A First Single-Photon Avalanche Diode Fabricated in Standard SOI CMOS Technology", ***Optics Express,*** Vol.23 No.10, 2015.
(DOI: 10.1364/OE.23.013200)

3. **Pengfei Sun**, Edoardo Charbon, and Ryoichi Ishihara, "A flexible ultra-thin-body single-photon avalanche diode with dual side illumination," ***IEEE Journal of Select Topics in Quantum Electronics,*** Vol.20 No.6, 2014.
(DOI: 10.1109/JSTQE.2014.2342193)

Conference Proceedings

1. **Pengfei Sun,** Ryoichi Ishihara and Edoardo Charbon, "A Flexible 32x32 SPAD Image Sensor with Integrated Microlenses" ***International Image Sensor Workshop (IISW),*** Session 11, Paper 3, Vaals, Netherlands, 2015.

2. Myung-Jae Lee, **Pengfei Sun,** and Edoardo Charbon, "Characterization of Single-Photon Avalanche Diodes in Standard 140-nm SOI CMOS Technology", ***International Image Sensor Workshop (IISW),*** Session 11, Paper 5, Vaals, Netherlands, 2015.

3. **Pengfei Sun**, Benjamin Mimoun, Edoardo Charbon and Ryoichi Ishihara, "A Flexible Ultra-Thin-Body SOI Single-Photon Avalanche Diode" ***IEEE International Electron Device Meeting (IEDM),*** Session 11, Paper 1, Washington, DC, USA, 2013.
(DOI: 10.1109/IEDM.2013.6724606)

4. **Pengfei Sun**, Michiel van der Zwan, Aslihan Arslan, Edoardo Charbon, and Ryoichi Ishihara, "Improvement on MIS Properties of Single-Grain Germanium by Pulsed-Laser Annealing" ***44th IEEE Semiconductor Interface Specialists Conference (SISC),*** Session 7, Paper 2, Arlington, USA, 2013.

Journal under Preparation

Pengfei Sun, Ryoichi Ishihara and Edoardo Charbon, "A Flexible CMOS Single-Photon Avalanche Diode Dual-Side Imaging Sensor."

About the Author

Mr. Pengfei Sun was born in Jingdezhen, China on February 14, 1986. He received his B.S. degree in Microelectronics from Nankai University, Tianjin, China, in 2007 and M.S. degree in 2010. He has broad interests in active matrix display, image sensor, flexible electronics and three-dimensional integration. From May, 2009, he worked as research assistant on high-performance poly-Silicon TFT circuit in Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST). In June, 2010, he joined in the Center for Display Research, HKUST as research and development staff. He worked on full-integration poly-Si AMOLED panel and OLED on silicon (Micro-display) technology. Since April, 2012, he joined in Delft Institute of Microsystems and Nanoelectronics, Delft University of Technology, the Netherlands, as PhD candidate. His PhD project is focused on novel single-photon avalanche diode technology for flexible electronics and three-dimension CMOS integration.

He is currently working on quantum device integration on CMOS platform within QuTech.