PHOTODIODES - WORLD ACTIVITIES IN 2011

Edited by **Jeong-Woo Park**

INTECHWEB.ORG

Photodiodes - World Activities in 2011
Edited by Jeong-Woo Park

Published by InTech
Janeza Trdine 9, 51000 Rijeka, Croatia

Copyright © 2011 InTech
All chapters are Open Access articles distributed under the Creative Commons
Non Commercial Share Alike Attribution 3.0 license, which permits to copy,
distribute, transmit, and adapt the work in any medium, so long as the original
work is properly cited. After this work has been published by InTech, authors
have the right to republish it, in whole or part, in any publication of which they
are the author, and to make other personal use of the work. Any republication,
referencing or personal use of the work must explicitly identify the original source.

Statements and opinions expressed in the chapters are these of the individual contributors
and not necessarily those of the editors or publisher. No responsibility is accepted
for the accuracy of information contained in the published articles. The publisher
assumes no responsibility for any damage or injury to persons or property arising out
of the use of any materials, instructions, methods or ideas contained in the book.

Publishing Process Manager Petra Zobic
Technical Editor Teodora Smiljanic
Cover Designer Jan Hyrat
Image Copyright Shebeko, 2010. Used under license from Shutterstock.com

First published July, 2011
Printed in Croatia

A free online edition of this book is available at www.intechopen.com
Additional hard copies can be obtained from orders@intechweb.org

Photodiodes - World Activities in 2011, Edited by Jeong-Woo Park
p. cm.
ISBN 978-953-307-530-3

INTECH OPEN ACCESS PUBLISHER

INTECH open

free online editions of InTech
Books and Journals can be found at
www.intechopen.com

Contents

Preface IX

Part 1 Generals of Photodetection 1

Chapter 1 **An Absolute Radiometer Based on InP Photodiodes 3**
Ana Luz Muñoz Zurita, Joaquin Campos Acosta and Alicia Pons Aglio

Chapter 2 **Physical Principles of Photocurrent Generation in Multi-Barrier Punch-Through-Structures 23**
A.V. Karimov, D.M. Yodgorova and O.A. Abdulkhaev

Chapter 3 **Photon Emitting, Absorption and Reconstruction of Photons 37**
Changjun Liao, Zhengjun Wei and Jindong Wang

Part 2 CMOS Related Topics 63

Chapter 4 **CMOS Photodetectors 65**
Albert H. Titus, Maurice C-K. Cheung and Vamsy P. Chodavarapu

Chapter 5 **Image Artifacts by Charge Pocket in Floating Diffusion Region on CMOS Image Sensors 101**
Sang-Gi Lee, Jong-Min Kim, Sang-Hoon Bae, Jin-Won Park and Yoon-Jong Lee

Chapter 6 **Active Pixel Sensor CMOS Operating Multi - Sampled in Time Domain 121**
Fernando De Souza Campos

Chapter 7 **Bandwidth Extension for Transimpedance Amplifiers 139**
Omidreza Ghasemi

Part 3 **APD and Single Photon Detection 157**

Chapter 8 **Avalanche Photodiodes in High-Speed Receiver Systems 159**
Daniel S. G. Ong and James E. Green

Chapter 9 **Silicon Photo Multipliers Detectors Operating in Geiger Regime: an Unlimited Device for Future Applications 183**
Giancarlo Barbarino, Riccardo de Asmundis, Gianfranca De Rosa, Carlos Maximiliano Mollo, Stefano Russo and Daniele Vivolo

Chapter 10 **Near-Infrared Single-Photon Detection 227**
Guang Wu, E Wu, Xiuliang Chen, Haifeng Pan and Heping Zeng

Chapter 11 **Geiger Avalanche Photodiodes (G-APDs) and Their Characterization 247**
Giovanni Bonanno, Massimiliano Belluso, Sergio Billotta, Paolo Finocchiaro and Alfio Pappalardo

Chapter 12 **Design of High Quantum Efficiency and High Resolution, Si/SiGe Avalanche Photodiode Focal Plane Arrays Using Novel, Back-Illuminated, Silicon-on-Sapphire Substrates 267**
Alvin G. Stern

Part 4 **Extended Topics of Photodiodes 313**

Chapter 13 **Single Crystal Diamond Schottky Photodiode 315**
Claudio Verona

Chapter 14 **GaN Based Ultraviolet Photodetectors 333**
D. G. Zhao and D. S. Jiang

Chapter 15 **Quantum Dot Composite Radiation Detectors 353**
Mario Urdaneta, Pavel Stepanov, Irving Weinberg, Irina Pala and Stephanie Brock

Chapter 16 **HgCdTe Heterostructures Grown by MBE on Si(310) for Infrared Photodetectors 367**
Maxim Yakushev, Vasily. Varavin, Vladimir Vasilyev, Sergey Dvoretsky, Irina Sabinina,Yuri. Sidorov, Aleksandr Sorochkin and Aleksandr Aseev

Preface

Photodiodes or photodetectors are prevailing technology in various fields giving many benefits to human race. Moreover, their application fields are extended in the world more and more. From photo camera application to medical application, they are pervading in the human life. They became an inevitable and valuable technology since its birth. Now we cannot imagine what is like our life without them. They would exist until end of human race even if we don't need our sight of today anymore. Even though we become a new race after gradual evolution, we would need suitable photodiodes and photodetectors. They are in one boat with our human race.

Many people in the world pursuit a birth of new photodiodes or photodetectors technology which are able to give better life to human. Their efforts are contained in this book. This book would be valuable to those who want to obtain knowledge and inspiration in the related area.

Jeong-Woo Park

Electronics and Telecommunications Research Institute,
161 Gajeong-dong, Yusong-gu, Daejeon,
Republic of Korea

Part 1

Generals of Photodetection

1

An Absolute Radiometer Based on InP Photodiodes

Ana Luz Muñoz Zurita[1], Joaquin Campos Acosta[2] and Alicia Pons Aglio[2]
[1]*Autonomous University of Coahuila*
Faculty of Engineering Mechanical and Electrical U Torreon
[2]*CSIC- Institute for Applied Physics*
[1]*Mexico*
[2]*Spain*

1. Introduction

Semiconductor photodetectors based on InP materials are the ones most often used in state of the art long wavelength optical fiber communication system. Mixed compounds such as InGaAs (P) and In(Al)GaAs lattice matched to InP are the materials responsible for detecting long wavelength light, specially the nondispersion wavelength (1.3 µm) and loss minimum wavelength (1.55 µm) of silica optical fibers. The characteristics of these InP-based photodectors are superior to those of conventional photodiodes composed of elemental Ge, which was the only material applicable for wavelengths below 1.55 µm. By using a heteroestructure, which hadn't been expected in group IV elemental semiconductors such as Si and Ge, new concepts and new designs for high performance photodetectors have been developed. For example, the absorption region can be confined to a limited layer and the InP wide bandgap layer can serve as a transparent layer for specific communication wavelength. Recently InGaAs/InP avalanche photodiodes (APDs) with a SAM (separation of absorption and multiplication) configuration have become commercially available. The SAM configuration is thought to be necessary for high performance APDs utilizing long wavelengths.

The photodiodes may be operated under reverse bias, high quality semiconductor layers need to be produced. To obtain photodiodes that operate at a low bias and have a low dark current, it is necessary to produce epitaxial layers that are pure and have few defects (such as dislocations, point defects, and impurity precipitates). To get stable and uniform gain in APDs, in which internal gain is achieved through the carrier avalanche process, the layers in the avalanche region must be uniform and free of dislocations. Furthermore, a planar device structure requires that a guard ring be used to keep the electric field around the photoreceptive area from increasing too much. Fabrication and processing technologies such as impurity diffusion, ion implantation, and passivation will also play important roles in the production of reliable photodetectors.

From a radiometric point of view, the photodetectors important characteristics are: Speed of (characterized by the bandwidth of the frequency response or the Full Width Half

Maximum (FWHM) of the pulse response), responsivity (determined as the ratio of current out the detector to the incident optical power on the device), sensitivity (defined as the minimal input power that can still be detected which, as a first approximation, is defined as the optical power which generates an electrical signal equal to that due to noise of the diode). One related characteristic is the quantum efficiency of the detector which is the ratio of the number of electron-hole pairs which contribute to current to the number of incident photons.

When the light radiation impinges on a detector, various physical processes occur; part of the incident light is reflected by the sensitive surface, while the rest passes inside the detector, where can be partially, because of losses due to absorption, converted into an electronic signal. The response of each photodetector is conditioned by a quantity of the converted light power, but for evaluating the incident power one has to know the ratios of the reflected, absorbed, and converted portions.

The InP photodetectors are chosen as the first device of interest because of their simple structure, and since their analysis is a natural extension, almost an example, of our discussion of p-n diodes. Whereas the field of photodetectors goes far beyond that of semiconductor photodetectors, we restrict ourselves here to such devices. It will be discussed p-i-n diodes, which are also referred to as photovoltaic detectors, photoconductors or solar cells photodetectors. The distinction between the different devices is somewhat artificial since many similarities exist between these devices but it enables to clearly separate the difference in structure, principle of operation and purpose of the devices.

InP-photodiodes from different manufacturers have got rather low noise level, a good response uniformity over the sensitive surface and a wide dynamic range. Therefore they are good devices to built radiometers in the NIR spectral region. As in any photodiode, the spectral short-circuit responsivity is determined by the wavelength and the photodiode's reflectance and internal quantum efficiency. Then if these quantities were known, the photodiode's responsivity would be known without being compared to another standard radiometer; i. e. the photodiode would be an absolute radiometer for optical radiation measurements.

This idea was firstly developed for silicon photodiodes in the eighties, once the technology was able to produce low defects photodiodes. Following this reference, the reflectance could be approached from a superimposed thin layers model. By knowing the thicknesses of the layers and the optical constants of the materials, it is possible to determine the device reflectance. However, this information is not completely available for InP photodiodes: the actual thickness of the layers is not known and optical constants of materials are only approximately known for bulk. Nevertheless it's possible to measure reflectance at some wavelengths and to fit the thicknesses of a layer model that would reproduce those experimental values.

The internal quantum efficiency cannot be determined.Since InP photodiodes are hetero-junctions rather than homo-junctions as silicon photodiodes are. In the other hand, since the internal structure is not accurately known, it is not possible to model the internal quantum efficiency without having experimental values for it.

Therefore the attainable scope at present is just to obtain a model to be able to calculate spectral responsivity values at any wavelength. To get this, a model has been developed to calculate reflectance values from experimental ones at some wavelengths and another

model has been developed to interpolate spectral internal quantum efficiency values from some values got from reflectance and responsivity measurements at some wavelengths. Both models will be presented in this chapter.

2. Pin-photodiode design, fabrication and modelling

Photodiodes with different lengths from 30 µm to 70 µm and width 8 µm were fabricated to analyse their RF behavior. A schematic view of the pin photodetector is depicted in figure 1. The structure of the pin-photodiode is based on a transverse twin-guide scheme. From the top down, the main parts of the pin-photodiode are a 400 nm p+ -InP contact layer, 100 nm InGaAs absorbtion layer, 100 nm non-intentionally doped and 400 nm n+ -InGaAsP(Q1.3) layer, acting as the top waveguide layer and as n-contact layer. With this structure layer and 8×60 µm2 area 0.25 pF can be achieved for a reverse bias voltage of 2 V and 0.12 pF for −5 V. The access waveguides consist of a 600 nm Q(1.3) layer with an InP cladding layer which is tapered from the normal width of 2 µm to 8 µm.

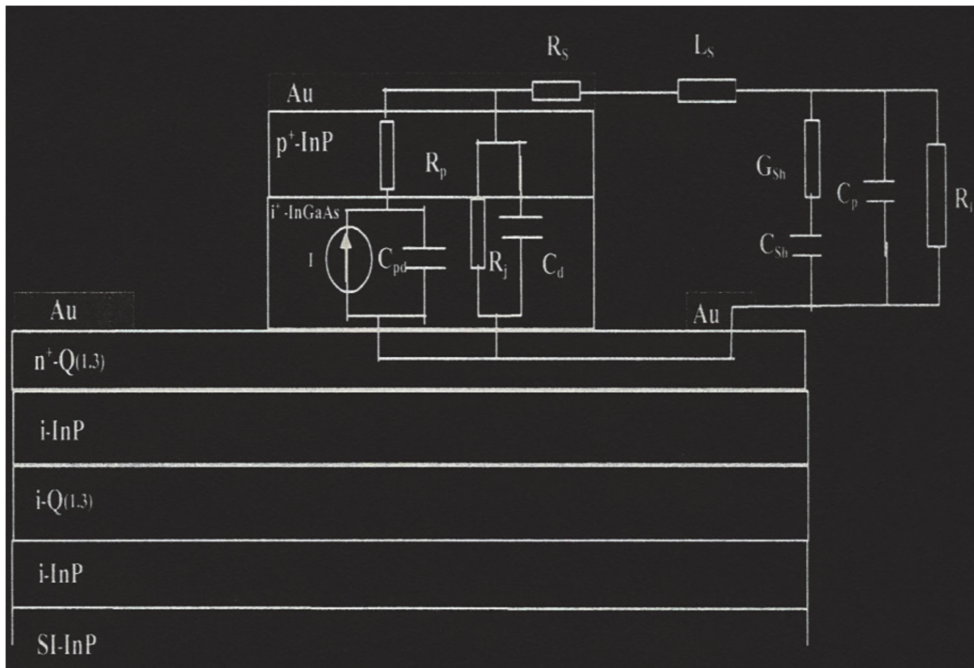


Fig. 1. Cross section of a side-illuminated twin-waveguide pin-photodiode. The equivalent circuit elements represent physical parameters of the photodiode and parasitic elements.

Electrically, the output of the pin photodiode is fabricated as a coplanar waveguide transmission line in order to facilitate hybrid integration with the GaAs travelling-wave amplifier and also for on-wafer high-frequency measurements. Figure 1 indicates how the parasitic components are connected and how they are related to the physical parameters of the pin photodiodes.

Here Rs is the p- and n-electrode contact resistance, R j is the resistance representing the dark current leakage, R p is the resistance of the bulk P+ -InP layer; Cpd is the reverse-bias junction capacitance, Cd capacitance between p- and n-contact metallization which is significant at low frequencies, Ls is the inductance of the p/n-contacts to signal/ground electrodes, Csh and Gsh are the shunt capacitance and conductance of coplanar waveguide which take into account the electric energy storage inside the transmission line and the losses caused by the transverse component, Cp is the open circuit at the end of the coplanar waveguide transmission line, and RL is the load resistance.
In the case of semiconductors photodetectors based on InP materials are the ones most often used in state of the art long-wavelength optical fiber communication systems. Mixed compounds such InGaAs(P) and In(Al)GaAs lattice matched to InP are the materials responsible for detecting long-wavelength light, especially the nondispersion wavelength (1.3 µm) and loss minimum wavelength (1.55 µm) of silica optical fibers.
The characteristic of these InP based photodetectors are superior to those of conventional photodiodes composed of elemental Ge, which was the only material applicable for wavelengths below 1.5 µm . By using a heterostructure, which had not been expected in group IV elemental semiconductor such as Si and Ge, new concepts and new designs for high-performance photodetectors have been developed. For example, the absorption region can be confined to a limited layer, and the InP wide-bandgap layer can serve as a transparent layer for specific communication wavelength. Recently InGaAs/InP avalanche photodiodes (ADPs) with a SAM (separation of absorption and multiplication) configuration have become commercially available.The SAM configuration is thought to be necessary for high performance APD´s utilizing long wavelenghts.
Because photodiodes operate under reverse-bias, high quality semiconductor layers need to be produced. To obtain photodiodes that operate at a low bias and have a dark current, it is necessary to produce epitaxial layers that are pure and that have few effects (such as dislocations, points defects and impurity precipitates).To get stable and uniform gain in APDs, in which internal gain is achieved through the carrier avalanche process, the layers in the avalanche region must be uniform and free of dislocations. Futthermore a planar device structure requieres that a guardring be used to keep the electric field around the photoreceptive area from increasing to much.
Fabrication and processing technologies such as impurity diffusion, ion implantation and passivation will also play important role in the production of reliable photodetectors. The photodiodes used in received circuits are required to traslate optical signals into electrical signals faitfully and efficiently.
The reflectance and the internal quantum efficiency are determined the photodiode spectral responsivity, which is the radiometric characteristic of interest in the fields where these devices can be used for optical radiation measurements. It presents the experimental set up for measuring the photodiode reflectance as well as the results of such measurements related to InGaAs/InP-photodiodes exploited in international laboratories.
The obtained experimental results show that some models of photodiodes have got an anti-reflecting coating on their sensitive facets and that reflectance does not change with varying the light polarization state within the measurement uncertain, when the angles of incidence are less that 7.4 ^{o}.

3. Applications of InP photodiodes

Future systems based on microwave and optics such as fiber optic radio communications will require very high speed InP photodetectors able to work in the microwave or millimeter

wave frequency range. It is well-known that PIN photodetectors limitations are due to transit time and capacitance.

To increase their cut-off frequency, it is necessary to reduce transit time, so thickness of absorbing layer, and capacitance. For top-illuminated PIN photodetectors, this will decrease dramatically the responsivity of the photodetector, and theses detectors are generally used up to around 20 GHz, as maximum bandwidth. For upper frequencies, the PIN waveguide photodetector is an attractive device, since it is possible to reduce transit time without decreasing responsivity, because of the absorbing core waveguide structure. Moreover, the optical signal is absorbed over a short length (5-10 µm), and the device , so the capacitance, can be very small. Recently, demonstrations with high responsivity up to 60 GHz at Thomson LCR1 and even at higher millimeter wave frequencies2 were performed with device grown on semi-insulating InP substrate. Microwave access is a coplanar line to reduce parasitics, and the waveguide is a multimode structure to improve the optical coupling with the optical fiber. Interdigitated Metal Semiconductor Metal (MSM) photodetectors on AlInAs/GaInAs/InP epilayers were also subject to attention because of their low capacitance.

Commercially available MSM (New Focus) were developed with 40 GHz cut-off frequency. Nevertheless, as for PIN photodetectors, high speed MSM photodetectors require thin absorbing layers together with short electrode spacing, so reduced responsivity. Solution would be waveguide devices, or MSM coupled with optical waveguide, but the technology is then more complicated, and the PIN waveguide photodetector remains the best solution.

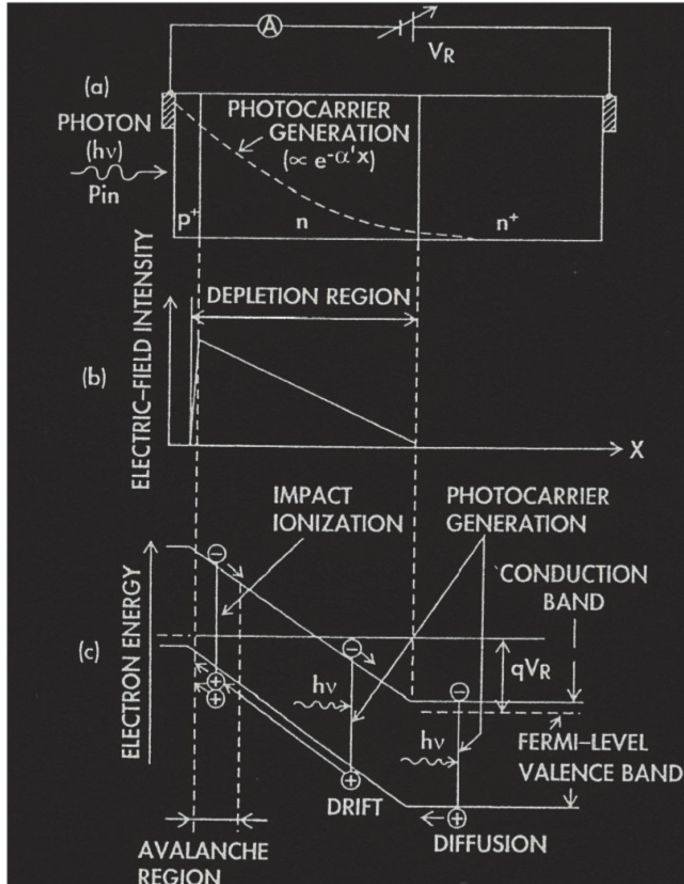
PIN photodiode (or MSM photodetector) does not exhibit internal gain. It is the reason why heterojunction phototransistor (HPT) in the InP/InGaAs material system has been studied for several years. HPTs could be a good alternative to top illuminated or edge-coupled PIN photodiodes since they exhibit an internal gain owing to transistor effect, without high bias voltage as required for avalanche photodiodes and without excess noise due to avalanching. If we do not consider doping of epilayers, epitaxial structure of HPT is very similar to the one of PIN waveguide photodiode. It means that waveguide technique could also be interesting to apply to HPTs for microwave or millimeter wave applications.

Photodiodes are semiconductor devices responsive to high energy particles and photons. Photodiodes operate by absorption of photons or charged particles and generate a flow of current in an external circuit, proportional to the incident power. Planar diffused silicon photodiodes are P-N junction diodes. A P-N junction can be formed by diffusing either a P-type impurity, such as Boron, into a N-type bulk or epitaxial silicon wafer, or a N-type impurity, such as Phosphorus, into a P-type bulk or epitaxial wafer. The diffused area defines the photodiode active area. To form an ohmic contact, another impurity diffusion into the backside of the wafer is necessary. The active area is coated with an Anti-Reflection coating to reduce the reflection of the light for a specific predefined wavelength. The P and N-sides of the junction have metal pads, which make an electrical contact through dielectric layers.

Due to the high absorption coefficient, the InGaAs absorption region is typically a few micrometers thick. The thin absorption layer enables the device to obtain high speed at a low reverse bias voltage, typically 2-5 volts. The InP window layer is transparent to 1.3µm – 1.55µm wavelengths, thus InGaAs/InP photodiodes do not have slow tail impulse response associated with the slow diffusion component from the contact layer.

4. Basic photodiode operation

Photodiodes operate under reverse vias to create a depleted region in which photogenerated electron-hole pairs. In the figure 2 show a schematic cross section of a photodiode (or avalanche) with a p(+),(-)n,n(+) structure. It also show the optical absorption, which is subject to the absorption coefficient of the material for incident light and decreases exponentially with increasing distance from the diode front p(+).



(a) Cross sectional sketch of p(+), n(-), n(+) diode under reverse bias.

(b) Electric field distribution

(c) Energy band diagram

Fig. 2. Basic operation of a Photodiode.

In the Figure 2(b) and (c) show electric field distribution and the energy band diagram, respectively. Most photocarriers are designed for use in the fully depleted n-region, so that they have a high-speed response: Electrons and holes generated within the depletion region are instantaneously separated by the elelectrical field and drift in the opposite direction,

inducing a photocurrent in the external circuit. On the other hand, minority-carrier holes excited within an average diffusion length in the undepleted n(+) region adjacent to the depleted region diffuse into the edge of the depleted junction with some recombination and are colleted across the high-field region, resulting in a diffusion photocurrent in the external circuit. Diffusion photocurrent is generally characterized by its slow response to the optical signal, since the speed of the response depends on the time it takes for the photogenerated minority carriers to diffuse from where they are generated in the neutral undepleted region into the edge of the depletion region.

The photodiode should therefore be designed in such a way that there is no optical absorption in the undepleted neutral region. For the same reason, as well as to reduce the recombination loss of photocarriers genereted in the P(+) region on the fron side of the diode, the p(+) region must be as thin as possible.

When the electric field of photodiode is elevated to several hundreds of kilocolts per centimeter by increasing reverse bias, as internal gain for a primary photocurrent can be obtained. This gain is a result of the electron-hole pair creation is denominated avalanche process, initiated by the photogenerated carriers, which creation is in turn governed by the relation between the strenght of the electric field and the electron and hole impact ionization rates of the material itself.

4.1 Basic InP PIN photodiodes

In making a simple planar structure the photodiode a double heterostructure consisting of InGaAs/InP with a inP capping layer is grown, and this growth is followed by selective impurity diffusion to form the p(+)n(-) junction. A window allowing light to pass through is formed on the front (the grown-layer superffice) or back surface of the InP substrate. The back-illumination type is of a structure often used to obtain the low capacitance desirable for high-speed operation.

A cross -sectional view of a front-illuminated planar structure InGaAs/InP – PIN photodiode is show in the figure 3.

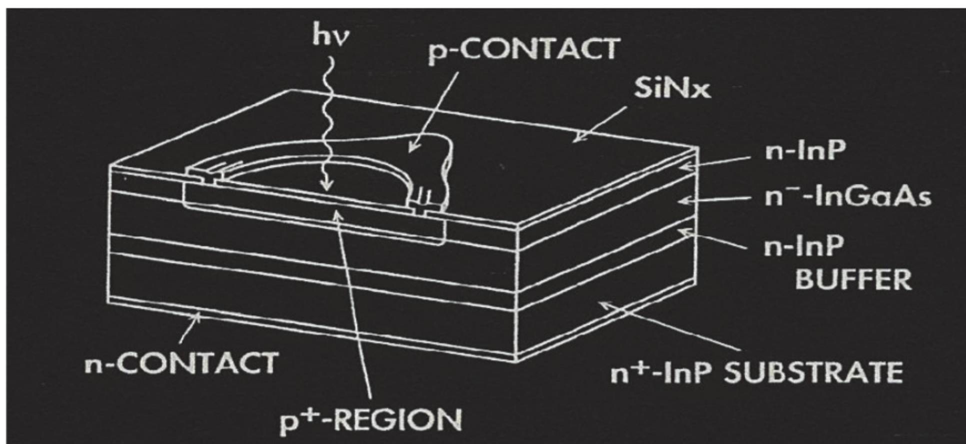


Fig. 3. Schematic cross section of a planar structure InP Photodiode.

In the case of InP Photodiodes we need to check something important parameters for example:

a. Receiver Sensitivity.
b. Response Speed.
c. Dark Current Reduction.
d. High Speed and high quantum efficiency.
e. Reduction of diode capacitance.
f. Large-ionization rate ratio materials for the avalanche layer.

5. InP-photodetectors

At the present time, the InGaAs/InP-photodetectors from different manufactures have rather low level of noise, a good uniformity of the surface response as well as a wide dynamic range and linearity.

For these reasons they are exploited in the instruments for measuring optical radiation within the near infrared (IR) range (800-1600 nm). Furthermore, the InGaAs/InP-photodetectors are used for maintaining the scale of spectral responsitivity in the same spectral range in many laboratories. It is presented this chapter devoted to studying the reflectance of photodiodes from different manufactures.

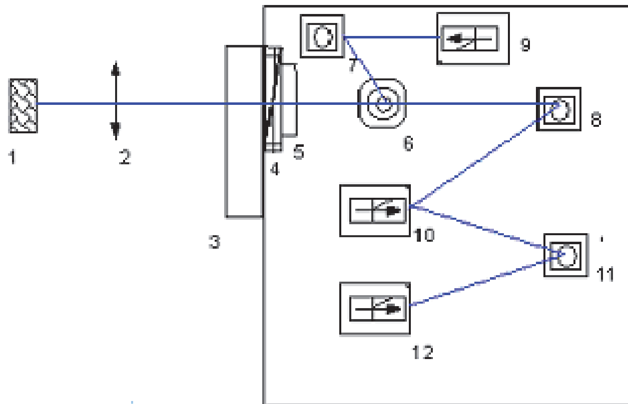
For this reason, measuring the reflectance of photodiodes is presented in this chapter as a preliminary step to finding the responsivity. The photodiode response depends on a set of parameters inherent in the incident light like the spectral distribution, polarization, modulation of frequency, angle of incidence, and radiant power.

Then, the response is determined by such characteristics of photodetector as the material refraction index and the structure of diode as well as by some environmental factors, for example, by the temperature.Photodiodes are semiconductor devices responsive to high energy particles and photons. Photodiodes operate by absorption of photons or charged particles and generate a flow of current in an external circuit, proportional to the incident power. Planar diffused silicon photodiodes are P-N junction diodes. A P-N junction can be formed by diffusing either a P-type impurity, such as Boron, into a N-type bulk or epitaxial silicon wafer, or a N-type impurity, such as Phosphorus, into a P-type bulk or epitaxial wafer. The diffused area defines the photodiode active area. To form an ohmic contact, another impurity diffusion into the backside of the wafer is necessary. The active area is coated with an Anti-Reflection coating to reduce the reflection of the light for a specific predefined wavelength. The P and N-sides of the junction have metal pads, which make an electrical contact through dielectric layers.

Due to the high absorption coefficient, the InGaAs absorption region is typically a few micrometers thick. The thin absorption layer enables the device to obtain high speed at a low reverse bias voltage, typically 2-5 volts. The InP window layer is transparent to 1.3μm – 1.55μm wavelengths, thus InGaAs/InP photodiodes do not have slow tail impulse response associated with the slow diffusion component from the contact layer.

6. Experimental procedure

To realize our experiments related to measuring the reflectance of InGaAs/InP photodiodes it have arranged the experimental set-up presented in Figure 4.



1. Source light (Incadescent lamp)
2. Lens
3. Monochromator
4. Polarizer
5.Stopper (Choper)
6. Beam splitter AR34
7,8,11. Mirrors
9,12. Germanium detectors
10. InGaAs/InP detectors with is exchanged with standart mirror BK47

Fig. 4. Experimental set-up for measuring the reflectance InGaAs/InP photodiodes

It has exploited an incandescence lamp is the source of white light imaged at the input slit of the monochromator. This lamp was able to cover the spectral range from 800 to 1600 nm and had appropriate blocking filters for second – order wavelengths. After the monochromator, it had placed a linear polarizer and a beam splitter, which serves to monitor temporal fluctuations of light power. A germanium photodiode was used as the monitoring reference photodetector.

The experimental set-up included an optical system of mirrors, which consists of two parts. An upper part (see mirror 7 and germanium photodiode 9) realized monitoring temporal fluctuations of light power. A bottom part (see mirrors 8, 11; InGaAs/InP-photodiode 10, and and germanium photodiode 12) formed an image of the monochromator's exit slit on

the sensitive surfaces of photodiodes. The angle of incidence was equal to 7.4 grades which was accepted as the normal incidence in this train of measurements. The method of measurement consists in comparing the response from a germanium photodiode to the radiation reflected by the InGaAs/InP photodiode with the response from an aluminium standard mirror whose reflectance is known, so that :

$$P(\lambda) \equiv \frac{I_p(\lambda)}{I_m(\lambda)} \rho_m(\lambda) \tag{1}$$

Here, $I_p(\lambda)$ is the response to the light reflected by the InGaAs/InP, $I_m(\lambda)$ is the response to the light reflected by the mirror, and $\rho_m(\lambda)$ is the reflectance of a standard mirror. With this method it has measured the reflectance of photodiodes from different manufacturers. One part of detectors had a round aperture of 5 mm in diameter and the other part had a rectangular aperture of 8 x 8 mm.

7. Analysis of reflectance

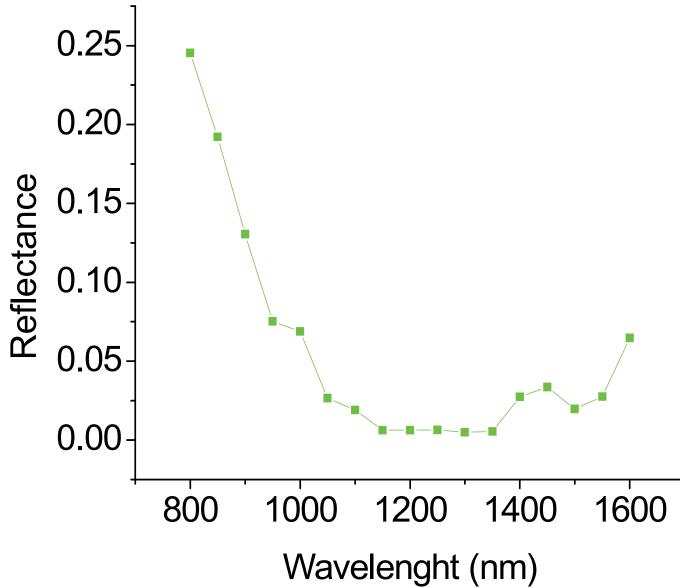
The grade of light polarization at the output the monochromator was different with varying the wavelength, figure 5.2 illustrates spectral dependences of the reflectance, which had been obtained from photodetectors belonging to three different manufacturers. In figures 5.2a and 5.2b that the reflectance of such detectors has a minimum in an area of 1000 – 1600 nm, and they both are related to a structure of layers providing maximal responses in the spectral interval of mayor utility of these detectors in near IR optics communication. The first photodiode, see Figure 5a, whose reflectance was minimized, is more efficient that the second one, see figure 5b.

One can see that this plot presents the other spectrum of reflectance in figure 6, is associated with a photodiode with rectangular aperture. In this case the reflectance has two minima at 1000 nm and 1600 nm, but the reflectance has a maximum between these minima. This photodiode is older than previous ones, and it was produced by the other manufacturer. One can remark that may be it was produced without good enough control, because the structure of layers on the sensitive surface modifies the reflectance.

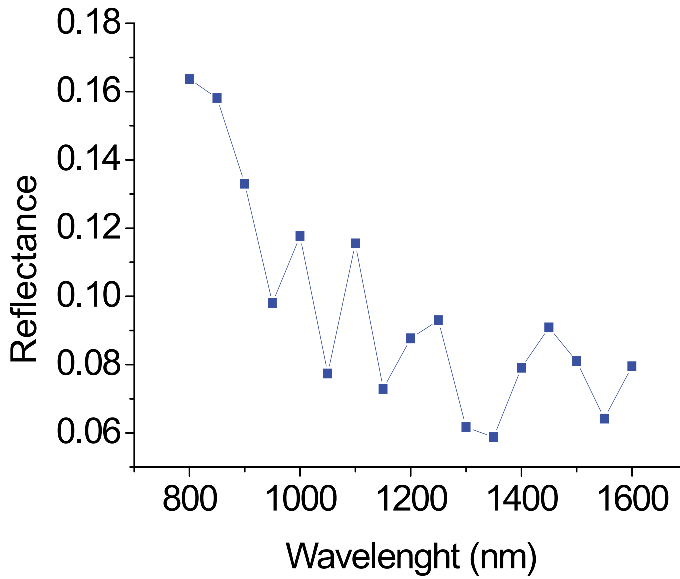
The spectrum of reflectance for photodiodes 1 and 4 is presents in figure 7 which belong to the same manufacturer. The reflectance was measured with linearly polarized and non-polarized lights, and these pair of measurements gives quite similar results. In fact, the difference was equal to approximately 2%. The same results are depicted for the photodiodes 2 and 5, by the second manufacturer. It is important that the results do not depend on the polarization state of the incident light when the angle of incidence is smaller 10 angular degrees.

All spectrums of reflectance for photodiodes 1-6 is presents in the figure 9, with linearly polarized and non polarized lights and is possible to see the different behavior of the photodiodes in the near infrared wavelength.

In fact, in this chapter is studying the behavior of the photodetectors in the near infrared with the linearly polarized and non polarized lights in the case of the polarized lights the angle of incidence is smaller 10 angular degrees and is possible observed it doesn't have changes in the behavior of the reflectance.



(a)



(b)

Fig. 5. Detector with a round aperture with the diameter of 5 mm

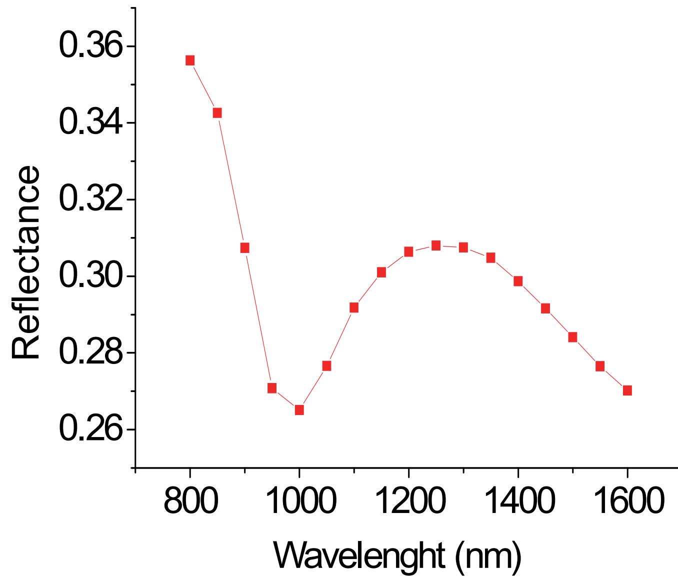


Fig. 6. Detector with a rectangular aperture of 8 x 8mm

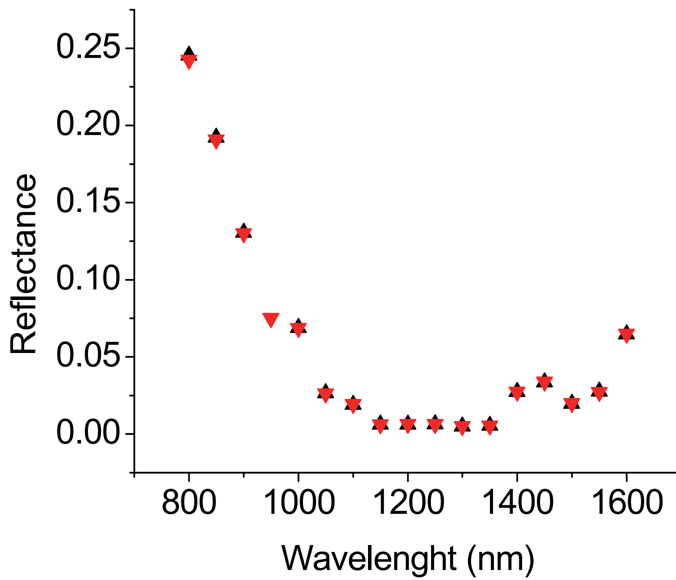


Fig. 7. Spectrum of reflectance for photodiodes 1 and 4

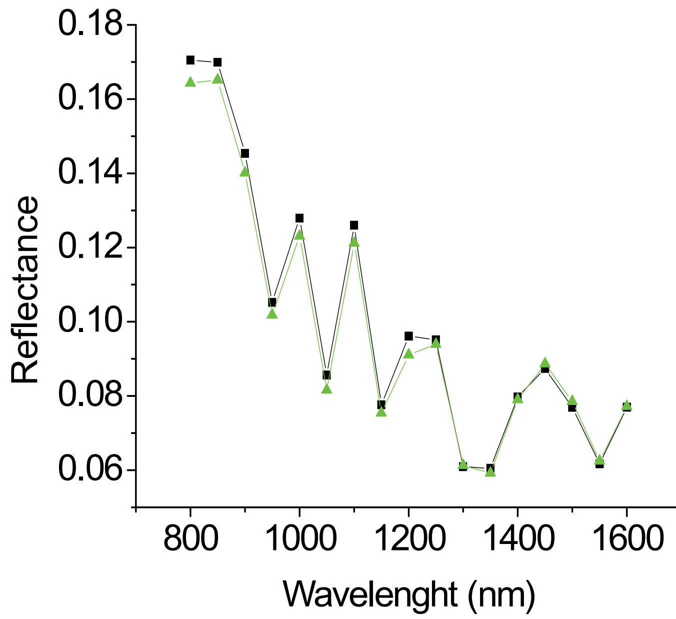


Fig. 8. Spectrum of reflectance of photodiodes 2 and 5.

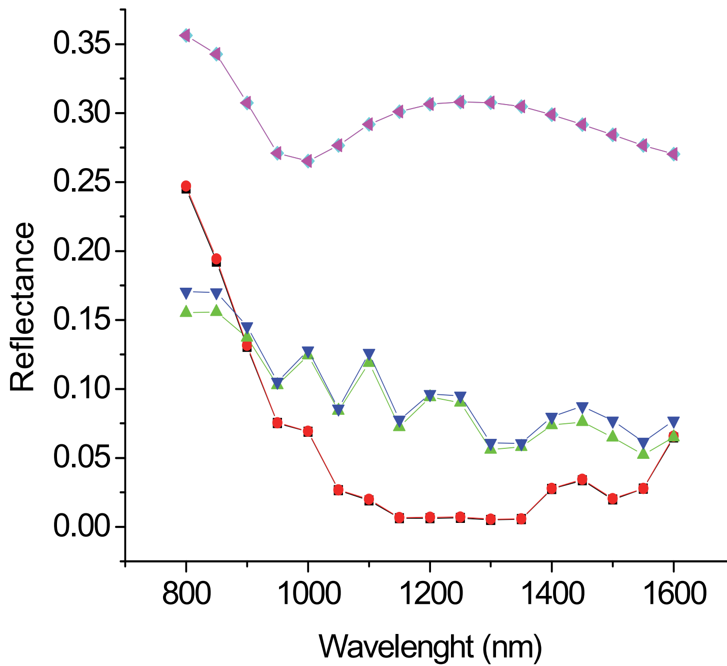


Fig. 9. Spectrum of reflectance of photodiodes 1-6.

7.1 External quantum efficiency

It is obtained from the responsivity values according to the equation:

$$Q(\lambda) \approx \frac{R(\lambda)hc}{\lambda e} \tag{2}$$

Where h, c and e are the usual physical constants and λ is the wavelength. Values obtained are presented in figure 10 for the same detectors as before. It can be clearly seen that the oldest detector (POL) presents a lower external quantum efficiency than the other and that detector GPD presents a higher external quantum efficiency than detector HAM, which starts to decrease its quantum efficiency at a shorter wavelength.

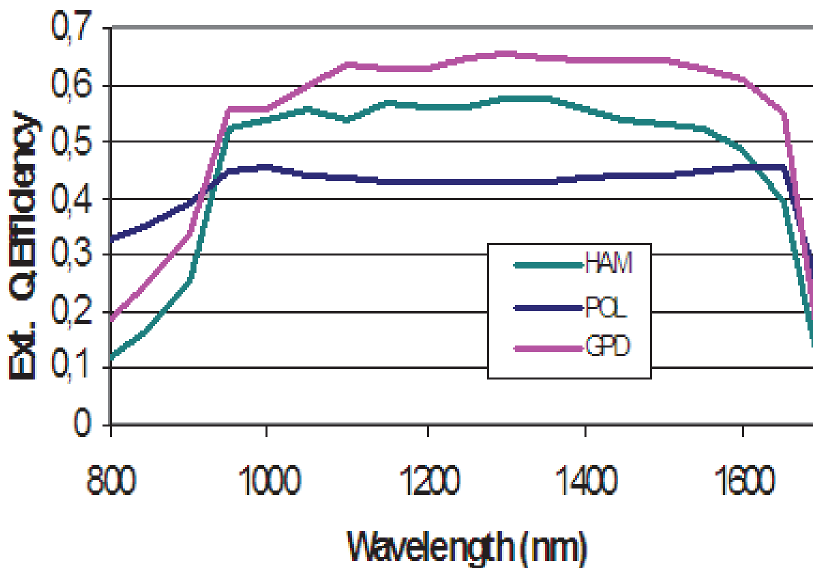


Fig. 10. Spectral external quantum efficiency from responsivity values.

7.2 Photodiode structure

Precise structure is not known. The assumed internal structure for photodiodes is shown in figure 11. It is more than likely that detector POL has got a different structure.

Considering for the photodiodes the structure and refraction index values shown (figure 11 and 12), the reflectance can be fitted by using a multilayer model. Results obtained are shown in figure 13, except for photodiode POL that gave a poorer result. The thickness values obtained from the fit are shown in the table.

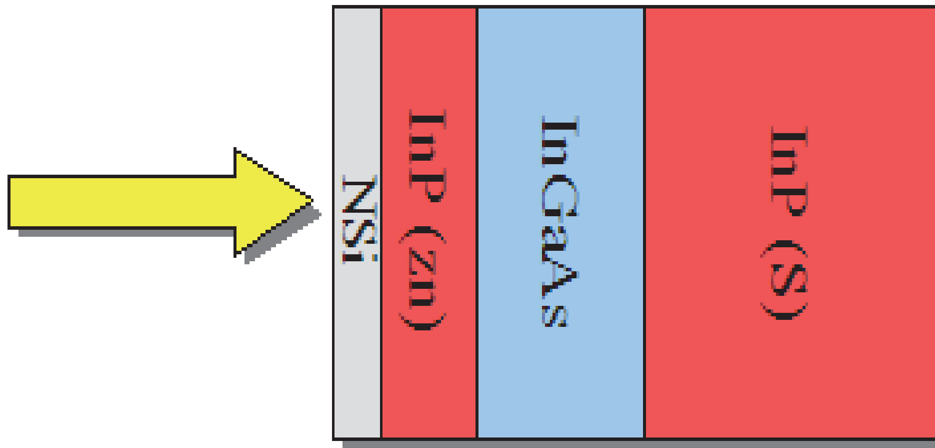


Fig. 11. Possible Structure of InP photodiode.

Spectral Reflectance measured and fitted values for photodiodes HAM and GPD.

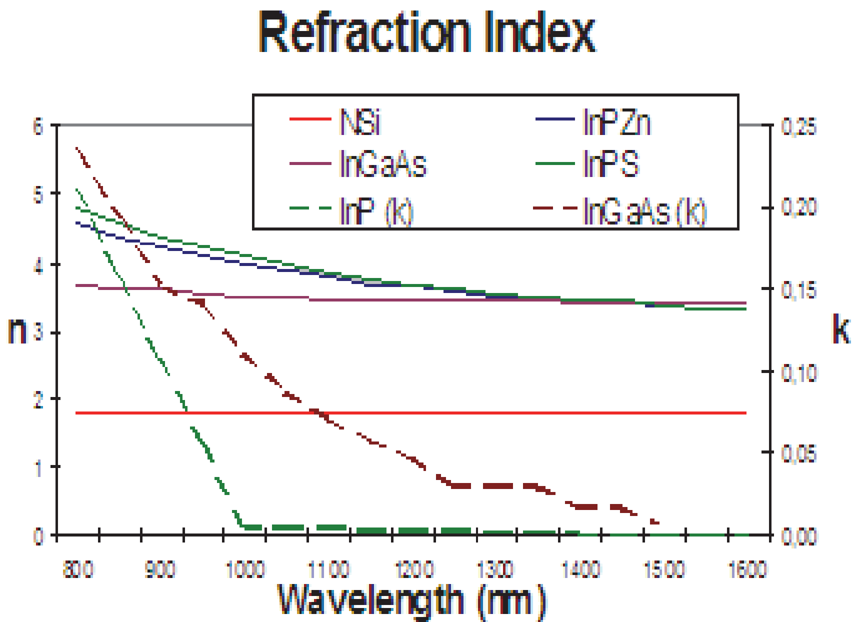


Fig. 12. Materials' Refraction Index

Photodiode	NSi	InP (Zn)	InGaAs
HAM	162.17nm	1203.35nm	1593.2nm
GPD	159.99nm	1200.54nm	1536.7nm

Table 1. The thickness values obtained of InP photodiodes.

7.3 Internal quantum efficiency

Internal quantum efficiency, e(l), is calculated as usual: e(l) = Q(l) / (1-r(l)). Values obtained are shown in figure 13 and 14 for photodiodes HAM and GPD.

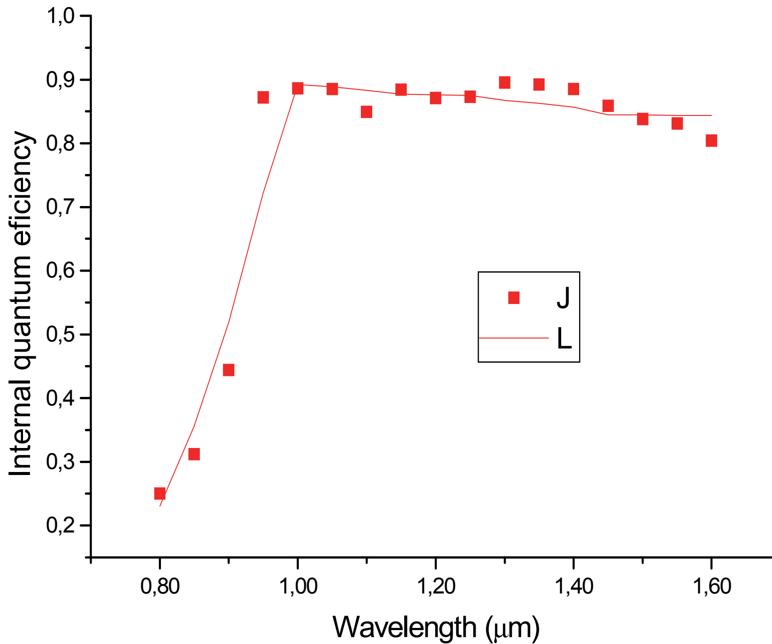


Fig. 13. Internal quantum efficiency of photodiodes HAM experimental values (dots) and fitted values according to the model shown below.

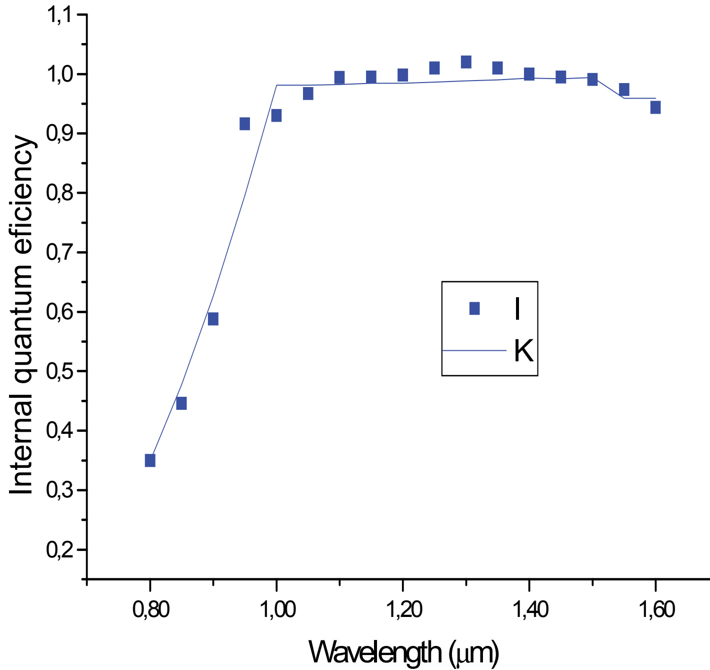


Fig. 14. Internal quantum efficiency of photodiodes GPD experimental values (dots) and fitted values according to the model shown below.

Considering a structure as shown before and a simple model for the collection efficiency of carriers in all regions given by a constant value, Pf,lower than 1 in the first region, 1 in the depletion region and Pb in the back region, and an "infinite" thickness for the diode, e(l) can be calculated by:

$$\varepsilon(\lambda) = P_f(1-\exp(-\alpha T)) + \exp(-\alpha T) - \exp(\alpha T') + \exp(-\alpha' T') - \\ -\exp(-\alpha' D') - \exp(-\alpha D') + \exp(1-P_b)\exp(-\alpha D) \quad (3)$$

Where T is the thickness at which collection efficiency becomes 1, T' is the thickness at which InGaAs region starts, D' is the the thickness at which the InP (S) starts and D is the thickness at which depletion region ends. By fitting the model to internal quantum efficiency value, the following parameters are obtained for every photodiode.

photodiode	Pf	T	T'	D'	D	Pb
HAM	0	0.44	2.19	2.19	11.96	0.844
GPD	0	0.32	1.65	1.62	4351.16	0.960

Table 2. Parameters obtained by photodiode

8. Conclusions

The reflectance was measured with linearly polarized and non-polarized lights, and these pair of measurements gives quite similar results. In fact, the difference was equal to approximately 2%. The same results are depicted for the photodiodes 2 and 5, by the second manufacturer. It is important that the results do not depend on the polarization state of the incident light when the angle of incidence is smaller 10 angular degrees.
In fact in this thesis it are studying the behavior of the photodetectors in the near infrared with the linearly polarized and non polarized lights in the case of the polarized lights the angle of incidence is smaller 10 angular degrees and is possible observed it don't have changes in the behavior of the reflectance.
Reflectance fittinis better for HAM than for GPD. Perhaps the structure model needs some refinement. Internal Quantum Efficiency fitting is not very good at the elbow region. The InP photodiodes' reflectance can be modeled by using a layered optical system having got a first transparent layer and three absorbing ones.
InP photodiodes' internal quantum efficiency can be adequately modeled by a three region structure, where the charge collection efficiency is constant in the two first regions and variable in the last one.

9. References

Trommer, R. Steingrüber, R. Löffler, A. Umbach, (2000).Ultrafast, High-Power WaveguidFed Photodetector with Integrated Spot Size Converter, *Proc. Of the 12th Int. Conf. On InP and Rel. Mat.* (IPRM 2000), Williamsburg, VA, USA, 2000, paper WA2.

A. Beling, H.-G. Bach, G. G. Mekonnen, T. Eckhardt, R. Kunkel, G. Jacumeit, M. Kroh, and J. Berger,(2003) "Fully Packaged InP-based Photodetector for 80/85 Gbit/s RZ Systems," *Proceedings 29th European Conference on Optical Communication (ECOC 2003)*,vol.6, pp. 82-83,September 1-25, 2003, Rimini,Italy.

A. Beling, D. Schmidt, H.-G. Bach, G. G. Mekonnen, R. Ziegler, V. Eisner, M. Stollberg, G. Jacumeit, E. Gottwald, C.-J. Weiske, A. Umbach. (2002). "High power 1550 nm twin-photodetector modules with 45 GHz bandwidth based on InP", Tech. Digest Optical Fiber Commun. (OFC 2002) , pp.274-275, March 17-22, 2002, Anaheim, CA, USA, paper WN4.

A. H. Gnauck, S. Chandrasekhar, J. Leuthold, and L. Stulz,(2003). "Demonstration of 42.7 Gb/s DPSK receiver with 45 photons/bit sensitivity", IEEE Photon. Technol. Lett., 15 (1), pp. 99-101.

A. Beling, H.-G. Bach, D. Schmidt, G. G.Mekonnen, M. Rohde, L. Molle, H. Ehlers, A. Umbach, "High-Speed Balanced Photodetector Module with 20 dB Broadband Common-Mode Rejection Ratio"(2003). Tech. Digest Optical Fiber Commun. (OFC 2003), vol. 1, pp. 339-340, Atlanta, GA, USA, March 23-28, 2003.

A. Beling, H.-G. Bach, D. Schmidt, G.G. Mekonnen, R. Ludwig, S. Ferber, C. Schubert, C. Boerner, B. Schmauss, J. Berger, C. Schmidt, U. Troppenz and H. G. Weber (2003). "Monolithically integrated balanced photodetector and its application in OTDM 160 Gbit/s DPSK transmission",Electron. Lett, 39 (16), pp. 1204-1205.

H.-G. Bach, A. Beling, G. G. Mekonnen, and W. Schlaak,(2002)."Design and fabrication of 60-Gb/s InP-based monolithic photoreceiver OEICs and Modules", IEEE J. Select. Topics Quantum Electron., vol. 8,pp. 1445-1450, Nov./Dec. 2002.
R.Thomas Hawkins II, M. D. Jones,S. H. Pepper, and J. H. Goll, (1991)."Comparison of fast photodetector response measurements by optical heterodyne and pulse response techniques," J. Lightwave Technol, vol. 9,pp. 1289–1294 , 1991.
S. Kawanashi, A. Takada, and M. Saruwatari,(1989). " Wide-band frequency response measurement of optical receivers using optical heterodyne detection," J. Lightwave Technol. , vol. 7, No. 1,pp. 92–98, Jan. 1989.
P. Debie and L. Martens,(1995)."Correction technique for on-chip modulation response measurements of optoelectronic devicees," IEEE trans. on Microwave Theory Tech., vol. 43, no. 6, pp. 1264–1269, June 1995.
G. Unterborsch, A. Umbach, D. Trommer, G. G. Mekonnen,(1997)."70GHz long-wavelength photodetector," in Proc.28th Eur. Conf. on Opt. (ECOC '97), pp. 25-28, Sep. 22–25 1997.
G.Wang, T. Tokumitsu,I. Hanawa, Y. Yoneda, K. Sato, and M. Kobayashi,(2003). " A time-delay equivalent-circuit model of ultrafast p-i-n photodiodes," IEEE Trans. on Microwave Tech., vol. 51, No. 4, pp. 1227–1233, April 2003.
A. Siefke, M. Dahlstrom, U. Westergren, X. J. Leijtens, N. van Melick, and M. K. Smit,"Integration of an 8×10 GHz polarization independentWDMreceiver with HBT-preamplifiers," proc. symp. IEEE/LEOS Benelux Chapter,pp. 25–28,Gent, Belgium.
A. Alping," Waveguide pin photodetectors: theoritical analysis and design criteria,(1989)procedins, vol. 136,Pt.J, No. 3, pp. 177–182, June 1989.
O.Wada, H.Hasegawa, K. Taguchi,(2002). "InP Based materials and devices" John Wiley, ISBN: 0-471-18191-9. Unietd States of America.
J. Campos, P. Corredera, A. Pons, A. Corróns J. L. Fontecha.(1998). "Reflectance dependencies of silicon trap detectors". ,IOP Metrología, 35, 455-460.
P Corredera, M L Hernanz, M González-Herráez,J.Campos(2003). "Anomalous non-linear behaviour of InGaAs photodiodes with overfilled illumination" IOP Metrología 40, S181-S184.
P Corredera, M L Hernanz, J Campos, A Corróns, A Pons and J L Fontecha.(2000). "Comparison between absolute thermal radiometers at wavelengths of 1300 nm and 1550 nm" IOP Metrología. 37. 237-247.
J.Castrellon, G.Paez, M.Strojnik.(2002) "Radiometric analisis of a fiber optic temperatura sensor" Opt. Eng, 41 (6),1255-1261.
R. Goebel, S. Yilmaz, R.Pello.(1996). "Polarization dependence of trap detectors" IOP Metrología, 33,207-13.
E.F.Zalewski, C.R.Duda, "Silicon photodiode device with 100% external quantum efficiency" Appl.Opt.,22, 2867-2873.
J.M.Coutin, F.Chandoul, J.Bastie. (2005)."Characterization of new trap detectors as transfer standards" Proceedings of the 9th international conference on new developments and applications in optical radiometry.
A.L.Muñoz Zurita, J.Campos Acosta, A.S.Shcherbakov, A.Pons Aglio.(2006)."Medida de reflectancia y de fotodiodos de InGaAs/InP". Proc. of the VIII National Meeting on Optics, (Optical Society of Spain), p.27-30.

A.L.Muñoz Zurita, J.Campos Acosta, A.S.Shcherbakov, and A. Pons Aglio.(2007). "Measuring the reflectance and internal quantum efficiency of InGaAs/InP photodiodes in near infrared range". Proc.of the IX International Congress on Advanced Infrared Technology and Applications, (AITA-9), p.1-6.

A.L.Muñoz Zurita, J.Campos Acosta, A.Pons Aglio. A.S.Shcherbakov.(2007). "Medida de la reflectancia de fotodiodos de InGaAs/InP" Óptica Pura y Aplicada, 40(1), 105-109.

A.L.Muñoz Zurita, J.Campos Acosta, A.S.Shcherbakov, and A.Pons Aglio.(2008) "Measuring the reflectance and the internal quantum efficiency of silicon and InGaAs/InP photodiodes in near infrared range". Proc of SPIE Photonics West Conference (January 22-26, 2008, San José, USA).

A.L.Muñoz Zurita, J.Campos Acosta, A.S.Shcherbakov, and A.Pons Aglio. (2008). "External Quantum efficiency of some InGaAs/InP photodiodes", Proc of NewRad2008 congress, MO_P41, (1-3).

2

Physical Principles of Photocurrent Generation in Multi-Barrier Punch-Through-Structures

A.V. Karimov, D.M. Yodgorova and O.A. Abdulkhaev
Physical-Technical Institute of the Scientific Association "Physics-Sun" of the Academy of Sciences of the Republic of Uzbekistan, Tashkent
Uzbekistan

1. Introduction

The reach-through effect representing close up the space charge regions of two adjacent oppositely biased junctions leads to a sharp exponential increase in current from the bias voltage (Sze et al., 1971). Therefore, this effect was originally found in transistor structures was undesirable. But in the further development of electronics, this effect has found many applications in electronic devices. For example, in barrier injection transit-time diodes as dc-current bias (Chu & Sze, 1973; Coleman & Sze, 1971; Presting et al., 1994), in static induction transistors as an extra advantageous current to increase the transconductance of the transistor (Nishizawa & Yamamoto, 1978), in low-voltage transient voltage suppressors as a clamp device (de Cogan, 1977; King et al., 1996; Urresti et al., 2005), in JFET optical detectors as a reset mechanism (Shannon & Lohstroh, 1974, as cited in Lohstroh et al., 1981), in IGFET tetrodes as a modulated current flow (Richman, 1969, as cited in Lohstroh et al., 1981), in punch-through insulated gate bipolar transistors (Iwamoto et al., 2002), in gate-field-controlled barrier-injection transit-time transistors and in light injection-controlled punch-through transistors (Esener & Lee, 1985).
Due to the predominance the diffusion processes in structures with reach-through effect (Lohstroh et al., 1981; Sze et al., 1971) characters of the generation-recombination processes in the space charge regions in these structures, as well as non-stationary processes caused by extraction of the majority carriers and formation of the uncompensated space charge in the base layer are still remain unexplored. To prevent the diffusion processes three-barrier structure was developed, in which the flow of both types of carriers in the structure is limited by rather high potential barriers (Karimov, 1991, 1994, 2002). This allowed us to research in such structures the generation-recombination processes in the space charge regions after reach-through, as well as the influence of illumination on these processes. In these structures is found the internal photocurrent gain (Karimov & Karimova, 2003; Karimov & Yodgorova, 2010), which can not be associated with an avalanche or injection processes. Thus, this section is devoted to disclosing the mechanisms of charge transport and the nature of the internal photocurrent gain in multibarrier reach-through-photodiode structures.
In this section, is given a brief overview of multibarrier photodiode structures, as well as the results of a comprehensive research of the dark and light characteristics of multibarrier reach-through-photodiode structures. On the basis of which is proposed model, which explains the mechanism of charge transport and internal photocurrent gain, as well as some future trends.

2. An overview of multibarrier photodiode structures

The sensitivity and the bandwidth of the photodetector is critical to the overall performance of the receiver. A higher sensitivity translates into a longer distance possible between the last repeater and the receiver, without loss of data. The bandwidth of the photodetector will define the overall upper bandwidth limit of the receiver. There are two major types of the photodetector used in the telecommunication systems today – a p-i-n photodiode and an avalanche photodiode. The sensitivity of the p-i-n photodiode by itself is often too low for long-haul applications, typically, as the bandwidth is increased, the sensitivity is decreased. The alternative to the p-i-n photodiode, the avalanche photodiode, improves the sensitivity of the p-i-n photodiode by an additional section of the chip (section with high enough electric field for the formation of the avalanche multiplication) that provides gain. Depending on the gain of the device the sensitivity can be varied over a few dB without severe penalty in the bandwidth of the device. However, there is an additional noise associated with the gain section of the device which will impact receiver sensitivity. Also, at high gains, the device bandwidth will be limited by the gain-bandwidth product (a typical value of this product is 100 GHz). A typical operating current gain of the gain section of the device is 3 to 10 without penalty in the device bandwidth. In this rang the device is usually RC-limited. One of the first multibarrier structures with internal photocurrent gain is a bipolar phototransistor (Campbell, 1982). A phototransistor can have high gain through the internal bipolar-transistor action, which was significantly improved by utilizing a wide-gap emitter (Chand et al., 1985), or by utilizing punch-through transistors (Esener & Lee, 1985). It should be noted that the inherent to transistors large areas degrades their high-frequency performance.

Semiconductor device with two metal-semiconductor rectifying junctions can also be attributed to multibarrier photodiode structures (Sugeta & Urisu, 1979). In these structures, high performance is ensured by non injecting metal-semiconductor junctions and low capacitance of planar barriers. Non injecting nature of the metal-semiconductor junction suppresses internal photocurrent gain. Presence of the carrier injection in one of the junctions allowed one to obtain photocurrent gain for low-frequency range. Internal photocurrent gain in the high-frequencies has been achieved only when avalanche multiplication is present. However, in case of variation of the parameters of the potential barriers may cause some amplification of the primary photocurrent. The mechanism of the observed internal photocurrent gain can be attributed to the formation of a nonuniform electric field distribution and the separation of light-generated carriers near the anode with simultaneous additional emission of electrons from the cathode (Klingenstein et al., 1994). However, the use of series-connected heterojunctions and metal-semiconductor junctions allows one to control the spectral range of responsivity.

Series connection of the three barriers to longer enough short base layers allows one to obtain the internal photocurrent gain as a photothyristor. However, it is having the S-shaped current-voltage characteristic leads to instability of its parameters, and therefore can only be used as an optical switch. By serial connection of the p-n-junction with a high resistive long-base layer were obtained the injection-based photodiodes whose characteristics are similar to photothyristor characteristics (Vikulin et al., 2008). At the same time these photodiodes are had sufficient internal gain in the prebreakdown region, but decreased high-frequency performance.

Thus, in most multibarrier photodiode structures are inconsistent the high-frequency performance and the photosensitivity, i. e. there is a competitive relationship between them, which leads to the constancy of their product. In this aspect, it would be appropriate to

create a new class of multibarrier photodiode structures that is an alternative to avalanche photodiodes and field phototransistors.

3. Three-barrier reach-through-photodiode structure

Investigated a three-barrier reach-through-photodiode structures on basis of gallium arsenide were produced on base of technology for obtaining abrupt p-n-junctions from epitaxial homolayer or heterolayer p(n)-type which was growth from a liquid phase on substrate n(p)-type (we used a substrates doped with shallow or deep impurities). The carrier concentration in the grown epi-layer (with thickness 1-2 microns) was $5\text{-}7{\cdot}10^{15}$ cm^{-3} and in the substrate $1\text{-}9{\cdot}10^{15}$ cm^{-3}. By evaporation in a vacuum of the translucent layers Ag (70 Å) on both surfaces of structure were obtained rectifying junctions (in some samples to obtain the barrier was used Au). Height of potential barriers were measured by a photoelectric method are 0.6-0.8 eV and are determined by fixing the Fermi level on surface states. As a result, were made three barrier m_1-p-n-m_2-structures with an area of 2-25 mm^2, in which m_1-p- and n-m_2-junctions are physically connected in series, and p-n-junction opposite. Due to the existence of a blocking junction at any bias polarity these structures are able to operate at both polarities of the bias voltage and a double-sided sensitive, i. e., the photocurrent can be taken under illumination from either side of structure. The total capacitance of the structure is close to the value determined by the geometric size of the entire structure and is about 0.2-0.5 pF/mm^2.
The proposed structure is similar to a thyristor, but differs from it in the larger thickness of one of the base regions (the thickness of the n-region is equal to 350 microns), while in the thyristor three barriers are separated by two base regions with a thickness, which is in one order with the diffusion length of minority carriers (Sze & Kwok, 2007). In this case, the smaller thickness of other base region contributes to close up the space charge regions of adjacent junctions before the onset of avalanche multiplication.

4. The dark characteristics

Investigated m_1-p-n-m_2-structure in the initial bias voltages has typical current-voltage characteristics for a structure with three successively connected barriers (i. e. the current transport is determined by the reverse-biased junctions and in the case of the generation mechanism the dependence is a power law with an index equal to 0.5), which are then in voltages above a certain voltage (U_o) changed to a linear dependence as resistors, Fig. 1. In this case, the resistance was determined from the slope of current-voltage characteristics is several orders of magnitude higher than determined by resistivity and the geometric sizes of the base regions, which indicates the existence of potential barriers in the modified structure. However, the observed linear character of current-voltage dependence can not be explained within the existing theories of the barriers.
The observed behaviors of the current-voltage characteristics are associated with the effects taking place in a three-barrier structure when the voltage is increased beyond the reach-through voltage. As is well known for reach-through-structures charge transport through these structures is determined by the minority carriers (in our case electrons) that are emerging from the forward-biased junction. However, in the m_1-p-n-m_2-structure the metal-semiconductor barriers restrict the flow of these carriers. Thus, in a three-barrier reach-through-structures the current density through the structure after the reach-through is determined by double-sided thermal electron emission, i. e. the flow of holes is limited by the left barrier, while the flow of electrons is limited by the right barrier.

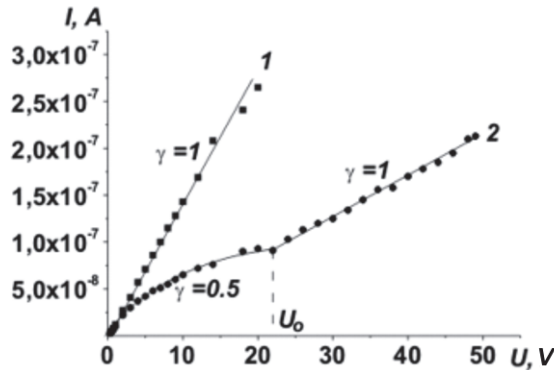


Fig. 1. Measured *I-V* characteristics of the three barrier structure at opposite bias polarities: 1 - (+)m-p-n-m(-); 2 - (-)m-p-n-m(+); γ – power index in $I \sim V^{\gamma}$.

In case of forward biased p-n-junction in the initial bias voltages the space charge regions of the metal-p and p-n-junctions are closed up which is caused by sufficiently thin base-layer. In the further increase of bias voltage the energy bands of the p-n-junction tend to become flat, which leads to a significant increase in current density of electrons from the n-region, Fig. 2. According to the research (Sze et al., 1971), the current density of electrons from the n-region is defined by:

$$j_n^{p-n} = A_n^* T^2 \exp\left(-\frac{q\left(V_{FB} - V\right)^2}{4kTV_{FB}} \right) \quad (1)$$

It should be noted that the current density of electrons incoming to the n-region is limited by the potential barrier of the n-metal-junction:

$$j_n^{m-n} = A_n^* T^2 \exp\left(-\frac{q\varphi_{m-n}}{kT} \right) \quad (2)$$

Therefore, in the n-region adjacent to the p-n-junction there is a strong depletion of the major carriers, which leads to the formation of an uncompensated positive space charge of ionized donors, which in turn attracts electrons from the nearby area leading to the formation of new uncompensated positive space charges, which contributes to the further development of non-stationary processes in n-region. These processes will continue until the establishment of equilibrium between the currents of electrons emerging from the n-region and incoming to the n-region. The required reduction in current density of electrons emerging from the n-region is given by the decreasing of equilibrium concentration of electrons in the n-region, which, while maintaining electrical neutrality of structure becomes possible when the donors go to the neutral state. The conductivity of the n-region becomes close to intrinsic conductivity, which leads to an increase in its resistivity and an increase in the incident in this area bias voltage. As a result, the current-voltage characteristic of the structure becomes close to linear. The degree of depletion of the n-region and thus its resistivity determines by the current density of electrons through the n-metal junction:

$$R_{n-region} = f\left(j_n^{m-n}\right) \tag{3}$$

In case of reverse biased p-n-junction in the initial part of the current-voltage characteristics the dependence is a power law with an index equal to 0.5, which is due to the predominance of the generation processes in the space charge region of reverse biased junction. Above a reach-through voltage the energy bands of the metal-p-junction tend to become flat, which leads to a significant increase in current density of holess from the p-region, which leads to the formation of an uncompensated negative space charge of ionized acceptors. Field of space charge reduces the built-in potential of the n-metal-junction, leading to an increase in

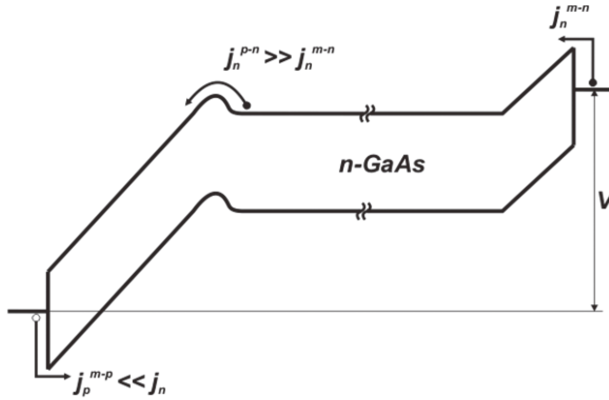


Fig. 2. Energy band diagram of a three barrier structure at bias polarity (+)m-p-n-m(-) after reach-through.

current density of electrons overcoming the barrier of n-metal while the current density of electrons incoming to the n-region is limited by barrier metal-p. As a result, just as in the above case, we have depleted n-region. Thus, the current-voltage characteristic of the structure is changed to a linear one. The degree of depletion of the n-region and thus its resistivity in this case are determined by current density of electrons through a metal-p:

$$R_{n-region} = f\left(j_n^{m-p}\right) \tag{4}$$

Thus, for both polarities the current transport is determined by an identical mechanism. Due to the fact that the barrier height of metal-p-junction is greater than the barrier of n-metal-junction resistance of the structure in the mode of blocking of the p-n-junction is of greater than another mode.

Temperature dependence of the resistance of the structure in the linear region in both modes is described by the function (Fig. 3.)

$$R_C T \propto \exp(f(\frac{1}{T})) \tag{5}$$

As noted above the current flowing through the structure is determined by the resistance of the depleted n-region, which in turn depends on the intrinsic carrier concentration in this

region and current density of electrons through the n-metal- (or metal-p-) junction, which explains the existence of two linear regions with different slopes in this relationship. Activation energy determined from these slopes at low temperatures corresponds to the energy band gap, and at high temperatures to the potential barrier's height. Accordingly, the change in slope with increasing the temperature is caused by prevalence of the thermionic electron current through the metal-semiconductor junction for high temperatures.

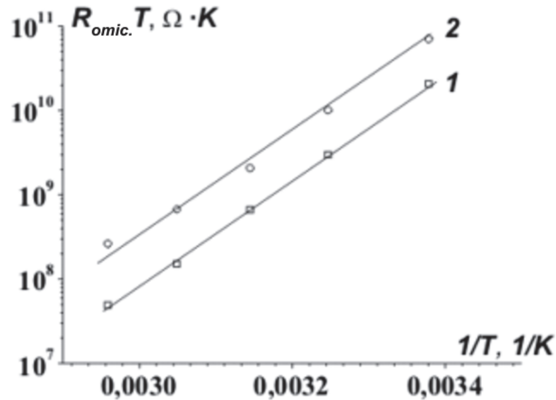


Fig. 3. The resistance of the three barrier structure as a function of temperature at opposite bias polarities in linear region of *I-V* characteristics: 1 - (+)m-p-n-m(-); 2 - (-)m-p-n-m(+)

Temperature coefficient of voltage break point of the current-voltage characteristic has a negative value with a coefficient of -0.098 V/K, so we can assume that the break point is uniquely determined by the reach-through of adjacent junctions of the structure.

Thus, despite the fact that the structure contains a number of series-connected barriers at voltages higher than a reach-through voltage its current-voltage characteristic becomes linear.

5. Light characteristics

Consideration of the structures in the photovoltaic mode, showed that in structure is generated the photo-EMF. The dependence of short circuit current on the intensity of light is nearly linear. Load characteristics in accordance with the current-voltage characteristics are linear, which leads to increased half-width of the maximum output power.

Light characteristics taken from the integral lighting (incandescent lamp) at 100 lux are shown in Fig. 4. In this figure solid line represents the data of the reference photodiode (single-barrier p-n-photodiode) without internal gain. For the researched structures at both polarities of the bias the photocurrent increases with bias voltage to much greater values than in the reference photodiode indicating the presence of internal photoelectric gain. In the reverse-biased p-n-junction at low voltages is taken a tendency to saturation of the photocurrent, as in conventional photodiode without amplification, but when the voltage is increased beyond the reach-through voltage the photocurrent begins steady with voltage.

Curves of light characteristics in case of forward-biased p-n-junction under illumination by side of the p-type layer with increasing light intensity move in parallel toward higher

currents. This can be explained by the fact that from the light emission increases the current density of electrons to the depleted n-region through the n-metal-junction, which leads to a decrease in the degree of depletion and resistance of this region

$$R_{n-region} = f\left(j_n^{m-n}\right) = f\left(j_{dark}^{m-n} + j_{photo}^{m-n}\right) \quad (6)$$

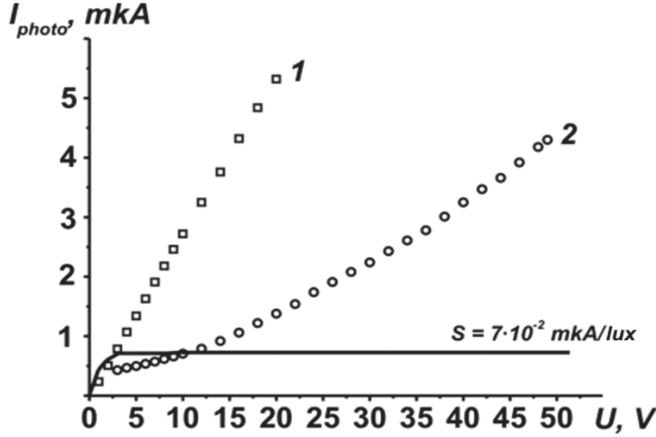


Fig. 4. Photocurrents of the three barrier structure and reference photodiode as a function of bias voltage at opposite bias polarities: 1 - (+)m-p-n-m(-); 2 - (-)m-p-n-m(+)

It is known that the light with energy in the region $\varphi < h\nu < E_g$, which is absorbed in the metal and excites photoemission of electrons from this metal is not absorbed in the bulk of the semiconductor, so changing the illuminated area does not affect the sensitivity of the photodiode.

Analysis of the spectral response of the three-barrier reach-throuch-structure (Fig. 5.) shows that in the both polarities of the bias and regardless from the illuminated surface (top or bottom surface) the photosensitivity is higher when the absorbed light excites photoemission of electrons from the metal than the case when the absorbed light excites intrinsic photogeneration. This agrees well above given mechanism of photosensitivity and due to the fact that the resistance of the depleted n-region is determined only by the current density through the metal-semiconductor junction.

It should be noted that in all the structures external quantum efficiency was greater than unity and indicates the presence of internal gain in these structures. In this case, the observed internal photocurrent gain in the structures does not fit into the framework of the avalanche and the transistor (injection) effects.

6. Mechanism of the internal photocurrent gain

The mechanism of charge transport, depending on the polarity of the operating voltage practically does not differ, which leads to the identity of the internal photocurrent gain in both modes, so we restrict ourselves to the case for direct mode, i. e. forward-biased p-n-junction mode.

In the forward-bias p-n-junction mode because of the narrowness of the p-region metal-p junction and p-n-junction interlock quickly, which leads to an exponential increase in current density of electrons from the n-region with the approach of the bias voltage to flat-band voltage. In this case, the current density in the n-region is limited to the saturation current density of n-metal junction. Since $j_n^{p-n} >> j_n^{m-n}$, the n-region is depleted of electrons, which leads to an increase in resistance of this region. Depletion of electrons continues until the current density of electrons emerging from the n-region decreases did not reach a current density of electrons incoming to the n-region. Thus, the degree of depletion is determined by current density of electrons incoming to the n-region.

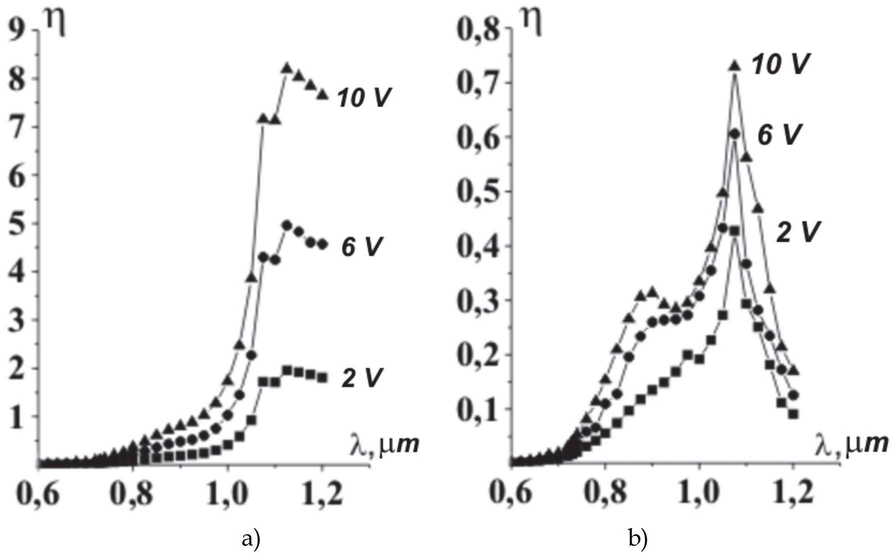


Fig. 5. Spectral response of the three barrier structure at opposite bias polarities and different bias voltages: a) (+)m-p-n-m(-); b) (-)m-p-n-m(+)

Thus, the structure represents a resistance whose value is controlled by a current density of n-metal junction, that is, by nature, similar to the FET, but is controlled by current density. Due to the fact that the intensity of light radiation directly controls the current density, which in turn controls the resistance of the structure, this structure has the internal photoelectric gain.

7. Structures with a heterojunction

Performing a three-barrier photodiode structure based heterojunction allows one to control its spectral respons: take a selective sensitivity or enhance the optical range, in the long or short waves. Increasing energy band gap of the base region can cover the short-wave part of the spectrum, while reducing the energy band gap of p-type region can reduce the sensitivity to shorter wavelengths until the completion of the selective sensitivity is determined by the potential barrier of n-metal. Reducing the potential barrier height of n-metal can expand the optical range to longer wavelengths.

In the case of a three-barrier photodiode Au-n$Al_{0.1}Ga_{0.9}As$-pGaAs-Ag-structure of the current transport mechanism similar to the mechanism of homojunction structure, with the difference that on case of reverse biased n-p-heterojunction with increasing voltage the current dependence is changed from the linear to quadratic, which can be explained by the mechanism of space-charge-limited current transport mechanism. Therefore, in this mode include not observed internal photocurrent gain. However, in a case of forward biased heterojunction the structure has the internal photoelectric gain (Fig. 6.), the magnitude of which increases with both increasing the operating voltage and intensity of light. The dependence of photocurrent on light intensity becomes superlinear character.
Spectral characteristics have also shown that the quantum efficiency in a direct displacement of n-p-heterojunction regardless of the surface to be illuminated in a broad spectral range from 0.95 to 1.3 eV (0.8 to 1.6 microns) unchanged, Fig. 7. By increasing the applied voltage to 65 volts at a radiation power of 0.2 mW/cm^2, the quantum efficiency of the structure increases to 2.77, i. e. there is an internal photoelectric gain.

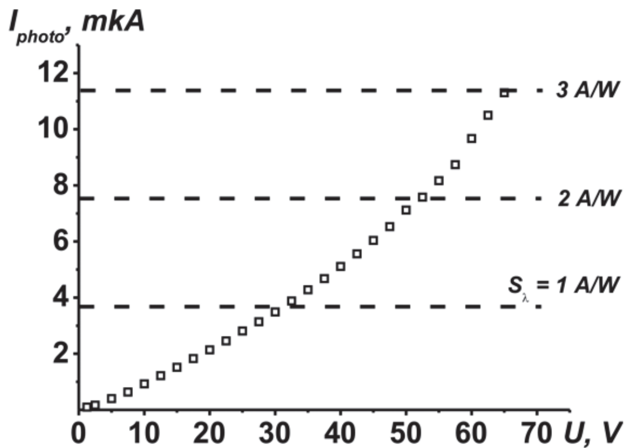


Fig. 6. Photocurrent of the three barrier structure with heterojunction as a function of bias voltage

As noted above, in the structures long base region is depleted of majority carriers and the impurity goes into a neutral state. To verify this situation have been investigated with the basic structure of the area containing deep impurity levels of oxygen. In this case, the compositions heterolayers and metals were chosen so that the height of the barriers have similar values and do not affect the current transport mechanism, and allowed us to determine the influence of deep impurity. In Au-p$Al_{0.05}Ga_{0.9}In_{0.05}As$-nGaAs:O-Ag-structure for both polarities of bias the current-voltage characteristics, spectral response and capacitance-voltage characteristics were identical. Spectral characteristics when excited by heterolayer show that the maximum photocurrent due to excitation of carriers from metal with a characteristic tail of the excitation of carriers from deep levels of oxygen, Fig. 8.
Raising the temperature leads to a broadening of the spectral characteristics of a clear manifestation of the photocurrent in the impurity region of the spectrum caused by oxygen levels (Fig. 9.), and for a given temperature increase of confining the illuminated metal-semiconductor junction bias voltage leads to a simultaneous increase in the photocurrent spectrum in the whole range with simultaneous spectral broadening.

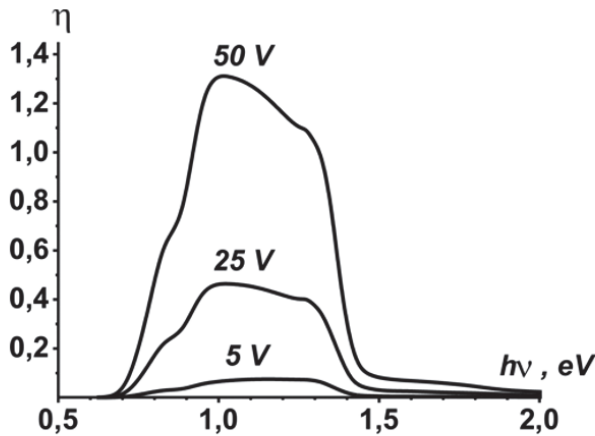


Fig. 7. Spectral response of the three barrier structure with heterojunction at different bias voltages and in bias polarity (+)m-p-n-m(-)

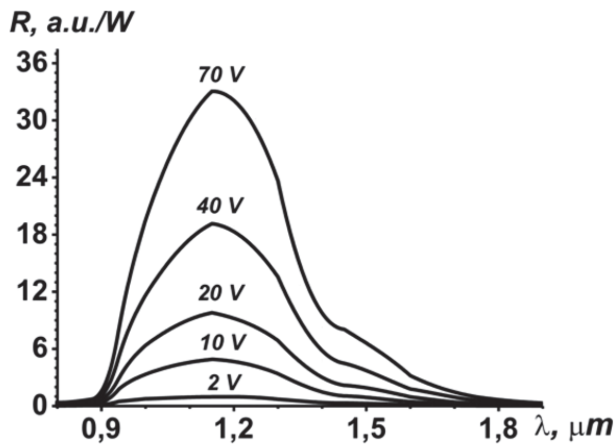


Fig. 8. Normalized spectral response of the three barrier structure with deep impurity levels of oxygen at different bias voltages and in bias polarity (+)m-p-n-m(-)

8. Some perspectives of multibarrier photodiode structures

Multibarrier photodiode structure with an appropriate choice of design parameters may provide a basis to create new structures with improved properties that are of interest for micro and optoelectronics.

Low-capacitance current-controlled transistor can be created can be created by forming a planar rectifying and ohmic contacts to the surface of a thick n-type region of the m_1-p-n-structure. The result will be m_1-p-n-m_2-structure with an ohmic contact to the base n-type region. Capacitance of the structure will be determined by the geometric dimensions of the structure. It creates a voltage forward bias p-n-junction is applied to the electrodes of the

potential barriers, and input to the ohmic contact and the contact potential barrier. By analogy with the field-effect transistor contact potential barrier, one might say, carry out the role of the drain and source, an ohmic contact – the role of the gate. However, in contrast to the FET output characteristics are controlled by the current of the gate. In the absence of the input signal through the structure is minimal and determined by the resistance base and lockable metal-semiconductor interface.

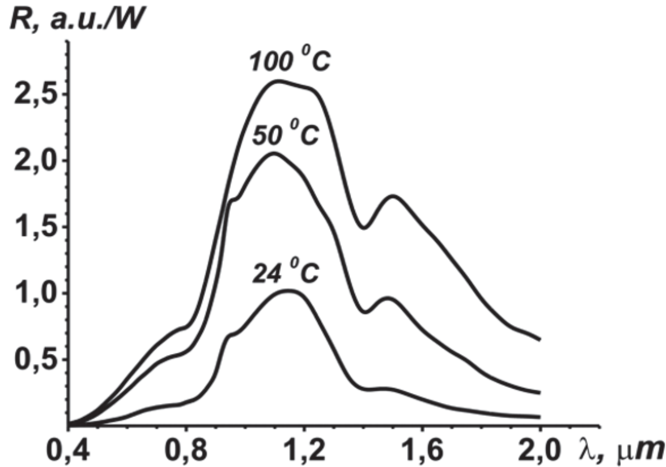


Fig. 9. Normalized spectral response of the three barrier structure with deep impurity levels of oxygen at different temperatures and in bias polarity (+)m-p-n-m(-) at 2 V.

Small change in the gate current, i.e. in the current flowing through the barrier n-m_2 leads to a stronger change in the total current through the structure. At the same time the output characteristics are obtained with characteristics similar to the static induction transistor.

In another embodiment, instead of the potential barrier is proposed to form a semiconductor n-p-junction. As a result multibarrier photodiode will be an m_1-p-n-p structure with an ohmic contact to n-base. The operating mode will create a voltage applied to the contacts of the barrier and the p-contact area. Useful signal will be fed to the resistance connected to the base and the p-region, or as a signal required for the gain can be a light signal fed to the n-p-junction, where you will create short-circuit current is proportional to the intensity of the radiation. As a result, the output is the amplified signal, ie, the proposed structure will have multibarrier reinforcing properties.

Highly sensitive photodetector can be created by the serial connection m-p-n-structure to p-n-junction. The result is a four barrier m-p-n-p-n-structure including a three-barrier m-p-n-p and bipolar n-p-n-structure. Operating voltage is applied to the external contacts m_1 and n-type region with a positive polarity to the contact potential barrier. In the dark current through the structure is determined by the electronic component of the collector p-n-junction, where the electrons have a significant barrier. Under illumination of the collector junction is created that matches the sign of the photocurrent to the dark current and summed. Coefficient of internal photoelectric amplification will consist of works of the gain on the part of a three-barrier transistor gain.

Multibarrier photodiode structures are sensitive to impurity and intrinsic emission can be created by the formation of a nonuniform distribution of deep impurities in a long base n-region in m-p-n-structure. The structure is a thin stripe of p and a thick n-type lightly doped layer (300 microns), which create internal barriers without illumination, did not exceed a few kT. However, if covering own light barrier height can prevail kT, leading to a reduction in current through the structure. When excited by light in the impurity region through the structure will increase.

9. Conclusion

In multibarrier mpnm-photodiode structures with the effect of closure of adjacent oppositely biased junctions, the mechanism of charge transport is determined by the depletion of the major carriers in the base, leading to the development of transitional processes in n-type region with the subsequent transition to a neutral donor state. The degree of depletion n-region and its specific resistance are determined by current density of electrons emerging from semiconductor-metal junction; the dependence of current on voltage obeys a power law with an exponent close to unity and is due by the day the main part of the external voltage on the depleted n-region.
Determining the noise and frequency properties of photodiode structures low capacitance and dark current distinguishes multibarrier structure compared with other types of detectors.
Performing a three-barrier photodiode structure based heterojunction allows you to control its spectral range, a selective sensitivity and enhance the optical range, in the long or short waves. Reducing the height of the barrier metal-semiconductor optical range can be extended to longer wavelengths.
In multibarrier photodiode structures of the internal photoelectric amplification controlled operating voltage and an order of magnitude more sensitive unijunction diode photodiode. For their work does not require any cooling at room temperature provides the required operating modes defined spectral regions (0.9 microns, 1.3 microns, 1.5 microns) with low values of capacitance of the order 0,2-0.5 pF/mm^2. In this case, the dark current for a voltage of 100 V was 40-100 nA. Internal photoelectric amplification of photocurrent is provided from the outset the applied voltage, ie, they possess sufficient sensitivity to record from low supply voltages (5 V). Due to the high input resistance are easily switched with the field-effect transistors and integrated circuits.
Thus, in the above material is presented original experimental data on the principles of creating improved multibarrier photodiode structures, some features of their photoelectric characteristics when exposed to light and heat radiation, the results of the analysis of processes of charge transport and photocurrent gain.

10. References

Campbell, J. C. (1985). Phototransistors for lightwave communications. *Semiconductors and Semimetals,* Vol.22D, p. 389-447.

Chand, N.; Houston, P.A. & Robson, P.N. (1985). Gain of a heterojunction bipolar phototransistor. *IEEE Transactions on Electron Devices,* Vol.32, No.3, p. 622-627.

Chu, J. L. & Sze, S. M. (1973). Microwave Oscillation in pnp Reach-Through BARITT Diodes. *Solid-State Electronics*, Vol.16, pp. 85.

Coleman, D. J., Jr. & Sze, S. M. (1971). A Low-Noise Metal-Semiconductor-Metal (MSM) Microwave Oscillator. *Bell Syst. Tech. J.*, Vol.50, pp. 1695.

de Cogan, D. (1977). The punchthrough diode. *Microelectronics,* Vol.8, No.2, pp. 20-23.

Esener, S. & Lee, S. H. (1985). Punch-through current under diffusion-limited injection: analysis and applications. *Journal of Applied Physics,* Vol.58, No.3, pp. 1380-1387.

Iwamoto, H.; Haruguchi, H.; Tomomatsu, Y.; Donlon, J. F. & Motto, E. R. (2002). A new punch-through IGBT having a new n-buffer layer. *IEEE Transactions on Industry Applications,* Vol.38, No.1, pp. 168-174.

Karimov, A. V. (1991). Karimov's three-barrier photodiode. *USSR Patent,* No.1676399.08.05.

Karimov, A. V. (1994). Karimov's three-barrier photodiode. *Uzb Patent*, No.933.15.04.

Karimov, A. V. (2002). Three-barrier photodiode structure. *15 International Workshop on Challenges in Predictive Process simulation,* pp. 71-72. Prague, Czech Republic, 13-17 October.

Karimov, A. V. & Karimova, D. A. (2003). Three-junction Au/AlGaAs(n)/GaAs(p)/Ag photodiode. *Materials Science in Semiconductor Processing*, Vol.6, No.1-3, pp. 137-142.

Karimov, A. V. & Yodgorova, D. M. (2010). Some features of photocurrent generation in single- and multibarrier photodiode structures. *Semiconductors,* Vol.44, No.5, pp. 647-652.

King, Y.; Yu, B.; Pohlman, J. & Hu, Ch. (1996). Punchthrough diode as the transient voltage suppressor for low-voltage electronics. *IEEE Transactions on Electron Devices,* Vol.43, No.11, pp. 2037-2040.

Klingenstein, M.; Kuhl, J.; Rosenzweig, J.; Moglestue, C.; Hülsmann, A.; Schneider, Jo. & Köhler, K. (1994). Photocurrent gain mechanisms in metal-semiconductor-metal photodetectors. *Solis-State Electronics*, Vol.37, No.2, pp. 333-340.

Lohstroh, J.; Koomen, J. J. M.; Van Zanten, A.T. & Salters, R. H. W. (1981). Punch-through currents in P+NP+ and N+PN+ sandwich structures—I: Introduction and basic calculations. *Solis-State Electronics*, Vol.24, No.9, pp. 805-814.

Nishizawa, J. & Yamamoto, K. (1978). High-frequency high-power static induction transistor. *IEEE Transactions on Electron Devices,* Vol.25, No.3, pp. 314-322.

Presting, H.; Luy, J.-F.; Schäffler, F. & Puchinger, J. (1994). Silicon Ka band low-noise BARITT diodes for radar system applications grown by MBE. *Solid-State Electronics,* Vol.37, pp. 1599.

Sze, S. M.; Coleman, D. J. & Loya, A. (1971). Current Transport in Metal-Semiconductor-Metal (MSM) Structures. *Solid-State Electronics*, Vol.14, pp. 1209.

Sze, S. M. & Kwok, K. Ng. (2007). Physics of Semiconductor Devices. John Wiley & Sons, Inc., Hoboken, New Jersey.

Sugeta, T. & Urisu, T. (1979). High-Gain Metal-Semiconductor-Metal Photodetectors for High-Speed Optoelectronics Circuits. *IEEE Trans. Electron Dev.*, Vol.26, pp. 1855.

Urresti, J.; Hidalgo, S.; Flores, D.; Roig, J.; Rebollo, J. & Mazarredo, I. (2005). A quasi-analyticl breakdown voltage model in four-layer punch-through TVS devices. *Solis-State Electronics*, Vol.49, No.8, pp. 1309-1313.

Vikulin, I. M.; Kurmashev, Sh. D. & Stafeev, V. I. (2008). Injection-based photodetectors. *Semiconductors,* Vol.42, No.1, pp. 112-127.

3

Photon Emitting, Absorption and Reconstruction of Photons

Changjun Liao, Zhengjun Wei and Jindong Wang
University Laboratory of Guangdong, School for Information and Optoelectronic Science and Engineering South China Normal University, Guangzhou China

1. Introduction

Photon cannot keep itself unchanged from emission to absorption. The information encoded on the photon is also changed due to interaction with environments. There has no definitely demonstration that the photon absorbed is the original one from ideal light source since the quantum mechanics itself is an indeterminate theory that the physical measurement is only the probability. Any operation on photon before detector involves unavoidably loss that means a quantum permutation with environments. Although a photon is detected with same energy the phase uncertainty exists. Section 2 describes the single photon sources and the questions about the conception of photons. The third section describes the quantization of electromagetic fields that makes the basis of the Fork states in that the number of the photon is considered. In Section 4, the representation of the photon in space is described by a complete set of eigenfunctions which represent the fundamental modes with different eigen values. Concept of optical modes is a result of quantization of electromagnetic fields. Optical modes can be occupied by photons in different way in comparison with levels in atomic system in which Coulomb interaction considered. Bunching and anti-bunching are considered as the fundamental properties of the optical modes in Section 5. Based on these theories, several applications are considered. Directional emission of single photons is considered in Section 6, study on single photon detectors is presented in Section 7, multipartite entanglement and its application in quantum key distribution is introduced in Section 8. The fluctuation in vacuum, dephase and decoherent are considered in Section 9. General consideration of reconstructions of photons including coherent combination and interfere coherently, resonant-enhanced density of photons are put in the last Section. Here in this chapter photons are considered as field quantum theory instead of that quantum field theory. Photons consist of mode fields which are quantized.

2. Sources of single photons

The concept of Photon has important contribution to the foundation of quantum mechanics which can be seen in any textbook of quantum mechanics (Greiner, 2001). The "particles" of light are called quanta of light or photons which are recognized to have wave-particle duality, a typical feature of a quantum system. The concept of photon was soon

demonstrated by photoelectric effect. This effect was interpreted by Einstein with his famous formula

$$E = \hbar(\omega - \omega_a) = h(\nu - \nu_a) \tag{1}$$

where E is a discrete quanta of light with energy of $\hbar\omega$, the Planck's constant $h = 2\pi\hbar$. The quantized energy was soon demonstrated with many experiments, including the Compton effects, the Ritz Combination principle, the Franck-Hertz experiment, etc.
Photons are ideal for quantum information applications due to its high transmission speed and easy to be coded with quantum information.
Single photons and entangled photon pairs are important for quantum information. Single photons are usually used in quantum key distribution (QKD) system as quantum information carriers to ensure the security of a key distribution system based on quantum mechanics principles (Gisin et al., 2002). The quantum mechanics assured that a single photon can not be divided and a single photon can not be cloned either (Wootters & Zurek, 1982). Furthermore, a single measurement is not enough for determining the quantum state with certainty, any measurement provides only the probability if the state is unknown and the original state changed after measurement and can never be recovered again. Therefore study on generation and detection of single photons is extremely important.
Light sources are everywhere. But real single photon sources are facing technical challenges. Among many methods to provide single photons, three kinds of single photon sources have attracted much attention: the faint optical pulses, spontaneous down conversion photon pairs and quantum dot.

2.1 Faint laser pulses

In practice, the single photons are usually produced by precisely controlled attenuation of laser pulses to a very weak level and assume at that level the photon follow Poisson distribution.

$$P_n = \frac{\mu^n}{n!} e^{-\mu} \tag{2}$$

where, P_n are the probability of the pulse containing n photons with mean photon number of μ (Walls & Milburn, 1994). The probability of containing more than one photon in faint pulses can be made arbitrarily small. For example, with mean photon number of 0.1 as quite usual, there is only 5% of the nonempty pulses contain photons more than one, most of the nonempty pulses contain only single photon. These single photon sources are pseudo-single-photons or correctly called faint laser pulses.

2.2 Down-conversion photon pairs

Single photon generators using correlated photon pairs generated by the spontaneous parametric down-conversion are widely reported (Mason et al., 2002; Migdall et al., 2002; Pittman et al., 2002; Walton et al., 2001). In particular, a periodically poled lithium niobate waveguide has high probability for generation of photon pairs at 1550 nm (Tanzilli et al., 2001; Yoshizawa et al., 2003; Mori et al., 2004). A photon pairs includes a signal photon and an idler photon correlated in time. Therefore the detection of idler photon can be used to control an optical switch so that the signal photon can surely emit in time. However, the

number distribution of the photon pairs follows also Poissonian statistics. A photon number resolving detector is needed for idler photon if single photon emitting should be guaranteed as required in quantum information applications. Photon number resolving detection faces also technical challenges although several kinds of photon number resolving detectors have been reported, they still far from commercially available (Kardyna et al., 2007; Miller et al., 2003).

2.3 Single quantum dot emission

In theoretical consideration, single photons should come from a single transition between two single state levels by one electron in single quantum cavity, such as single atom, quantum dot, etc. It might be properly called "turn style" or "photon gun".

A single photon generation using semiconductor quantum dot has been reported (Santori et al., 2001). Electrically driven single-photon source has been demonstrated experimentally (Zhiliang Yuan et al., 2002). In their experiments, at low injection currents, the dot electroluminescence spectrum reveals a single sharp line due to single exciton recombination (one electron and one hole within a quantum dot), while another line due to the biexciton emerges at higher currents. The second order correlation function of the diode demonstrated anti-bunching under a continuous drive current with Hanbury-Brown and Twiss arrangement (Hanbury-Brown & Twiss, 1956). But the efficiency of collecting the emitted photons is low since the emission from single point diverges in all directions. The reported collection efficiency is about 0.014, and the emitting photons are not at the communication wavelengths. Furthermore, the device emitting single photons with quantum dot is technically complex so that the faint laser pulses are considered as practical single photon sources (Zbinden, 2002).

Optical patch antenna has been proposed for directional emission of single photons and experimentally demonstrated (Esteban et al., 2010; Curto et al., 2010). In their experiment, a single quantum dot emitter is coupled to a nanofabricated Yagi-Uda antenna that resulting quantum-dot luminescence is strongly polarized and highly directed into the high-index glass substrate. Questions are how is the photons reformed or reconstructed so that the divergence changed?

In practical application, the quantum state of a photon is formed by encoding phase information on part of the photon and then recombining the partite so that can be detected at a specified detector (Muller et al. 1997 & Hughes et al., 2000). That means a dividable photon.

2.4 Questionable properties of the single photon

The single photon emitting from single quantum dot can be predicted by the spectrum of the luminescence where only one spectrum line from single exciton exists. The single exciton contains only one electron and one hole so that their recombination can only emit one photon. The results of Hanbury-Brown and Twiss measurement indeed demonstrated no two photons emitted at the same time. However, it can be considered as due to Pauli principle of Fermion, can not be taken as a demonstration of photon antibunching.

The collection efficiency of the single photons needs to be explained. Even the collection efficiency has been increased to about 80%, what is the mean of loss in the photon collection? Is part of the photon lost or the collection is only complete photons but with very low probability. How to increase the efficiency of collecting photons?

There once more appear the conceptual and philosophical problems of quantum mechanics. There is not clear that if the photons are robust enough so that the collection operation obtains single complete photons by probability, or each of the single photons has lost part of their energy and the detected single photon is somehow reconstructed that is follow the measurement theory developed by Niels Bohr and his colleagues in Copenhagen, saying that it is impossible to separate the quantum mechanical system from the measuring apparatus.

Same question about transmission loss appears: is there exist some single photons they are robust enough until being detected, that the transmission loss of the photons should be quantized based on single photon or all the single photon loses part of their energy gradually until too weak to be detected. The photon has been encoded to an eigen-state so that it should be definitely detected in according to the protocol. The practical results are that the loss increased with transmission distance, and the error bit rate increased also with the transmission distance. For example, in a report, bit error rate of 2.3% with communication distance of 15 km rises to 4.1% with distance of 65 km (Namekata et al., 2007). This phenomenon can not be simply explained only due to detectors detected more empty pulses after more photon lost in longer transmission distance. This is due to an effect well known as dephase that the transmission photon interaction with environment or quantum permutation. The photon lost part of its energy and combined with equal part of energy with phase unknown from vacuum fluctuation. In fact, there has not definitely been demonstrated that the absorbed photon at the detector is exactly the original one.

Many experiments show that photons are sensitive to environment and the absorption at the detector is a complex process. Many experiments show also that photon emission and absorption usually contain multi-photon interaction, especially in nonlinear process. For example, photons with higher energy down converted to Twin photons in form of entangled states (Neves et al., 2005). Multi-photon absorption has been successfully used for imaging with high resolution and micro-fabrication (Yi et al., 2004). Photons being scattered to emitting a photon in different wavelength such as Raman scattering or Stockes scattering are well-known optical phenomena and extensively been used. For example, single-shot measurement of revival structures of molecules by sequential stimulated Raman transition (Zumuth et al., 2005). Photon emission and absorption are usually in company with the emission or absorption of phonons, for example, photo-acoustic topography has been successfully used for imaging of nanoparticle-containing object (Zh. Yuan, 2005). Multi-photon absorption has been taken as nonlinear phenomenon. The theoretical calculation of three photon absorption is quite agreed with the experiments (Cronstrond & Jansik, 2004).

Photons are too fragile to be trapped to be study. Nevertheless, the structure of the photon has been considered by Gong Zutong in 1980. a English version of his paper (Gong, 1999) was published for his centennial, but his ideal was from even earlier study in 1933 and based on the elementary quantum theory (Chao, 1933). A photon was thought to consist of a positive and a negative photinos so that can explain many properties of the photon.

The nature of the photon has attracted much attention in recent years. A special issue of Optics & Photonics was published named "The nature of Light: What is a photon?" in which six feature papers are included (Roychoudhuri & Roy, 2003).

In the field of quantum information including quantum communication and quantum computation, photons are ideal information carriers that they can be easily coded into different quantum state and transmit a long distance. The most serous problems of the quantum information transmission and processing are due to the dephase and/or

depolarization that result in quantum bit error rate of more than a few percent in comparison with classical communication where the bit error rate has decreased to less than 10^{-9}. The quantum bit error rate has its special physical origin and need to be studied in more detail. The preparation, transmission and detection of quantum states of the photons in the quantum information system with high fidelity are essential (Combes &Torner, 2005; Torres & Torner, 2005; Eiseman et al. 2004).

A concept of reconstruction of photons is introduced in this chapter to discuss quantum state of photon and its detection, to analyze the origin of the bite error, including some technical detail in the single photon detectors.

3. Quantization of electromagnetic field

There are mainly two kinds of particles in according to their statistical properties, the Fermion and boson. Photons possess only electromagnetic energy, and mediate electromagnetic interaction. Therefore, photons belong to the type of bosons and are the most ubiguitous bosons. Photons possess discrete energy which can also be deduced from quantization of electromagnetic fields (Yariv, 1988). The total energy of the classical electromagnetic fields, the Hamiltonian

$$H = \frac{1}{2}\int_V (\mu \mathbf{H}\cdot\mathbf{H} + \varepsilon \mathbf{E}\cdot\mathbf{E})dV \tag{3}$$

where, μ is the magnetic permeability, ε is the dielectric constant, $\mathbf{H}$ and $\mathbf{E}$ are the magnetic vector and electric vector respectively. The integral performs over volume V in consideration. The normal mode expansions of $\mathbf{H}$ and $\mathbf{E}$ are

$$\mathbf{E}(\mathbf{r},t) = -\sum_a \frac{1}{\sqrt{\varepsilon}} p_a(t)\mathbf{E}_a(\mathbf{r}) \tag{4}$$

$$\mathbf{H}(\mathbf{r},t) = \sum_a \frac{1}{\sqrt{\mu}} \omega_a q_a(t)\mathbf{H}_a(\mathbf{r}) \tag{5}$$

respectively, where ω_a is the radian oscillation frequency of the ath mode. The normal modes are normalized to meet the orthogonal condition:

$$\int_V \mathbf{H}_a \cdot \mathbf{H}_b dV = \delta_{a,b} \tag{6}$$

$$\int_V \mathbf{E}_a \cdot \mathbf{E}_b dV = \delta_{a,b} \tag{7}$$

The magnetic vector and electric vector $\mathbf{H}$ and $\mathbf{E}$ in Eq.(3) are substituted with their normal mode expansions Eq.(4) and Eq.(5) leading to

$$H = \sum_a \frac{1}{2}\left(p_a^2 + \omega_a^2 q_a^2\right) \tag{8}$$

Therefore, the Hamiltonian of the electromagnetic fields equals to a sum of harmonic oscillator Hamiltonians. The variable p_a canonically considered as momentum, and q_a is the canonical coordinate. They are conjugate variables of a quantum mechanical harmonic oscillator, connected by the commutator relations:

$$[p_a, p_b] = [q_a, q_b] = 0 \tag{9}$$

$$[q_a, p_b] = i\hbar\delta_{a,b} \tag{10}$$

Similar to the quantum mechanical harmonic oscillator, the creation operator a_l^+ and the annihilation operator a_l for the electromagnetic fields can be defined:

$$\begin{aligned} a_l^+(t) &= \left(\frac{1}{2\hbar\omega_l}\right)^{1/2} \left[\omega_l q_l(t) - i p_l(t)\right] \\ a_l(t) &= \left(\frac{1}{2\hbar\omega_l}\right)^{1/2} \left[\omega_l q_l(t) + i p_l(t)\right] \end{aligned} \tag{11}$$

Solving these equations for p_l and q_l which are inserted into Eq.(8), the result is

$$H = \sum_l \hbar\omega_l \left(a_l^+ a_l + \frac{1}{2} \right) \tag{12}$$

There has an important Hermitian operator $\hat{n}_{\mathbf{k}s} = a_l^+ a_l$, its eigen value is an positive integer (Mandel & Wolf, 1995).

$$\hat{n}_{\mathbf{k}s} |n_{\mathbf{k}s}\rangle = n_{\mathbf{k}s} |n_{\mathbf{k}s}\rangle . \tag{13}$$

where, the subscript **k** represents the wave vector, and s represents the polarization. $n_{\mathbf{k}s} = 0,1,2,\cdots,\infty$. Therefore, the state vector $|n_{\mathbf{k}s}\rangle$ is number state, or Fock state. The ground state $|0\rangle$ is called as vacuum state. The energy in the vacuum state is not really zero,

$$\langle 0|H|0\rangle = \frac{1}{2}\sum_l \hbar\omega_l \tag{14}$$

It is an average value over all possible frequencies. It actually represents the vacuum fluctuation, important in the quantum information theory. The annihilation and the creation operators acting on the Fock states result in one photon change in the states that

$$\begin{aligned} a_l |n_k\rangle &= n_l^{1/2} |n_k - 1\rangle \\ a_l^+ |n_l - 1\rangle &= n_l^{1/2} |n_l\rangle \end{aligned} \tag{15}$$

The probability distribution of photons in a coherent state is the Poisson distribution as shown in Eq.(2).

4. Optical modes

The detector detects energy so that especially suitable for Fock state study. In an idealized case absorbing one photon leads to one transition of the electron that releases one pair of electron and hole charge carriers. The quantum efficiency is said to be 100%. The actual emission and absorption are more complex as described in the section 2. While in theoretical study, it is usual to consider single mode light field interaction with two level atomic systems. However, compared with electron levels, optical modes have much more different features.

4.1 Interaction of electron with optical fields

In study on the interaction between the radiation fields and the atom system, the Hamiltonian describing interaction between the electromagnetic fields and an electron (neglecting the electron spin) is

$$H = H_{Elec} + H_{Inter} + H_{Field} \tag{16}$$

Where, H_{Elec} refers to electron motion without the external electromagnetic field. H_{Inter} is interaction of the electron with the light field. H_{Field} is the Hamiltonian of the light fields. The interaction of the electron with the light fields may be written in two parts as (Walls & Milburn, 1994)

$$\hat{H}_{Elec,1} = \int \Psi^{+}(x)\left(-\frac{e}{m}\mathbf{A}\cdot\mathbf{p}\right)\Psi(X)d^{3}x \tag{17}$$

$$\hat{H}_{Elec,2} = \int \Psi^{+}(x)\left(-\frac{e^{2}}{m}A^{2}\right)\Psi(x)d^{3}x \tag{18}$$

where, $\mathbf{A}$ is the vector potential of the electromagnetic field, $\mathbf{p}$ is the momentum of the electron, e and m are the charge and mass of the electron respectively. The electromagnetic field operator expressed as a superposition of the unperturbed wave functions that

$$\Psi(x) = \sum_{j} a_{j}\varphi_{j}(x) \tag{19}$$

Though the Fock state and number operator can be used to study photon emission and absorption, explain quantum collapses and revivals for interaction of a two-level atom with a single mode field. It is not good enough for practical situation or from a rigorous theoretical treatment in consideration of the reasons:

1. The size of a quantum dot is very small so that in consideration of the uncertainty principle of quantum mechanics the light emitting is omnidirectional, it has a large divergence angle.
2. The absorption happens in a very short time so that the emitted photon has a large bandwidth in according to the uncertainty principle of the quantum mechanics.
3. Even for the polarization, it is always considered as combination of two parts or two orthogonal modes so that a horizontal linearly polarization light is a sum of two linear polarizations orthogonal each other making a 45° angle to the horizontal or a sum of right and left circularly polarized light.

4. Optical modes should be considered as a position similar to levels for electron in an atomic system. The different is that a level allows only one electron to occupy, while an optical mode can be occupied by many photons, and the number in one optical mode has no limitation yet. One photon can also be shared by several optical modes as doing in quantum information where information is coded to part of the photon in different mode. One photon shared by four optical modes shown as multipartite entanglement was experimentally demonstrated. Therefore, a part of one photon exists in one optical mode should not considered as "probability of finding one photon", that is indeed a component of the photon (Papp et al., 2009).

Therefore, we should consider that one photon consists of multi-components existing in different modes, or shared by multi-modes. It should be noticed that the number operator is obtained from the integral of a set of normal modes which describe field distribution. One photon should be expresses as a sum of a complete set orthogonal normalized eigen functions. We therefore prefer to start from the studying the fundamental modes of the electromagnetic fields.

4.2 Quantization conditions of planar waveguides

The fundamental modes are characterized by the quantization of the vectors appeared in the wavefunction including polarization and the wave vector. To find the eigen value, Einstein proposed a generalized quantization rule (Stone, 2005, as cited in Einstein, 1997)

$$\oint_{C_i} \mathbf{p} \cdot d\mathbf{q} = n_i h, \quad i = 1,2,3,\cdots \tag{20}$$

where, $\mathbf{p}$ is momentum, $\mathbf{q}$ is coordinate, h is the Planck's constant, C_i is a closed independent loop. This formula is correct in deal with angular momentum. Though the Einstein quantization condition may need a small modification in dealing with practical system including no central forces, this formula is fundamental correct if the boundary condition is properly considered. For example, in dealing with a symmetrical planar dielectric waveguide, the integral form of the quantization condition changes to summation form that phase changes at the boundary have to be added to the summation (Saleh & Teich, 1991). The quantization condition is

$$2k_y d - 2\varphi_r = 2\pi m, \quad m = 1,2,3,\cdots \tag{21}$$

where k_y is the y-component of the wave vector, d is the geometrical thickness of the waveguide in the y-direction, the wave is guided to the z-direction. The Einstein's formula holds still if we use the effective optical length $d_{eff} = d - \varphi_r$. The phase change of reflection at boundary can be obtained from Fresnel Equations

$$r_{TE} = \frac{n_1 \cos\theta_1 - n_2 \cos\theta_2}{n_1 \cos\theta_1 + n_2 \cos\theta_2} \tag{22}$$

$$r_{TM} = \frac{n_2 \cos\theta_1 - n_1 \cos\theta_2}{n_2 \cos\theta_1 + n_1 \cos\theta_2} \tag{23}$$

These formulae show different polarization has different reflection coefficient. r_{TE} is the reflection of the TE waves. The direction of the electric component of the TE modes keeps

unchanged at the boundary while the magnetic components and wave vector change their directions in the incident plan so that called transverse electric light waves. TM is called the transverse magnetic light waves or TM polarized light waves for the same reason. It is convenient to define $\xi_j = n_j k_0 \cos\theta_j$, $j = 1,2$, to rewrite Eg.(22) as

$$
\begin{aligned}
r_{TE} &= \frac{\xi_1 - \xi_2}{\xi_1 + \xi_2} \\
&= \frac{(\xi_1 - \xi_2)^2}{\xi_1^2 - \xi_2^2} = \left[\frac{\xi_1 + i(i\xi_2)}{\sqrt{\xi_1^2 + (i\xi_2)^2}} \right]^2
\end{aligned} \tag{24}
$$

The same calculation can be performed for TM waves that leads the reflection from medium *a* at *a/b* boundary can be written as

$$
r_{ab} \equiv \exp\{-2i\varphi_{ab}\} \tag{25}
$$

where,

$$
\varphi_{ab} = \arctan\left(\frac{\eta_a \xi_b}{i\eta_b \xi_a} \right) \tag{26}
$$

where the coefficients of the polarization

$$
\eta_k = \begin{cases} 1, & TE \\ n_k^2, & TM \end{cases} \tag{27}
$$

Therefore, not only the wave vector quantized, the polarization modes are also non-degenerated.

4.3 The fundamental modes in free space

Photons mediate electromagnetic interactions which also lead to quantization of the physical observables similar to Coulombic interaction in an atomic system. The electromagnetic interactions include coherent interference and coherent combination. The electromagnetic interaction happens during their superposition.

It is reasonable to start from consideration of polarized monochromatic optical waves. We consider TM mode that the direction of the magnetic vector of the optical wave will not change while the electric vector and the wave vector can change their directions in a plan containing them. The resonant condition considered in the phase space is

$$
r \frac{dk}{d\theta} d\theta = -k dr \tag{28}
$$

The photons survived in resonance. This is why a lens can focus an optical beam: the relative phase delays are compensated by direction change of the wavevector so that the curvature of the wavefront changes as shown in Fig.1. There exists angular momentum due to strong electromagnetic interaction.

$$\mathbf{J} = \mathbf{r} \times \frac{\partial \mathbf{k}\hbar}{\partial \theta} \tag{29}$$

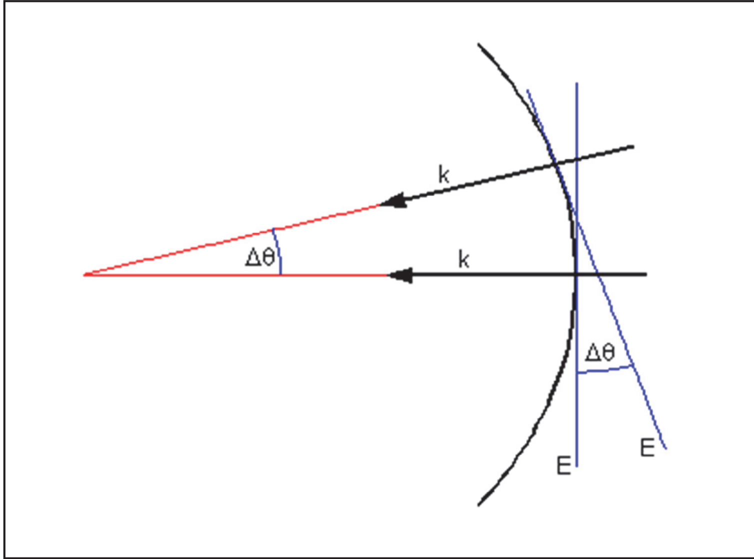


Fig. 1. Angular momentum requires the wavevectors changing their direction in a plane perpendicular to the angular momentum to meet the resonance condition. The relative phase delay along a distance is compensated by electric field vector rotation of an angle equal to that of the k vector rotation of $\Delta\theta$.

To deal with angular momentum, the Einstein's generalized quantization rule is correct, used here leading to

$$\oint \mathbf{p} \cdot d\mathbf{q} = n_i h, \quad n_i = 0, \pm 1 \tag{30}$$

These results indicate three fundamental modes in free space. $n_1 = 0$, corresponding to plane waves. The modes with nonzero angular momentum have curved phase front that the phase delayed monotonically clockwise or anti-clockwise indicating the direction of the angular momentum. This is a result that the coherent superposition of the electromagnetic fields appeared as self-confined field, a condition called self-consistency.

The fundamental modes with non-zero-angular momentum have attracted much attention in recent years. The integral closed with a phase difference of 2π for a complete mode with angular momentum equal to +1 or -1. A complete mode includes all directions of the waves in the plan needed for the loop integral. This mode contains always two oppositely propagating waves. A relative phase delay of π exists between the oppositely propagating waves so that the wave function with non-zero angular momentum should be

$$\psi_{\pm 1} = \frac{1}{\sqrt{r}}\left(e^{-ikr} + e^{i(kr + n\pi)}\right), \quad n = \pm 1, \tag{31}$$

where the time-dependent factor $e^{i\omega t}$ being omitted. The intensities of the mode have cylindrical symmetrical distribution as shown in Fig.2

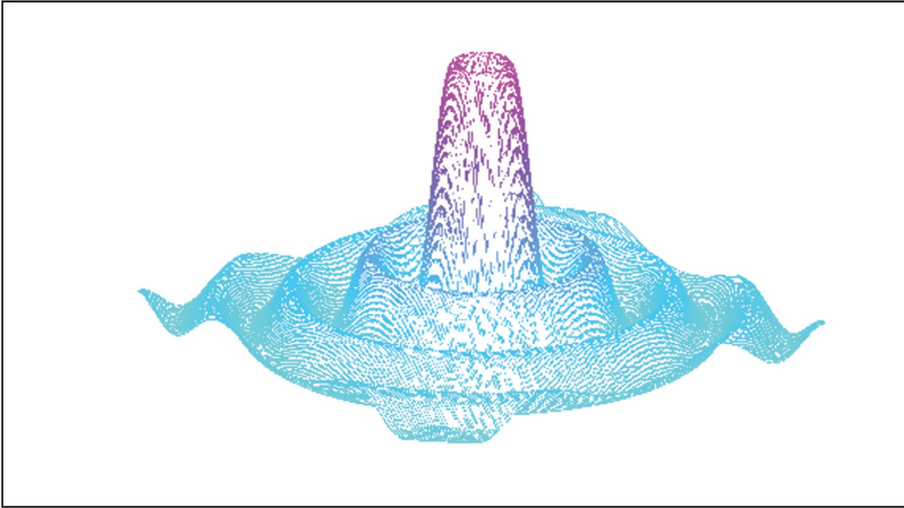


Fig. 2. A complete fundamental mode with angular momentum equal to ±1 has cylindrically distribution in space. The intensity vanished at center.

The further quantization is the generation of vortex lattice (Ghai et al. 2009; Wei et al. 2009).

4.4 Mode structure of an optical beam

One of the solution paraxial Helmholtz equation is Gaussian beam. In comparison with Hermite-Gaussian beams or Laquerre-Gaussian beams, the Gaussian beam is taken as the fundamental mode or single mode. The intensity distribution of a Gaussian beam is a single spot that is similar to the intensity distribution of the lowest order mode (0,0) of the Hermite-Gaussian beam. A Gaussian beam appears under paraxial waves assumptions.
The phase of a Gaussian beam may be expressed (Saleh & Teich, 1991) as

$$\varphi(\rho, z) = kz - \zeta(z) + \frac{k\rho^2}{2R(z)} \tag{32}$$

where, the beam propagates to z-direction, $\rho^2 = x^2 + y^2$. On the beam center, $\rho = 0$, the Eq.(32) represents a plan wave. $R(z)$ is the wavefront radius of the curvature. However, the radius of the curvature changes with distance z that the radius is infinite at $z = 0$ where is called beam waist, having a planar wavefront. The radius decreases with z to a minimum and then increases. The wavefront far from beam waist is approximately the same as that of a spherical wave. The Gaussian beam has a intensity distribution similar to that of the lowest order mode (0,0) of the Hermite-Gaussian beam, and therefore be called as single mode beam. However, the variation of the radius of the wavefront curvature could not be explained by mode theory. A single mode requires a constant eigen value for the

momentum. A fundamental mode based on self-consistency guarantees shape invariance. The variation of the wavefront curvature could only be explained by interaction between different modes.

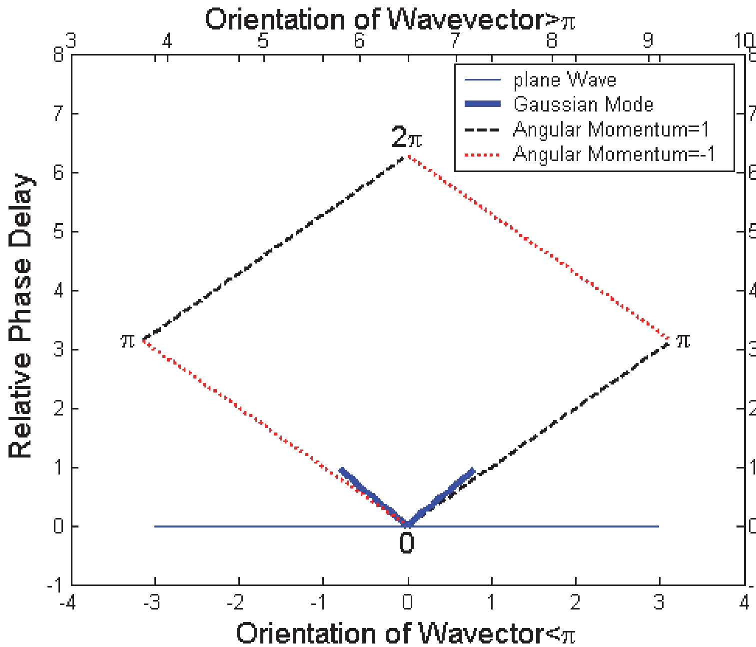


Fig. 3. The phase distribution of a Gaussian beam at a position far from its beam waist where it has a spherical phase front. Compared with the fundamental spatial phase which monotonically delayed clockwise or anti-clockwise indicating the direction of the angular momentum, the Gaussian beam (sick line) starts at its center increased its spatial phase all the same in both direction and limited to a small area surrounding the beam axis due to the paraxial assumption. The thin line represents plan wave.

The phase distribution of a Gaussian beam at a position far from the beamwaist has approximately spherical wavefront. The spatial phase of the Gaussian beam, fundamental modes with angular momentum equal to 0, ±1 are shown in Fig.3. The fundamental modes with non-zero angular momentum increase or decrease their phase monotonically to close with a phase difference of 2п. A Gaussian beam increases its spatial phase all the same at both sides each belongs to different angular momentum. It is the interaction between the two modes decides the variation of the wavefront of the Gaussian beam. The interaction between the two fundamental modes in a Gaussian beam is depicted in Fig. 4.
A Gaussian beam is usually produced by a cavity formed by spherical mirror resonators with a small aperture as output mirror. The output beam has a wavefront similar to spherical waves. A spherical wave output from the mirror is that phase delayed at the center and then increasing its spatial phase with distance in the radius direction. In the consideration of fundamental modes, if the output is polarized light, it contains nonzero angular momentums belong to both electric vector and the magnetic vector respectively so

that there are four fundamental modes at the least. Under the paraxial assumption for the optical beam, the light within a neighborhood of size λ is locally like a plan wave, while the Gaussian beam actually posses paraboloidal wave at far from the beam waist.
Interaction of theses modes result in the paraboloidal wavefront. We consider one polarization in a plan containing wavevector and the wavefront of two fundamental modes with angular momentum equal to ±1 as depicted in Fig.4. The two modes keep in phase at the center, while the phase difference increases with the distance to the beam center. That is the destructive interference increases with this distance. This is the reason a Gaussian beam can confine spatially and transport light energy through free space to a long distance. However, two or more Gaussian beams can not coherently combine all together, instead, they produce interference pattern.
Optical beams that belong to same fundamental mode can topologically combine to form optical vortices that are promise to provide extreme high power density needed for fundamental study at extreme conditions. It is also a possible scheme for efficient laser fusion.

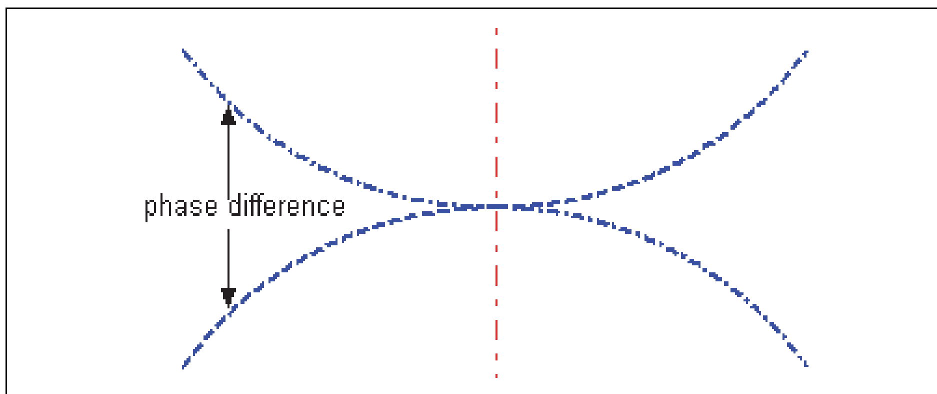


Fig. 4. A Gaussian beam has nearly a spherical wavefront far from its waist where two modes combine at center keeping in phase and then the phase increases in both directions clockwise and anti-clockwise, the phase difference between the two fundamental modes increased also. Dashed line and dot-dashed line represent different modes respectively.

Two or more light beams each belongs to different fundamental mode may be coherently combined via unitary transformation. The unitary transformations are actually mode transform. However, for the Gaussian beams, their complex mode structures prevent this kind of transformation. The number of the modes increases with the number of the Gaussian beams, only components belong to single fundamental mode can be combined while the others have both constructive and destructive interference that lead to complex pattern.
An analysis of a focusing Gaussian beam indicates that it has wavevector on the beam axis carrying plane waves surrounded with declined wavevectors. Looking at a plane perpendicular to the beam axis, there are components of these vector pointing to the beam center from all possible direction. With a proper topological charge, these vectors are topologically combined to form optical vortex. However, the optical vortex is appeared in

first order beam. In an experiment, the formed optical vortex has to be separated from 0th order beam for measurements. The 0th order beam still carries the component of the plane wave and it is much stronger (Yang, 2009).

5. Bunching and antibunching of photons

Photons have a trend to have collective behavior such as in the stimulated amplification. Even in a faint laser pulse, the multi-photon events could not disappear. However, in the quantum statistics, there exists indeed antibunching and sub-Poissonian statistics. Anti-bunching predicts that photons come one after the other that are single photons. Single photons are extreme important for quantum key distribution that guarantees the security of the communication. Photon anti-bunching effect is also important for understanding photons since the photon anti-bunching is regarded as a fully quantum mechanical phenomenon without a classical analogue (Pathak & Mandal, 2003).
The photon anti-bunching is the quantum statistical property. It is usual to consider the second order correlation function of a single-mode field for the photon statistics. The second order correlation $g^{(2)}(0)$ factor with zero-time delay is defined as

$$g^{(2)}(0)=\frac{\langle a^+a^+aa\rangle}{\langle a^+a\rangle^2}=1+\frac{V(n)-\bar{n}}{\bar{n}^2} \tag{33}$$

where, $V(n)=\langle(a^+a)^2\rangle-\langle a^+a\rangle^2$. For a coherent state, $g^{(2)}(0)=1$ and $V(n)=\bar{n}$, the number state has a Poissonian distribution. A classical analysis predicts always $g^{(2)}(0)\geq 1$, there is a tendency of photon bunching. For $g^{(2)}(0)<1$, it is the anti-bunching which is a feature peculiar to the quantum mechanical nature of the electromagnetic field (Walls & Milburn, 1994).

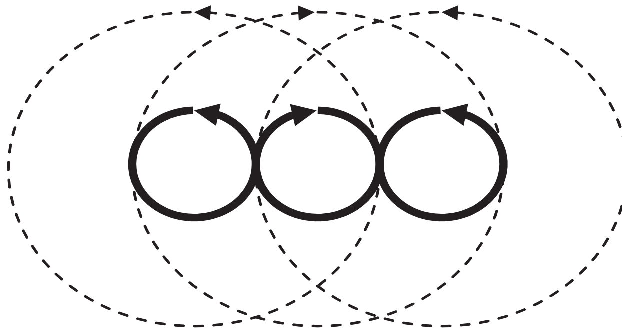


Fig. 5. Optical vortices with angular momentum =±1 refuse to share same position in space that can be considered as spatial antibunching.

The explanation of the photon anti-bunching is challenge. The macroscopic electromagnetic interaction show coherent combination and coherent interference phenomena. Therefore the anti-bunching were explained as the particle natures of the photons. Many models have been proposed to possess anti-bunching effect. Among several experiments containing

interaction of light field with electron or atomic systems that demonstrated anti-bunching light, the most promising is the electrically driven single-photon source (Zhiliang Yuan et al., 2002). In their experiment, the HBT measurement demonstrated indeed sub-Poissonian photon statistics and anti-bunching. Most of the experiments reported to date demonstrate antibunching using radiation field interaction with atomic or ion system which limit the number of the photon emitting. Is there any reason that from the property of the photons by themselves that should result in antibunching? Why the photons can distinguish each other? Destructive two-photon interference demonstrated photon antibunching with calculated $g^{(2)}(0)<1$. Actually, they obtained both the bunching and antibunching by controlling the phase difference between two input beams so that the production process is difficult to understanding by a naïve photon picture (Kaoshi & Matsuoka, 1996).

To answer these questions, one should also consider the physical observables that are quantized wavevectors or polarization vectors. Electromagnetic fields never refuse to interfere with each other. There are two cases that should be considered. One is the coherently combination, the other is the interference coherently. If we consider the monochromatic wave as single mode, the two linearly polarization orthogonal each other to form a new state only when their wave vectors lay on line. For the linearly polarized light with wave vector in different direction in same plan, they can coherently interfere to produce fringes when across each other, while keeps their independent so that keep their polarization and wave vector unchanged after crossover. However, if these waves topologically charged they can combine to share a same angular momentum. The optical vortices formed under strong electromagnetic interaction where the wavevectors topologically combined. The optical vortex with angular momentum equal to 1 refuses to take same position with the optical vortex with angular momentum equal to -1 in the space so that vortex lattice formed as shown in Fig.5. This phenomenon can be considered as spatial anti-bunching. However, there should be collective behavior of large amount of photons since the optical vortices are the results of strong electromagnetic interaction provided by photons themselves. Here, the optical vortices with different angular momentum can distinct each other. The optical vortices are now mostly considered due to orbital angular momentum.

6. Directional emission of photons

The divergence of photon emitting from a cavity can be easily estimated from uncertainty principle of the quantum mechanics.

$$\Delta x \Delta p_x \geq \frac{\hbar}{2} \tag{34}$$

Supposing the size of the antenna is $R=\Delta x$, the divergence angle $\theta=\Delta p_x / p_x$ and $p_x=\hbar k=2\pi\hbar/\lambda$, it is immediately to have

$$\theta=\frac{\lambda}{4\pi R} \tag{35}$$

Therefore, the divergence angle increases with decreasing the size of the antenna. The divergence angle of photon emitting from quantum dots is very large that a photon shared by all possible modes. As a result, the collection efficiency of the photon is very low.

Put a quantum dot in cavity or waveguide that may limit the photon to emit only to desired modes. But that is difficult in technical realization. Using surface plasmon resonances to form optical patch antennas is promise (Esteban et al., 2010). The unidirectional emission of a quantum dot coupled to a nanoantenna was experimentally demonstrated (Curto et al., 2010). In their experiment, a quantum dot was placed in the near field of one of the five-element gold Yagi-Uda antennas for operation wavelengths of ~800 nm. The total length of the antennas is 830 nm. The antennas emission a beam with an divergence angle at half maximum of $12.5^o\times37^o$ pointing into the glass substrate. The simulations indicated that as much as 83.2% of the QD emitted light were collected.
This coupling demonstrates a mode selection. The surface plasmon can only be excited by TM waves. As a result, quantum dot emission is transformed onto TM mode via surface plasmon resonance. The interaction of the five-element emission decided the direction of the optical beam. The mode selection is from two physical reasons:

1. Photons are survived in resonance. Once the TM mode resonant with surface plasmon, it absorbs all energy of the photon possess, TE mode is depressed.
2. The interaction of the radiation field emitted by the antennas decided the emission direction and divergence. That is decided the mode structure of the photon.

Divergence from an emitter can not be avoidable. The detector requires convergence of the beam. There is still a need to study the wavefronts and the mode structure, if one considers the reciprocity: that is if a next similar five-element Yagi-Uda can sense the emitted beam so that results an absorption in a quantum dot under the assumption adiabatic unitary transformation?
There has no evidence that an adiabatic unitary transformation is technical realistic. Loss and dephase are unavoidable. However, many experiments have demonstrated with single photon detector now available commercially, many quantum information experiments can be successfully performed.

7. Single photon detector

Single photon detectors are essential for quantum information applications. Single photon detectors at the communication wavelengths have attracted much attention in recent years. Here we introduce some of the results on developing single photon detector at communication wavelengths in our laboratory.

7.1 Characterizing an APD for single photon detection

Single photondetectors used in quantum key distribution ask for very high sensitivity and extreme low noise. The InGaAs avalanche photodiodes (APD) are usually chosen for single photon detection at the infrared communication wavelengths. This APD has a structure of separate absorption, grading, and multiplication (SAGM). The SAGM APD has been studied extensively and successfully used for single photon detection in the infrared communication wavelengths. The structure of this APD is depicted in Fig.6. The absorption layer InGaAs is designed to have a bandgap of 0.73 eV so that the sensitivity can extend to about 1650 nm. A grading layer between the absorption layer and the multiplication layer facilitates the holes induced by absorption of photons to transit into the multiplication layer (Hiskett et al., 2000).

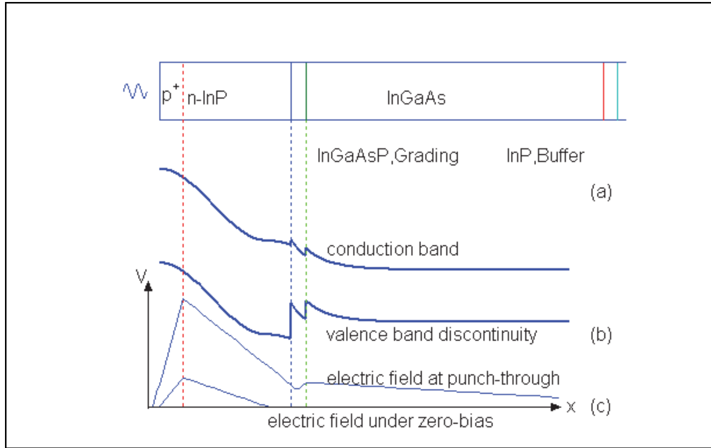


Fig. 6. (a) The structure of the SAGM APD, (b) the band diagram of the APD, (c) The inner build electric field under zero bias and under punch-through.

This APD has to be operated in Geiger mode to exploit the extreme sensitivity. A Geiger mode means the APD is operated at a bias higher than its breakdown voltage that any carrier in the multiplication layer will initial self-sustained avalanche. As a result, the APD should have extreme low dark current. Therefore, there are only a few choices of APDs commercially available for operation in Geiger mode. The APDs have to be operated at low temperature. The APD will be damaged if there is not quenching voltage immediately after the avalanche to stop the self-sustained current. Therefore, the Geiger mode is usually realized in gated mode operation. The gated pulses applied on the APD are synchronized with the arriving of the signal photons.

However, the breakdown voltage itself is not very clearly defined. Theoretically, breakdown voltage is said at that voltage the multiplication factor goes to extreme large, or self-sustaining avalanche appeared. But in experiment, it usually included various guess work. For example, the breakdown voltage is defined as the bias voltage at which dark current is 100 μA (Maruyama et al., 2002). Or, the voltage when the first pulse with peak value of 100 mV (0.5 mV at the APD) was detected (Rarity et al., 2000). The voltage higher than the breakdown voltage is called excess voltage or the relative excess bias. Exploring the use of the excess voltage for higher sensitivity has also been reported. Therefore, the characterization of the APD at voltage higher than the breakdown is also needed. However, the characterization of the I-V curves is usually stopped at the guessed breakdown voltage to prevent APD from being damaged. Here we introduce the I-V characterization including the excess voltage with breakdown voltage well defined by actual measured value.

A passive quench circuit with a 200 KΩ quench resistor is used to characterize an APD of type C307645E from EG&G. The APD was cooled to a temperature of -25℃ by Peltier effect. A 1.31 μm pigtailed DFB diode laser attenuated to -45dBm was used as input signals which switches on to measure photon-current-voltage curves, and switches off to measure dark-current-voltage curves. The measured results are shown in Fig.7 which is very similar to other corresponding reports.

The punch through voltage and the breakdown voltage are not clearly indicated in the I-V curve shown in Fig.7. We define a parameter called relative current gain

$$G_r = 1 + \frac{dI}{dV} \tag{36}$$

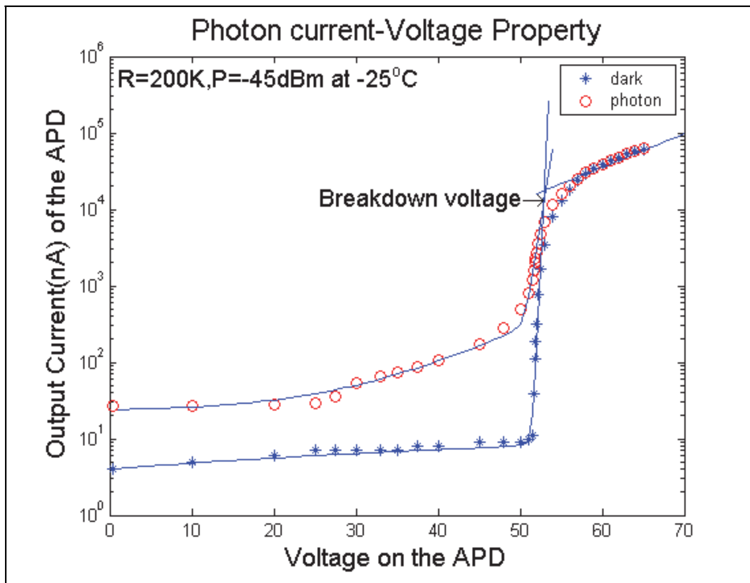


Fig. 7. The breakdown voltage and the bias were decided from these I-V relations experimentally.

The reason is obvious. The multiplication factor has reached it maximum after breakdown, the gain is saturated. The detector performs as a linear device before avalanche and after breakdown. The relative current gains are plotted with the applied voltage both for the photon-current and the dark current as shown in Fig.8 where the data are the same as in Fig.7. The breakdown voltage and the punch through voltage are much more clearly indicated.

The breakdown voltage makes no difference between the photon induced carriers and the dark carriers. But the dark carriers start to avalanche at higher bias. This features a depletion region exists at the vicinity of the heterojunction of the InGaAs/InGaAsP.

7.2 Operation parameters of APD for better performance

It is usual in designing a single photon detector, the operating voltage and the temperature should be carefully considered. The operating voltage should higher than the breakdown, that is an excess voltage is needed. With increasing the excess voltage, the sensitivity seems to be increased, but the error bits or dark counts increased also. There is still uncertain that how high the excess voltage is the best. Thermal excitation decreased with cooling the temperature. It seems the lower temperature is the better. But from the consideration of practical application and the phenomenon the dark carriers not start to avalanche immediately after punch-through, it is reasonable to optimize the operation temperature.

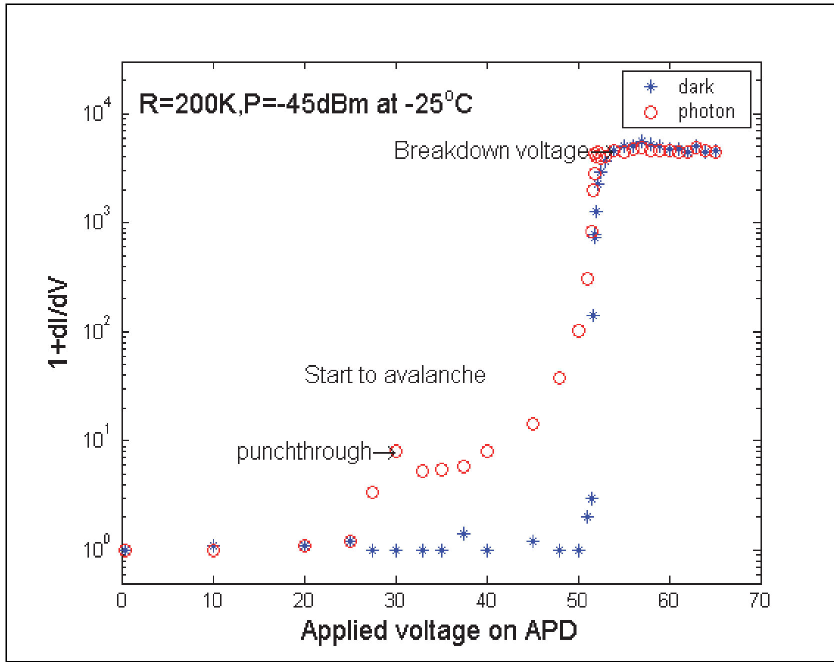


Fig. 8. The relative current gain versus bias characterization curves clearly indicating the punch through voltage, breakdown voltage and the avalanche.

In according to the I-V property shown in the Fig.8, one should consider both the bias and the operation temperature at the same time. Because the breakdown voltage decreases with temperature, the breakdown voltage can be adjusted by cooling the APD so that it is larger than the punch through voltage but not too high to avoid the breakdown initiated by dark carriers. In the Fig.8, the breakdown voltage of about 50 V is a reasonable choice. The excess voltage needed is very limited.

It is well known that the photon absorption follows an exponential law while the thermo-generated carriers follow Gaussian distribution. A calculation analysis is shown in Fig.9 where we calculated the distribution for incident of average 0.1, 0.3, 0.5 photons per pulse and suppose the thermo-generated carriers in the pulse duration are 1, 0.75, and 0.5, corresponding to the curves (a), (b), and (c) respectively. The S/N ratio is defined as the photon-induced carriers divided by thermal carriers that can drift into the multiplication layer. It is clear that the S/N is fairly high if the bias voltage do not penetrate into the absorption layer too much.

The excess voltage can be controlled by temperature since the breakdown voltage is a function of the temperature. That is the basis the temperature control could be used to adjust the breakdown voltage. The bias larger than that of the punch-through is necessary and the depth of punch through should be carefully chosen.

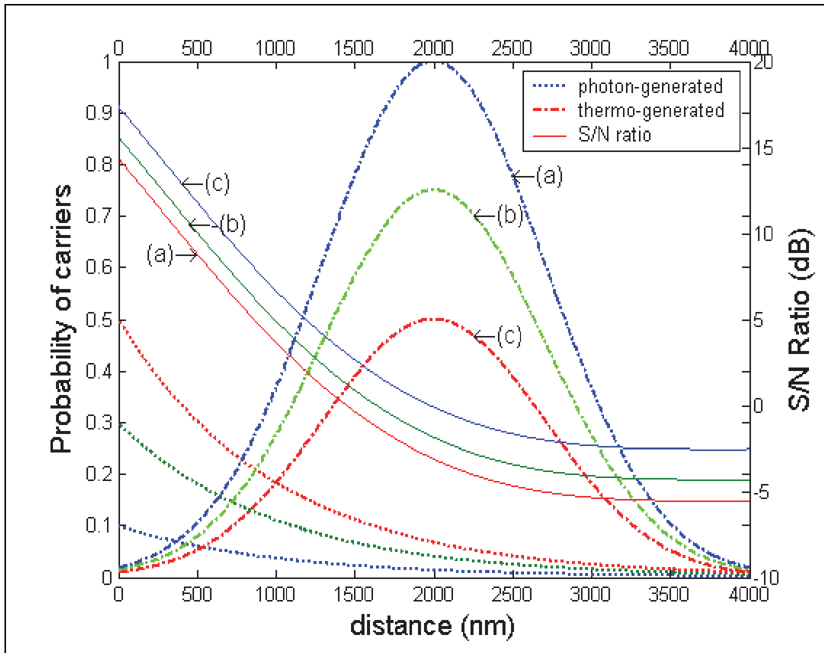


Fig. 9. The spatial distribution of the photon-induced carriers and thermal carriers in the absorption layer calculated on the basis of average number per pulse duration indicating that the depth of the punch-through voltage penetrate into the absorption is very limited for a reasonable S/N.

7.3 Integral detection

The gated electric pulses may produce electric spikes that would result in error counts. Therefore, single photon detector with balanced two APDs has been reported that S/N ratio improved by more than one order of magnitude in compared to the conventional usage of APDs (Tomita & Nakamura, 2002; Kosaka et al., 2003). Although various proposals had added to the balancing structures, they are not only technically complex, the spikes can not be cancelled completely. The integral gated mode single photon detection is much promising for use in quantum key distribution (Wei et al., 2007).

In the method of the integral gated mode detection, an integral capacitor stores the charges of the avalanche current and gives a negative feedback to the APD bias that leads to quench the avalanche at a fixed level. The integral capacitor and a charge amplifier compose as integrator so that the detected signals are static charge on the capacitor. There are no spikes at all and easy for digital processing afterwards.

In the experiments with the integral gated mode single photon detector, the single photon source was attenuated faint pulses with width of 50 ps at 1550 nm from a gain-switched laser (Sepia PDL808, Picoquant). The APD used in the experiments was from JDS Uniphase (ETX 40 APD BA, ETX00408052-005). The temperature of the APD was stabilized at 224±0.1 K. The static bias was 43.1 V which is below the punch through voltage. The gated pulses of 5.13 ns in FWHM and 4.4 V peak-to-peak were added to DC bias. The breakdown voltage of

this APD measured at 224 K is 46.6 V. Therefore, there is only an excess voltage of 0.9 V. A gate pulse frequency of 100 kHz was chosen in the measurement.
The traces of the APD avalanche recorded by oscilloscope TDS1012 show clearly the transient spike cancellations. A single photon detection efficiency of 29.9% at dark count probability of 5.57×10^{-6} per gate or 10.11×10^{-7} has been achieved.

8. Multipartite entanglement of photons

The single photon detector records classical information. It records only the energy. In quantum information, there should have records of quantum bits. Therefore, two or more detectors have to be used, and these detectors should be entangled each other. This can be realized with some kind of interferometers. For example, the Mach-Zehnder interferometer as shown in Fig.10, the record at the detector A should be entangled with that at detector B. If detector A records 1, then detector B should record 0. That is, only record of (10) is correct, while all other records including (01), (00), (11) are error bits.

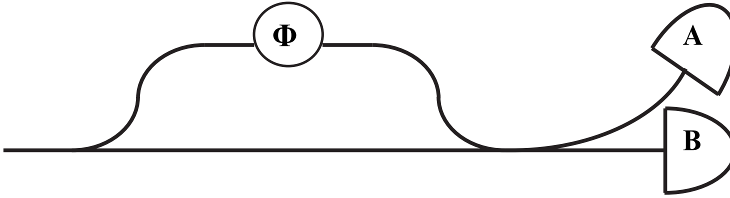


Fig. 10. Mach-Zehnder implementation of quantum key distribution

To obtain a correct record, Bob has to decode by using phase modulation in according to their protocol so that the maximum of the interference is at detector A corresponding to the input quantum signals:

$$\begin{pmatrix}1\\0\end{pmatrix}=\frac{1}{2}\begin{pmatrix}1 & e^{-i\varphi}\\-1 & e^{-i\varphi}\end{pmatrix}\begin{pmatrix}1\\e^{i\varphi}\end{pmatrix} \tag{37}$$

One photon can be coherently shared among N spatially distinct optical modes to form multipartite entanglement, a quantum state being called W state. A W-state with N=4 can be expressed in the form

$$|W\rangle=\frac{1}{2}\Big[\big(|1000\rangle+e^{i\varphi_1}|0100\rangle\big)+e^{i\varphi}\big(|0010\rangle+e^{i\varphi_2}|0001\rangle\big)\Big] \tag{38}$$

The multipartite entanglement with N=4 where the partite is formed by beamsplitter has been detected and characterized in more detail (Papp et al., 2009).
A genuine N-partite entanglement is realized only with simultaneous participation of all N of the constituent systems. There is also a similar case where is N time-distinct partite entanglement. There are N time-distinct faint pulses to share one photon. This kind of multipartite entanglements is realized in differential phase shift key distribution system where the simultaneous participation of the constituent systems is due to the nonlocality of the photons.

However, the utilization ratio of the traditional differential phase shift is low. Their key creation rate under ideal condition can only reach (1-1/N) if a single photon pulse is split into N sequential ones. The utilization ratio of photons can reach 1 under ideal condition by discrimination and controlled delay of the first pulse (Wang et al., 2009). This scheme can also result in a genuine N-partite entanglement with N time-distinct constituents. In the proposed scheme of 6-partite entanglement, the two pulse trains containing three time-distinct pulses in each formed by beam splitter and controlled time delay. The multiplex at Bob side are designed such that with the first single pulse of the first train two-bit delayed the rearranged pulse train can recombine with the next pulse train in three time slots exactly. Their coherent superposition represents an entangled state of a photon in three-dimensional Hilbert space with four non-orthogonal states:

$$\frac{1}{\sqrt{3}}\left(|V\rangle|1\rangle_A|0\rangle_B|0\rangle_C \pm |H\rangle|0\rangle_A|1\rangle_B|0\rangle_C \pm |H\rangle|0\rangle_A|0\rangle_B|1\rangle_C\right) \tag{39}$$

Where, $|V\rangle$ is vertical polarization state, $|H\rangle$ is the horizontal polarization state. All the constituents of a photon have involved in the key creation so that is a genuine multipartite entanglement state. Not only the key create ratio is increased, the security is also enhanced.

9. Dephase and decoherence

Dephase and decoherence are unavoidable even in the case under idealized condition without loss if one consider a rigorous representation of the wavefunction for single photons. In according to quantum mechanics, a state of a particle is represented as a complete set of eigen functions. Therefore, the number state of N=1 should be expressed in a two dimensional Hilber space that

$$|\psi(1)\rangle = \alpha|0\rangle + \beta|1\rangle \tag{40}$$

Even in pure vacuum, the fluctuated electromagnetic fields exist that make up the zero-point energy. Quantum theory predicts that empty space is not truly empty. In an electromagnetic field, virtual photons created and annihilated constantly that make contribution to a small renormalization of the energy of a quantum system, known as the Lamb shift. The experimental observation of the Lamb shift in a solid system has been reported (Frabner et al., 2008). A scheme including vacuum state that can be used to demonstrate the nonlocality of a single photon experimentally has also been proposed (Dunningham & Vedral, 2007).

In their scheme, classical faint pulse incident on beamsplitter has been expressed as two input ports and two output ports where a state of Eq.(40) and a vacuum state $|0\rangle$ are incident on the two input ports of a 50:50 beamsplitter, and two output state of U1 and U2 as shown in Fig.11. It is obvious that vacuum fluctuation in the output state is increased. In their model, they have chosen the particular values that $\alpha = 1/\sqrt{3}$, and $\beta = \sqrt{2/3}e^{i\varphi}$ to simplify the analysis without losing the generality of their arguments. The state after the beamsplitter is

$$|\psi\rangle = \frac{1}{\sqrt{3}}\left[|0\rangle|0\rangle + e^{i\varphi}\left(|0\rangle|1\rangle + i|1\rangle|0\rangle\right)\right] \tag{41}$$

Where the first ket in each term represents the number of the particles on path U1 and the second ket represents the number of particles on path U2. It is general the case that any operation on the state of the photon leads to lose and combining vacuum fluctuation. That is the quantum permutation.
The phase is important while modes are distinguished only with their polarization and momentum. In quantum mechanics, states with only a pure phase difference are taken as the same state also. Now that the constituents of the photon and the vacuum fluctuations with phase unknown come together in the same mode to superposition that are the reason the dephase come from.

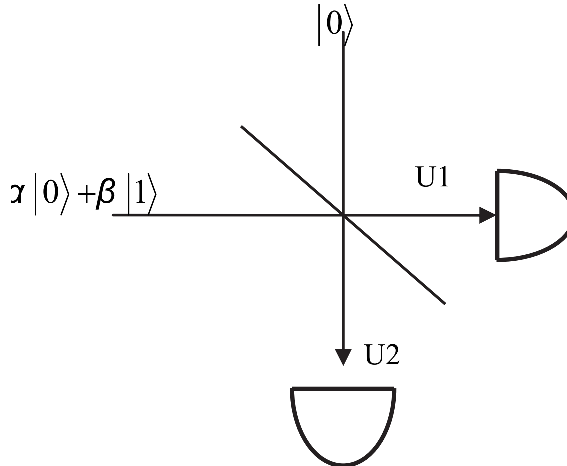


Fig. 11. Single photon is operated by a beam splitter, the vacuum states are involved.

10. Reconstruction of photons

Vacuum supports all optical modes and each of the modes can contain constituent of any photon with phase and amplitude arbitrarily defined which are relay on the initial condition. Therefore a photon emitted from a quantum dot or from a single atom will be coupled to all possible modes with equal probability. The interaction of all the emitting mode fields decides the Emitting pattern that has large divergence. The divergent light will soon be scattered to become a part of the fluctuation in the vacuum if it is not immediately collected and focused to a detector and absorbed. Any procedure or operation on the emitted light means mode change and mode structure reorganization.
The actual probability of detecting a photon is decided by the collecting and focusing as much as possible the constituents of the photon. The highest detection efficiency of a photon is always at where the superposition of the mode fields has constructive coherence to the maximum while detection probability of zero may indicate a destructive superposition. Therefore one should consider the mode structure of a photon and the interaction of the mode fields. It is usual to limit the mode number so that can make the modes controllable.
It is essential to control the mode selection so that to control the mode structure of a photon. There are two mechanisms to collect energies of the photon. One is the mode competition. The other is control the mode volume geometrically such as in waveguides or in cavities where only limited modes can be excited.

In case the quantum dot coupled to the surface plasmon in an element of the Yagi-Uda antenna, the emitting energy concentrated in resonant TM mode due to excitation of the surface plasmon so that the TE mode is depressed. This is the resonant enhancement effect. The direction of emitted light by the Yagi-Uda is decided by interaction of all elements composing the antenna. Here only one direction has coherent maximum of superposition.
Reconstructions of photons for quantum information are quite usual. The photons are divided into two or more part and encoded with phase information. On recombining these constituents of the photon, the coherent superposition decides where the maximum detection probability should appear to a detector as predicted. The coherent maximum and the coherent minimum appeared at the same time and in different places showing an entanglement. A successful unitary transformation will guarantee a photon appeared to a detector with probability of 1 while the probability is zero at all other places. However, this transformation can not prevent a detector to sense the vacuum fluctuation.
The vacuum fluctuation has important role in the detection. Since the vacuum contain all possible modes that deserve the ergodic assumption. That is why the quantum permutation with vacuum exists always. In fact, the photon reconstruction is decided eventually under the choice of the detector where the herald mode fields combining necessary energy from the vacuum fields can form a photon the mode structure of which matches the needs for resonant absorption. There has not a photon detector that is quantum state sensitive, for example, a polarization dependent detector. One may consider a detector which is state sensitive so that can decrease the quantum error bits.
In conclusion, photons are different from other particle with static mass. Photons compose by themselves of electromagnetic field modes which are quantized by the electromagnetic interaction. Therefore, mode structure of a photon should be considered so that nonlocality and entanglement of photons could be explained.

11. References

Chao, C. Y., & Kung T.T. (Gong Zutong), (1933). Interaction of hard γ-ray with atomic nuclei. *Nature*, 4, (1933) 709

Combes, J. & Torner, L., (2005). States for phase estimation in quantum interferometry. *Journal of Optics B: Quamtum and Semiclassical Optics*. 7, (2005) 14-21

Cronstrond, P. & Jansik, B., (2004). Density functional response theory calculation of three-photon absorption. *Journal of Chemical Physics*, 121, (2004) 9239-9346

Curto, A. G.; Volpe, G., Taminiau, T.H., *et al.* (2010). Unidirectional emission of a quantum dot coupled to a nanoantenna. *Science*, 329, 5994, (20 August 2010) 930-933

Dunningham, J. & Vedral, V., (2007). Nonlocality of a single particle. *Physical Review Letters,* 99, (2 November 2007) 180404

Einstein A., (1997). On the quantum theorem of Sommerfeld and Epstein. A translation of the paper appears in "*The collection Papers of Albert Einstein*", Vol.6, Engel, trans., Princeton U. Press, Princeton, NJ (1997), pp.434, see also: A. Douglas Stone, Physics Today, (August 2005) 37-43

Eiseman, M.D., L. Childress, A., Andre, *et al*. (2004). Shaping quantum pulses of light via coherent atomic memory. *Physical. Review Letters,* 93, (2004) 233602

Esteban R., Teperik, T.V., & Greffet, J. J., (2010). Optical patch antennas for single photon emission using surface plasmon resonances. *Physical Review Letters,* 104, (15 January 2010) 026802

Frabner, B.,Göppl. M., Fink, J..M., *et al.* (2008). Resolving vacuum fluctuations in an electrical circuit by measuring the Lamb shift. *Science*, 322, 5906,(28 November 2008) 1357-1360

Ghai, D. P., Vyas, S., Senthilkumaran, P., & Sirohi, R.S., (2009). Vortex lattice generation using interferometric techniques based on lateral shearing. *Optics Communications*, 282, (2009) 2692-2689

Gisin, N., Ribordy, G., Tittel, W., & Zbinden, H., (2002). Quantum cryptography. *Reviews Modern Physics*, 74(1), (January 2002)145-195

Gong Zutong, (1999). Structure theory of photon, *PHOTONICS SINICA*, 28(1), (1999) 1-10

Greiner W., (2001).*Quantum Mechanics*, Fouth Ed. Springer-Verlag Berlin Heidelberg, ISBN 7-5062-7263-6/O·530, New York

Hanbury Brown R. & Twiss, R. Q. (1956). Correlation between photons in two coherent beams of light. *Nature* (London) 177, (1956) 27-32

Hiskett, Ph. A., Buller, G. S., Loudon, A. Y. *et al.* (2000). Performance and design of InGaAs/InP photodiodes for single-photon counting at 1.55 µm. *Applied Optics,* 39(36), (December 2000) 6818-6829

Hughes,R. J., Morgan, G. I., & Perterson, C. G., (2000). Quantum key distribution over a 48 km optical fibre network. *Journal of Modern Optics* 47, 2/3,(2000) 533-547

Kardyna B.E. , Hees, S.S., & Shields,A.J., (2007). Photon number resolving detector based on a quantum dot field effect transistor. *Applied Physics Letters,* 90, (2007) 181114

Koashi, M. & Matsuoka,M., (1996). Photon antibunching by destructive two-photon interference. *Physical Review* A53, (5), (May 1996) 3621-3624

Kosaka, H., Tomita, A., Namhu, Y., *et al.* (2003). Single-photon interference experiment over 100 km for quantum cryptography system using balanced gated-mode photon detector. *Electron. Lett.* 39(16), (2003) 1199-1201

Mandel, L., & Wolf, E., (1995). *Optical Coherence and Quantum Optics*, Cambridge, ISBN 0 521 41711 2, New York. pp.476

Maruyama, T., Narusawa, F., Kudo, M., *et al.* (2002). Development of a near-infrared photon-counting system using an InGaAs avalanche photodiode. *Optical Engineering,* 41(2), (2002) 395-402

Mason E.J., Albota, M. A., Konig, F., & Wong, F.N.C., (2002) Efficient generation of tunable photon pairs at 0.8 and 1.6 µm. *Opt. Lett.* 27, (2002) 2115-2117

Migdall A.L., Branning, D., Castelletto, S., & Ware, M., (2002). Single photon source with individualized single photon certifications. *Proc. SPIE* Vol.4821 (2002) 445-465

Miller A. J. , Sae Woo Nam, and J. M. Martinis, (2003). Demonstration of a low-noise near-infrared photon counter with multiphoton discrimination, *Appl. Phys. Lett.* 83(4), (2003) 791-793

Mori S., Motoya, M., Namekata, N., & Inoue, S., (2004). Generation of correlated photon pairs at 1550 nm by periodically poled Lithium Niobate. *Trans IEICE* , J87-C(2004)675-685

Muller, A., Herzog, T., Huttner, B. *et al.* (1997). "plug and play" systems for quantum cryptography. *Appl. Phys. Lett.* 70(7), (1997) 703-795

Namekata, N., Fujii, G., & Inoue, S., (2007). Differential phase shift quantum key distribution using single-photon detectors based on a sinusoidally gated InGaAs/InP avalanche photodiode. *Appl. Phys. Lett*. 91, (2007) 011112

Neves,I., Lima, G., Gomez, A. J. G. *et al.* (2005). Generation of entangled states of Qbits using twin photons. *Phys. Rev. Lett*. 94(10), (2005) 100501

Papp, S. B., Choi, K. S., Kimble, H. J. *et al.* (2009). Characterization of multipartite entanglement for one photon shared among four optical modes, *Science,* 324, 5928, (2009) 764-771

Pathak A. & Mandal, S., (2003). Photon-bunching, photon-antibunching and nonclassical photon statistics of coherent light couped to a cubic nonlinear medium. *Modern Physics Letters* B, 17(5&6),(2003) 225-233

Pittman T.B., Jacobs, B.C., & Franson, J.D., (2002). Single photon on pseudodemand from stored parametric down-conversion, *Phys. Rev. A* 66 (2002)042303

Rarity, J. G., Ridley, T.E., Ridley, K.D., *et al.* (2000). Single photon counting for the 1300-1600-nm range by using of Peltier-cooled and passively quenched OnGaAs avalanche photodiodes. *Applied Optics,* 39(36), (2000) 6746-6753

Roychoudhuri Ch. & Roy, R., Guist. Ed. (2003). OPN Trends: The nature of light, What is a photon. *Optics & Photonics News,* 3(1)

Saleh, B. E. A. & Teich, M. C.,(Eds.). (1991). *Fundamentals of Photonics*, John Wiley & Sons, ISBN 0-471-83965-5, New York, pp.87-88, 249

Santori C., Peton, M., Solomon, G., Dale, Y., *et al.* (2001). Triggered single photons from a quantum dot. *Phys. Rev. Lett.* 86, (2001) 1502-1505

Tanzilli S., Riedmatten, H.D., Zbinden, H., *et al.* (2001). Highly efficient photon pair source using periodically poled lithium niobate waveguide. *Electron. Lett.* 37 (2001) 26-28

Tomita, A. & Nakamura, K. (2002). Balanced gated-mode photon detector for quantum-bit discrimination at 1550 nm. *Opt. Lett.* 27(20), (2002) 1827-1829

Torres, J. P. &Torner, L. (2005). Quantum state engineering for spatial control of entangled photon pairs. *Proceeding of SPIE* Vol.5736, (2005) 173-184

Walls D.F., & Milburn, G.J., (1994). *Quantum Optics,* Springer-Verlag, ISBN 3-540-58831-0, Berlin Heidelberg, pp.41-58, 199

Walton Z., Sergien, A.V., Atature, M., Saleh, B.E.A., & Teich. M.C. (2001). Performance of photon-pair quantum key distribution system. *J. Mod. Opt.* ; 48 (2001) 2055-2063

Wang Jindong., Qin Xiaojuan, Zhang Huani, *et al.* (2009). A free-space-based differential phase shift quantum key distribution scheme with higher key creation efficiency. *Optics communications,* 282, (2009) 3379-3381

Wei, G. X., Lu, L-L., Guo, Ch-Sh., (2009). Generation of optical vortex array based on the fractional Talbot effect. *Optics Communications,* 282, (2009) 2665-2669

Wei Zhengjun, Zhou Peng, Wang Jindong, *et al.* (2009). An integral gated mode single photon detector at telecom wavelengths. *Journal of Physics, D: Applied Physics*, 40(22), (2007) 6922-6925

Wootters W.K. & Zurek, W.H., (1982). A single quantum cannot be cloned, *Nature*, 299, (1982) 802-803

Yang, Sh., Powers, P. E., & Zhan, Q., (2009). Experimental verification of focus tailoring using circularly polarized vortex beams with annular pupil mask. *Optics Communications,* 282, (2009) 4657-4659

Yariv, A., (1988). Quantum Electronics, third edition, John Wiley & Sons, ISBN 0-471-60997-8, New York, pp.96-99

Yi, S. W., Lee, S.K., Yang, G.Y. *et al.* (2004). Three dimensional micro-fabrication using two-photon absorption by femtosecond laser. Procedings of SPIE, Vol.5342, (2004) 137-145,

Yoshizawa, A., Kaji, R., & Tsuchida, H., (2003). Generation of polarization-entangled photon pairs at 1550 nm using two PPLN waveguides. *Electron. Lett.* 39, (2003) 621-622

Yuan, Zhiliang, Beata E. Kardynal, R. M. Stevenson, *et al.* (2002). Electrically driven single-photon source. *Science*, 295,(4 January 2002) 102-105

Yuan, Zh., Wu. Ch., Zhao, H., Zhao, & Jiang, H., (2005). Imaging of small nanoparticle-containing objects by finite-element-based photoacoustic tomography. *Opt. Lett.* 30, (2005) 3054-3056

Zbinden H., Pasel, S., Gisin, N. *et al.* (2002). Practical quantum key distribution. *Quantum Optics In Computing and Communications, Proceedings of SPIE,* Vol.4917, (2002) 40-44

Zumuth, S., Ansari, Z., Lepine, E. & Vrakking, M. J. J. (2005). Single-shot measurement of revival structures in femtosecond laser-induced alignment of molecules. *Opt. Lett.* 30, (2005) 2326-2328

Part 2

CMOS Related Topics

4

CMOS Photodetectors

Albert H. Titus[1], Maurice C-K. Cheung[2]
and Vamsy P. Chodavarapu[2]
[1]Department of Electrical Engineering, University at Buffalo, The State University of New York
[2]Electrical and Computer Engineering, McGill University, Montreal
[1]USA
[2]Canada

1. Introduction

The inclusion of cameras in everything from cell phones to pens to children's' toys is possible because of the low cost and low power consumption of the imaging arrays that form the core of the cameras. However, these arrays are low cost and low power because they are CMOS-based; this allows for the devices to be made with the same processes and facilities that are used to make memory and computer chips. Yet, the continued surge in CMOS imager popularity goes beyond the lower cost to other factors such as ability to integrate the sensors with electronics, and the ability to achieve fast, customizable frame rates [1].

In this chapter, we will cover the design and layout of the various types of CMOS photodetectors and fundamentals of photo-conversion processes in these devices, including a brief review of CMOS photodetector history. We will then describe the emerging CMOS based technologies in photodetector design optimization to tune device responsivity, integration of micro-optics to achieve enhanced detection in low-light conditions, integration of photonic grating structures on photodetectors for spectral response selectivity, and bio-inspired CMOS imaging devices. We will conclude the chapter with some examples of applications of these technologies.

2. CMOS photodetector history

Since the mid-1960s, combinations of p-n (or n-p-n) junctions have been used to convert light into electronic signals [2, 3]. Work not only focused on the conversion of photons to electrons, but also on the ability to read the signals out from arrays of pixels. For example, Shuster and Strull reported in 1965 that they had developed a 2500 pixel array of phototransistors with 100 leads for readout [4, 5]. For these earliest devices, high gain had to be used at the pixels because no integration of light (or charge) was used. A year later, Weckler demonstrated how to operate a p-n junction photodiode in "photo flux integrating mode" which enabled pixels to be much simpler and ultimately smaller [6].

Photon integration required that the photodiode be turned on for a fixed amount of time to raise (or lower) the voltage level on a charge storage device; thus, the amount of charge remaining (or charge that was removed) is proportional to the light intensity over that integration time.

In the 1970s, CMOS detectors and imaging arrays began to lose popularity to Charge-Coupled Device (CCD) -based imagers because CCDs could achieve a higher fill-factor; the fill-factor was lower for CMOS imagers because of the need for transistors at the pixels for read out and gain. The ability to produce CCD imagers with the necessary number of pixels for applications (such as TV) gave CCDs a large advantage over CMOS imagers [7]. Noise in CCDs was also considerable less than in CMOS devices; it was generally regarded that fixed pattern noise in CMOS imaging devices was worse than in CCDs, which remained true into the late 1980s and early 1990s. However, improvements in CMOS fabrication technology and increasing pressure to reduce power consumption for battery operated devices began the re-emergence of CMOS as a viable imaging device.

It is generally regarded that the first all-CMOS sensor array to produce acceptable images is the active pixel sensor (APS) imager [8-10]. The APS design used the linear integration method [6] for measuring light because of the large output signal generated, as opposed to the logarithmic method [11-16]. The active column sensor (ACS) [17] is similar to the APS but has lower fixed pattern noise (FPN). As of 2011, we have CMOS image sensors that have 14.6 Megapixels, and higher.

While CMOS and CCDs continue to compete for a share of the image array sensor market, the ability to design custom integrated circuits (ICs) with photodetectors to perform specific functions is an enormous advantage over CCD arrays. These ICs are often used in applications that have specific requirements such as extremely low power consumption [18] or variable read-out (frame) rates [19] or very fast read-out rates [20]. Normal video rates of 30 or 60 frames per second are fine for standard definition videos, but for some applications, frame rates of more than a thousand frames per second are needed to capture extremely fast occurring events (for example see [20]).

CMOS photodetectors are the technology of choice in smart focal plane arrays. In the mid-1980s, Carver Mead and Misha Mahowald introduced the Silicon Retina that used a "vertical bipolar transistor" as the light detecting element [21]. This spawned a significant amount of research into bio-inspired vision chips that used CMOS photodetectors combined with CMOS signal processing circuitry [22-32]. Generally, these chips are arrays of smart pixels with significantly more transistors per pixel than the three or four found in typical APS-based arrays. However, the relative low cost of fabrication for prototyping CMOS ICs enables chips with one, two, dozens or thousands of detectors to be designed, fabricated and tested. For chips with few photodetectors, the remaining silicon die area can be used for signal processing, read-out, or digital interface logic, so there is no wasted space. Applications of these custom detector and imaging chips range from sub-retinal implant imagers [33] to glare detection [34] to fluorescence imaging [35, 36] and x-ray imaging [37], to name a few. A study of three common photodiode structures available in non-imager/standard CMOS processes provides valuable benchmark data for designers looking to use CMOS photodetectors [38]. While not an exhaustive study of all possible CMOS photodetectors, this chapter provides a useful starting point for selecting the best structure for the application.

3. Operation of CMOS photodetectors

The core of the sensing element of a CMOS detector is the photosensitive element of the circuit. Photogates, phototransistors, and photodiodes all can be used as the sensing element. In this section, the use of a photodiode is discussed. As its name implies, the photodiode is simply a junction between a p-type and an n-type semiconductor, commonly known as a p-n junction. Although a simple p-n junction can be used for light detection, the more sophisticated p-i-n junction with an intrinsic region between the p-type and n-type region is often used to improve the device efficiency. In this section, the basic working principle of a photodiode will be discussed, followed by a discussion on the p-i-n photodiode, and a method of signal amplification resulting in the avalanche photodiode.

3.1 Photogeneration and recombination in semiconductors

When a semiconductor is illuminated, a photon that has higher energy, $h\nu$, than the bandgap energy, E_g, may cause an excitation of an electron from the lower energy valence band to the higher energy conduction band. This results in a pair of the mobile charge carriers - electrons and holes. This process known as photogeneration can occur if the total energy and total momentum among the photogenerated electron-hole pair and the incoming photon is conserved.

The probability of photogeneration by a single photon is a property of the material. Macroscopically, this is described by the absorption coefficient, α. As a light beam propagates through a piece of homogeneous semiconductor, its power decreases exponentially as the semiconductor absorbs some of the power for photogeneration. The power that remains in the light beam after propagating through a depth of z is given by:

$$P = P_0 \exp[-\alpha z] \tag{1}$$

where, P_0 is the intensity at zero depth. Note that the rate of absorption, $-dP/dz$, decreases exponentially with depth. Therefore, more photogeneration is expected to occur near the surface.

Because the photogenerated carriers exist in an excited state, the excess electrons and holes will recombine after a short period of time (~ picoseconds on average) to release the excess energy. This process is known as recombination, and it returns the carriers distributions to thermal equilibrium condition. These excess carriers are lost if they are not captured to create an electrical signal for light detection. Therefore, a semiconductor device structure is needed to facilitate the capturing of the photogenerated carriers. The simplest and most commonly used structure for this purpose is a diode structure known as photodiode (PD).

3.2 Quantum efficiency and responsivity

Before the discussion on photodiode, two important parameters that are used to characterize the effectiveness of detection by a photodetector should be discussed; these are quantum efficiency and responsivity. Quantum efficiency is defined as the probability that an incident photon will generate an electron-hole pair that will contribute to the detection signal, so it can be expressed as,

$$\eta = (1 - \mathcal{R})\zeta[1 - \exp(-\alpha d)] \tag{2}$$

where, $\mathcal{R}$ is the surface reflectance, ζ is the probability that the generated electron-hole pair will become a contribution to the detection signal, and d is the depth of the photo-absorption region. Therefore, quantum efficiency is affected by material properties and device geometry. The captured carriers are used to generate a signal either as a voltage or a current. The measure of signal strength to incident power of the device is called responsivity. If the output is a current, then it is related to the quantum efficiency by the following expression,

$$R = \frac{q}{h\nu}\eta = \frac{q}{hc}\lambda\,\eta = \frac{\lambda}{1.24\ [\mathrm{nm{\cdot}W/A}]}\,\eta \tag{3}$$

where, q is the electron charge and λ is the wavelength in nanometers.

4. Photodiode

Many photodetectors utilizes the formation of p-n junctions, and the simplest of these is a photodiode, because a photodiode is simply a p-n junction that is designed for capturing the photogenerated carriers. In CMOS sensing, a photodiode is usually made by forming an n-type region on a p-type semiconductor substrate, or vice-versa. This can be done by epitaxial growth, diffusion or ion implantation.

4.1 A quick review of P-N junction

Figure 1 shows the carrier distributions, charge distribution, built-in electric field and band-diagram of a typical p-n junction. The inhomogeneous charge and carrier distributions are the result of a state of equilibrium between diffusion, drift and recombination. The following is a review of the key features of a p-n junction,

1. An absent of the carriers in a region known as the depletion region or space charge region. It has layer width, W, and ionic space charge of the donors and acceptors are exposed in this region.

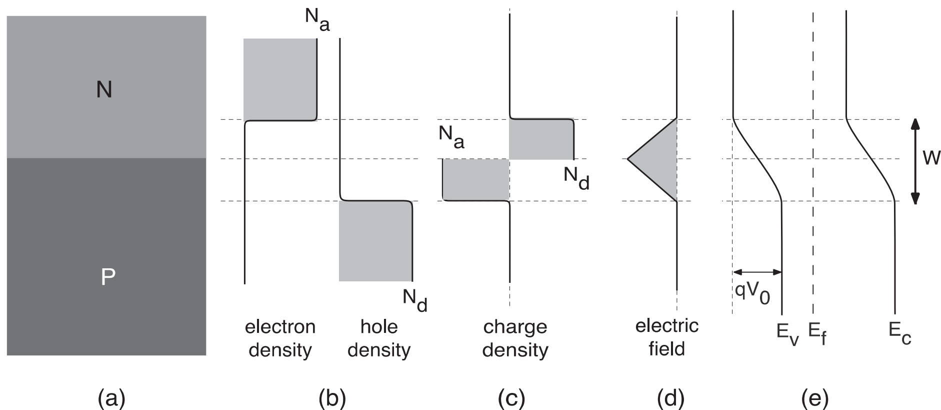


Fig. 1. (a) Diffusion of carriers at the p-n junction, (b) the resulting electron and hole density distribution, where N_a and N_d are the acceptors and donors densities, (c) the resulting charge density distribution, (d) the resulting electric field, and (e) the band diagram of a p-n junction showing the alignment of the Fermi level, E_f, and shifting of the bands. E_v is the top of the valence band and E_c and the bottom of the conduction band.

2. Due the charge distribution, a built-in electric field, **E**, has emerged.
3. Alignment of E_f between the p-type region and the n-type, as the conduction and valance bands shift in the depletion region.
4. There is a potential difference of V_0 between the p-side and the n-side.

4.2 Operating rinciple

The operation of a photodiode relies upon the separation of the photogenerated carriers by the built-in field inside the depletion region of the p-n junction to create the electrical signal of detection. Under the influence of the built-in electric field, the photogenerated electrons will drift towards the n-side, and photogenerated holes will drift towards the p-side. The photogenerated carriers that reach the quasi-neutral region outside of the depletion layer will generate an electric current flowing from the n-side to the p-side; this current is called a photocurrent. The generation of the photocurrent results in the shift of the I-V characteristic of the photodiode as shown in Figure 2. Therefore, the I-V characteristic of a photodiode is expressed as,

$$I_L = I_s\left[\exp\left(\frac{qV}{kT}\right) - 1\right] - I_{ph} \tag{4}$$

where, the first term is the Schottky Equation that described the ideal I-V characteristics with I_S being the saturation current, k the Boltzmann constant and T the operating temperature, and the second term, I_{ph}, is the photocurrent.

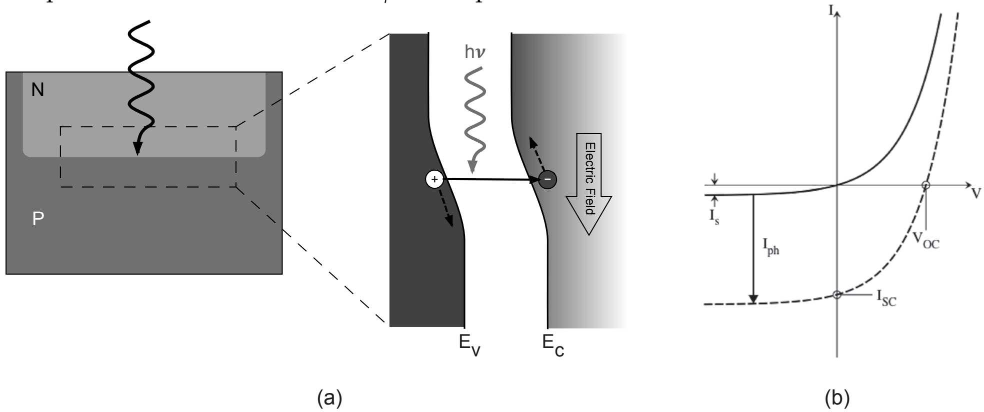


Fig. 2. Photogeneration in a p-n junction: (a) the built-in electric field driving the photogenerated carriers from the depletion region away from the junction, and (b) shifting of the I-V characteristic due to the photogenerated current, I_{ph}.

4.3 Basic modes of operation

A photodiode can be operated in three basic modes: open circuit mode, short circuit mode, and reverse bias (or photoconductive) mode. The circuit diagrams of the three different basic operating modes are shown in Figure 3.

Open circuit (OC) mode is also known as photovoltaic mode. As the name implies, in this mode, the terminals of the photodiode is connected an open circuit. In this mode, there is no net current flowing across the photodiode, but due to the photogenerated current, a net

voltage is created across the photodiode, called the open circuit voltage, V_{OC}. In reference to Figure 2(a), the photodiode is operating at the point where the I-V characteristic curve intersects the x-axis.

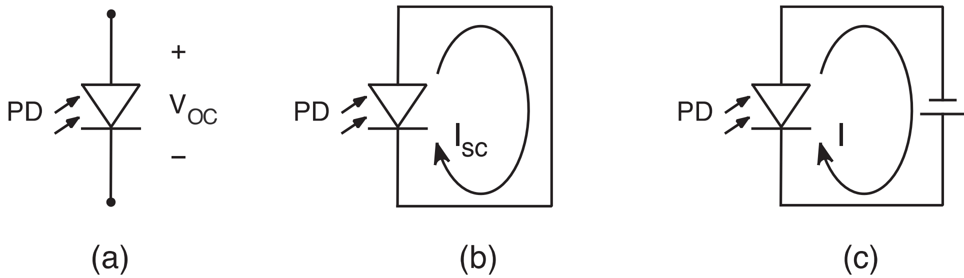


Fig. 3. Basic operating mode of a PD: (a) open circuit mode, (b) short circuit mode, and (c) reverse-bias mode.

In contrast, in the short circuit (SC) mode, the terminal of the photodiode is short-circuited. This allows the photogenerated current to flow in a loop as illustrated in Figure 3(b). In Figure 2(b), this is represented by the point at which the I-V characteristic curve intersects the y-axis. The current that flows in the loop in SC mode is also known as the short circuit, I_{sc}, and it has the same magnitude as I_{ph}.

In reverse bias mode, a reverse bias is applied across the photodiode as shown in Figure 3(c). Therefore it is operated in the lower left quadrant of Figure 2(b). Note that by applying a bias voltage, the potential difference across the p-n junction changes to $V_0 - V$, and the balance between drift and diffusion in the p-n junction also changes. This will affect the depletion width (W) and **E** as well. The dependence of W on the bias voltage can be described by,

$$W = K(V_0 - V)^{m_j} \tag{5}$$

where, K is a constant, and m_j depends on the junction geometry ($m_j = 1/2$ for a step junction and $m_j = 1/3$ for a linear junction). Therefore, operating in reverse bias has the effect of increasing W. The increase in W is not as great as the change in the potential difference, because $m_j < 1$, so **E** should also increase. From the point of view of charge distribution and Gauss's Law, a wider depletion region exposes more of the ionic space charge, which in turn increases the electric field.

The widened depletion region under reverse bias creates a greater photogeneration region, while the stronger **E** increases the drift velocity of the photogenerated carriers. In principle, the drift velocity increases in proportion with **E**, so even with an increase in W given by Equation (5), the transit time (the average time that a drifting carrier to reach the end of the depletion region) is reduced. Therefore signal loss due to recombination in the depletion region is reduced. Because of these beneficial effects, reverse bias operation is often preferred.

4.4 Dark current

4.4.1 Saturation current – diffusion of minority carriers

As shown in Figure 2(b) there exist a small current under revers bias in the I-V characteristic even in dark condition. This dark current is caused by saturation current, I_S. On the boundaries of the depletion region, minority carriers (electrons on the p-side and holes on the n-type side) can diffuse into the depletion region. Because of the built-in electric field, these diffused minority carrier may drift across to the depletion region; this is a source of the saturation current. Therefore, in photodiodes there exist a dark current; however, the diffusion process is not the only contribution the dark current.

4.4.2 Generation-recombination current

Apart from the diffusion contribution to the dark current, carriers that are generated by thermal excitation inter-band trap (defect) states in the depletion region can also have a contribution to the dark current. This trap-assisted process is essentially the reverse of Shottky-Read-Hall (SRH) recombination. Just like the photogenerated carriers, carriers created by trap-assisted generation in the depletion region can also be swept away from the depletion region by drift before they can recombine and form part of the dark current. This dark current contribution is known as the Generation-Recombination (G-R) current, and it can be more significant than the diffusion contribution.

4.4.3 Tunneling currents

Under sufficient reverse bias, the E_C on the n-side can fall below E_V on the p-side. In this condition, there is a finite possibility that an electron in the valence band of the p-side can tunnel through the bandgap into the n-side conduction band. This process is call *direct tunneling* or *band-to-band tunneling*. If sufficient numbers of direct tunneling events occur, its contribution to the dark current will be measurable.
Tunneling can also occur through an inter-band trap (defect) state. Due to thermal excitation, a carrier can be trapped in one these states. If this state exists in the depletion region, and sufficient reverse biased is applied, a trapped electron from the valence band can have energy higher than E_C, and tunneling into the conduction band can occur. This is called trap-assisted tunnel.

4.4.4 Surface leakage current

Due to the interruption of the crystal lattice structure, there can be a high density of surface charge and interface states at the physical surface of a device. The surface charge and interface states can affect the position of the depletion region, as well as behaving as generation-recombination centers. Therefore, the surface of a device can introduce another contribution to the dark current called surface leakage current. Passivation of the surface can be used to control the surface leaking current. This is usually achieved by adding a layer of insulator such as oxide to the surface.

4.4.5 Frankel-poole current

When a sufficiently large electric field is applied to an insulator, it will start to conduct. This is called the Frankel-Poole Effect. This effect is caused by escape of electrons from their localized state into the conduction band. Therefore, the Frankel-Poole Effect can occur in a semiconductor as well. When a sufficiently large reverse bias is applied across a p-n

junction, electrons generated by the Frankel-Poole Effect can also have a contribution to the dark current.

4.4.5 Impact ionization current

Under reverse bias, the motion of carriers in the depletion can be described as a drift where the carriers are repeatedly accelerated by the electric field and collide with the atoms in the crystal lattice. Under strong reverse bias, the acceleration between collisions can be large enough for a carrier to obtain the energy required to dislodge a valance electron to create a new electron-hole pair. This process is known as impact ionization and it can generate new carriers that contribute to the reverse bias current. When the applied reverse bias is beyond the breakdown voltage, V_{bd}, impact ionization becomes a dominant factor of the photodiode behavior, and the photodiode is said to be operating in avalanche mode.

4.4.6 Summary of dark current

Table 1 summarizes the different dark currents and their dependence. When these dark currents are taken into consideration, the photodiode no longer follows the ideal diode characteristic. A detailed discussion on dark current can be found in [39].

Process	Dependence
Diffusion	$\propto \exp\left(-\frac{E_g}{kT}\right)$
G-R	$\propto \sqrt{V}\exp\left(-\frac{E_g}{2kT}\right)$
Band-to-band tunneling	$\propto V^2\exp\left(-\frac{a}{V}\right)$
Trap-assisted tunneling	$\propto \exp\left(-\frac{a\prime}{V}\right)^2$
Surface Leakage	$\propto \exp\left(-\frac{E_g}{2kT}\right)$
Frankel-Poole	$\propto V\exp\left(-\frac{b}{T}\right)$
Impact Ionization	$\propto \exp\left(-\frac{c}{T}\right)$

Table 1. Summary of dark current dependence, where a, a', b and c are constants.

4.5 Noise

There are two basic noise generating mechanisms in a photodiode: the statistical fluctuation in the number of carriers and photons, and the random motion of the carriers. The statistical fluctuation in particle numbers is the cause of shot noise. The root mean square of the current fluctuation due to shot noise is

$$i_{sh,\ rms} = \sqrt{2\ q\ I_{avg}\ \Delta f} \tag{6}$$

where, I_{avg} is the average signal current and Δf is the bandwidth of the signal. The signal-to-noise ratio (SNR) of shot noise is given by,

$$\mathrm{SNR} = \frac{\sqrt{\mathrm{I_{avg}}}}{2\,\mathrm{q}\,\Delta\mathrm{f}}. \tag{7}$$

The random movement of carriers produces thermal noise also known as Johnson-Nyquist noise. The root mean square of the current fluctuation due to Johnson-Nyquist noise is

$$i_{sh,\ rms} = \sqrt{\frac{2\,k\,T\,\Delta f}{R_L}} \tag{8}$$

where, R_L is the load resistance.

4.6 Capacitance and dynamic esponse

In high-speed light detection application of the photodiode, the dynamic response of the photodiode is of the utmost importance. The dynamic response of the photodiode depends on the drift velocities of the photogenerated carriers, the junction capacitance that is associated with the space charge in the depletion region, and the diffusion of photogenerated carriers from the quasi-neutral regions into the diffusion region. The delay related to the drift can be characterized by a transit time, the amount of time that it takes a photogenerated carrier to reach the quasi-neutral region. It is simply given by,

$$t_{\mathrm{tr}} = v_{\mathrm{drift}}\,x = \mu\,\mathrm{E}\,x \tag{9}$$

where, v_{drift} is the drift velocity, x is the distance from the point of photogeneration to the quasi-neutral region, and μ is the mobility of the carrier. Then, the longest possible transit time is,

$$t_{\mathrm{tr}}(\max) = \mu\,\mathrm{E}\,W. \tag{10}$$

Another delaying factor is due to the bias voltage dependence of the depletion layer width. A change in bias voltage will change the depletion region width as described by $W=K(V_0 - V)^{m_j}$ Equation (5), which in turn changes the amount of exposed space charge. This change in the amount of space charge due to a change in the bias voltage can simply be modeled by the junction capacitance. The junction capacitance is given by,

$$C_{\mathrm{PD}} = \frac{\varepsilon A}{W} \tag{11}$$

where, ε is the dielectric constant, A is the cross-sectional area of the p-n junction.

Photogenerated carriers in the quasi-neutral region are normally lost by recombination. However, on occasion, the minority species of the photogenerated electron-hole pair can diffuse into the depletion region and contribute to the photogenerated current. Although the contribution to the overall signal by these carrier diffusion is small, a delay due to this diffusion process is observable [40, 41]. The time that takes a minority carrier in the quasi-neutral region to diffuse into the depletion region is approximately,

$$t_{\mathrm{diff}} = \frac{x^2}{4D} \tag{12}$$

where, x is the carrier's distance to the depletion region boundary, and D is the diffusion constant.

5. Accumulation mode for signal integration in imaging

The basic operating modes are very useful for real time measurement of light intensity that falls on a photodiode. However, in imaging applications, an integrated signal from the photodiode is often preferred. The integrated signal from the accumulation of photogenerated charge provides a form of statistical binning. This statistical binning not only provides a stronger accumulative signal, but the accumulative signal is a more faithful representation of light intensity that falls on the pixel within the time period of signal integration than an instantaneous measurement, especially for weak or noisy signals. Moreover, in imaging applications, the integration period of the pixels can be synchronized using a shutter. In this case, the stored charge can then be read one by one after the exposure. The result is an image.
Charge accumulation can be achieved in the accumulation mode, and it is used in CCD and CMOS imaging.

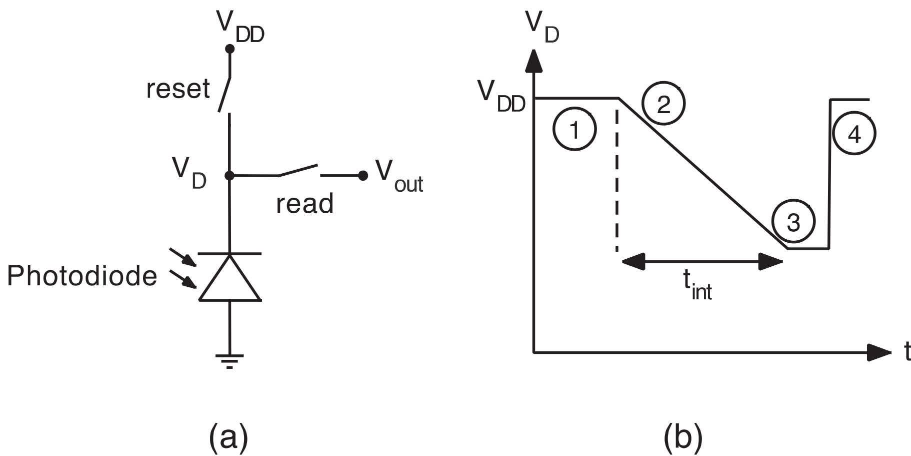


Fig. 4. An abstract representation of a circuit for accumulation mode operation of a photodiode

Figure 4(a) shows an abstract representation of a circuit for operating a photodiode in accumulation mode. This circuit represents the circuit element of a pixel in an imaging array. The goal of the circuit is to accumulate the photogenerated carriers (as charge) within a set period of time and readout the amount of charge collection as the voltage V_{out}. Here is a description of an operation cycle, assuming that there is an optical shutter for exposure control:

1. To begin, the reset switch is closed. The photodiode is in reverse bias, and $V_D = V_{DD}$.
2. The reset switch opens, follow by the shutter. V_D starts to drop.
3. After t_{int}, the shutter closes, followed by the read switch, and $V_{out} = V_D$.
4. Read switch opens, and then reset switch closes. Circuit returns to initial state.

To measure the light intensity of the pixel for single image, only one cycle of operation is needed. To measure a series of images (e.g. in video recording), the operation cycle can be repeated continuously. During the integration phase, the rate of voltage drop depends on I_{ph}, I_d and C_{PD}, which can be described by the following differential equation,

$$\frac{dV_{\mathrm{D}}}{dt} = -\frac{I_{\mathrm{ph}}+I_d}{C_{\mathrm{PD}}} \tag{13}$$

Note that C_{PD} varies with the bias voltage. As discussed in [42], this voltage drop is usually linear for a wide range of values below V_{DD}. Therefore, the voltage readout at the end of the integration period can be used as a measurement of the amount of light that has fallen on the pixel during the integration phase.

6. CMOS active pixel sensing

There are many ways to implement CMOS pixel sensing using accumulation mode. The simplest active pixel sensing implementation is the T3-APS, as shown in Figure 5. In this implementation, a MOSFET is used to control the reset current. Another MOSFET is used as a source follower (SF) to keep the output voltage the same as V_{D}, and the control of the readout is control by yet another MOSFET (Select). The SF-FET prevents discharge during the readout; therefore it is possible to re-read the same pixel twice without losing V_{D}. Because of difficulty in suppressing thermal noise and other design limitations [43-45], the T3-APS has been superseded by other implementation such as T4-APS [43, 44, 46].

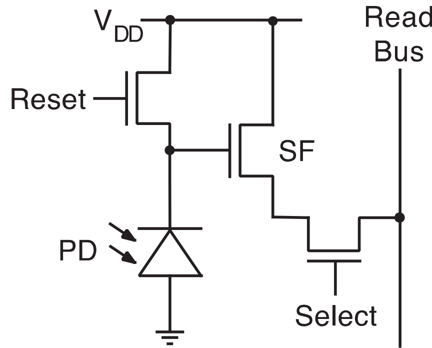


Fig. 5. T3-APS implementation of CMOS pixel sensing.

7. P-i-N photodiode

Increasing the active region where the signal generating photogenerated carriers originated from should in principle increase the collection efficiency. Previously, increasing the active region by increasing W through reverse biased was discussed. To further increase the active region, the device geometry can be altered to include an intrinsic region between the p-type and n-type regions, as shown in Figure 6, this results is a p-i-n junction.

Figure 6(e), shows a band diagram of a p-i-n junction under reverse bias. Under reverse bias, certain assumptions can be made because the external field has driven almost all the carriers from the intrinsic region [47]. These assumptions are,

- No net charge in the intrinsic region.
- A very narrow depletion region on the doped side of each of the doped-intrinsic junctions.

With these assumptions, the depletion layer width is simply the width of the intrinsic region; the electric field in the intrinsic region is simply,

$$E = \frac{qV_0 - V}{W}, \tag{14}$$

and the junction capacitance is given by Equation 11.
Consequently, the p-i-n photodiode under reverse bias can have a C_{PD} smaller than a p-n junction photodiode, but the wider depletion region also implies a longer transit time. Therefore, transit time becomes the dominant limiting factor in the speed the device, and operating the device under sufficient reverse bias is essential, and quick estimation in [41] shows that the transit time of a carrier to cross the intrinsic layer is on the order 0.1 ns. Moreover, to optimize for quantum efficiency without sacrificing speed, it is common practice to design the intrinsic layer thickness to be larger than absorption length given by α-1, but not much more [41]. Further discussion on p-i-n photodiodes can be found in references [40, 41].

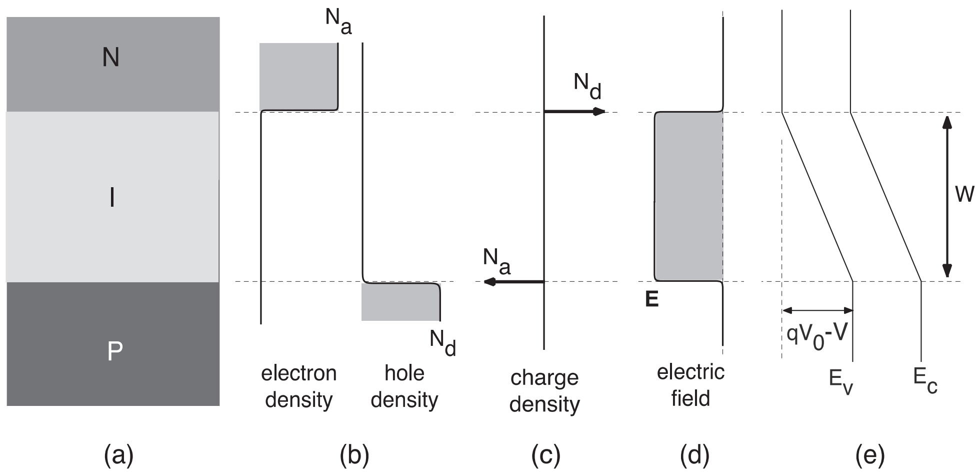


Fig. 6. (a) Structure, (b) carrier distribution, (c) charge distribution, (d) electric field, and (e) band diagram of P-I-N junction under reverse bias.

8. Avalanche photodiode

The avalanche photodiode (APD) is the solid state equivalent of a photomultiplier tube. Its main application is for detection of weak optical signal such as single photon events. The APD exploits the impact ionization of carriers by photogenerated carriers under extremely high reverse bias. During its operation, carriers that trigger impact ionization and the carriers that are generated by impact ionization continue to drift to cause more impact ionization events to occur. The result is a cascade of impact ionization events that produces an avalanche effect to amplify the photogenerated current. The amplification of the current is given by a multiplication factor, $M = I_{\text{mph}} / I_{\text{ph}}$. It is fundamentally related to the ionization coefficients of the carrier by the follow Equation,

$$1 - \frac{1}{M} = \int_0^W \alpha_n \exp\left[-\int_x^W (\alpha_n - \alpha_p) dx'\right] dx \tag{15}$$

where, α_n and α_p are the ionization coefficients of the electrons and holes respectively [48]. Empirically, it can be approximated by,

$$M = \frac{1}{1-\left(\frac{V}{V_{BD}}\right)^n} \quad (16)$$

where, n is material dependent parameter [49]. When taking the dark current into account, it becomes

$$M = \frac{1}{1-\left(\frac{V-IR'}{V_{BD}}\right)^n} \quad (17)$$

where, I is the total current following through the APD, and R' is the differential resistance observed in junction breakdown [50]. Because the random nature of impact ionization, APD suffers from another form of statistical noise call excessive noise [41]. It is given by,

$$F = M\left[1-(1-k)\left(\frac{M-1}{M}\right)^2\right] \quad (18)$$

where, k is the ratio α_n/α_p.
An APD can also be operated in Geiger mode under reverse bias beyond the breakdown voltage. In this case, an electron-hole pair generated by a single photon will trigger the avalanche effect that generates a large current signal. Therefore, photon counting (like particle counting with a Geiger counter) can be achieved in Geiger mode. An APD that can operate in Geiger mode is also known as a single photon avalanche diode (SPAD).

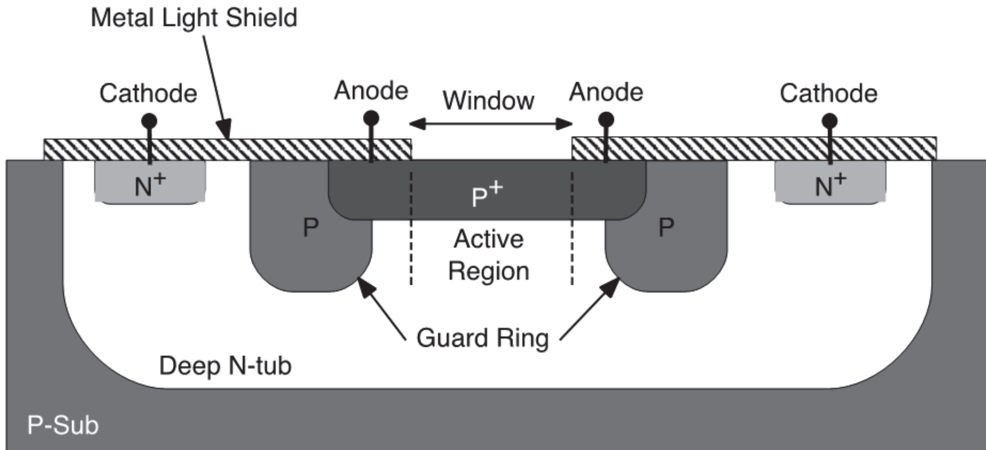


Fig. 7. The cross-section of a SPAD CMOS sensor [51] showing the guard ring surrounding the active region.

Because APDs operate in or near the breakdown region, a physical feature known as a guard ring around the active region is used prevent surface breakdown, as shown in Figure 7. Moreover, the high reverse bias voltage required to produce the avalanche effect had hindered the incorporation of CMOS technology with APD. In 2000, Biber et al. at Centre Suisse d'Electronique et de Microtechnique (CSEM) produced a 12×24 pixel APD fabricated

in standard BiCMOS technology [52, 53]. Since that pioneering work, there has been a steady growth in the development of CMOS APD [54-56] and CMOS SPAD [51, 57-61] for application such as fluorescence sensing [51, 62, 63] and particle detection [64].

9. Mask-level layout and structure

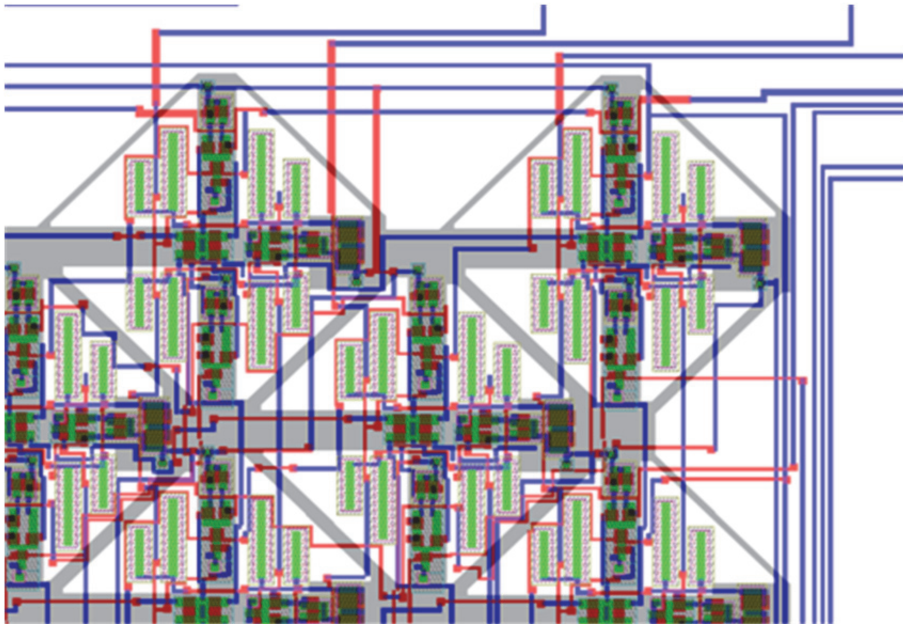
Mask-level layout is the process in which integrated circuits (IC) are defined in terms of their fabrication layers. In this process, a circuit is defined in terms of the functional layers that are later fabricated at a foundry. The designer specifies sizes, locations, connections and spacings of all of the devices in a circuit. The term “mask-level” means that the design results in specifications for the lithographic masks used to make the IC.

Mask-level layout is done using specialized computer-aided design (CAD) tools, such as CADENCE, L-Edit, or the venerable Magic. The process is essentially one of drawing colored boxes or other shapes, and “connecting” them by drawing other boxes. The view of design is as one looks at an IC held in one’s hand; for example, see Figure 8 to compare the mask-level layout and the fabricated circuit itself.

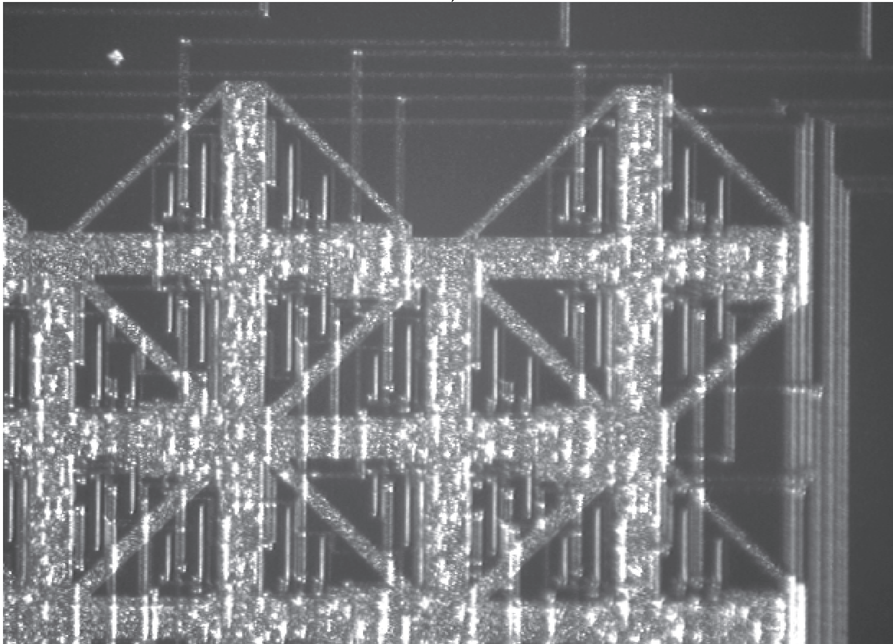
The layout of a circuit involves the placement of shapes, where the shapes are predefined layers; the layers are represented as different colors or shadings or fill. For example, a P-type MOSFET is shown with the layers labeled in Figure 9. These layers, in addition to other metal layers and, in some processes, another poly layer are available for circuit design. The designer controls the sizes of all of the components (wires, capacitors, transistors, etc.), how they are connected and their location. How large or small these components can be, how close to other layers, and whether layers can overlap are all determined by the design rules for the particular technology. The CAD tools used for layout will have these rules available and can aid the designer by indicating when these rules are violated. The rules help to ensure that the circuit can be fabricated without the fabrication process itself causing the circuit to fail. These failures are a result of limitations in the lithography process that can create undesired short circuits or open circuits. However, the layout tools by themselves do not guarantee proper functionality; the designer must perform a circuit verification step in which the layout is compared to the original circuit to verify that the two match. See [65] for additional information on layout.

What is clearly missing in this process is the third dimension, which is the depth into the substrate or height above the substrate. Since designers have no control over depths or heights, it is “hidden” from the designer in the CAD tools and ignored. While, for integrated photodetectors, the depths of the p-n junctions are critical, the designer still does not have control over these in a standard CMOS fabrication process. Also, it should be noted that most processes are an n-well/p-substrate process, and we will assume that for the discussion of photodetector devices.

The simplest structure is the vertical p-n photodiode; it can be formed as a p+ region in an n-well (Figure 10) or as an n+ region in a p-substrate (Figure 11). The uncovered active area is the region that is intended to be the photon collection area. To prevent unwanted charge carrier generation, other regions of the IC should be covered in a metal layer (see Figure 12). It is also possible to create a p-n photodiode using n-well/p-substrate; the difference with this type of device is that the p-n junction is quite a bit deeper than the junction for the p+ active or n+ active devices. As discussed previously, this can affect the wavelength sensitivity of the device.



a)



b)

Fig. 8. (a) Mask-level layout of a circuit. (b) Microphotograph of the fabricated circuit shown in (a).

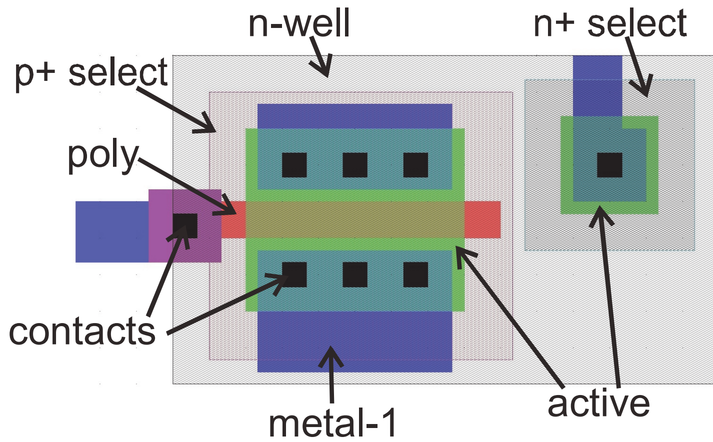


Fig. 9. Mask level layout with layers labeled.

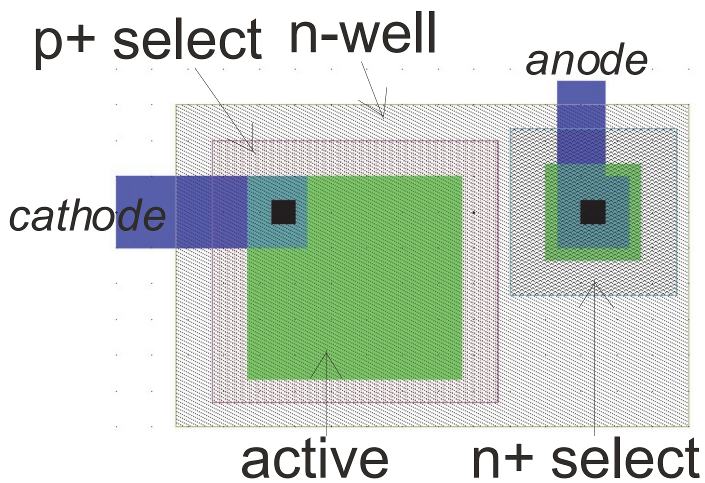


Fig. 10. CMOS-based photodiode: p+/n-well.

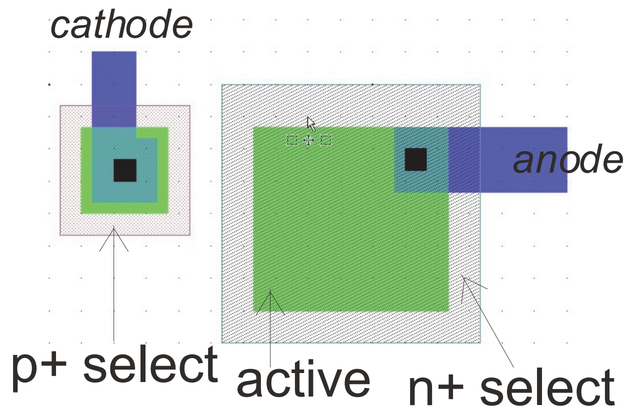


Fig. 11. CMOS-based photodiode: n+/p-substrate.

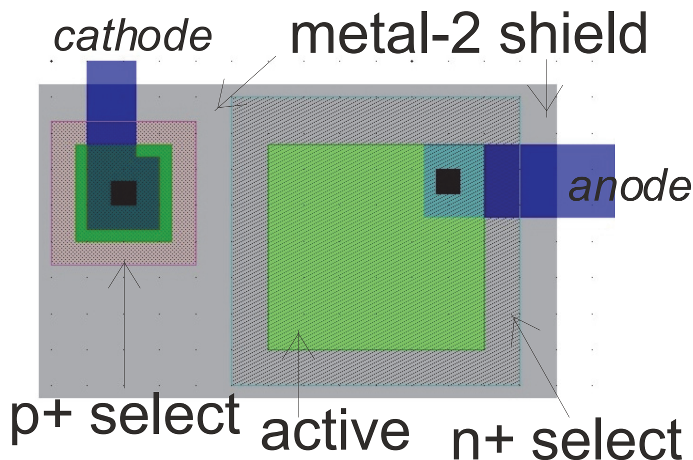


Fig. 12. Photodiode with metal-2 layer as a shield to block photons from reaching the substrate.

In order to create a dense array of photodiodes, as needed for a high-resolution imaging device, the ratio of the area designated for the collection of light to the area used for control circuitry should be as high as possible. This is known as the fill-factor. Ideally, this would be unity, but this is not possible for an imaging device with individual pixel read-out. Thus, actual fill-factors are less than one. A layout of the APS pixel as shown in Figure 5 is shown in Figure 13 . The fill factor of this 3 transistor pixel is 41% using scalable CMOS design rules. The metal shielding of the circuitry outside of the photodetector is not shown for clarity; in practice, this shielding would cover all non-photoactive areas and can also be used as the circuit ground plane.

In order to create an array of imaging pixels, the layout not only requires maximizing the active photodetector area, but also requires that the power (VDD), control and readout wires be routed so that when a pixel is put into an array, these wires are aligned. An example of this is shown in Figure 14.

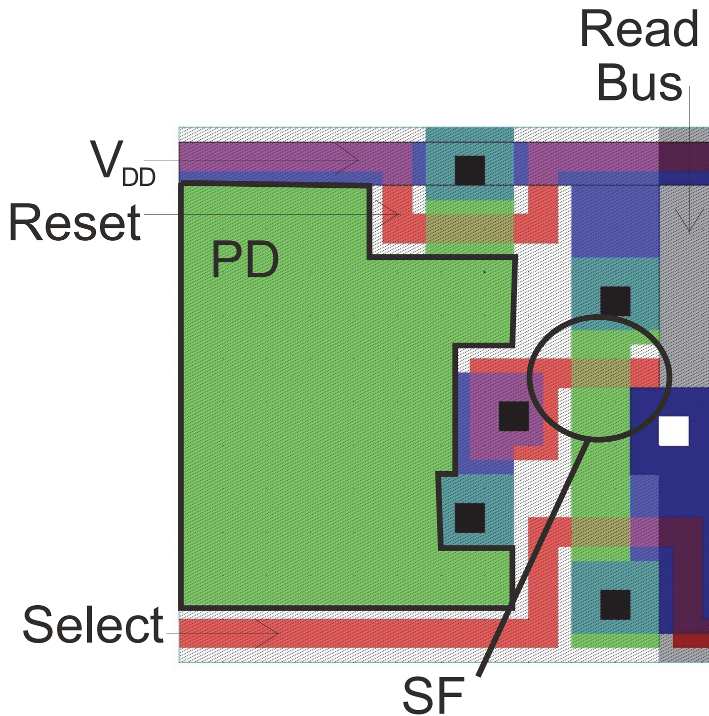


Fig. 13. Layout of the T3-APS pixel as shown in Figure 5.

A slightly more complex structure is the buried double junction, or BDJ, photodiode [66]. The BDJ is formed from two vertically stacked standard p-n junctions, shown in Figure 15. The shallow junction is formed by the p-base and N-well, and the deep junction is formed by the N-well and P-substrate. As discussed previously, the depth of each junction is determined by the thickness of the p-base and n-well. Incident light will be absorbed at different depths, so the two junctions will produce currents based on the wavelength of the incident light. The current flow through two junctions is proportional to the light intensity at the junction depth. An example layout of the structure is shown in Figure 16.

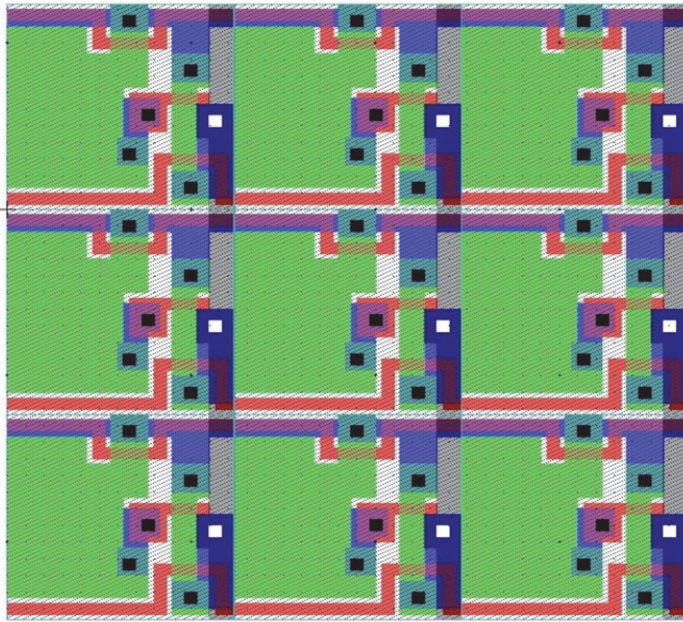


Fig. 14. Simple 3x3 arrays of pixels shown in Figure 13. Notice that the wires align vertically and horizontally.

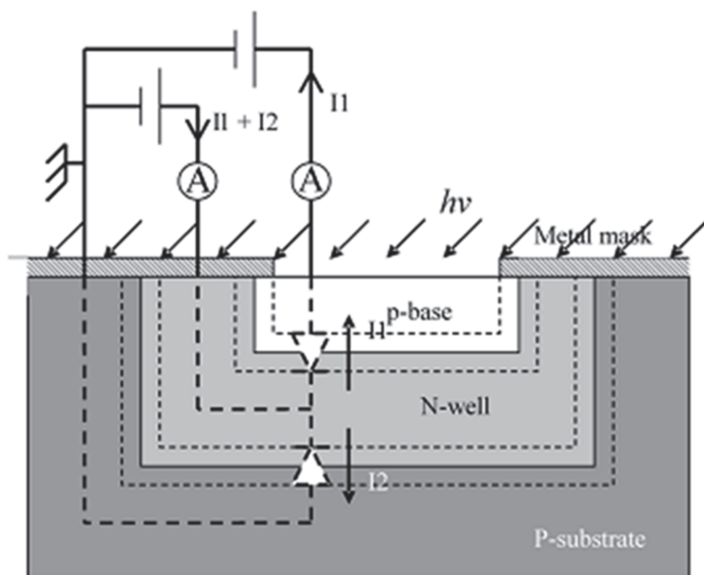


Fig. 15. Cross-sectional view of the BDJ (not to scale).

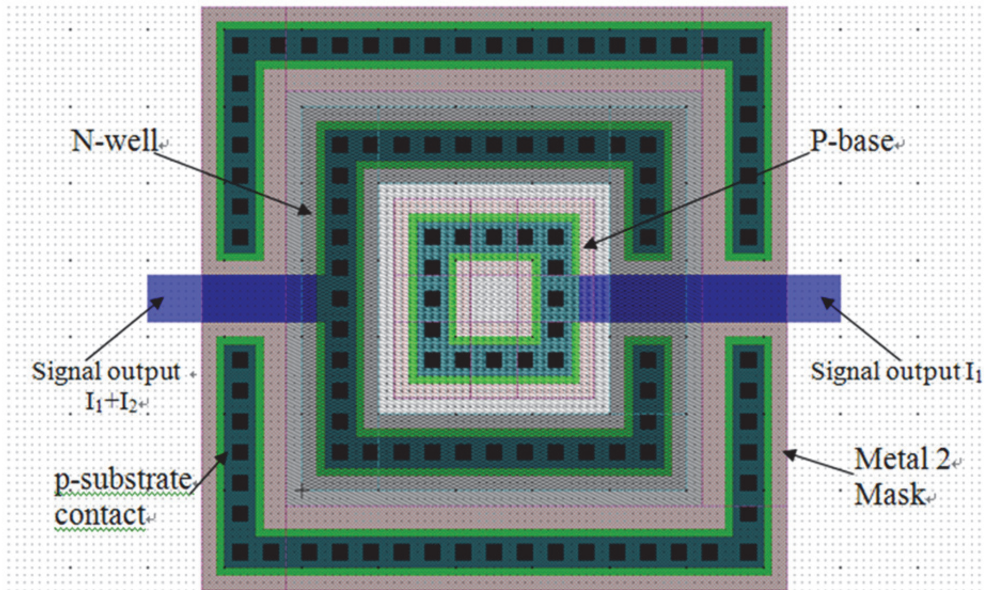


Fig. 16. Layout of a 24µm x 24µm BDJ photodiode in the AMI 1.5µm process.

The final structure we will discuss is the phototransistor. Typically, a phototransistor can produce a current output which is several times larger than a same area size photodiode due to the high gain of the transistor. However, a major drawback of these phototransistors is their low bandwidth, which is typically limited to hundreds of kHz. Additionally, the current-irradiance relationship of the phototransistor is nonlinear, which makes it less than ideal to use in many applications. Like the photodiode, there are a number of possible configurations for the phototransistor in a standard CMOS process, such as the vertical p-n-p phototransistor and lateral p-n-p phototransistor [67-71].

A cross-section of a vertical p-n-p phototransistor is shown in Figure 17 and an example layout is provided in Figure 18.

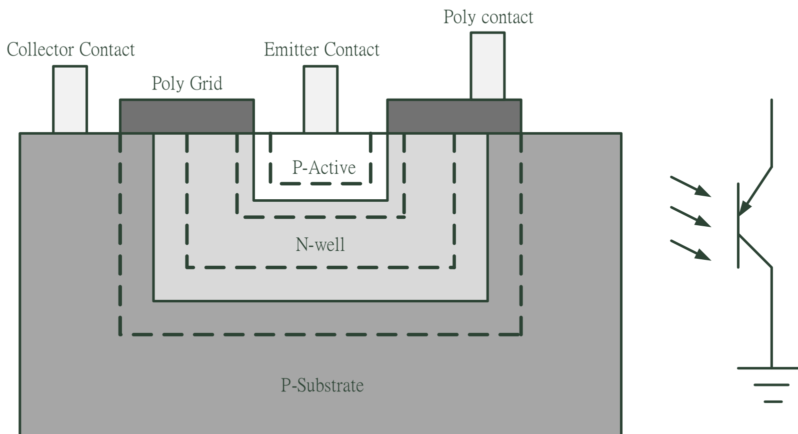


Fig. 17. Cross-sectional view of a vertical p-n-p phototransistor (not to scale).

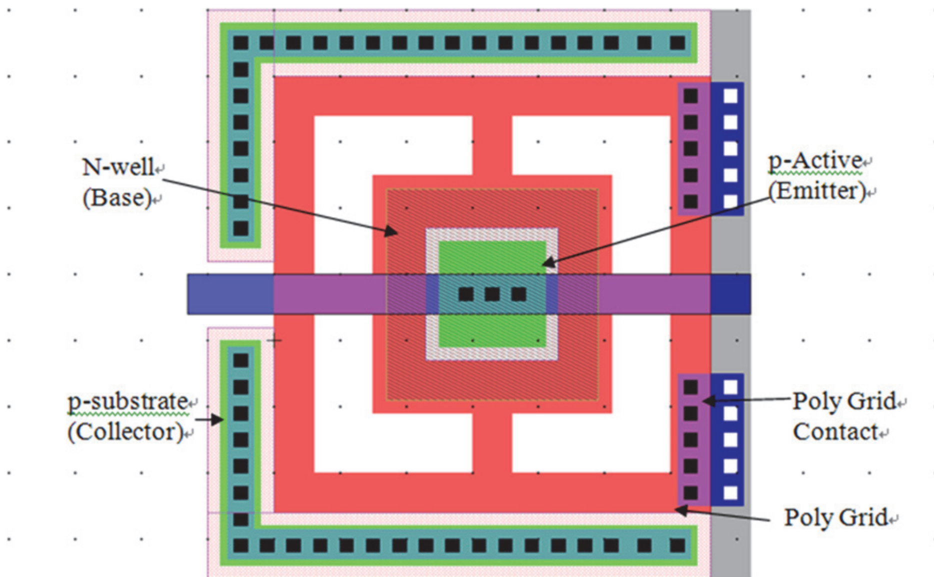


Fig. 18. Layout of a 60 x 60 um vertical p-n-p phototransistor.

10. Current trends in performance optimization

10.1 Tune device responsivity

In a standard CMOS technology, a photodiode can be formed using different available active layers, including n-active/p-substrate, p-active/n-well and n-well/p-substrate, to form a p-n junction. In a photodiode, the photo-conversion mostly takes place in the depletion region where an incident photon creates an electron and hole pair with the

electron passing to the n-region and hole to the p-region. Hence, varying the depth at which the depletion region forms in the silicon wafer would control the performance of the photodiodes in terms of responsivity and quantum efficiency. Also, varying the width of the depletion region by appropriately applying a reverse bias to the photodiode, one could control the response time of the detector. A wider depletion region reduces the junction capacitance of the p-n-junction and improves the response time of the detector.

Here, we will aim to understand the effect on responsivity and external quantum efficiency on the design of photodiode structures. Given that all materials in a standard CMOS process are set by the manufacturer, the external quantum efficiency, which takes into account only the photon-generated carriers collected as a result of the light absorption or, in other words, the useful portion of signal generated by interaction of light and photodetector, is more relevant. The external quantum efficiency depends on the absorption coefficient of the material, a (units: cm^{-1}) and thickness of the absorbing material. Assuming that the entire incident light is absorbed by the detector, if the photon flux density incident at the surface is Φ_o, then the photon flux at depth, x, is given by Beer's law (Equation 1) [72].

The external quantum efficiency is also a function of wavelength of the incident light. Thus in a CMOS photodiode, one can strategically chose the depth of the location of the depletion region to which photons are likely to penetrate and thereby optimize the photodetector to provide high absorption for particular spectrum of wavelengths. In practical optoelectronic systems development, *responsivity*, that is defined as the output current divided by the incident light power, may be a more relevant performance metric. Responsivity is related to quantum efficiency by a factor of $h\mu/q$, where, q is the electron charge, h is Planck's constant, and μ is the frequency of the incident photon. The spectral response curve is a plot of responsivity as a function of wavelength.

Thus, to optimize a silicon photodiode structure for detecting blue wavelengths, the depletion region should be near to the silicon surface. For red wavelengths, the depletion region should be placed deeper in the silicon substrate. Based on this idea, Yotter et al. [73] have compared photodiode structures (p-active/n-well and n-well/p-substrate) to develop photodiodes better suited for blue or green wavelengths for specific biosensing applications. The blue-enhanced structure used interdigitated p+-diffusion fingers to increase the depletion region area near the surface of the detector, while the green-enhanced structure used n-well fingers to increase the depletion region slightly deeper within the substrate. Bolten et al. [74] provided a thorough treatment of the photodiode types and their properties. They reported that in a standard CMOS process *n*-well/*p*-substrate structure provides relatively better quantum efficiency for biosensors operating in visible electromagnetic spectrum.

Using the properties that external quantum efficiency varies as a function of wavelength of the incident light and Beer's law, many research groups reported the use of buried double p-n junction (BDJ) and buried triple p-n junction structures, which can be implemented with a standard CMOS process, for monochromatic color detection [75, 76]. The BDJ structure has two standard p-n junctions (p-base/n-well/p-substrate) are stacked vertically in the CMOS chip. For the BDJ detector, we obtain I_{top} (only from top p-n junction) and I_{bottom} (sum of currents from top and bottom p-n junctions) from the detector. The current ratio, $I_{top}/I_{top-}I_{bottom}$ can be used for the color/ wavelength measurements. The CMOS BDJ detector has been used for fluorescence detection in microarrays [77], and for the detection and measurement of ambient light sources [78]. The BDJ color detectors have been used in many chemical and biological sensors such as seawater pH measurement [79] and volatile organic compounds detection [80].

10.2 Monolithic integration of photonic devices on photodetectors
10.2.1 Microlens and microfilters

Most CMOS image sensors are monochrome devices that record the intensity of light. A layer of color filters or color filter array (CFA) is fabricated over the silicon integrated circuit using a photolithography process to add color detection to the digital camera. CFA is prepared by using color pigments mixed with photosensitive polymer or resist carriers. Many recent digital color imaging systems use three separate sensors to record red, green, and blue scene information, but single-sensor systems are also common [81]. Typically, single-sensor color imaging systems have a color filter array (CFA) in a Bayer pattern as shown in Figure 19**.** The Bayer pattern was invented at Eastman Kodak Company by Bryce Bayer in 1976 [82]. This CFA pattern has twice as many green filtered pixels as red or blue filtered pixels. The spatial configuration of the Bayer pattern is tailored to match the optimum sensitivity of human vision perception. Imager sensors also include microlenses placed over the CFA to improve the photosensitivity of the detection system and improve the efficiency of light collection by proper focusing of the incident optical signal over the photodetectors [83]. A microlens is usually a single element with one plane surface facing the photodiode and one spherical convex surface to collect and focus the light. Thus, as photons pass through the microlens and through the CFA filter, thus passing only wavelengths of red, green, or blue color and finally reach the photodetectors. The photodetectors are integrated as part of an active pixel sensor to convert the incident optical signal into electrical output [84]. The analog electrical data from the photopixels are then digitized by an analog-to-digital converter. To produce a full color image, a spatial color interpolation operation known as demosaicing is used. The image data is then further processed to perform color correction and calibration, white balancing, infrared rejection, and reducing the negative effects of faulty pixels [85, 86].

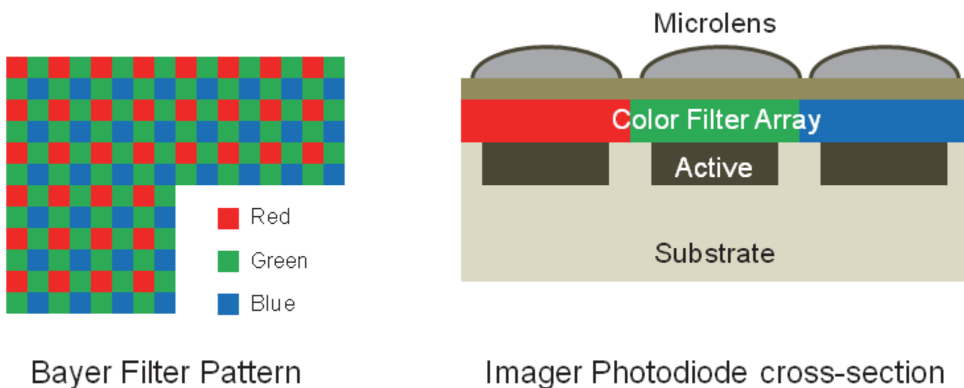


Fig. 19. Bayer film pattern and microlenses integrated onto a device.

One of the first example of a monolithic microlens array fabricated on the MOS color imager was done using photolithography of a polymethacrylate type transparent photoresist [87].

In commercial camera production, glass substrates are typically used as carrier and spacer wafers for the lenses and are filled with an optical polymer material which is photolithographically patterned to form the microlenses. A fairly straightforward method used in many microlens implementations is to photolithographically pattern small cylinders of a suitable resin on a substrate. The small cylinders are then melted in carefully controlled heating conditions. Hence, after melting they tend to form into small hemispheres due to surface tension forces. However, molten resin had a tendency to spread such that lens size and spatial location is difficult to control. A well-defined spherical surface for the microlens is required to achieve high numerical aperture which improves the image sensor efficiency. Different techniques are used to control the spherical shape and spatial location of the microlens including pre-treatment of the substrate to adjust the surface tension to control the reflow of the microlens [88] and use of microstructures such as pedestals to control the surface contact angle [83]. In more recent processes the glass substrates are eliminated and instead microlenses are made with polymer materials that are molded using master stamps. The molded polymer microlenses are cured with ultra violet exposure or heat treatment. By replacing the glass substrates, wafer-level system manufacturers face fewer constraints on the integration optics and imager integrated circuit enabling the production of compact and efficient imager sensors.

10.3 Waveguides, gratings, and couplers

In this section, we will concentrate on understanding device architectures that deal with monolithic integration of photonic waveguides, gratings and couplers with CMOS photodetectors for applications in optoelectronics to improve quantum efficiency, spectral response selectivity, and planar coupling and guiding of light signals to on-chip photodetectors. CMOS photodetectors operate only in visible and near infra-red region between 400nm and 1.1μm of the electromagnetic spectrum. There are applications in sensing and optical communications in this wavelength region where silicon or CMOS photodetectors can offer low-cost and miniaturized systems. Monolithic integration of photonic components with silicon/CMOS photodetectors started as a major research area since early 1980's [89-92]. It is advantageous that a monolithic integrated optoelectronic system on silicon use materials typically employed in CMOS-technology. The dielectrics available in CMOS are favorable as the light guiding layer for wavelengths in the visible and near infrared region. The available materials in CMOS processing technology to develop the photonic devices include layers such as silicon nitride [3], Phospho-Silicate Glass (PSG) —SiO_2 doped with P_2O_5 [4] or silicon oxynitride layers deposited as insulating and passivation layers. Confinement of light is achieved by an increased refractive index in the light guiding film, compared to silicon oxide. The first proposed CMOS compatible devices were based on using silicon oxynitride waveguides sandwiched with silicon dioxide (SiO_2) layers [93-95].

System-level integration is commonly used in compact spectrometers with Lysaght et al. [96] [97] first proposing a spectrometer system in the year 1991 that would integrate silicon photodiode array with microfabricated grating structures for diffraction of the incident light signals and subsequent detection of the optical spectrum components by the silicon photodiode array. More recent and commercial available compact spectrometers use a CMOS line array. Csutak et al. [98] provided an excellent background for related work done

prior to their research article. After considering the absorption length of silicon and required bandwidth for high-speed optical communications, the improvement of quantum efficiency of the photodetectors remains an important challenge.

10.4 Biosensors on CMOS detectors

Many research groups are working on the idea of contact imaging systems for imaging or detection of a biological specimens coupled directly to the chip surface which was first proposed by Lamture et al. [99] using a CCD camera. As the photodetector components in biosensors, CMOS imagers are preferable to convert the optical signals into electrical signals because of monolithic integration of photodetection elements and signal processing circuitry leading to low cost miniaturized systems [100, 101]. In 1998, a system termed as bioluminescent-bioreporter integrated circuit (BBIC) was introduced that described placing genetically engineered whole cell bioreporters on integrated CMOS microluminometers [102]. In a more recent implementation of BBIC system includes sensing low concentrations of a wide range of toxic substances such as salicylate and naphthalene in both gas and liquid environments using genetically altered bacteria, *Pseudomonas fluorescens* 5RL, as the bioreporter [103]. BBIC system operates on the basis of using a large CMOS photodiode (1.47 mm^2 area using *n*-well/*p*-substrate structure) for detection of low levels of luminescence signals by integration of the photocurrent generated by the photodiode over time and a current-to-frequency converter as signal processing circuit to provide a digital output proportional to the photocurrent.

Recent implementations of contact imaging include using custom-designed CMOS imagers as platform for imaging of cell cultures [104] and DNA sequencing [105, 106]. Now researchers are working on the integration of molded and photolithographically patterned polymer filters and microlenses with CMOS photodetectors and imagers towards complete development of miniaturized luminescence sensors. Typically, luminescence sensors require an optical excitation source for exciting the sensor materials with electromagnetic radiation and a photodetector component for monitoring the excited state emission response from the sensor materials at a higher wavelength electromagnetic spectrum that is filtered from the excitation input. The next step towards convenient monolithic integration of filters, biological support substrates, and microfluidic interfaces create interesting challenges for engineers and scientists. A recent report discusses the approach of using poly(acrylic acid) filters integrated with custom-designed CMOS imager ICs to detect fluorescent microspheres [107]. Polydimethylsiloxane (PDMS) could offer a more versatile material to fabricate lenses, filters, diffusers and other components for optical sensors [108]. PDMS is a silicone-based organic polymer that is soft, flexible, biocompatible and optically transparent and well amenable to various microfabrication techniques. PDMS can be doped with apolar hydrophobic color dyes such as Sudan-I, -II or -III to form optical filters that work in different regions of visible electromagnetic spectrum [109]. The Authors group recently proposed a prototype compact optical gaseous O_2 sensor microsystem using xerogel based sensor elements that are contact printed on top of trapezoidal lens-like microstructures molded into PDMS that is doped with Sudan-II dye as shown in Figure 20 [110]. The molded PDMS structure serves triple purpose acting as immobilization platform, filtering of excitation radiation and focusing of emission radiation onto the detectors. The PDMS structure is then integrated on top of a custom design CMOS imager to create a contact

imaging sensor system. The low-cost polymer based filters is best suited for LED excitation and may not be able to provide optimum excitation rejection performance when laser radiation is used for excitation. As a more traditional alternative, Singh et al. [111] proposed micromachining a commercially available thin-film interference filter and gluing it to the CMOS imager die.

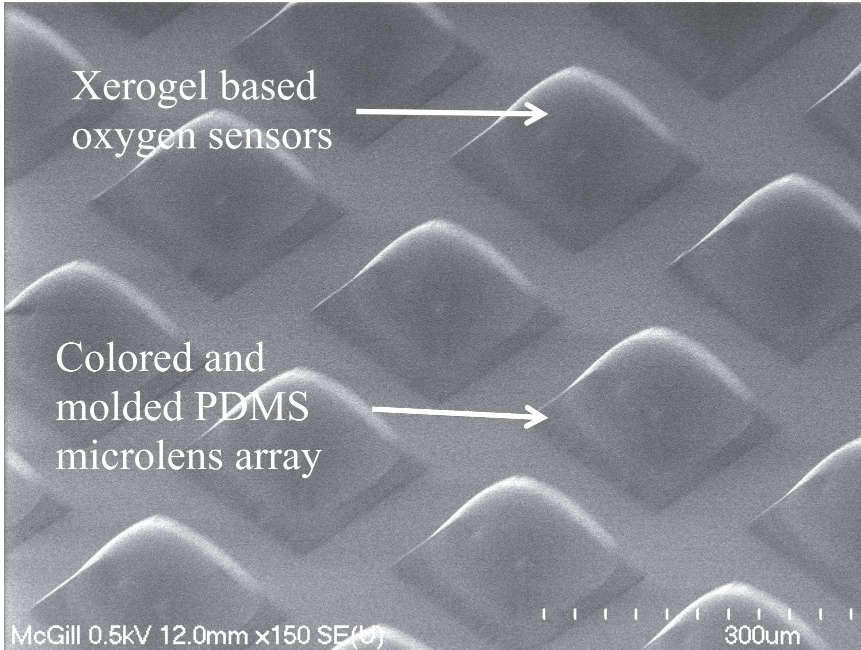


Fig. 20. Fabricated microlenses and xerogel sensors.

11. Optical sensor chip with Color Change-Intensity Change Disambiguation (CCICD)

In this section we present a CMOS-sensor chip that can detect irradiance and color information simultaneously. Compared with other BDJ-based systems [112-114], this system includes an irradiance detection pathway that can be used in combination with the color information to provide color change -intensity change disambiguation (CCICD). The irradiance detection pathway is based on the work by Delbruck and Mead [23] with a single standard CMOS photodiode (see Figure 21). Thus, this pathway can function ambient light conditions without an additional light source, is more robust to background light changes, has a higher bandwidth for time-varying signals, and has the ability to emulate adaptation to background light levels, which is an important phenomena found in biological visual systems. The color detection pathway consists of a BDJ photodetector and subsequent processing circuitry to produce a single voltage as the chip output *without* additional external circuitry. The BDJ produces two currents which are used as inputs to individual logarithmic current to voltage converter circuits whose outputs are converted to a voltage

difference using a differential amplifier. The output of this pathway is a single voltage that represents the color of the input signal, with better than 50nm resolution. Both pathways are integrated on the same IC.

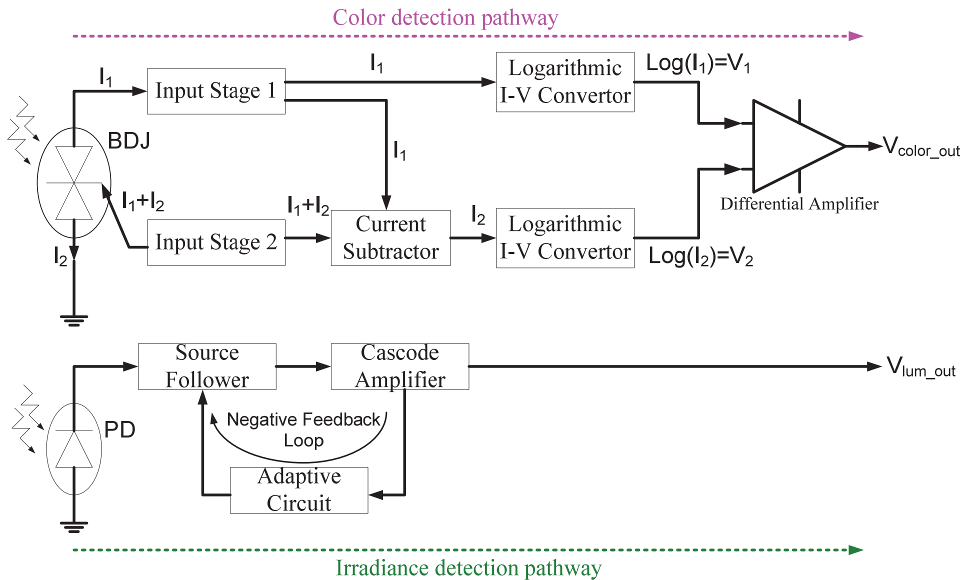


Fig. 21. Block diagram of the sensor chip pathways (from [115]).

The irradiance detection pathway out is shown in Figure 21. The response of the irradiance detection pathway circuit is logarithmic over the measured irradiance range spanning nearly 3 orders of magnitude.

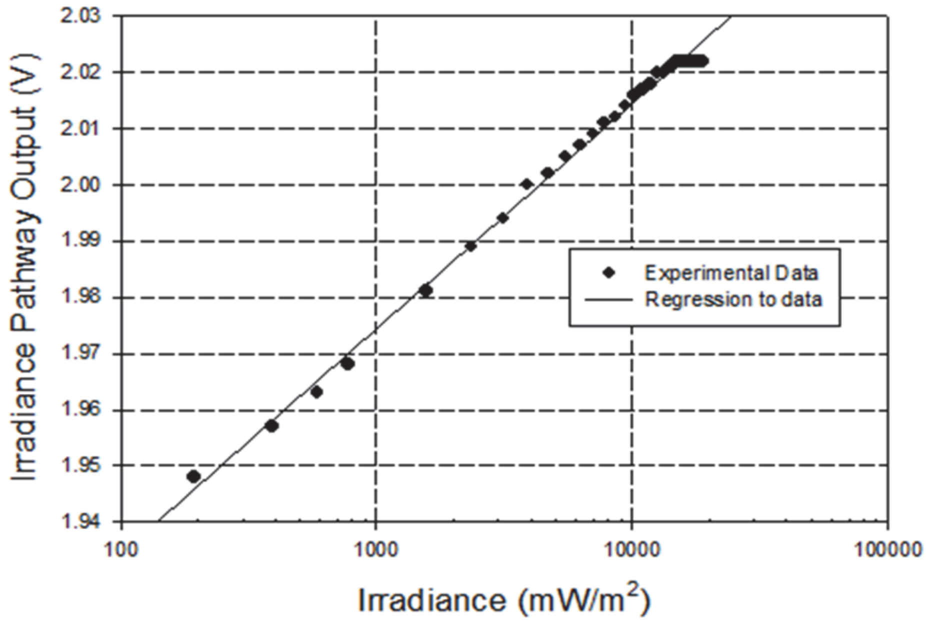


Fig. 22. Output of the irradiance detection pathway (from [115]).

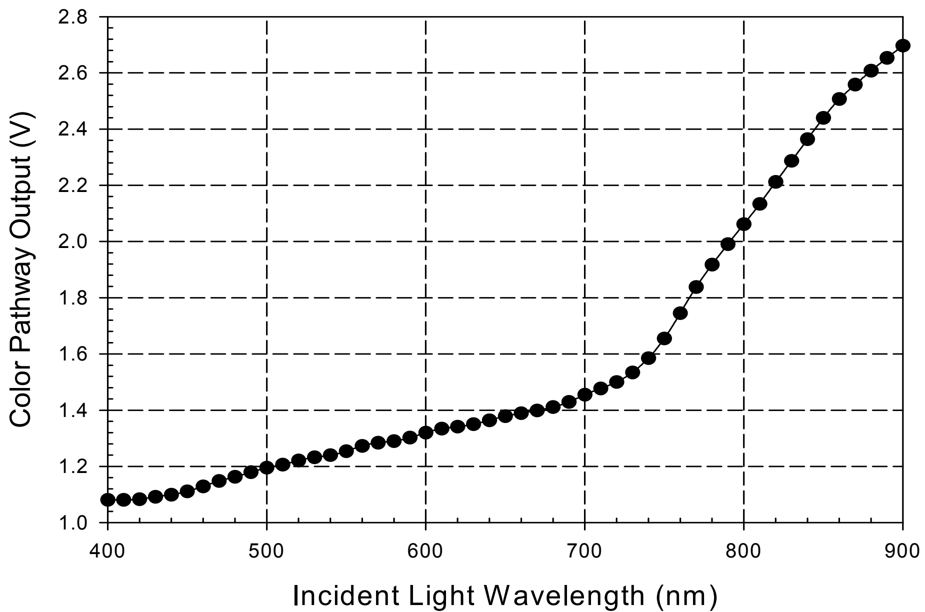


Fig. 23. Output of the color detection pathway as a function of incident light wavelength (from [115]).

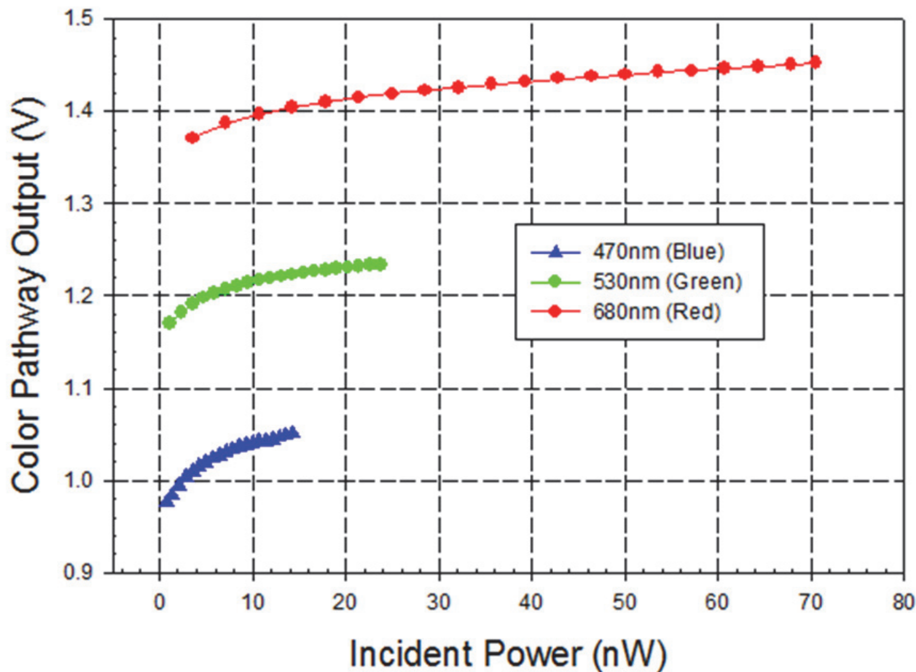


Fig. 24. Output of the color detection pathway as a function of incident power (from [115]).

From the experimental results, we can see that the output voltage is larger for longer incident light wavelengths (see Figure 23). So, for a practical implementation based on this chip, a look-up table can be used to map the output voltage to the incident wavelength. Moreover, from Figure 24, the changes in the output voltage caused by irradiance change will not cause confusion between which color (primary wavelength) is detected; the R, G and B curves will not overlap for a normal operating range of irradiance. The reason for this performance is because the I2/I1 ratio from the BDJ is (ideally) independent of light intensity.

12. References

[1] A. R. Thryft. (2011, February 1) CCD vs. CMOS image sensors. *Test & Measurement World*. Available: http://www.tmworld.com/article/512499-CCD_vs_CMOS_image_sensors.php

[2] J. W. Horton, *et al.*, "The scanistor-A solid-state image scanner," *Proceedings of the IEEE*, vol. 52, pp. 1513-1528, 1964.

[3] S. R. Morrison, "A new type of photosensitive junction device," *Solid-State Electronics*, vol. 6, pp. 485-494.

[4] M. A. Schuster and G. Strull, "A monolithic mosaic of photon sensors for solid state imaging applications," in *Electron Devices Meeting, 1965 International*, 1965, pp. 20-21.

[5] M. A. Schuster and G. Strull, "A monolithic mosaic of photon sensors for solid-state imaging applications," *Electron Devices, IEEE Transactions on,* vol. 13, pp. 907-912, 1966.

[6] G. P. Weckler, "Operation of p-n Junction Photodetectors in a Phton Flux Integrating Mode," *IEEE Journal of Solid State Circuits,* vol. SC-2, pp. 65-73, Sept 1967.

[7] R. Melen, "The tradeoffs in monolithic image sensors; MOS vs CCD," *Electronics,* vol. 46, pp. 106-11, 1973.

[8] S. Mendis, *et al.*, "CMOS active pixel image sensor," *IEEE Trans. Electron Devices,* vol. 41, pp. 452-453, 1994.

[9] S. K. Mendis, *et al.*, "CMOS Active Pixel Image Sensors for Highly Integrated Imaging Systems," *IEEE J. Solid-State Circuits,* vol. 32, pp. 187-196, 1997.

[10] E. R. Fossum, "CMOS Image Sensors: electronic camera-on-a-chip," *IEEE Transactions on Electron Devices,* vol. 44, pp. 1689-1697, 1997.

[11] B. Dierickx, *et al.*, "Random addressable active pixel image sensors," in *Proceedings of the SPIE* vol. 2950, ed. Berlin, Germany: SPIE, 1996, pp. 2-7.

[12] D. X. D. Yang and A. El Gamal, "Comparative analysis of SNR for image sensors with enhanced dynamic range," in *Proceedings of the SPIE* vol. 3649, ed. San Jose, CA, USA: SPIE, 1999, pp. 197-211.

[13] O. Yadid-Pecht, "Wide-dynamic-range sensors," *Optical Engineering,* vol. 38, pp. 1650-60, 1999.

[14] S. G. Chamberlain and J. P. Y. Lee, "A novel wide dynamic range silicon photodetector and linear imaging array," *Electron Devices, IEEE Transactions on,* vol. 31, pp. 175-182, 1984.

[15] N. Ricquier and B. Dierickx, "Pixel structure with logarithmic response for intelligent and flexible imager architectures," *Microelectronic Engineering,* vol. 19, pp. 631-634, 1992.

[16] N. Ricquier and B. Dierickx, "Random addressable CMOS image sensor for industrial applications," *Sensors and Actuators A (Physical),* vol. A44, pp. 29-35, 1994.

[17] M. A. Pace and J. J. Zarnowski, "Complementary metal oxide semiconductor imaging device," USA Patent 6 084 229, 2000.

[18] S. Hanson, *et al.*, "A 0.5 V Sub-Microwatt CMOS Image Sensor With Pulse-Width Modulation Read-Out," *Solid-State Circuits, IEEE Journal of,* vol. 45, pp. 759-767, 2010.

[19] M. Furuta, *et al.*, "A high-speed, high-sensitivity digital CMOS image sensor with a global shutter and 12-bit column-parallel cyclic A/D converters," *IEEE Journal of Solid-State Circuits,* vol. 42, pp. 766-74, 2007.

[20] S. Kleinfelder, *et al.*, "A 10000 frames/s CMOS digital pixel sensor," *Solid-State Circuits, IEEE Journal of,* vol. 36, pp. 2049-2059, 2001.

[21] C. A. Mead and M. A. Mahowald, "A Silicon Model of Early Visual Processing," *Neural Networks,* vol. 1, pp. 91-97, 1988.

[22] T. Delbruck, "Silicon retina with correlation-based, velocity-tuned pixels," *Neural Networks, IEEE Transactions on,* vol. 4, p. 529, 1993.

[23] T. Delbruck and C. A. Mead, "Adaptive Photoreceptor with Wide Dynamic Range," in *1994 IEEE International Symposium on Circuits and Systems*, London, England, 1994, pp. 339-342.
[24] R. A. Deutschmann and C. Koch, "An analog VLSI velocity sensor using the gradient method," *Proceedings of the 1998 IEEE International Symposium on Circuits and Systems,*, vol. 6, p. 649, 1998.
[25] T. J. Drabik, *et al.*, "2D Silicon/Ferroelectric Liquid Crystal Spatial Light Modulators," *IEEE Micro,* vol. 15, pp. 67-76, 1995.
[26] R. Etienne-Cummings, *et al.*, "Hardware implementation of a visual-motion pixel using oriented spatiotemporal neural filters," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on [see also Circuits and Systems II: Express Briefs, IEEE Transactions on],* vol. 46, pp. 1121-1136, 1999.
[27] R. Etienne-Cummings, *et al.*, "A foveated silicon retina for two-dimensional tracking," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing,* vol. 47, pp. 504-517, 2000.
[28] C. M. Higgins, *et al.*, "Pulse-based 2-D motion sensors," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on [see also Circuits and Systems II: Express Briefs, IEEE Transactions on],* vol. 46, pp. 677-687, 1999.
[29] R. M. Philipp and R. A. Etienne-Cummings, "A 1V current-mode CMOS active pixel sensor," in *Proc. IEEE Int.Symp. Circ. Syst.*, Kobe, Japan, 2005, pp. 4771-1774.
[30] A. Sartori, *et al.*, "A 2D photosensor array with integrated charge amplifier," *Sensors and Actuators A: Physical,* vol. 46, pp. 247-250, 1995.
[31] B. E. Shi, "A One-Dimensional CMOS Focal Plane Array for Gabor-Type Image Filtering," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications,* vol. 46, pp. 323-327, 1999.
[32] A. H. Titus and T. J. Drabik, "An Improved Silicon Retina Chip with Optical Input and Optical Output," in *Proceedings: Tenth Annual IEEE International ASIC Conference*, Portland, OR, USA, 1997, pp. 88-91.
[33] H. G. Graf, *et al.*, "High Dynamic Range CMOS Imager Technologies for Biomedical Applications," *Solid-State Circuits, IEEE Journal of,* vol. 44, pp. 281-289, 2009.
[34] A. H. Titus, *et al.*, "Autonomous low-power glare sensing chip," *Electronics Letters,* vol. 47, pp. 508-509, 2011.
[35] K. Murari, *et al.*, "A CMOS In-Pixel CTIA High-Sensitivity Fluorescence Imager," *Biomedical Circuits and Systems, IEEE Transactions on,* vol. PP, pp. 1-10, 2011.
[36] Y. Lei, *et al.*, "CMOS Imaging of Pin-Printed Xerogel-Based Luminescent Sensor Microarrays," *Sensors Journal, IEEE,* vol. 10, pp. 1824-1832, 2010.
[37] K. Kwang Hyun and K. Young Soo, "Scintillator and CMOS APS Imager for Radiography Conditions," *Nuclear Science, IEEE Transactions on,* vol. 55, pp. 1327-1332, 2008.
[38] K. Murari, *et al.*, "Which Photodiode to Use: A Comparison of CMOS-Compatible Structures," *Sensors Journal, IEEE,* vol. 9, pp. 752-760, 2009.
[39] A. Rogalski, *et al.*, *Narrow-gap semiconductor photodiodes*. Bellingham, Wash.: SPIE Press, 2000.

[40] G. Ghione, *Semiconductor devices for high-speed optoelectronics*. Cambridge; New York: Cambridge University Press, 2009.

[41] K. K. Ng, *Complete guide to semiconductor devices*. Hoboken: Wiley-Interscience, 2002.

[42] G. P. Weckler, "Operation of p-n Junction Photodetectors in a Photon Flux Integrating Mode," *IEEE Journal of Solid State Circuits,* vol. SC-2, pp. 65-73, Sept 1967.

[43] J. Ohta, *Smart CMOS image sensors and applications*. Boca Raton, Fla.; London: CRC ; Taylor & Francis [distributor], 2006.

[44] A. Hoffman*, et al.*, "CMOS Detector Technology," *Experimental Astronomy,* vol. 19, pp. 111-134, Feb 01 2005.

[45] P. Magnan, "Detection of visible photons in CCD and CMOS: A comparative view," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 504, pp. 199-212, Jun 21 2003.

[46] A. Gamal*, et al.*, "CMOS Sensors for Optical Molecular Imaging," in *CMOS Biotechnology*, H. Lee*, et al.*, Eds., ed: Springer US, 2007, pp. 331-379.

[47] B. Sapoval and C. Hermann, *Physics of semiconductors*. New York: Springer-Verlag, 1995.

[48] K. G. Mckay and K. B. Mcafee, "Electron Multiplication in Silicon and Germanium," *Physical Review,* vol. 91, pp. 1079-1084, 1953.

[49] S. Miller, "Ionization Rates for Holes and Electrons in Silicon," *Physical Review,* vol. 105, pp. 1246-1249, Mar 1957.

[50] H. Melochior and W. Lynch, "Signal and Noise Response of High Speed Germanium Avalanche Photodiodes," *Ieee Transactions On Electron Devices,* vol. ED13, pp. 829-&;, 1966.

[51] D. Stoppa*, et al.*, "Single-Photon Avalanche Diode CMOS Sensor for Time-Resolved Fluorescence Measurements," *IEEE Sensors Journal,* vol. 9, pp. 1084-1090, 2009.

[52] A. Biber*, et al.*, "Avalanche photodiode image sensor in standard BiCMOS technology," *Ieee Transactions On Electron Devices,* vol. 47, pp. 2241-2243, Nov 01 2000.

[53] A. Biber*, et al.*, "Avalanche photodiode image sensor in standard silicon BiCMOS technology," *Sensors And Actuators A-Physical,* vol. 90, pp. 82-88, 2001.

[54] A. Rochas*, et al.*, "Low-noise silicon avalanche photodiodes fabricated in conventional CMOS technologies," *Ieee Transactions On Electron Devices,* vol. 49, pp. 387-394, Apr 01 2002.

[55] G. F. Marshall*, et al.*, "Avalanche photodiode-based active pixel imager," *Ieee Transactions On Electron Devices,* vol. 51, pp. 509- 511, Apr 01 2004.

[56] L. Pancheri*, et al.*, "Low-Noise Avalanche Photodiode in Standard 0.35-µm CMOS Technology," *Ieee Transactions On Electron Devices,* vol. 55, pp. 457-461, 2008.

[57] C. Niclass*, et al.*, "Design and characterization of a CMOS 3-D image sensor based on single photon avalanche diodes," presented at the Ieee Journal of Solid-State Circuits, 2005.

[58] C. Niclass*, et al.*, "A 4 µs integration time imager based on CMOS single photon avalanche diode technology," presented at the Sensors And Actuators A-Physical, 2006.

[59] C. Niclass, *et al.*, "A single photon avalanche diode implemented in 130-nm CMOS technology," *Ieee Journal Of Selected Topics In Quantum Electronics,* vol. 13, pp. 863-869, 2007.

[60] D. Stoppa, *et al.*, "A CMOS 3-D imager based on single photon avalanche diode," presented at the Ieee Transactions on Circuits and Systems I-Regular Papers, 2007.

[61] N. Faramarzpour, *et al.*, "Fully integrated single photon avalanche diode detector in standard CMOS 0.18-mu m technology," *Ieee Transactions On Electron Devices,* vol. 55, pp. 760-767, 2008.

[62] D. E. Schwartz, *et al.*, "A Single-Photon Avalanche Diode Array for Fluorescence Lifetime Imaging Microscopy," presented at the Ieee Journal of Solid-State Circuits, 2008.

[63] B. R. Rae, *et al.*, "A Vertically Integrated CMOS Microsystem for Time-Resolved Fluorescence Analysis," presented at the Ieee Transactions on Biomedical Circuits and Systems, 2010.

[64] C. J. Stapels, *et al.*, "CMOS-based avalanche photodiodes for direct particle detection," presented at the Nuclear Instruments & Methods in Physics Research Section a-Accelerators Spectrometers Detectors and Associated Equipment, 2007.

[65] R. J. Baker, *CMOS Circuit Design, Layout, and Simulation*, 3rd ed. Hoboken, N.J.: IEEE Press/Wiley, 2010.

[66] G. de Graaf and R. F. Wolffenbuttel, "Smart optical sensor systems in CMOS for measuring light intensity and colour," *Sensors & Actuators A-Physical,* vol. A67, pp. 115-119, 1998.

[67] H. Zimmermann, "Silicon Optoelectronic Integrated Circuits," *Berlin, Germany: Springer-Verlag,,* 2004.

[68] M. Davenport, "Detection and Processing Strategies for Integrated Xerogel-based Sensor Systems," *in Ph.D. Dissertation, Department of Electrical Engineering: University at Buffalo, The State University of New York,,* 2004.

[69] M.P. Vidal, *et al.*, "A bipolar photodetector compatible with standard CMOS technology," *Solid-State Electronics,* vol. 34, pp. 809-814, 1991.

[70] R. W. Sandage and J. A. Connelly, "Producing phototransistors in a standard digital CMOS technology," in *Circuits and Systems, 1996. ISCAS '96., 'Connecting the World'., 1996 IEEE International Symposium on*, 1996, pp. 369-372 vol.1.

[71] R. W. Sandage and J. A. Connelly, "A 128x128 imaging array using lateral bipolar phototransistors in a standard CMOS process [fingerprint detection]," in *Circuits and Systems, 1998. ISCAS '98. Proceedings of the 1998 IEEE International Symposium on*, 1998, pp. 641-644 vol.6.

[72] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*. New York: Wiley, 1991.

[73] R. A. Yotter, *et al.*, "Optimized CMOS photodetector structures for the detection of green luminescent probes in biological applications," *Sensors and Actuators B-Chemical,* vol. 103, pp. 43-49, Sep 29 2004.

[74] E. K. Bolton, *et al.*, "Integrated CMOS photodetectors and signal processing for very low-level chemical sensing with the bioluminescent bioreporter integrated circuit," *Sensors and Actuators B-Chemical,* vol. 85, pp. 179-185, Jun 20 2002.

[75] M. Ben Chouikha*, et al.*, "Colour detection using buried triple pn junction structure implemented in BiCMOS process," *Electronics Letters,* vol. 34, pp. 120-122, Jan 8 1998.
[76] G. N. Lu*, et al.*, "Colour detection using a buried double p-n junction structure implemented in the CMOS process," *Electronics Letters,* vol. 32, pp. 594-6, 1996.
[77] G. N. Lu*, et al.*, "Investigation of CMOS BDJ detector for fluorescence detection in microarray analysis," *1st Annual International Ieee-Embs Special Topic Conference on Microtechnologies in Medicine & Biology, Proceedings,* pp. 381-386, 2000.
[78] G. de Graaf and R. Wolffenbuttel, "Optical CMOS sensor system for detection of light sources," *Sensors and Actuators a-Physical,* vol. 110, pp. 77-81, Feb 1 2004.
[79] M. Sedjil*, et al.*, "A colorimetric method with the use of BDJ detector for seawater pH measurement," *Analytica Chimica Acta,* vol. 377, pp. 179-184, December 31 1998.
[80] X. Y. Fang*, et al.*, "Colorimetric porous photonic bandgap sensors with integrated CMOS color detectors," *Ieee Sensors Journal,* vol. 6, pp. 661-667, Jun 2006.
[81] R. Palum, "Image sampling with the Bayer color filter array," *Pics 2001: Image Processing, Image Quality, Image Capture, Systems Conference, Proceedings,* pp. 239-245, 2001.
[82] P. L. P. Dillon*, et al.*, "Color Imaging-System Using a Single Ccd Area Array," *Ieee Journal of Solid-State Circuits,* vol. 13, pp. 28-33, 1978.
[83] Z. D. Popovic*, et al.*, "Technique for Monolithic Fabrication of Microlens Arrays," *Applied Optics,* vol. 27, pp. 1281-1284, Apr 1 1988.
[84] A. El Gamal and H. Eltoukhy, "CMOS image sensors," *Ieee Circuits & Devices,* vol. 21, pp. 6-20, May-Jun 2005.
[85] E. R. Fossum, "CMOS image sensors: Electronic camera-on-a-chip," *Ieee Transactions on Electron Devices,* vol. 44, pp. 1689-1698, Oct 1997.
[86] J. Adams*, et al.*, "Color processing in digital cameras," *Ieee Micro,* vol. 18, pp. 20-30, Nov-Dec 1998.
[87] Y. Ishihara and K. Tanigaki, "A High Photosensitivity IL-CCD Image Sensor with Monolithic Resin Lens Array," in *Proceedings of International Electron Devices Meeting*, Washington, DC, 1983, pp. 497-500.
[88] A. Schilling*, et al.*, "Surface profiles of reflow microlenses under the influence of surface tension and gravity," *Optical Engineering,* vol. 39, pp. 2171-2176, Aug 2000.
[89] D. G. Hall, "Survey of Silicon-Based Integrated-Optics," *Computer,* vol. 20, pp. 25-32, Dec 1987.
[90] S. Valette*, et al.*, "Si-Based Integrated-Optics Technologies," *Solid State Technology,* vol. 32, pp. 69-75, Feb 1989.
[91] U. Hilleringmann and K. Goser, "Optoelectronic System Integration on Silicon - Wave-Guides, Photodetectors, and Vlsi Cmos Circuits on One-Chip," *Ieee Transactions on Electron Devices,* vol. 42, pp. 841-846, May 1995.
[92] R. A. Soref, "Silicon-Based Optoelectronics," *Proceedings of the Ieee,* vol. 81, pp. 1687-1706, Dec 1993.
[93] S. Wunderlich*, et al.*, "Integration of Sion Wave-Guides and Photodiodes on Silicon Substrates," *Applied Optics,* vol. 31, pp. 4186-4189, Jul 20 1992.

[94] D. Peters, *et al.*, "Integrated-Optics Based on Silicon Oxynitride Thin-Films Deposited on Silicon Substrates for Sensor Applications," *Sensors and Actuators a-Physical,* vol. 26, pp. 425-431, Mar 1991.
[95] C. H. Henry, *et al.*, "Low-Loss Si3n4-Sio2 Optical Wave-Guides on Si," *Applied Optics,* vol. 26, pp. 2621-2624, Jul 1 1987.
[96] M. J. Lysaght, *et al.*, "Laptop Chemistry - a Fiberoptic, Field-Portable, near-Infrared Spectrometer," *Review of Scientific Instruments,* vol. 62, pp. 507-515, Feb 1991.
[97] H. Bezzaoui and E. Voges, "Integrated-Optics Combined with Micromechanics on Silicon," *Sensors and Actuators a-Physical,* vol. 29, pp. 219-223, Dec 1991.
[98] S. M. Csutak, *et al.*, "CMOS-compatible planar silicon waveguide-grating-coupler photodetectors fabricated on silicon-on-insulator (SOI) substrates," *Ieee Journal of Quantum Electronics,* vol. 38, pp. 477-480, May 2002.
[99] J. B. Lamture, *et al.*, "Direct-Detection of Nucleic-Acid Hybridization on the Surface of a Charge-Coupled-Device," *Nucleic Acids Research,* vol. 22, pp. 2121-2125, Jun 11 1994.
[100] J. P. Golden and F. S. Ligler, "A comparison of imaging methods for use in an array biosensor," *Biosensors & Bioelectronics,* vol. 17, pp. 719-725, Sep 2002.
[101] L. Yao, *et al.*, "CMOS Imaging of Pin-Printed Xerogel based Luminescent Sensor Microarrays," *Ieee Sensors Journal,* vol. 10, pp. 1824-1832, 2010.
[102] M. L. Simpson, *et al.*, "Bioluminescent-bioreporter integrated circuits form novel whole-cell biosensors," *Trends in Biotechnology,* vol. 16, pp. 332-338, Aug 1998.
[103] S. K. Islam, *et al.*, "A bioreporter bioluminescent integrated circuit for very low-level chemical sensing in both gas and liquid environments," *Sensors and Actuators B-Chemical,* vol. 123, pp. 922-928, May 21 2007.
[104] J. Honghao, *et al.*, "CMOS contact imager for monitoring cultured cells," in *Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on*, 2005, pp. 3491-3494 Vol. 4.
[105] H. Eltoukhy, *et al.*, "A 0.18-/mu m CMOS bioluminescence detection lab-on-chip," *Ieee Journal of Solid-State Circuits,* vol. 41, pp. 651-662, Mar 2006.
[106] Y. J. Wang, *et al.*, "A CMOS image sensor utilizing opacity of nanometallic particles for DNA detection," *Ieee Transactions on Electron Devices,* vol. 54, pp. 1549-1554, Jun 2007.
[107] Y. Dattner and O. Yadid-Pecht, "Low Light CMOS Contact Imager with an Integrated Poly-Acrylic Emission Filter for Fluorescence Detection," *Sensors,* vol. 10, pp. 5014-5027, 2010.
[108] T. K. Shih, *et al.*, "Fabrication of PDMS (polydimethylsiloxane) microlens and diffuser using replica molding," *Microelectronic Engineering,* vol. 83, pp. 2499-2503, Nov-Dec 2006.
[109] O. Hofmann, *et al.*, "Monolithically integrated dye-doped PDMS long-pass filters for disposable on-chip fluorescence detection," *Lab on a Chip,* vol. 6, pp. 981-987, 2006.
[110] D. S. Daivasagaya, *et al.*, "Contact CMOS Imaging of Gaseous Oxygen Sensor Array," Sensors and Actuators: B Chemical, vol 157, pp. 408-416, 2011

[111] R. R. Singh, *et al.*, "A CMOS/Thin-Film Fluorescence Contact Imaging Microsystem for DNA Analysis," *Ieee Transactions on Circuits and Systems I-Regular Papers,* vol. 57, pp. 1029-1038, May 2010.
[112] G. D. Graaf and R. F. Wolffenbuttel, "Optical CMOS sensor system for detection of light sources," *Sensors & Actuators A-Physical,* vol. 110, pp. 77-81, 2004.
[113] X. Fang, *et al.*, "Colorimetric porous photonic bandgap sensors with integrated CMOS color detectors," *IEEE Sensors Journal,* vol. 6, pp. 661 - 667, 2006.
[114] G. N. Lu, *et al.*, "Design and testing of a CMOS BDJ detector for integrated micro-analysis systems," *Microelectronics journal,* vol. 32, pp. 227-234, 2001.
[115] F. ZhenHong and A. H. Titus, "CMOS neuromorphic optical sensor chip with color change-intensity change disambiguation (CCICD)," *IEEE Sensors Journal,* vol. 9, pp. 689-96, 2009.

5

Image Artifacts by Charge Pocket in Floating Diffusion Region on CMOS Image Sensors

Sang-Gi Lee[1], Jong-Min Kim[1], Sang-Hoon Bae[2],
Jin-Won Park[1] and Yoon-Jong Lee[1]
[1]Division of Mixed Signal Process Development, Dongbu HiTek
[2]Foveon, Inc. Santa Clara, California
[1]Korea
[2]USA

1. Introduction

Complementary metal-oxide-semiconductor (CMOS) image sensors (CIS) are widely used in applications such as mobile equipment, PC cameras, and portable digital cameras due to their low power consumption and low cost. Today, CIS pixel sizes are being continuously reduced, down to the 1.0 μm level in 90-nm CMOS, as industrial applications require higher pixel density. The industry requests, however, continued good performance even when the pixel dimensions are further reduced. In particular, dark current in CIS is an important parameter that determines the image performance in low light. The dark current component can change the charge capacity in photo diodes (PD) and hence the output signal as a function of location and time (J. P. Albert, 2001 and H.-S. Philip, 1998). An increase in dark current can also affect the dynamic range due to an increase in shot noise (H.Y.Cheng, 2003). By suppressing the dark current, the fixed pattern noise of the imager and the white pixel-defects can be reduced (K. A. Parulski, 1985). Photo diodes and pixel architecture of image sensors have been optimized to reduce image artifacts and hence dark current. (N. V. Loukianova, 2003; H.I.Kwon, 2004).

The key technology feature for high image-quality CMOS image sensor is the formation of a low-leakage buried photodiode with a transfer gate (TG). The buried PD is a promising concept to reduce the dark current. Many researchers reported that the vicinity of the shallow-trench isolation (STI) to the PD is the main source of dark leakage. By designing and characterizing special diode test structures, H. I. Kwon et al. (2004) demonstrated that the dark leakage decreases as the distance between STI and PD increases. Takashi Watanabe et al. (2010) showed that dark current can be improvefd by avoiding charge diffusion injection from the PD to the substrate using an optimized well structure under the PD. The effects of area, perimeter, and corner on lekage were, however, not separately investigated.

The buried floating diffusion (FD), on the other side of the pass transistor, was also studied by H.I.Kwon et al. (2004) and K. Mabuchi, et al. (2004). H.I.Kwon et al. (2004) explained that

the FD can be the source of increased dark currents in the single frame capture mode, even though dark signal in the normal operation mode is negligible because the signal processing time in the FD is very short compared to the integration time of the transfer gate. K. Mabuchi, et al.(2004) explained that zero lag and zero noise can be achieved when the signal electrons in the buried FD are completely transferred. There was no analysis done, however, on image artifacts, called dark spots, originating from the buried FD in CMOS image sensors.

The following two sections of this chapter deal with technologies to reduce dark leakage in buried photo diodes (Richard Merrill), and to reduce image artifacts in floating diffusions. The discussion begins with the design of new diodes with surface and buried configurations for photo diode and floating diffusion region. In the first section, the mechanism accountable for leakage in diode structures and the dependence of leakage current on electric field are discussed, explaining the importance of introducing a buried floating diffusion to improve dark leakage. In the second section, the mechanism for image artifact in floating diffusions is explained and a new method to avoid artifacts is suggested.

In this paper, leakage current was measured at diode structures with and without boron in Si surface to analyze the difference between surface and buried PD by comparing activation energy. Pinch off voltage and the charge pocket as a function of buried FD types are examined. Simulation and image characterization are done to find out the source of the dark spot. Finally generation mechanism for dark spot in CMOS Active Pixel Sensor (APS) is explained.

2. Diode design and fabrication

The design and fabrication of photo diodes and floating diffusion are discussed separately.

2.1 Design of surface and buried photo diodes

New test structures are designed to investigate potential sources of leakage at the area, perimeter, and corner of surface and buried photo diodes.

A schematic cross-section of the surface and buried test structures is shown in Figure. 1(a). The buried structure minimizes the surface leakage component as opposed to the surface diode where interface-state generation increases leakage. Fig. 1(b) shows the top view of the test structure with island cell, block, periphery (peri), and island-types. Here, island cell includes the transfer gate and island type does not. The structures are designed to have same area but different perimeter and number of corners to analyze edge and corner leakage components and compare them with standard n+/pwell diodes of logic devices (Fig. 2). To analyze the contribution from the area, perimeter, and corners, leakage current is measured using test patterns with large area. Large area block diodes (area 60000 μm^2, peri. : 910 μm, corner : 4 ea) are designed to monitor the leakage current due to bulk and surface of the PD, and perimeter finger type diodes (area : 60,000 μm^2, peri., : 12,750 μm, corner : 300 ea) are designed to measure the leakage current by the sidewall junctions. Corner intensive type diodes (area : 60,000 μm^2, peri.: 24,000 μm, corner : 4,800 ea) also are designed to monitor the corner effect on photo diode. To compare the image quality with and without Si surface effects, CMOS image sensors with 3.3μm x 3.3μm pixel size are fabricated using the standard 0.18μm CMOS logic process.

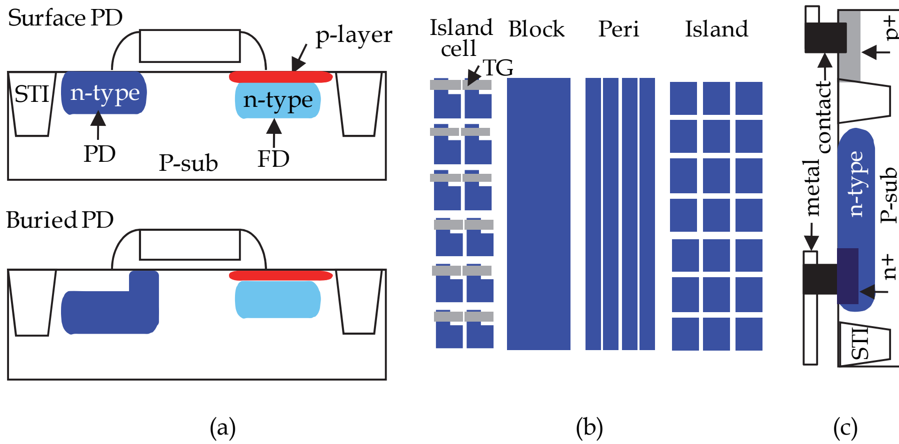


Fig. 1. (a) Schematic cross-section of CMOS image sensor for surface and buried diode. (b) (b) Top view for buried diode structure with island cell, block, peri, and island-types. (c) Concept for test pattern to measure the leakage of diode.

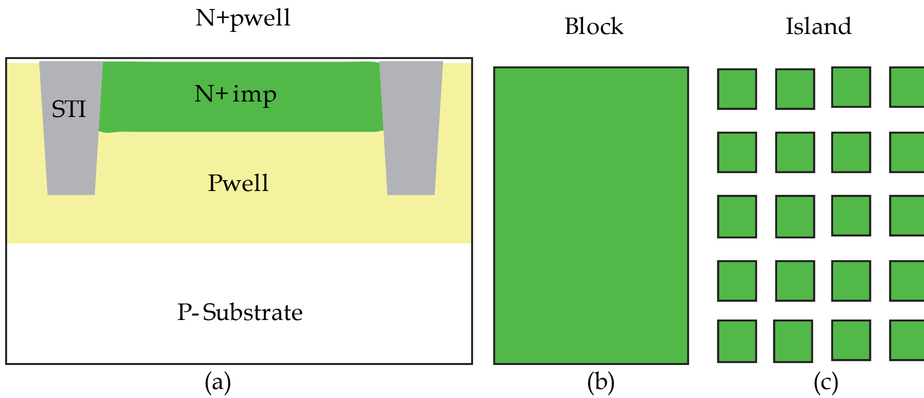


Fig. 2. (a) Shows the cross sectional view for standard n+/pwell diode structure. Fig. 2(b) and 2(c) show top view of standard diode structure for block and island types, respectively. Standard n+/pwell diodes constitute a reference for junction leakage characteristics.

2.2 Diode designs of floating diffusion

To evaluate the leakage characteristics of floating diffusion, two different structures are designed, one with a top p-layer (Fig. 3a), and the other without a top p-layer (Fig. 3b). Both floating diffusions are placed at a safe distance from the STI boundaries. The purpose of adding a top p-layer to one of the buried diffusions is to suppress the surface leakage component by reducing the spread of the junction depletion region into the surface.

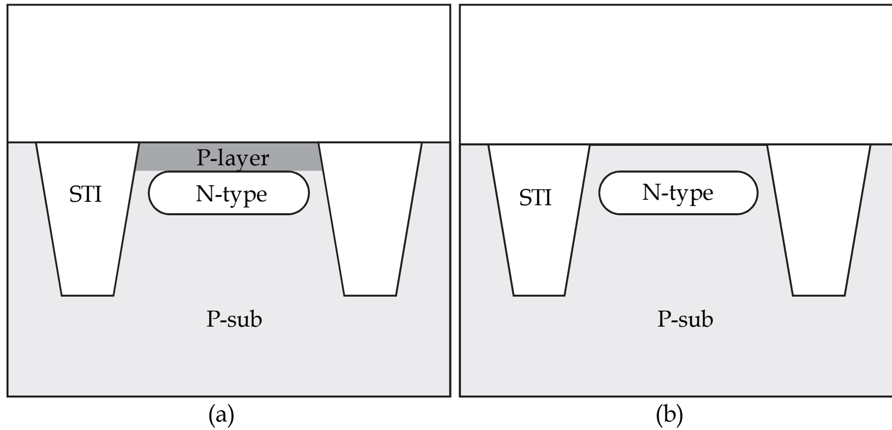


Fig. 3. Schematic cross-section of floating diffusion test structures. (a) with top p-layer. (b) without top p-layer. The p-layer reduces the spread of junction depletion into the surface and hence suppresses the surface leakage component.

2.3 Fabrication

The surface and buried photo diodes in Fig. 1 are fabricated in a typical 0.18um CMOS logic technology. This includes shallow trench isolation, retrograde channel doping, dual gate oxide (70Å/32Å for 3.3 V/1.8 V) and self-aligned gate and source/drain. The diode is formed by implanting an n-region at a phosphorus dose of 5.0E12 cm^{-2} with 75keV while the standard N+ source/drain is implanted with arsenic at a dose of 5E15 cm^{-2} with 50keV. After completion of the process, micro lenses are formed to evaluate the image quality by focusing more light into pixels. Fig.4(a) shows the process sequence for photo diode and floating diffusion region. Photo diode is fabricated after gate patterning. Floating diffusion region as a buried type is applied prosperous implant after photo diode formation and then processed boron implant before and after sidewall etching. In order to control doping profile under the overlap region of transfer gate, two different kinds of concept are processed as shown in Fig. 4(b). Process-1 means floating diffusion applying phosphorus and boron in FDN and boron in FDP to control capacitance in FD region. This concept is applied a large amount of Ph dose under the overlap region. But, in process-2, phosphorus is applied separately in FDN and FDP step. Ph dose is applied smaller dose in FDN step to control doping profile under the sidewall spacer region as a lower dose than process-1 and is added at FDP step for the remaining dose to make same amount with process-1 which is to control capacitance as the same target with process-1 at the non-overlap region. Thus, boron dose is applied same dose on process-1 and process-2 at FDP step. P-layer dose has to be decided to minimize the leakage and control total capacitance in floating diffusion on both processes. Only difference between process-1 and process-2 is Ph dose under gate overlap in FD side, which is to manage potential profile.

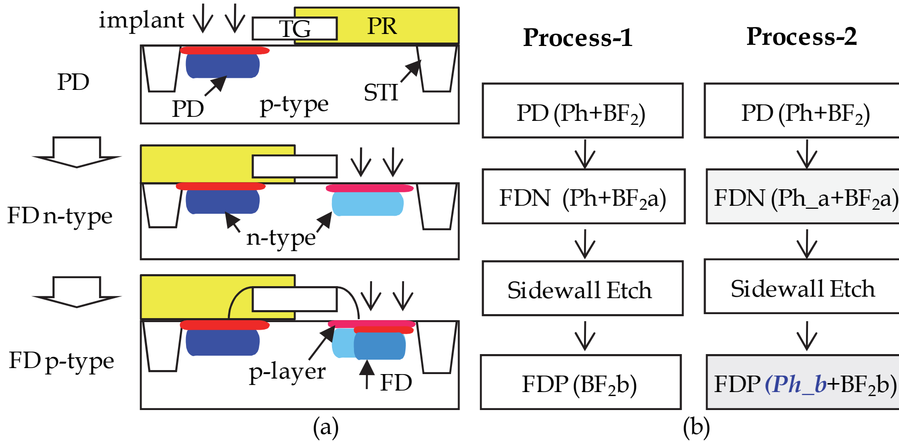


Fig. 4. (a) Process sequence for photo diode and floating diffusion region. Floating diffusion region was fabricated after photo diode process and p-layer region on FD was applied to reduce leakage current. (b) Process condition with process schemes was compared. Total dose for Ph_a and Ph_b in process-2 is the same amount with Ph in process-1

3. Measurements and results

This section discusses the electrical characteristics with test pattern types and temperature dependence that have become critical elements affecting image performance with PD and FD structures in scaled CIS technology.

3.1 Photo diode

The new diode structures to monitor silicon surface effects and STI edge effects on PD are evaluated. Characteristics such as dark leakage and activation energy are analyzed. Also, the PD image performance is characterized to confirm the surface effects.

3.1.1 Junction leakage characteristics

The reverse leakage current was measured at room temperature 25℃. The reverse dark current characteristics are shown for different diode configurations in Fig. 5(a). Good leakage characteristic is found for standard n+/pwell diodes of block type (Fig. 1). Surface diodes exhibit the highest leakage current. For a given reverse applied bias of 2.0V, the standard n+/pwell diode of block type shows around one order of magnitude lower leakage current than the buried diode even though the buried diode was formed at a lower dose than the standard n+/pwell junction. Fig. 5(b) shows the reverse dark current characteristics measured in photo diodes of different area, perimeter, and number of contacts. The block type without the STI edge-effect shows the lowest leakage compared to other structures. As shown in Fig. 5(b), current density of PD for perimeter and corner is higher than that of the area type. Diode like pixel array shows highest leakage due to the more corners when it compares with normal diode types.

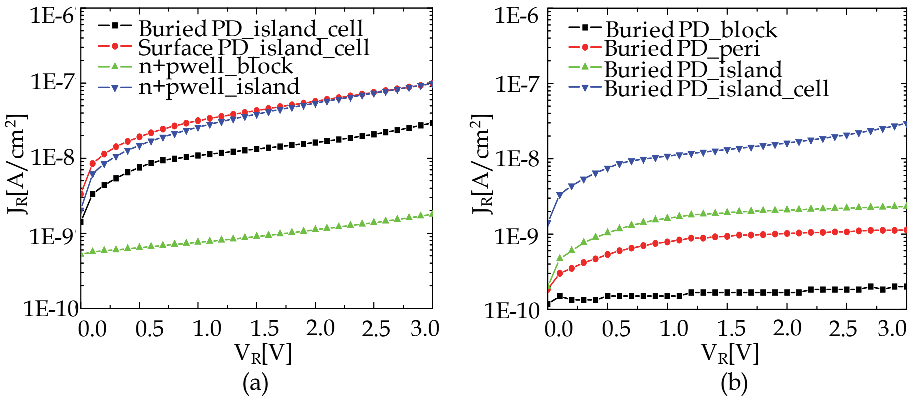


Fig. 5. Measured dark current. (a) Comparison of buried, surface, and standard diode types, (b) Comparison of buried PD of different geometry (or of different area, perimeter and number of corners).

To understand the difference between buried photodiodes and standard n+/pwell diodes, the reverse leakage current was measured as a function of temperature from 25℃ to 200℃ and analyzed with an Arrhenius plot by plotting log I_R/T^3 versus 1/T. Arrhenius plots are shown in Fig. 6 for a reverse voltage of 2.0V. Standard n+/pwell diodes with block type show the lowest temperature dependence of reverse current. The total diode area and extracted activation energy (Ea) at two temperature ranges are given in Table 1. Activation energies for all diodes are close to half energy gap ($E_g/2$) at low temperature range, which suggests that the reverse current is dominated by generation-recombination current (H.D.Lee, 1998). For standard block-type n+/pwell diodes, at high-temperature range, where the reverse current is dominated by diffusion, the extracted activation energy is 1.13 eV, which is approximately the energy gap, E_g. However, the high-temperature activation energy of surface diode is 0.80 eV, which is close to that of a standard n+/pwell diode of island type. Also the high-temperature activation energy for a pixel-type buried diode is around 0.83 eV, lower than Eg. As shown in Fig. 1(c), pixel-type buried diodes have a large number of corners. It is therefore necessary to analyze corner effects in buried diodes.

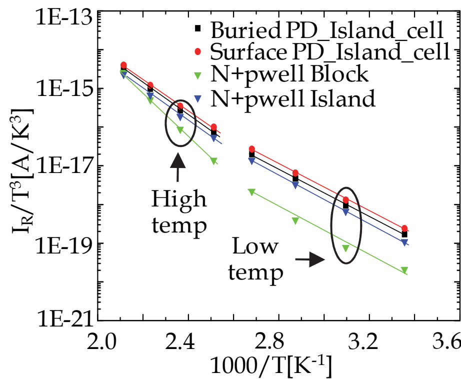


Fig. 6. Arrhenius plot of log I_R/T^3 versus 1/T for island-type diodes and standard n+/pwell diodes.

Diode type	Area (μm^2)	Ea [eV] @ 2.0V	
		Low temp	High temp
Buried PD Island_cell	36,000	0.61	0.83
Surface PD Island_cell	18,000	0.60	0.80
N+Pwell block	49,580	0.59	1.13
N+Pwell island	5,560	0.62	0.81

Table 1. Extracted activation energy.

To understand effects for block, perimeter, and corner with buried PD types, the reverse leakage current is also measured as a function of temperature from 25℃ to 200℃ and analyzed with an Arrhenius plot by plotting log I_R/T^3 versus 1/T. Fig.7 shows an Arrhenius plots (log I_R/T^3 versus 1/T) for the leakage current at 2.0V reverse bias with buried PD types. As shown in Table 2, E_a at high temperature range shows a strong dependence of diode structures. That is, E_a decreases as a length of perimeter and a number of corners increase. Buried PD of island-type and buried PD island-cell types show similar activation energies. This means that the increase in leakage could be attributed to an increase in perimeter and number of corners since surface and STI components is suppressed.

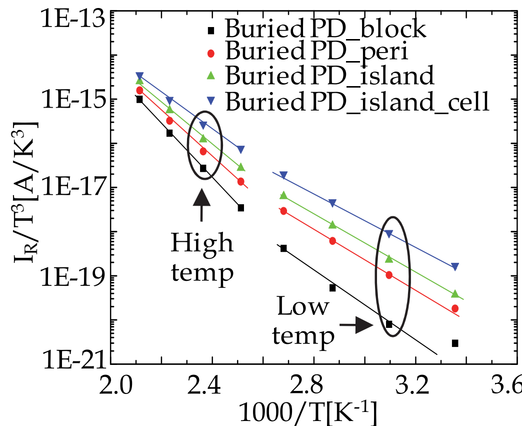


Fig. 7. Arrhenius plot (log IR/T^3 versus 1/T) as a function of pattern types of buried PDs.

Buried PD Type	Area (μm^2)	Peri (μm)	Corner (ea)	Ea [eV] (L/H-temp)
Block	60,000	910	4	0.82 / 1.12
Peri	60,000	12,750	300	0.65 / 1.02
Island	60,000	24,000	4,800	0.66 / 0.97
Island _cell	36,000	14,400	2,880	0.61 / 0.83

Table 2. Activation energy as a function of pattern types of buried PDs

Components of area, perimeter, and corners from each test pattern are defined in Table 2. Three simultaneous equations are used to separately extract the contributions of area, perimeter, and corners to the total leakage current. Fig. 8 shows the results of three components which are extracted from leakage current data for the test pattern with block, perimeter and corner structure.

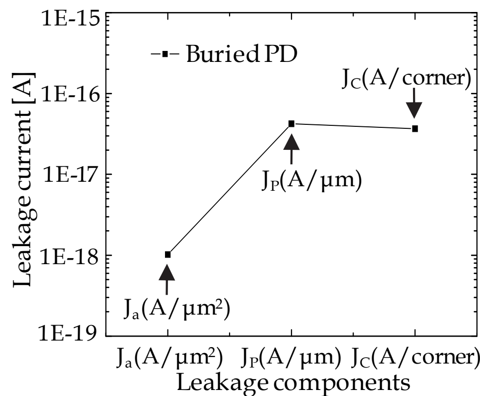


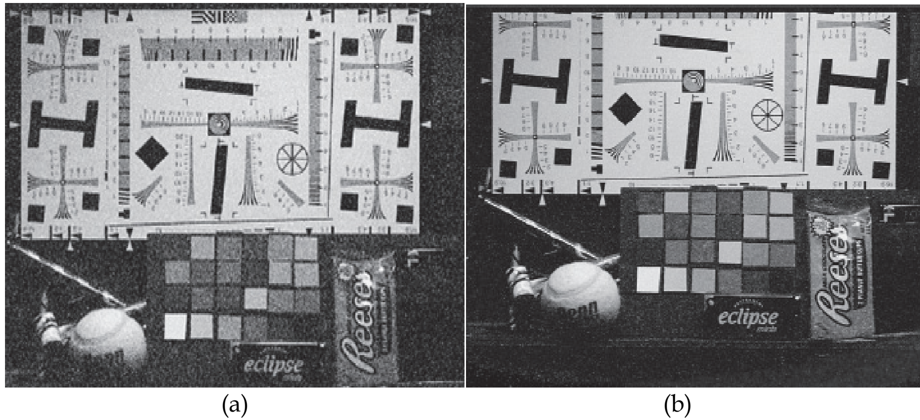
Fig. 8. Extracted leakage components for area (J_a), perimeter (J_p), and corner (J_c) from measurements on buried diodes of different geometry.

The contribution for perimeter current (J_p) and corner current (J_c) are higher than that of the area current (J_a).

3.1.2 Image performance

The impact of dark leakage on image quality is an important part of this experiment. Image quality is evaluated for the different diode configurations.

Figure 9 shows images captured in the normal mode for the surface photodiode and the buried photodiode. The image quality of buried photodiode in Fig. 9(b) is superior to that of the surface photodiode shown in Fig. 9(a). This shows that image quality can be drastically affected by surface induced noise.



(a) (b)

Fig. 9. (a) Image quality for surface PD. (b) Image quality for buried PD. The buried PD results in a superior image quality than the surface PD.

3.2 Floating diffusion

New diode structures to monitor the silicon surface effect and STI edge effect on floating region are evaluated. The characteristics such as dark leakage and activation energy in these diode structures are analyzed. Finally image performance is characterized to understand the mechanism of dark spots.

3.2.1 Floating diffusion leakage characteristics

To determine the leakage caused by surface effects, the activation energy is extracted from measurements of leakage current on floating diffusion of different configuration as a function of reverse bias in the temperature range 25℃ to 200℃. Fig. 10 shows the Arrhenius plot used to extract activation energy under constant reverse bias Vr of 2.0V. E_a is extracted from the slopes in the high and low temperature ranges

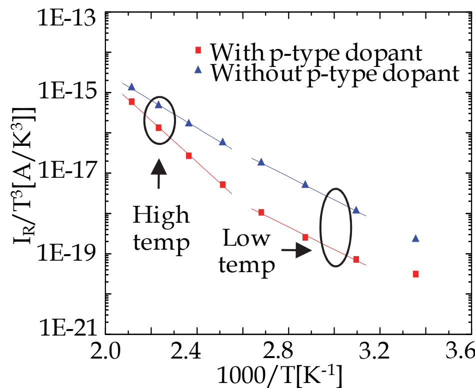


Fig. 10. Arrhenius plot (log I_R/T^3 versus 1/T) as a function of pattern types with and without boron is extracted from the measurements of large area test patterns as a function of temperatures. The diode structure without a p-layer shows higher leakage components than that with a top p-layer.

Type	Area (um^2)		Peri (um)		Corner (ea)		Ea [eV]	
	A_b	%	P_b	%	C_b	%	Low temp	High temp
With p-layer	30000	51	24000	41	4800	8.2	0.56	1.02
Without p-layer	30000	51	24000	41	4800	8.2	0.57	0.68

Table 3. Extracted activation energy from measurements on floating diffusions of identical dimensions with and without a top p-layer

Table 3 compares the activation energy extracted from measurements on floating diffusions of identical geometries with and without a top p-layer.
The activation energy for both floating diffusions is close to half of the energy gap at the low temperature, which suggests that the reverse current is dominated by generation-recombination current. For the floating diffusion with top p-layer, the extracted high-temperature activation energy is about 1.02 eV, which is close to the energy gap. In this temperature range, the reverse current is dominated by diffusion current. The activation energy of the buried diffusion without top p-layer is, however, about 0.68 eV. Thus, it can be assumed that without a top p-layer the surface generation component becomes dominant, exhibiting an activation energy below E_g. The top p-layer was introduced to suppress the surface leakage component in photo diodes (I. Inoue at al., 2003). The above results demonstrate that the top p-layer can also be implemented for floating diffusions to suppress surface leakage.

3.2.2 Pinch off voltage

The gate-drain overlap region of the pass transistor must be optimized to fully transfer the integrated charge from the photodiode to the floating diffusion. To confirm the charge transportation in this overlap region, floating diffusion designed 2-different concepts by changing of doping profile
Figure 11(a) shows the top view for photo diode with transfer gate Tr. and floating diffusion. Floating diffusion separates two diode regions with and without gate overlap. Diode with gate overlap is the type with overlap under sidewall spacer in floating diffusion, and diode without gate overlap is the type with STI effect. In order to evaluate both effects, it is necessary to design new JFET (Junction Field Effect Transistor) structure to monitor pinch off voltage in floating diffusion. Test patterns have been designed to see effect for the overlap under sidewall spacer and the STI edge as shown in Fig. 11(b) and Fig. 11(c). Fig.11(d) shows the concept for test pattern to measure pinch off voltage. Ph under sidewall spacer in Fig. 11(b) was applied to the lower dose than that of center diode region. Pinch off voltage between Fig. 11(b) and Fig. 11(c) will be different due to JFET structure schemes.

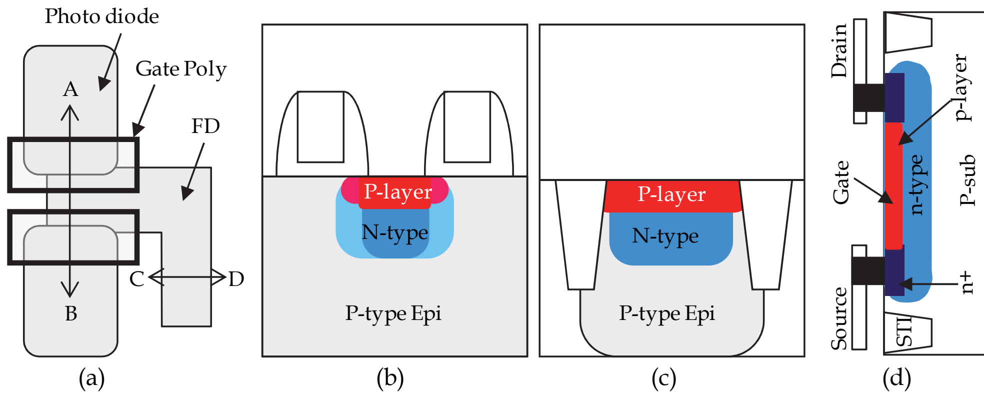


Fig. 11. The region of photo diode, pass transistor, and floating diffusion for four-transistor type active pixel. (a) Layout. (b) Tape-A : Cross sectional view for diode on line AB region (c) Type-B : Cross sectional view for diode on line CD region. (d) Concept for test pattern to measure the pinch off voltage. Source direct frontside and drain direct backside on (b) and (c). Also gate connected with p-sub.

In order to confirm the changing of the doping profile, pinch off voltage is measured in JFET (junction field effect transistor) structure as shown in Fig. 12(a). Pinch off is defined as a voltage when junction is fully depleted by applying reverse voltage at the p-layer and p-sub region. Here pinch off voltage can be defined the ratio of n- and p-type doping concentration. Doping profile with FD regions can be easily analyzed by using the JFET structure. Dependences of process and test pattern are analyzed to understand doping profile by measuring pinch off voltage. Fig. 12(b) shows the pinch off voltage for gate poly-bounded FD (type-A) and STI-bounded FD (type-B) as a function of test patterns for process-1 and -2. Gate poly-bounded FD (Type-A) shows the different pinch off voltage with process-1 and process-2. Clearly, the increase of pinch off voltage in process-2 leads to the changing of doping profile under sidewall spacer in gate poly-bounded FD (type-A), which could explain the electric field improved in the sidewall spacer overlap region. But STI-bounded FD (type-B) shows similar pinch off voltage without processes, which means pn diode controlled to the same dose. For the test pattern types, gate poly-bounded FD (type-A) with sidewall spacer overlap has higher pinch off voltage than that of STI-bounded FD (type-B). This means that pattern geometric can be effected the difference of pinch off voltage during biasing. From these results, it is observed that process-2 has lower n-type doping concentration than process-1 on gate poly-bounded FD (type-A) as expected. This is interpreted as a concept to reduce pinch off voltage by controlling doping profile under sidewall overlap region. Therefore, it is concluded that the controlling of dopant ratio to have same electric field both sidewall overlap and center region is possible to transfer integrated charges without loss in sidewall overlap.

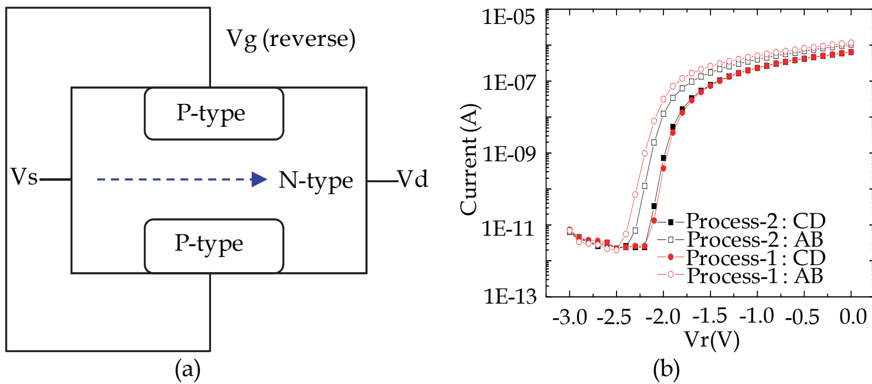


Fig. 12. (a) Concept for JFET to measure pinchoff voltage. (b) Comparison pinch off voltage with process schemes and regions for floating diffusion.

3.2.3 Image performance

From these experiments, we can suspect that a charge pocket in the floating diffusion regions can change the output voltage. To confirm this, image quality has been analyzed by using the camera system. Fig. 13a shows artifact in the pixel area randomly distributed, which means that due to the confined signal charge under sidewall spacer in process-1 described in Fig. 4, the output voltage is low in dark spot regions compared to defect free pixels. Fig. 13b shows good image, without dark spots for process-2.

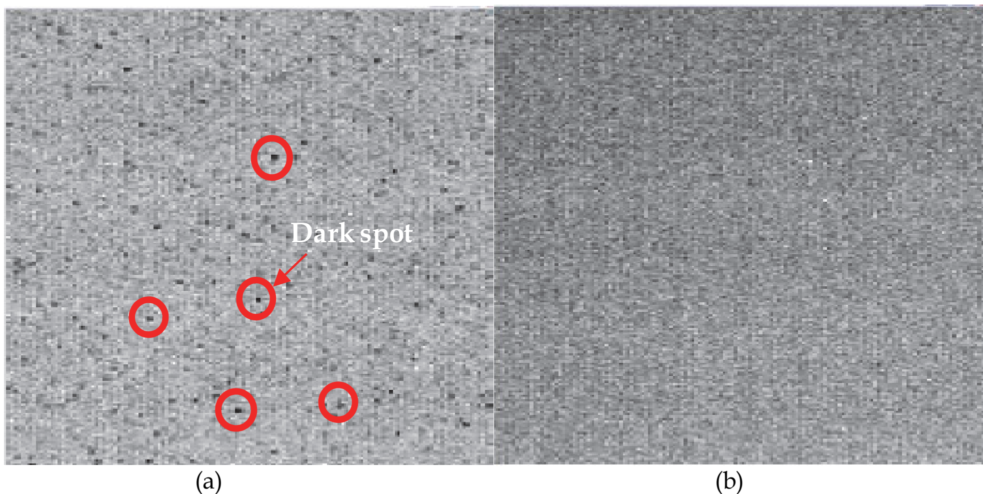


Fig. 13. Image performance analyzed by the camera system. (a) Image for process-1. Dark spot means pixel with low output voltage. (b) Image for process-2. Dark spots were removed.

4. Simulations

Simulations are done to relate electrical test results to photodiode and floating diffusion types. The electric field distribution is compared for photodiodes of different types, and charge pockets in floating diffusions are identified by comparing the potential profiles for different floating diffusion types.

4.1 Photo diode

Electric field both surface and buried PD is compared to find weak point for leakage characteristics. Also doping profile is analyzed at the surface PD to confirm the STI edge effect.

4.1.1 Electric field

Junction leakage current is the highest in the surface diode test structure, even if area is smaller than buried diode. Electric field and doping profile are compared between diode structures by simulation to see the difference. Electric field for surface diode is concentrated at the edge region of junction near the Si surface as shown in Fig. 14(a). Moreover, the junction is located near the Si surface region where the largest electric field is. Fig. 14(b) shows buried diode profile and electric field. As shown in Fig. 14(b), buried diode has two junctions, bottom and top, and has an additional junction at the plug region. Electric field is highest at the corner region of diode and plug. Also buried diode junction has a lower electric field than surface diode. From the simulation result, electric field shows pattern dependence independent of surface effects. It is suspected, from comparison of electric fields, that the higher leakage current at surface diode is induced from edge and surface.

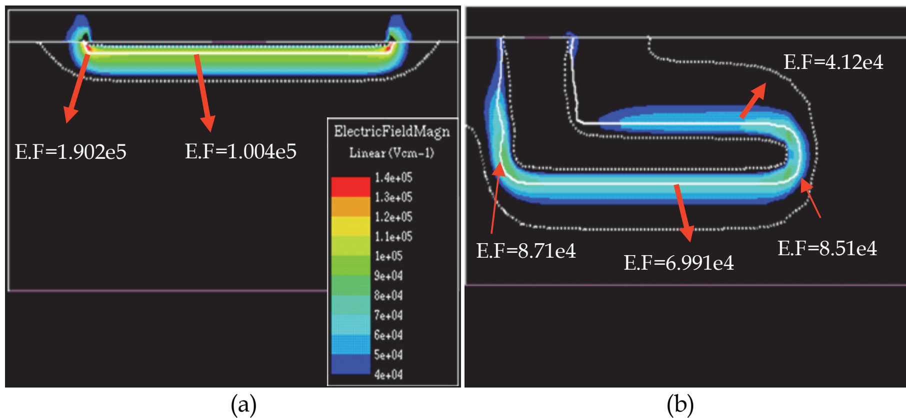


Fig. 14. (a) Electric field contour at V_R = 2.0 V for blue PD. (b) Electric field contour at V_R = 2.0 V for buried PD.

4.1.2 Doping profile for surface PD

In order to evaluate the defective sidewall effect of STI and the interface traps of surface on CIS, an image sensor with 3 Mega pixels is fabricated where edge of surface diode is

controlled to isolate it from the STI using boron implantation. Two-dimensional boron profile simulations with SUPREM-4 are run on an image sensor where boron is implanted to separate the photo diode from the STI edge. The results in Fig. 15 confirm that a p-region of adequate concentration and width is formed, electrically separating the photodiode from the STI boundary

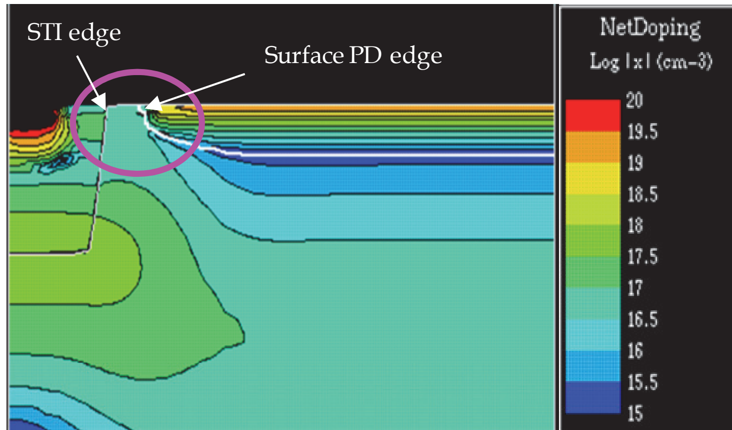


Fig. 15. 2D doping profile for surface PD at STI edge. Boron is applied between edge of STI and Surface PD.

4.2 Floating iffusion

Potential profile and pinch off voltage with FD types are compared to find the location for charge pocket. Also pinch off voltage is analyzed with FD types to confirm the potential profile.

4.2.1 Potential profile

Figures 16a and 16b show the simulated potential profiles for the gate poly-bounded and STI-bounded FD shown, respectively, in Fig. 11b as type-A, and in Fig. 11c as type-B. The structures are fabricated in process-1 (Fig. 4). Simulation is done for an applied reverse bias of 3.3V. The potential profile at gate poly-bounded FD shows the higher than that of STI-bounded FD. Higher potential under the sidewall spacer indicate charge pocket because higher voltage is needed to do pinch off. Thus, the integrated charge can’t be fully transported in without loss under sidewall spacer overlap region.
Figures 17a and 17b show the simulated potential profiles for the gate poly-bounded and STI-bounded FD shown, respectively, in Fig. 11b as type-A, and in Fig. 11c as type-B. The structures are fabricated in process-2 (Fig. 4). Simulation is done for an applied reverse bias of 3.3V. The potential profile under the sidewall spacer does not create a charge pocket. Also, the potential profiles are similar for type A and type B FD, suggesting that the impurity profiles are also similar. Thus, the integrated charge can be transported in both types without loss under sidewall overlap region.

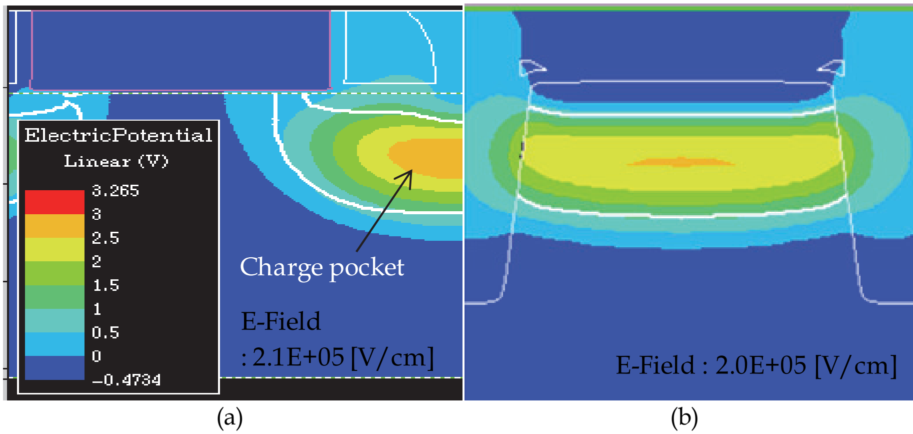


Fig. 16. Simulated potential profile for process-1. (a) Profile for gate poly-bounded FD. (b) Profile for STI-bounded FD. Gate poly-bounded FD shows the highest potential than STI-bounded FD.

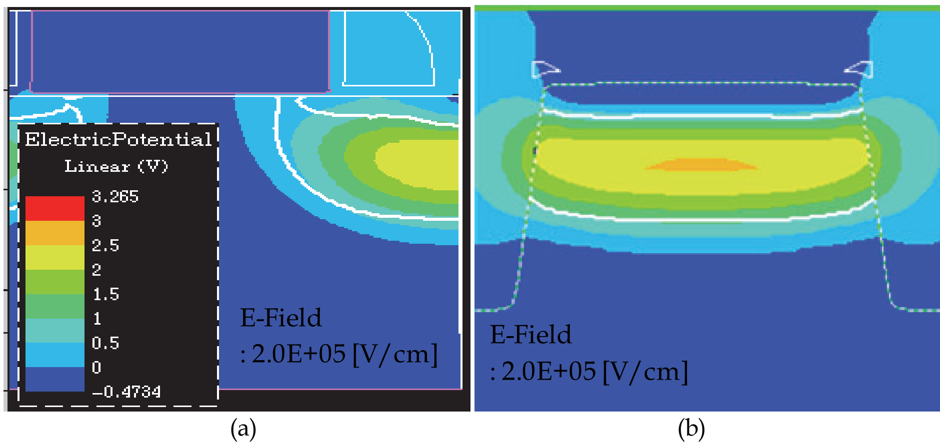


Fig. 17. Simulated potential profile for process-2. (a) Profile for gate poly-bounded FD. Here area for the highest electrical potential decreases than that of process-1. (b) Profile for STI-bounded FD. Here high potential region shows very small area and similar profile with process-1.

4.2.2 Charge pocket and delta vout

I. Inoue at al (2003) explained charge pocket model on local region under sidewall within photo diode and focused image lag in terms of potential barrier and potential pocket in the buried photo diode. Fig. 18(a) shows the current path from photo diode to floating diffusion during signal processing. In the present experiment, the FD is constructed as a buried diode covered with a p-top layer. When a potential pocket is generated in the FD in a local region under sidewall, it can become a source for output voltage variation on the APS. Fig. 18(b)

shows a schematic of the potential distribution causing a charge pocket between the photo diode and the gate of the source follower (SF). To completely transfer the signal electrons from photo diode to the gate of the SF, the potential under sidewall overlap region (AB line) has to be higher than or equal to that in FD center (CD line). Otherwise, a fraction of the signal charge would be confined in the sidewall overlap region and the integrated signal would not be completely transferred through the gate of the SF. The output signal would then be smaller than expected.

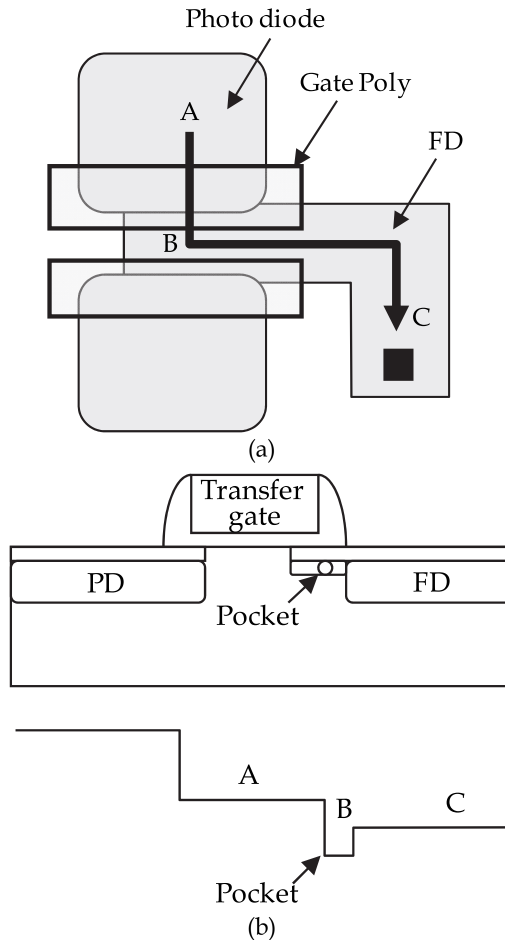


Fig. 18. (a) The potential pocket for the schematic cross section. (b) The diagram for potential distribution.

To establish the relationship between charge pocket and output voltage in the APS, transient simulation is done on the APS circuit as shown in Fig. 19a. Output voltage is calculated from simulation as a function of charge pockets in FD region, whereby the charge pockets is changed intentionally to see the difference in the output voltage. Delta Vout means the

voltage difference between $V_{out\,1}$ and $V_{out\,2}$ during the readout interval in the timing diagram for an APS pixel as shown in Fig. 19b. V_{out} is measured before and after charging electrons transfer from PD to FD region. In the presence of a local charge pocket on the path through FD during pixel operation, delta Vout would be reduced from the expected value.
As shown in Fig. 19c, delta V_{out} decreases as the amount of pocket charge in FD increases. Pocket-free FD structure can be achieved by controlling the dopant ratio between n- and p-type to the same value throughout the FD regions.

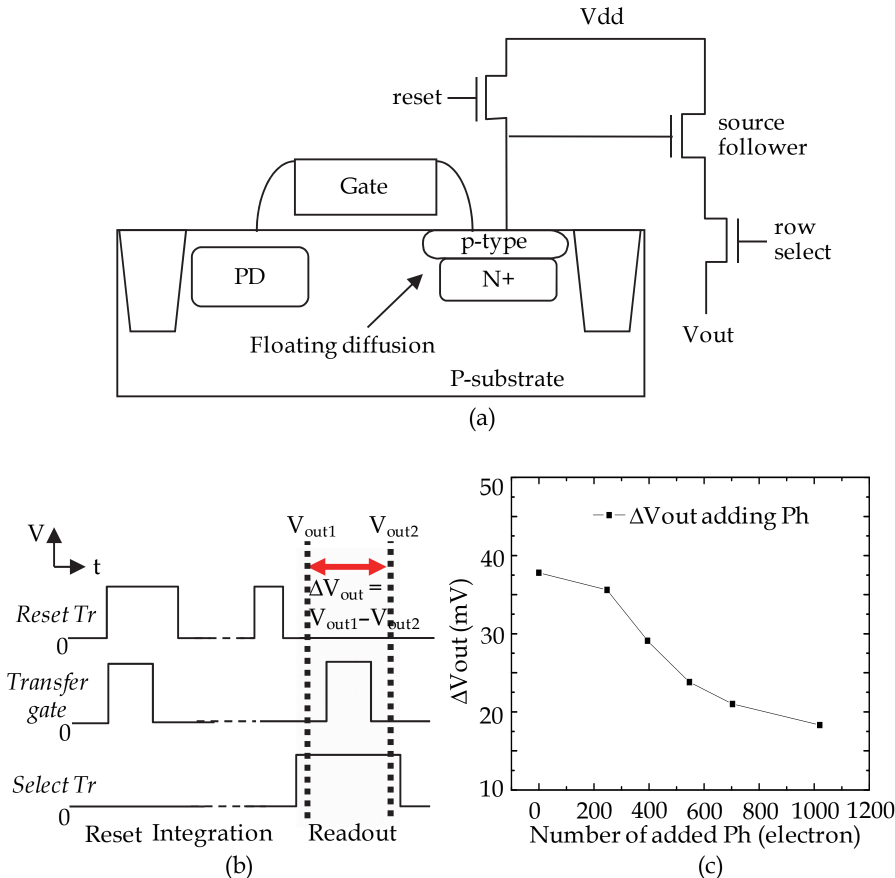


Fig. 19. (a) Circuit for four-transistor type active pixel. (b) Timing diagram for APS circuit. (c) ΔVout as a function of charge pockets.

5. Discussion

The floating diffusion is designed to transfer the integrated charge from the photo diode to the source follower without time delay in the active pixel sensor (APS). The floating diffusion is shared with the drain of the transfer transistor and reset transistor, and the gate of the source follower in APS which consists of photodiode, reset transistor, transfer gate,

source-follower transistor, and select transistor. Photodiode and floating diffusion are depleted during the reset period by turning on reset the transistor and transfer gate. During integration time, electron charge generated by an incident optical signal is integrated in the photodiode. After integration, the floating diffusion is reset at a reference voltage (V_{out1}) by turning on the reset transistor only. The reference voltage is sampled in the readout period between turning on the select transistor and turning on the transfer gate. Charge in photodiode is transferred into floating diffusion by turning on the transfer gate and converting into voltage signal. A voltage V_{out2} is sensed on the floating diffusion after turning off the transfer gate. The optical signal is interpreted as the voltage difference between the reference voltage V_{out1} and the sensing voltage V_{out2}. Conversion gain of charge to voltage depends on the capacitance of photodiode and that of floating drain node. Dark current in the read-out process influences image parameters such as dark signal, conversion gain, noise, and signal to noise ratio (SNR). On the other hand, fill factor, i.e., the ratio of light-sensitive (photodiode) area to pixel total size, decreases as the shrink of pixel pitch shrinks. This reduces the sensitivity and SNR due to the reduction in photodiode size. To improve the fill factor, 2 or 4 shared pixel architectures, sharing both the floating diffusion and the source follower transistor are needed (J. Bogaerts, 2006 and Young Chan Kim, 2009). Therefore, a larger floating diffusion area is needed, however, at the cost of increasing the floating diffusion capacitance and hence decreasing the conversion gain. To reduce the capacitance of the floating diffusion, a buried floating diffusion should be implemented. The control of capacitance and potential profile in a buried floating diffusion is therefore very important.

6. Summary and conclusions

The leakage current and activation energy are compared for diodes of different configurations, using a standard n+/pwell diode as a reference. The temperature dependence of leakage yields an activation energy which depends on area, perimeter and number of corners for the buried photodiode (PD) with a top p-layer. For the first time, leakage characteristics are analyzed for a buried PD, taking into account area, perimeter, and corner effects.

In addition, leakage current and activation energy are analyzed for a buried floating diffusion (FD) with and without a top p-layer using a new diode structure. It is confirmed that the dark current can be reduced by implementing a buried floating diffusion rather than a surface FD. **A** charge pocket under the sidewall spacer can change the output voltage and cause a dark spot on the image. This is predicted by TSUPREM 4 simulation.

It is shown that the charge pocket can be generated by a higher doping concentration under the sidewall at the drain-side of the transfer gate, including FD region. This charge pocket is an image artifact that causes the output voltage to drop.

In summary, the mechanism for dark spot**s** has explained by investigating pinch off voltage and potential profile on buried FD. Dark spot can be controlled by removing charge pocket under the sidewall spacer in the buried FD. The buried FD is a good candidate to control the capacitance and reduce dark leakage in future designs.

7. Acknowledgement

This work was performed during the development of a pixel-project with Foveon.

8. References

Dun-Nian Yaung, Shou-Gwo Wuu, Yean-Kuen Fang, Chung S. Wang, Chien-Hsien Tseng, and Mon-Song Liang. (2001). "Nonsilicide Source/Drain Pixel for 0.25-um CMOS Image Sensor," *IEEE Electron Device Lett.*, vol. 22, pp. 538–540.

David X. D. Yang, Abbas El Gamal, Boyd Fowler, and Hui Tia. (1999). "A 640 512 CMOS Image Sensor with Ultrawide Dynamic Range Floating-Point Pixel-Level ADC," *IEEE J. Solid-State Circuits*, vol. 34, NO.12, pp. 1821–1834.

H.-D.Lee, S-G Lee, S-H Lee, Y-G Lee, and J-M Hwang. (1998). Characterization of Corner-Induced Leakage Current of a Shallow Silicided n+/p Junction for Quarter-Micron MOSFETs, *Jpn. J. Appl.* Vol. 37, pp.1179-1183.

Hyuck In Kwon, In Man Kang, Byung-Gook Park, Jong Duk Lee, and Sang Sik Park. (2004). "The Analysis of Dark Signals in the CMOS APS Imagers From the Characterization of Test Structures," *IEEE Trans, Electron Devices*, vol. 51, NO. 2, pp. 178-184.

H.-S. Philip Wong, T. Chang, E. Crabb´e, and P. D. Agnello. (1998). CMOS Active Pixel Image Sensors Fabricated Using a 1.8-V, 0.25- m CMOS Technology, *IEEE Trans. Electron Devices*, vol. 45, pp.889–894.

H.-Y. Cheng and Y.-C. King. (2003). "A CMOS Image Sensor With Dark-Current Cancellation and Dynamic Sensitivity Operations," IEEE Trans. Electron Devices, vol. 50, pp.91–95.

H.-Y. Cheng and Y.-C. King. (2002). An ultra-low dark current CMOS image sensor cell using n ring reset, *IEEE Electron Device Lett.*, vol. 23, pp. 538–540.

I. Inoue, H. Ihara, H. Yamashita, T. Yamaguchi, H. Nazaki, and R. Miyagawa. (1999). Low dark current pinned photo diode for CMOS image sensor, in *Proc. IEEE Workshop on CCD's and AIS*, pp. 25–28.

Ikuko Inoue, Nagataka Tanaka, Hirofumi Yamashita, Tetsuya Yamaguchi, Hiroaki Ishiwata, and Hisanori Ihara. (2003), Low-Leakage-Current and Low-Operating-Voltage Buried Photodiode for a CMOS Imager, *IEEE Trans, Electron Devices*, vol. 50, NO. 1, pp. 43-47.

J. Bogaerts, G. Meynants, G. Lepage, G. Vanhorebeek, B. Ceulemans, and K. Ruythooren. (2009), CMOS image sensor with two shared pixel and staggered readout architecture, *Int. Image Sensor Workshop*.

J. P. Albert Theuwissen. (2001). CCD or CMOS image sensors for consumer digital still photography, presented at the 2001 *Int. Symp. on VLSI Technology, Systems, and Applications*, Hsinchu, Taiwan, R.O.C., pp. 168–171.

K. A. Parulski. (1985). Color filters and processing alternatives for one-chip cameras, *IEEE Trans. Electron Devices*, vol. ED-32, pp. 1381–1389.

Keiji Mabuchi, Nobuo Nakamura, Eiichi Funatsu, Takashi Abe, Tomoyuki Umeda, Tetsuro Hoshino, Ryoji Suzuki, and Hirofumi Sumi. (2004). CMOS Image Sensors Comprised of Floating Diffusion Driving Pixels With Buried Phtodiode, *IEEE journal of solid-state circuits*, Vol. 39, pp. 2408-2416.

Keiji Mabuchi, Nobuo Nakamura, Eiichi Funatsu, Takashi Abe, Tomoyuki Umeda, Tetsuro Hoshino, Ryoji Suzuki, and Hirofumi Sumi. (2004). CMOS Image Sensors Comprised of Floating diffusion Driving Pixels With Buried Photodiode, *IEEE J. Solid-State Circuits*, vol. 39, NO.12, pp. 2008–2416.

N. V. Loukianova, H. O. Folkerts, J. P. V. Maas, D. W. E. Verbugt, A. J. Mierop, W. Hoekstra, E. Roks, and A. J. P. Theuwissen.(2003). Leakage current modeling of test structures for characterization of dark current in CMOS image sensors, *IEEE Trans. Electron Devices,* vol. 50, pp.77–83.

Natalia V. Loukianova, Hein Otto Folkerts, Joris P. V. Maas, Daniël W. E. Verbugt, Adri J. Mierop, Willem Hoekstra, Edwin Roks, and Albert J. P. Theuwissen. (2003). Leakage Current Modeling of Test Structures for Characterization of Dark Current in CMOS Image Sensors, *IEEE Trans, Electron Devices*, vol. 50, NO. 1, pp. 77-83.

Richard Merrill, Shri Ramaswami, & Glenn Keller. "CMOS pixel sensor with depleted photocollectors and a depleted common node", US patent 7834411.

Takashi Watanabe, Jong-Ho Park, Satoshi Aoyama, Keigo Isobe, and Shoji Kawahito. (2010). Effects of Negative-Bias Operation and Optical Stress on Dark Current in CMOS Image Sensors, *IEEE Trans, Electron Devices*, vol. 57, NO. 7, pp. 77-83.

Young Chan Kim, et al.(2006), "1/2-inch 7.2Mpixel CMOS Image Sensor with 2.25um Pixels using 4-Shared Pixel Structure for Pixel-Level Summation", *ISSCC Dig*. Tech. Papers.

Active Pixel Sensor CMOS Operating Multi - Sampled in Time Domain

Fernando De Souza Campos
São Paulo State University
Brazil

1. Introduction

CMOS image systems have receiving great attention from industry and academy due to the growing demand for compact and low power image systems. Compared to charge-couple device CCD, CMOS image sensors presents as advantage higher integration capability. In general, CCD achieve better performance due to its particular fabrication process, however, they require high operation voltage and cannot be easy integrated with CMOS circuits that compound cameras. In last decades, the CMOS imager sensor technology has been improving and they are being used in several applications as multimedia and biomedicine (Fossum, 1997, Hosticka, 2003, Sandage, 1995).
Dynamic range is one of the most important merit figure of image sensors. It is defined as the ratio between the maximum and minimum signal acquired. External scenes present dynamic range higher than 100dB but conventional CMOS image sensors and CCDs shows dynamic range about 60dB. Therefore, they are not able to capture properly external images. However, several researchers proposed different CMOS image sensors architectures with high dynamic range (>80 dB) (Stoppa, 2002, Trepanier, 2002, Yadid-Pecht, 2003, Yang, 2002, Yasuda, 2003, Saffih, 2007).
An attractive high dynamic range architecture approach is the digital pixel sensor (DPS) (Kleinfielder, 2001). This architecture is composed by a ramp digital converter and an 8 bit memory integrated per pixel. The main advantage of this approach is the high frame rate operation however, it presents as disadvantage low fill factor. Different architectures based on DPS were proposed (Doge, 2002, Kitchen, 2004, Qi, 2004). Time-domain DPS were proposed in (Bermak, 2006) and (Chen, 2006). They are characterized by the measurement of fall time of photodiode´s voltage. In general, time-domain DPS architectures integrate a comparator and a 8 bit counter in each pixel. The main disadvantage of this approach is the low fill-factor due to the great number of transistors integrated per pixel. A time-domain imager with only 10 transistors per pixel was proposed in (Lai, 2006). A pipeline operation is proposed in order to achieve high dynamic range. However this approach requires the use of two ramp, one at beginning and other at the end of the integration time reducing the sensitivity at middle range of illumination.
A sampled time-domain CMOS imager was proposed in (Campos, 2008). This pixel architecture is composed by a clocked comparator and a dynamic D flip-flop integrated per pixel. The number of transistor integrated per pixel is still significantly and the fill-factor is low. However the sampling in time-domain concept proposed suggest that the comparison

can be performed outside pixel lowering at maximum the number of transistors integrated per pixel. In this chapter the multisampling time-domain CMOS imager is described.

2. Principle of operation

Most of CMOS imagers operate in voltage domain. However, one disadvantage of this operation mode is that the dynamic range small (<60dB). In this operation mode, the photodiode is charge reversely to V_{dd} voltage by controlling a reset transistor as shown in Fig. . This operation is called reset. After the reset time, the reset transistor "open" and the photodiode, in high impedance, starts to discharged. The period of discharge is called integration time. After certain integration time the photodiode´s voltage is sampled and digitalized.

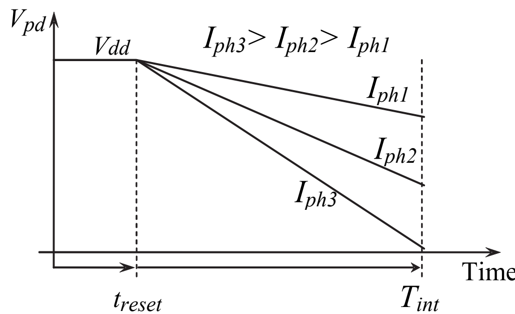


Fig. 1. Photodiode operating in integration mode

During the integration time, the photodiode´s voltage is usually linearized and given by

$$V_{pd} = V_{reset} - S \cdot L_I \cdot t \tag{1}$$

where V_{pd} is the photodiode voltage, V_{reset} is the initial reverse voltage, S is the sensitivity, L_I is the light intensity and t is time.

Time-domain CMOS imagers were proposed as a technique to obtain a CMOS imager with high dynamic range. Time-domain operation is based on fall time of photodiode voltage. The photodiode voltage is compared to a reference constant voltage and the fall time is measured from integration time beginning to the instant of comparison as shown in Fig. 2a. Each fall time is related to a different light intensity.

Fig. 2b shows typical pixel architecture of time-domain imagers. A comparator and a counter are integrated per pixel. The counter starts the count in the beginning of integration time. The comparator output signal goes high in the comparison instant stopping the counting. Assuming that the photodiode voltage decrease linearly, the instant of comparison or comparison time is given by

$$t_d = \frac{V_{reset} - V_{ref}}{S \cdot L_I} \tag{2}$$

where V_{reset} is the initial photodiode voltage, V_{ref} is the reference constant voltage, S is the sensitivity and L_I is the light intensity.

In time-domain the dynamic range is given by

$$. DR = 20log\left(\frac{L_{Imax}}{L_{Imin}}\right) = 20log\left(\frac{t_{dmax}}{t_{dmin}}\right) \tag{3}$$

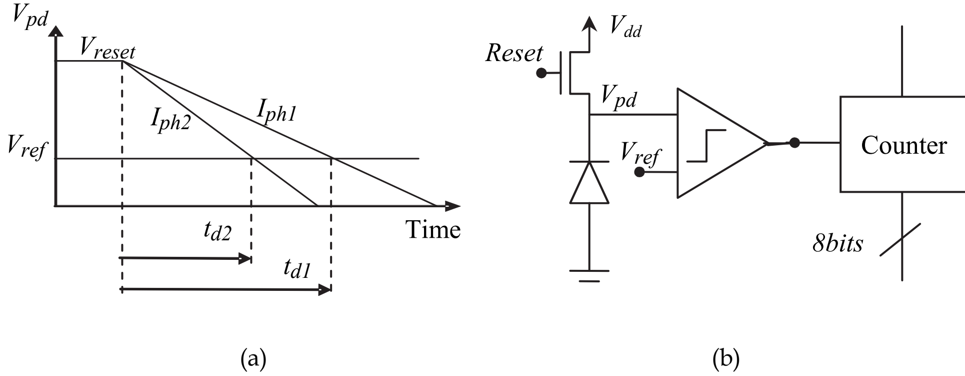


Fig. 2. Time-domain imager (a) main signals and (b) typical pixel architecture

According equation (3), operation with high dynamic range (>100dB) requires long maximum integration time (t_{dmax}) leading to low frame rates. In order to reduce the maximum time it has been proposed to vary the reference voltage, usually as a ramp voltage (Fig. 3). As one can see in Fig. 3 the comparison time is reduced when the reference voltage is varied as a ramp.

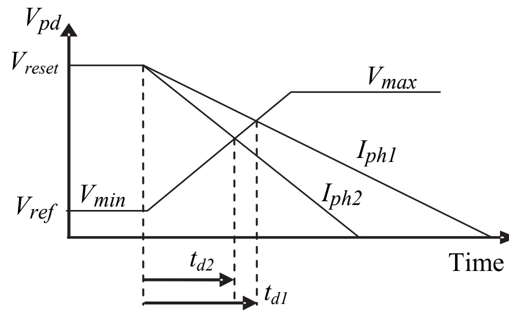


Fig. 3. Time-domain imager main signals using ramp reference

For ramp voltage reference the discharge time is given by

$$t_d = \frac{\left(V_{reset} - V_{min}\right) \cdot T_{int}}{\left(V_{max} - V_{min}\right) + S \cdot L_I \cdot T_{int}} \tag{4}$$

Fig. 4 shows the transfer curve td versus L_I for constant reference voltage and ramp reference voltage. It was assumed V_{reset}=3.3, V_{ref}=1.5, S=3.6·10^{+6}, V_{max}=3V,V_{min}=0.3V and T_{int}=1s. One can see that the ramp reference voltage reduce the time discharge at lower light intensities. However, the compression becomes higher at low light intensities making difficult to discrete.

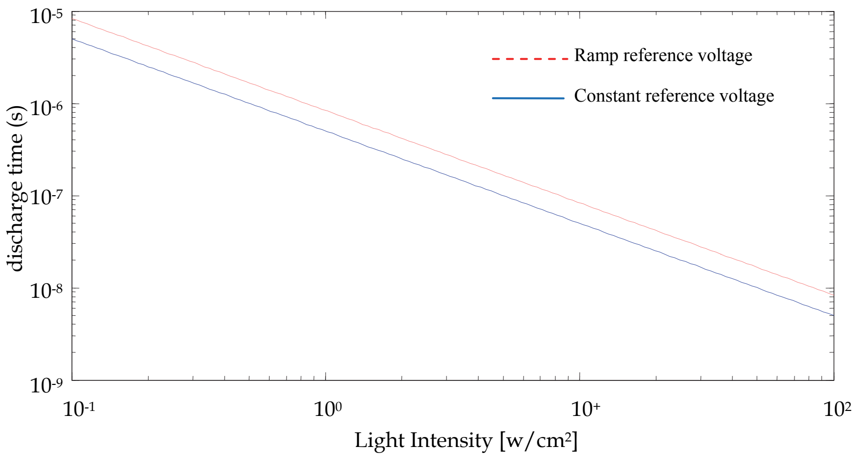


Fig. 4. Discharge time characteristic

Multisampled time-domain CMOS imagers are time-domain imagers in which the comparison result is sampled. The comparison result is sampled where each instant of sample is coded. The first time in which the sample indicated that a comparison occurs determines the code related to that comparison time or light intensity incident. Fig. 5(a) shows the main signals for regular interval sampling time T_i. For a given integration time T_{int} the regular interval sampling time is given by $\Delta T_s = T_{int}/2^N$ where N is the total number of bits.

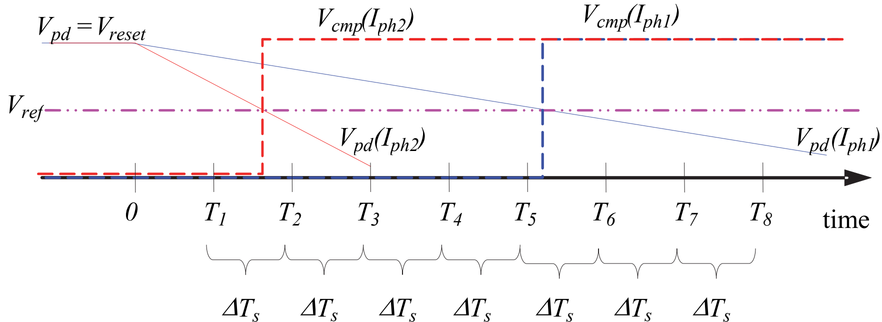


Fig. 5. Sampling in regular interval time.

In this case the dynamic range is given by

$$DR = 20log\left(2^N\right) \tag{5}$$

In this case the dynamic range for N=8 is 48.16dB, for N=10 is 60 dB and for N=12 is 72dB. As one can see, in order to achieve high dynamic range (100dB) is needed more than 16 bits. Therefore, the high number of the bits required by this approach is a disadvantage. However, the sampling can be non linear as shown in Fig. 6. The interval time can be varied logarithmic (Fig. 6a) or a combination of linear and log (Fig. 6b). The number of bits required can be reduced applying non-linear sampling intervals.

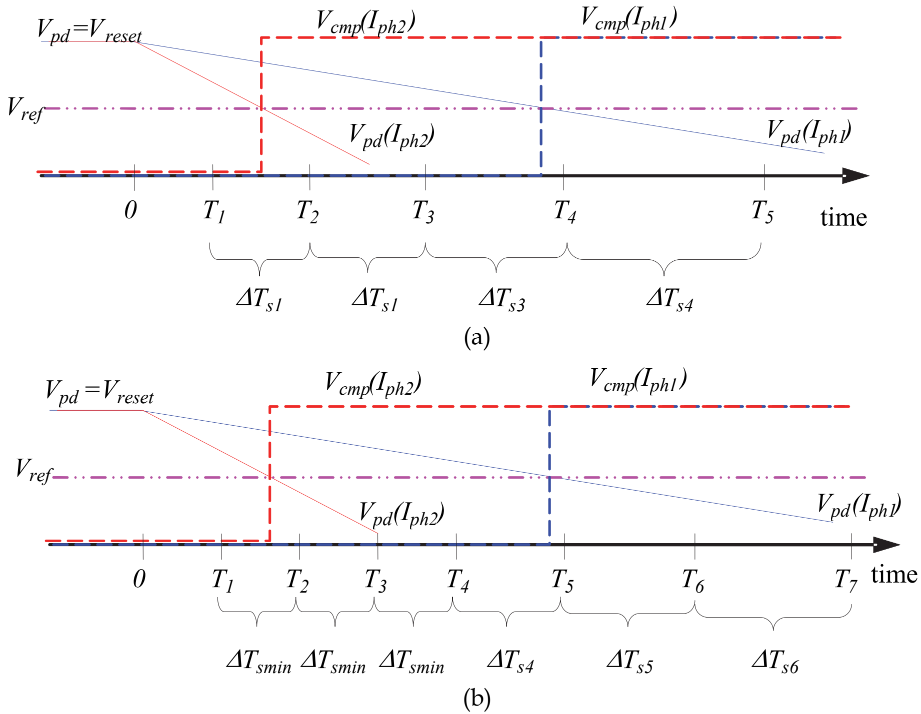


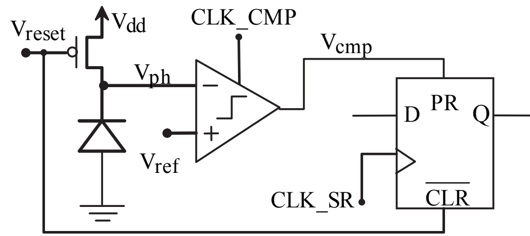
Fig. 6. Sampling in (a) logarithmic interval times (b) linear-logarithmic interval times

3. Pixel architecture

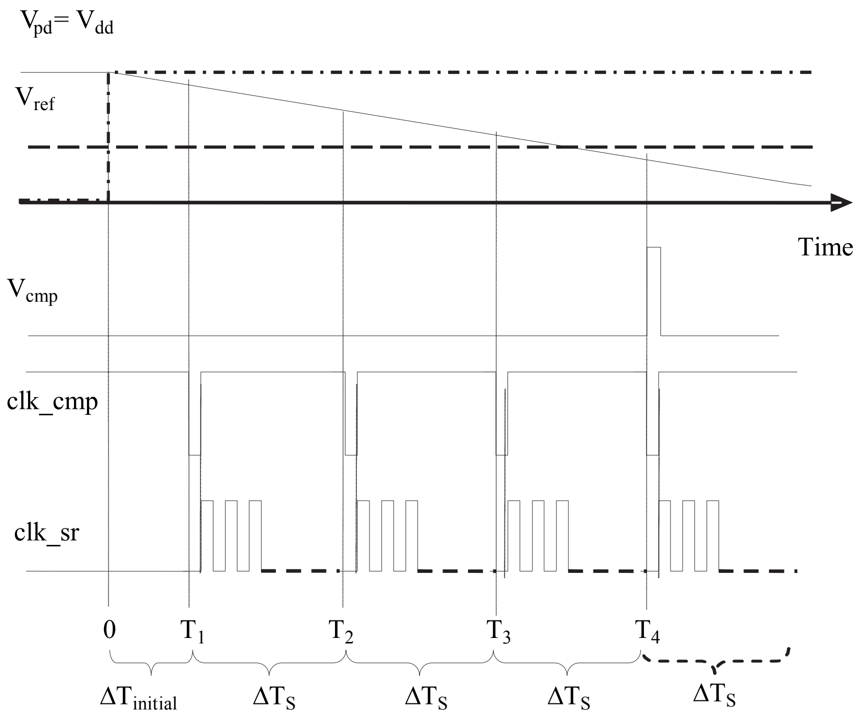
Fig. 7a shows the block diagram of the pixel architecture proposed in (Campos 2008). The pixel is composed of a photodiode, the reset transistor, a comparator type of clocked flip-flop and a D-type with asynchronous inputs (PR and CLR). The clocked comparator type offers operating speed and timing between the pixels of the array. The D-type flip-flop are used to store the comparison result and they are connected together forming a shift register for serial line per row. The serial shift register allows reading outside of the comparison result by means of shifting the data using the clock signal (clk_sr) of the shift register. Fig. 7b shows the timing diagram of the main control signals of the system. The reset signal is responsible for activating the reset transistor and loading the initial photodiode voltage V_{reset} and simultaneously activates the CLR input initializing the flip-flop with Q = 0. The clock signal (clk_cmp) of the comparator determines the moment of performing the comparison, while the sampling interval (T_s). Note that the comparison occurs simultaneously in all pixels in the array.

In the sampling instants (T_s) in which $V_{fd} < V_{ref}$, the output signal of the comparator (V_{cmp}) normally remains low and the value in the flip-flop remains at 0 because in this case PR = 0. In the sampling instants in which $V_{fd} < V_{ref}$, the output signal of the comparator (V_{cmp}) determines PR = 1 and hence the state in the flip-flop to go high (Q = 1). Thus, as a result of the comparison, the state Q = 0 indicates $V_{fd} \geq V_{ref}$ while the state Q = 1 indicates $V_{fd} \leq V_{ref}$.

The D flip-flops of the pixels on each row are connected together to form a shift register on line. The shift register for serial line allows you to read outside of the comparison results by means of shifting the data using the clock signal (clk_sr) of the shift register.



(a)



(b)

Fig. 7. Architecture´s (a) block diagram and (b) control signals

Fig. 8 shows the block diagram of the system architecture proposed for an imaging system in which the sampling results and subsequently its encoding is performed externally. The data from the shift registers are provided in the input multiplexer whose number of output bits is PO.

The discharge time of the voltage signal from the photodetector is associated with the first moment of the comparison sample (T_S) at which the comparison result is positive (Q = 1). Offering only a memory 1bit per pixel, the array reading should be performed after each sample without exceeding the time interval between samples (ΔT_S). Whereas the operating frequency of the multiplexer ($1/T_{mux}$) is greater than the operating frequency of shift registers, the total time for reading the array (T_m) can be defining as:

$$T_m = T_{clk_cmp} + \left(\frac{n_L \cdot n_C}{PO}\right) \cdot T_{mux} \quad (6)$$

where $n_L x n_C$ is the dimension of the array of pixels, PO is the number of bits in the output of the multiplexer, T_{clk_cmp} is the time to compare and t_{mux} is the period of operation of the multiplexer.

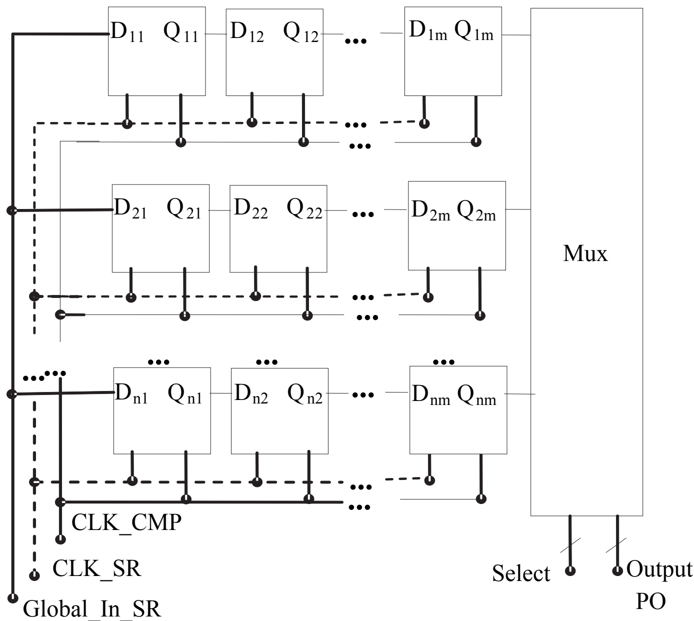


Fig. 8. Array block diagram

The total read time of the matrix (T_m) must be less than the intervals between samples (ΔT_S). The dimensions of the matrix and the operation speed of the circuit determine the array time reading. Thus, the limit of operation of the system depends on the size of the array, the speed of circuit operation, the number of bits that determines the time interval between samples (number of bits of the image) and the number of bits read in parallel at the output multiplexer. Assuming that sampling is performed at regular intervals, the interval between samples will be $T_{int}/2^N$ where T_{int} is the maximum integration time and N is the number of

bits of the image. Fig. 9 shows the total time given by equation (6) depending on the size of a square matrix DXD for different number of output PO, considering typical times of 100ns to 5ns and operating times of the comparator (T_{clk_cmp}) and multiplexer (t_{mux}) respectively. Fig. 9 also shows the intervals between sampling (ΔT_S) for 8-bit image and 10bits. It can be seen in Fig. 9 the maximum dimensions of the matrix in which to operate within the limit $T_m < \Delta T_S$.

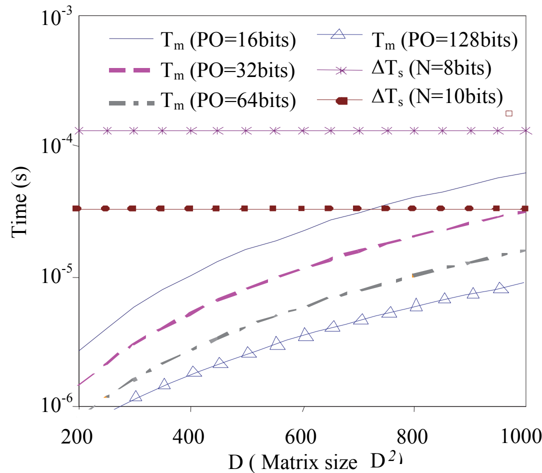


Fig. 9. Time Reading

Assuming operation rate of 30 frames/second or a total time of integration (T_{int}) of 33.33 ms, the interval between samples of linear $\Delta Ts = T_{int}/2^N$ is approximately 130 μ s for 8 bit images and 32 μ s for 10bits images. Note that once defined the technology and dimensions of the array of pixels, the choice of the number of output bits (DB) has a fundamental role in limiting the maximum frame rate. The CMOS imaging system proposed could be constructed with VGA resolution (580x640) capturing images of 10 bits considering output of 16bit. However in order to achieve dynamic range higher than 90dB it is need operates with images of 16bits or higher.

4. Fill-factor

Fig. 10 shows the comparator topology. The comparator is a typical clocked comparator composed by a low gain first stage followed by a decision circuit as a second stage. The clocked comparator make possible to sample the comparison result at sampling instant.
There are two stages of operation, the track and latch phases. In track stage, the switch M13 is closed and the output has low gain. The role circuit operates as a differential gain stage. Fig. 11 shows simulations results of the output voltage (drain of M09 and M10) as a differential voltage is applied to input. As one can see, the resolution is about 2mV and the maximum differential output voltage is about 0.6V. Simulations were performed using model of 0.35 um technology. The transistors sizes are show in table 1.

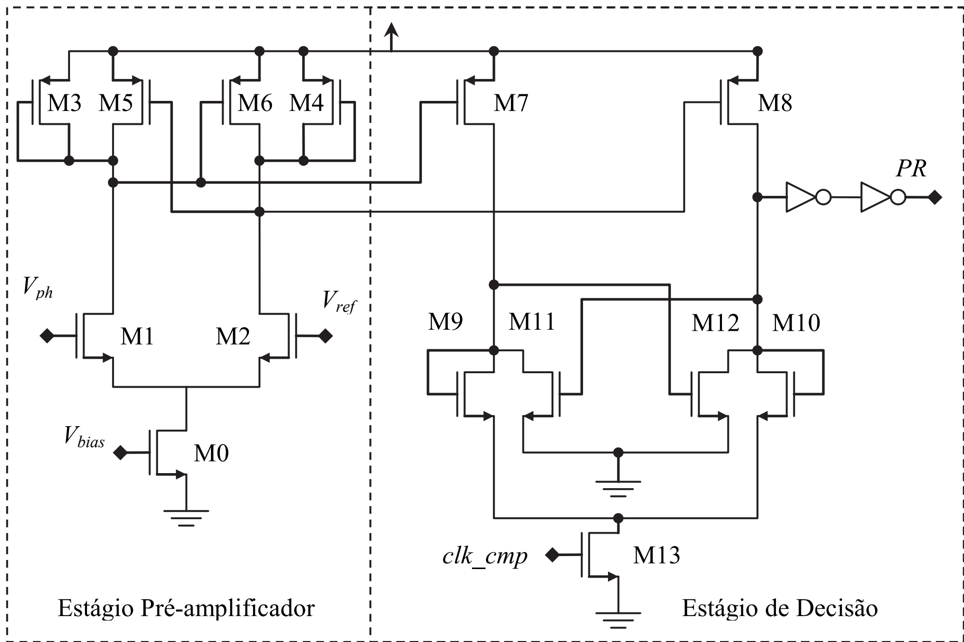


Fig. 10. Topology of clocked comparator

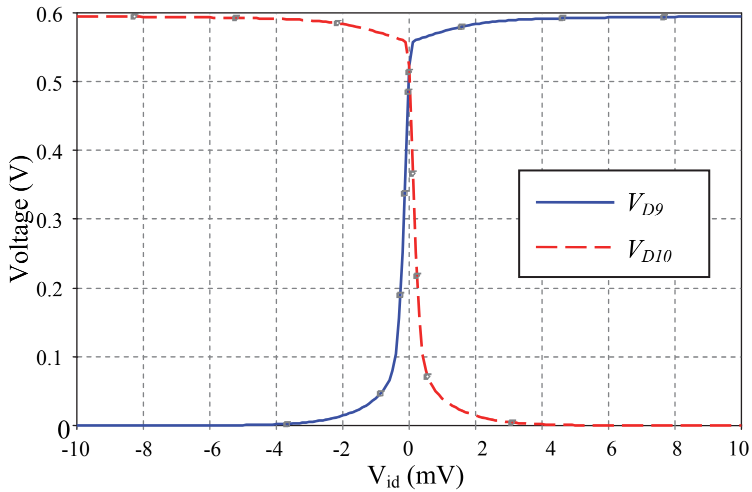


Fig. 11. Output voltage of comparator in track stage.

M0	2µm/2µm
M1= M2	2µm/1µm
M3=M4=M5=M6=M7=M8	0,8µm/0,4µm
M9=M10=M11=M12=M13	0,4µm/0,35µm

Table 1. Transistors sizes used in simulations.

After the track stage, the M13 operating as switch is opened beginning the latch stage. In latch stage M9 and M10 are disconnected and M7, M8, M11 and M12 compose a latch circuit. The initial differential voltage (0,6V on Fig. 11) is then amplified by latch circuit to logic voltage levels. Two inverters were used at output to ensure logic levels even during track stage.

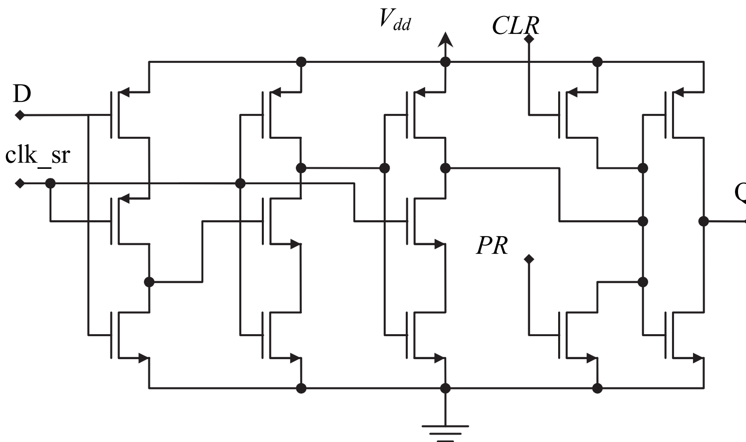


Fig. 12. Topology of dynamic D flip-flop with preset and clear

Fig. 12 shows the D flip-flop topology. The D flip-flop topology is a dynamic flip-flop with preset and clear. The dynamic topology was chosen instead the static topology in order to reduce the total number of transistor required. Simulations results in Fig. 13 show the main propagation delays using load of 50fF. All transistor of D flip-flop are minimum sizes.
As one can see there are 22 transistors integrated per pixel (Figs. 7(a), 10 and 12). The prototype presented in (Campos, 2008) shows only 16% of fill-factor. However, it is possible to implement other pixel architecture using the multisampling in time-domain technique in order to reduce the number of transistor integrated per pixel and to increase the fill-factor.
The pixel architecture showed in Fig. 14 is proposed here as an alternative to pixel architecture showed previously proposed in (Campos, 2008). The pixel is the typical 3T architecture. It is composed by a buffer (source follower) between the photodetector and the transistor select. The transistor select isolates the pixel from bus column and make possible to select the pixel line that must be read outside matrix. Also, there is a reset transistor. The comparator is integrated per column outside pixel. In this case the fill-factor becomes maximized.

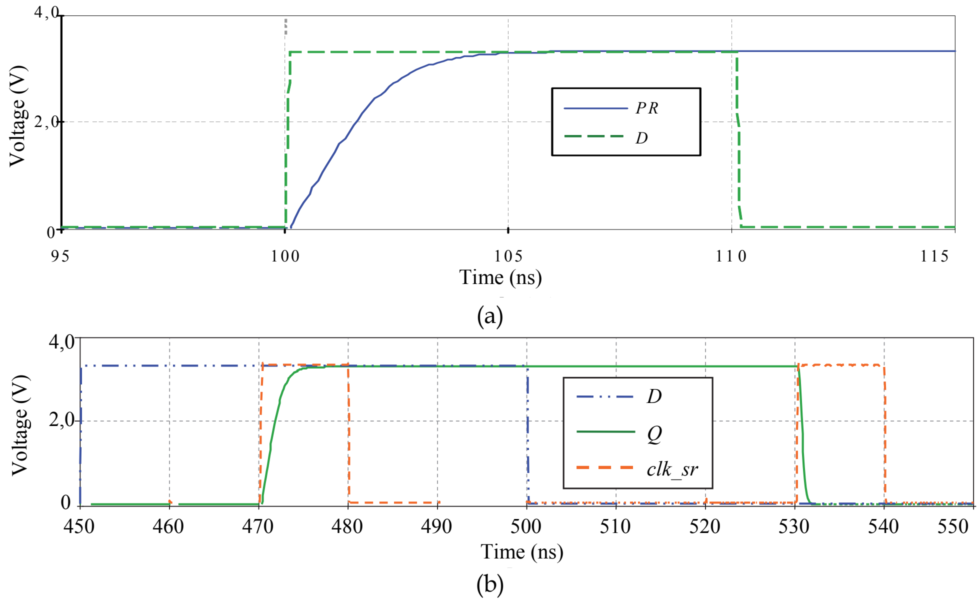


Fig. 13. D flip-flop propagation delays with 50fF load (a) related to the preset (PR) and (b) related to the input (D)

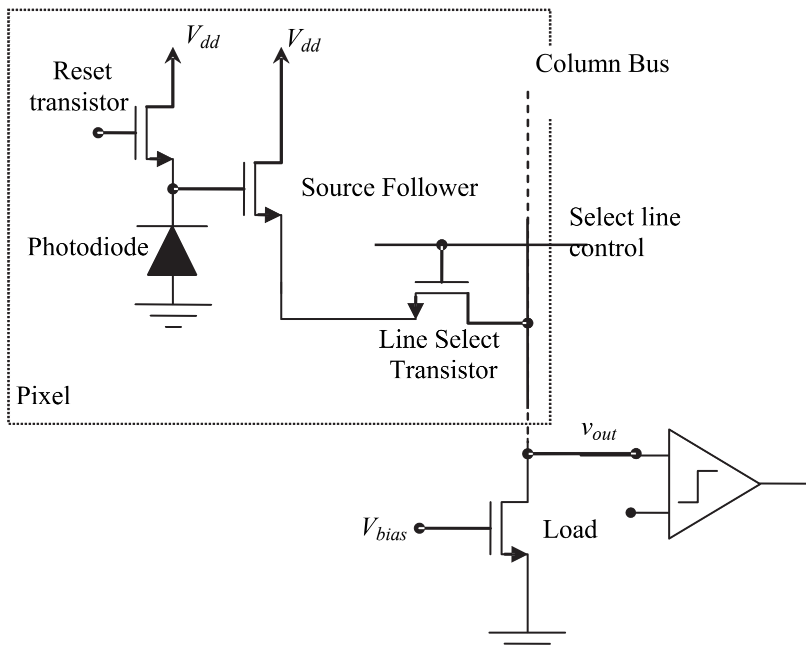


Fig. 14. A different architecture with high fill-factor for multisampling in time technique.

5. Fixed-pattern noise

The fixed pattern noise (FPN) is defined as the non-uniformity resulting from the image signal variation from pixel to pixel when a beam of light intensity is applied uniformly. The FPN in CMOS imaging systems is a major disadvantage compared to CCDs. Two major sources contribute to FPN in CMOS imaging systems in time domain: (i) the variation of initial voltage (V_{reset}) and (ii) the variation of the offset voltage of the comparator. Pixels in which the reset transistor is NMOS type, the maximum initial voltage is V_{dd}-V_{thn}. Different values of the threshold voltage cause different initial voltage (reset voltage) from pixel to pixel introducing non-uniformity at image. The conventional method to eliminate non-uniformity due to the reset transistor is using as reset transistor PMOS type MOSFET which provides maximum initial voltage V_{dd} regardless of the value of the threshold voltage.
The offset voltage of the comparator is modeled as a voltage source in series with inputs of the comparator. Considering the offset voltage as a voltage source in series with the reference voltage, the equations of discharge time (2) and (4) can be rewritten as

$$t_{dct} = \frac{(V_{dd} - V_{refct} \pm V_{offset})}{S \cdot I_L} \qquad (7)$$

$$t_{drp} = \frac{(V_{máx} - V_{mín} \pm V_{offset}) \cdot T_{\text{int}}}{(V_{máx} - V_{mín}) + S \cdot I_L \cdot T_{\text{int}}} \qquad (8)$$

where t_{dct} and t_{drp} are the time of discharge using a constant reference voltage and a ramp reference voltage respectively. Manipulating equations (7) and (8), we obtain the absolute error in the time of discharge using a constant reference voltage and a ramp respectively as:

$$Absolute\ error = \pm \frac{V_{offset}}{S \cdot I_L} \qquad (9)$$

$$Absolute\ error = \pm \frac{V_{offset} \cdot T_{\text{int}}}{(V_{\max} - V_{\min}) + S \cdot I_L \cdot T_{\text{int}}} \qquad (10)$$

According to the equations (9) and (10), the absolute error in the time of discharge varies with the incident light intensity. Manipulating equations (7), (8), (9) and (10), the relative error is given by:

$$Relative\ error = \pm \frac{V_{offset}}{(V_{dd} - V_{ref})} \text{ for constant reference voltage} \qquad (11)$$

$$Relative\ error = \pm \frac{V_{offset}}{(V_{dd} - V_{mín})} \text{ for ramp reference voltage} \qquad (12)$$

The results of equations (11) and (12) suggest that it is possible to implement methods of correcting the error introduced by the offset voltage using multiplicative factors. Error introduced by the offset voltage can be graphically interpreted as a steady shift in the t_d

photoresponse features shown in Fig. 4. Fig. 15 show the relative error given by equation (11) as function of reference voltage (V_{ref}) for V_{offset}=50mV and V_{dd}=3.3V. The relative error is lower for lower values of reference voltage and it is constant and independent of light intensity.

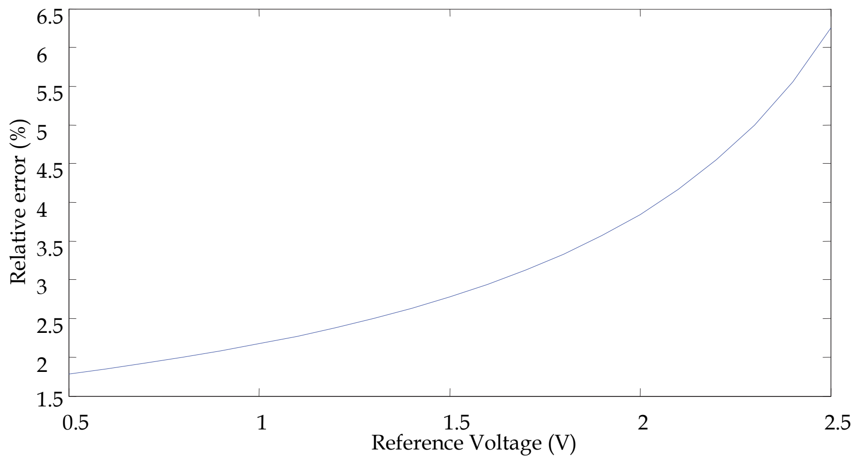


Fig. 15. Relative error introduced by offset voltage of comparator.

6. Experimental results

This section presents the measurements results reported in (Campos, 2008). The prototype integrated circuit containing an array of size 32x32 pixels was fabricated in 0.35 μ m CMOS standard process of C35O Austriamicrosystems. A Xenon lamp type with a peak luminous intensity at 830 nm and an optical filter of 825nm were used to perform the measurements of the discharge time. All measurements were performed using the nominal supply voltage of 3.3 V and bias voltage of the differential pair comparator fixed at 0.7 V. Fig. 16 shows the result of measuring the photoresponse td versus light intensity when the pixel operates with constant reference voltage of 1.5 V. The points in Fig. 16 represent the average discharge time measured for the same light intensity and the continuous curve corresponds to the mean curve fitted. According to the result of curve fitting the photodiode has a sensitivity of 3.4 μV-cm^2/s-W. The measurements showed that the discharge time of darkness is 37segundos.

Fig. 17 shows three different photoresponse in time domain resulting from the use of three different ramp reference voltages. Reducing the integration time of the ramp the time discharge time is reduced. By comparing the discharge times (time comparison) using a constant reference voltage and the ramp one can see that there is a reduction in discharge time at lower light intensities when a ramp reference voltage is used. Note that the photoresponse is approximately flat in the range of irradiation below 10^{-1} (μ W/cm^2) (Fig. 17). In this region the discharge time becomes almost constant becoming harder to indentify the light intensity.

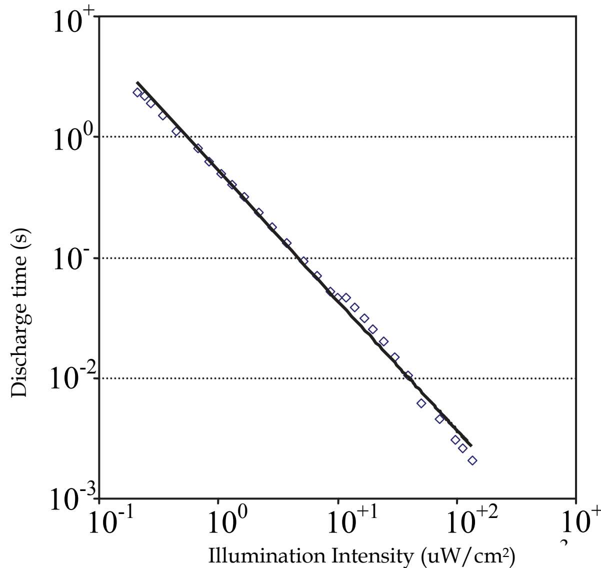


Fig. 16. Time discharge characteristic for constant reference voltage

The experimental measure of the normalized spectral response is shown in Fig. 18. The wide range in which the response will be approximately constant due to the fact that although the junction depth is shallow, low doping concentration of the substrate causes the width of the depletion region extends significantly into the substrate (greater depth). Documents supplied by the factory also report approximately flat spectral response in most of the visible spectrum.

Fig. 19 shows the relationship obtained from SNR measurements as a function of light irradiation. The ratio remains constant SNR value of 54dB in most of the range of irradiation showing good agreement with the theoretical analysis presented in subsection 3.2.5 (Figure 3.25). The theoretical results indicate that using the constant reference voltage SNR is constant throughout the range of light irradiation. It is considered that the discrepancy observed in the SNR measured in the upper level of irradiation was due to the process of determining the average using the oscilloscope. At the time of the measurements was not obtained proper adjustment of the oscilloscope to perform the average. A constant feature of the SNR shows very attractive and important when compared to the results reported in the literature so far. The results reported in other studies shows that the SNR varies from 0 to 54dB, where 54 dB is suggested as a maximum.

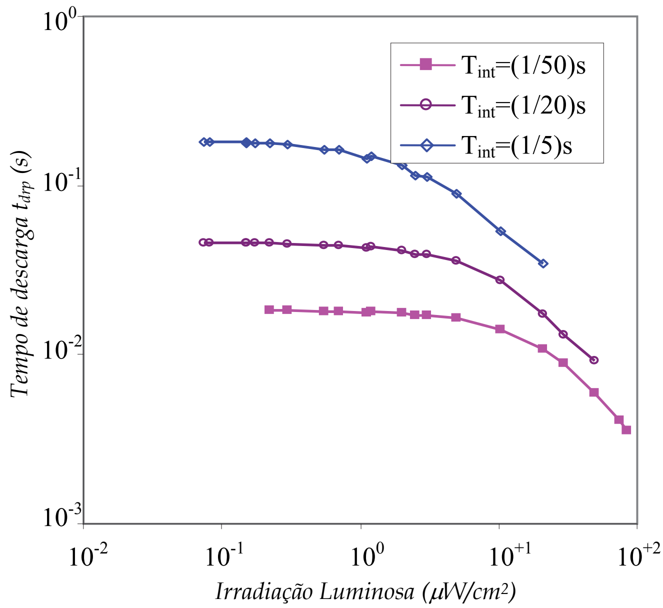


Fig. 17. Spectral photoresponse in time-domain using constant reference voltage

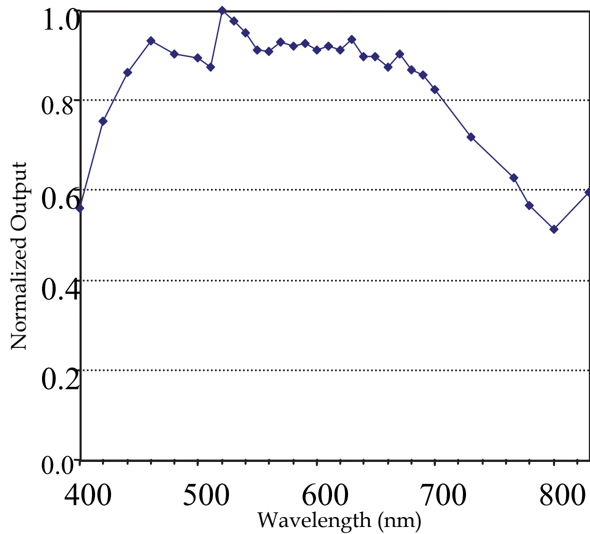


Fig. 18. Spectral photoresponse in time-domain using constant reference voltage

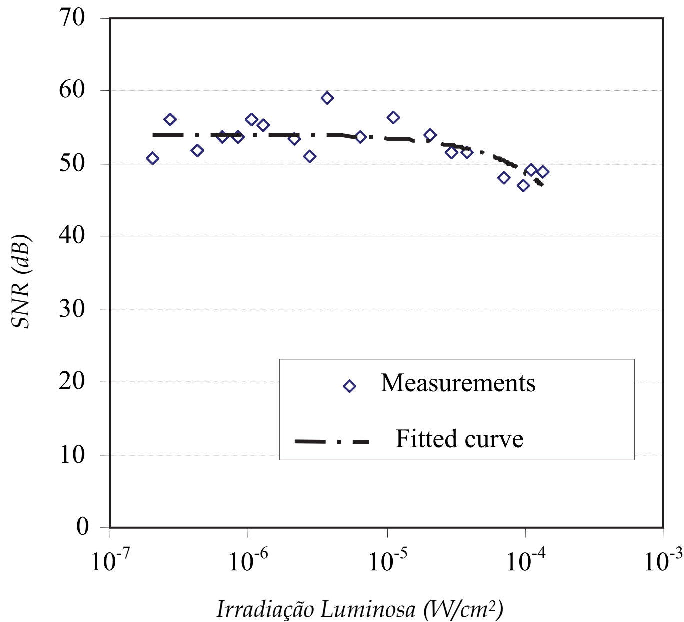


Fig. 19. Noise in time-domain for constant reference voltage (V_{ref}=1.5V)

7. Conclusions

Multisampling time-domain CMOS imagers were presented. Sampling can be linear, logarithmic and linear-logarithmic. Pixel architecture is composed only by a comparator and a D flip-flop reducing the total transistors integrated per pixel compared to others approach. Analysis shows that the matrix size limits the total bit number that represents the image. High dynamic range is achieve only for operation in which the image representation is higher or equal to 16 bits. The FPN introduced by voltage offset of comparator presents constant relative error. This constant relative error might be particularly interesting for lower light intensities.

8. References

Bermak, A., Yung, Y-F. (2006). A DPS Array with Programmable Resolution and Reconfigurable Conversion Time. *IEEE Transactions Very Large Scale Integration Systems*, 14, 1, 15-22.

Campos, F. S., Marinov, O., Faramarzpour, N., Saffih, F., Deen, J.M., Swart, J.W.. (2008). A multisampling time-domain CMOS imager with synchronous readout circuit. *Analog Integrated Circuit and Signal Processing*, 57, 151-159.

Chen, S., Bermak, A., and Boussaid, F. (2006). A Compact Reconfigurable Counter Memory for Spiking Pixels. *IEEE Electron Device Letters,* 27, 4, 255-257

Doge, J., Schonfelder, G., Streil, G. T., and Konig, A. (2002). An HDR CMOS Image Sensor with Spiking Pixels, Pixel Level ADC, and Linear Characteristics. *IEEE Transactions Circuits Systems II, Analog Digital Signal Processing,* 49, 2, 155-158.

Fossum, E. (1997). CMOS Image Sensors: Electronic Camera-on-Chip. *IEEE Transactions Electron Devices,* 44, 10, 1689-1698.

Hosticka, B. J., Brockherde, W., Bussmann, A., Heimann, T., Jeremias, R., Kemma, A., Nitta, C., and Schrey, O. (2003). CMOS Imaging for Automotive Applications. *IEEE Transactions on Electron Devices,* 50, 1, 173-183.

Kitchen, A., Bermark, A., and Bouzerdoum, A. (2004). PWM Digital Pixel Sensor Based on Asynchronous Self-Reset Scheme. *IEEE Electron Device Letters,* 25, 7, 471-473.

Kleinfielder, S., Lim, S. H., Liu, X., and Gamal, A. E.. (2001). A 10000 Frames/s CMOS Digital Pixel Sensor. *IEEE Journal Solid-Stated Circuits,* 36, 12, 2049-2059.

Lai, C.-H., King, A.-C., and Huang, S.-Y. (2006). A 1.2-V 0.25-μm Clock Output Pixel Architecture with Wide Dynamic Range and Self-Offset Cancellation. *IEEE Sensors Journal,* 6, 2, 398-405

Qi, X., Guo, X., and Harris, J. (2004). A Time-to-First Spike CMOS Imager. In *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS 2004)* (Vancouver, BC, May 23-26). IEEE Press, 4, 824- 827.

Saffih, F., and Hornsey, R.. (2007). Foveated Dynamic Range of the Pyramidal CMOS Image Sensors, *IEEE Transaction on Electron Devices*, 54, 12, 3422-3425.

Sandage, R., and Connelly, J. (1995). A Fingerprint Opto-Detector Using Lateral Bipolar Phototransistor in a Standard CMOS Process. In *IEDM Digest of Technical Paper*, Washington, DC, USA, December 10-13. IEEE Press, 171-174.

Stoppa, D., Simoni, A., Gonzo, L., Gottardi, M., and Betta, Dalla Betta, G.-F. (2002). A 138 dB Dynamic Range CMOS Imager Sensor with New Pixel Architecture. In *Proceedings of the IEEE International Solid-State Circuits Conference, Digest of Technical Papers*, San Francisco, USA, February 3-7. IEEE Press, 1, 440-442.

Trepanier, J.-L., Sawan, M., Audet, Y., and Coulombe, J.. (2002). A Wide Dynamic Range CMOS Digital Pixel Sensor. In *Proceedings of the 2002 45th Midwest Symposium on Circuits and Systems (MWSCAS-2002)*, Tulsa, USA, August 4-7. IEEE Press, 2, 437-440.

Yadid-Pecht, O., and Belenky, A.. (2003). In-Pixel Autoexposure CMOS APS. *IEEE Journal of Solid-State Circuits,* 38, 8, 1425-1428.

Yang, S.-H., and Cho, K.-R.. (2002). High Dynamic Range CMOS Image Sensor with Conditional Reset. In *Proceedings of the IEEE Custom Integrated Circuits Conference,* Orlando, USA, May 12-15. IEEE Press, 265-268.

Yasuda, T., Hamamoto, T., and Aizawa, K.. (2003). Adaptive-Integration-Time Image Sensor with Real-Time Reconstruction Function. *IEEE Transactions on Electron Devices,* 50, 1, 111-120.

Bandwidth Extension for Transimpedance Amplifiers

Omidreza Ghasemi
Department of Electrical and Computer Engineering
Concordia University, Montreal
Canada

1. Introduction

This chapter aims to introduce optical receivers in general and application of electronic circuits in these receivers then some important parts of these electronic circuits which are the amplifiers are discussed. An attempt has been made in this chapter to provide some useful information about different aspects of the optical communication in our life and the importance of high speed wideband aspects in such systems. Eventually the role of transimpedance amplifiers in a typical optical communication receiver is discussed.

1.1 Role of fiber optic systems in present-day communication

By growing the number of Internet nodes, the volume of the data transported on the backbone has increased. The load of the global Internet backbone will increase to tens of terabits per second very soon. This indicates that the backbone bandwidth will increase by a factor of 100. Handling of such volumes of data requires suitable media with low loss and high bandwidth. Among the available transmission media, Optical fibers have the best performance for loss and bandwidth. High speed data can be transported over hundreds of fiber without significant loss in signal integrity. These fibers benefit from reduction of cost and performance. The number of the Internet nodes increase with a fast pace, leading to bit rate of a few terabits per second. The bandwidth requirements are growing with fast pace. Applications such as virtual reality will require data rates that are 10,000 times higher than currently available ones [1].

The arrival of cheaper and more powerful personal computers has not only expanded the user base but also created a demand for greater transmission capacity among the telecom networks. In state-of-the-art technology, fiber optic devices and systems are widely employed to satisfy this need for more data throughput economically [2]. Fiber optic communication is an economic solution because its physical nature lets network providers broaden capacity by increasing the transmission bit rate.

Because fiber optics are only communication medium capable of handling such high data rates, there is a widespread demand for high-speed optical and electronic devices, circuits, and systems.

Today possibility of high levels of integration on a single chip enables higher performance. VLSI technologies such as CMOS can now take over the territories of GaAs and InP devices.

1.2 Review of some important optical communication systems

In the past two decades, CMOS technology has dominated the analog integrated circuit design arena, providing low-cost, high performance solutions and. Around 90% of the analog and mixed-signal products in today's semiconductor industry are designed and fabricated in CMOS technologies [1].
Use of CMOS process for fabrication of the electronic system in the optical system lets integration of high-speed front-end circuits and low-speed framers on the same chip. This integration can reduce the package count, board size, and cost of the system.
The two widely accepted commercial systems, SONET OC-192 and OC-768, operate at 10 and 40 Gb/s. The 10-Gb/s CMOS transceiver has already been introduced by a few companies and an extensive amount of research has been performed to improve the design of these systems. However, implementations of the 40- Gb/s CMOS transceivers is behind the 10 Gb/s receivers by a few years because these systems have only become realizable in relatively advanced technologies. In modern fiber optic transmission system, the synchronous optical network (SONET) and synchronous digital hierarchy (SDH) standard define a technology for carrying many signals of different capacities. The basic transmission bit rate is OC-1 at 51.8Mbit/s, and higher bit rates offered by SONET/SDH are summarized in Table 1.1 [2].

SONET	SDH	Bit Rate
OC-1	-	51.84 Mbit/s
OC-3	STM-1	155.52 Mbit/s
OC-12	STM-4	622.08 Mbit/s
OC-48	STM-16	2.4883 Gbit/s
OC-192	STM-64	9.9533 Gbit/s
OC-768	STM-196	39.8131 Gbit/s

Table 1. 1-standard bit rates for optical communication

1.3 A typical optical communication system

Optical communication systems become more important because of the increasing demands for high-speed and large-capacity data communication. The Fig 1.1 and 1.2 show the block diagram of the optical transmitter and receiver system. In this system data are received in the digital form in the optical transmitter side and should be delivered in the digital form in the receiver side. In general in the transmitter data are converted to light using the Laser Diode and delivered to the Optical fiber and in the receiver the data coming from the optical fiber are converted to the electrical signal (current) using the Photo Diode (PD) and amplified using the amplifiers in the receiver.
The transimpedance amplifier (TIA) which converts the photodiode current into a voltage requires high gain, wide bandwidth, low noise and low input impedance with low power consumption.

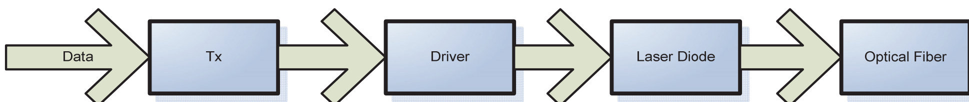


Fig. 1.1 Optical Transmitter Block Diagram

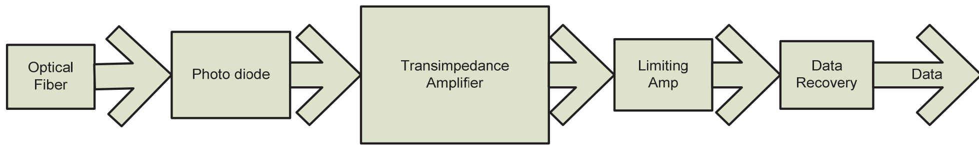


Fig. 1.2 Optical Receiver Block Diagram

An optical receiver must convert a μ A–input current into a digital signal. Furthermore the receiver should use a standard commercial digital CMOS process with little external overhead. In this way the optical receiver can be integrated with a DSP into a single VLSI device. An optical receiver can not be characterized only by its maximum bit rate. The transimpedance of the first stage is an important parameter as well. A high gain transimpedance is necessary when low input currents (a few μ A) must be detected. This is necessary to achieve a high output voltage in the first stage in order to reject noise from sources, such as the digital environment integrated on the same IC [2].
Transimpedance amplifiers play a vital role in optical receivers. Trade-offs between speed, gain, noise and supply voltage exist in TIA design. As TIAs experience a tighter performance envelope with technology scaling at the device level and speed scaling at the system level, it becomes necessary to design the cascade of the TIA, the limiter, and the decision circuit concurrently [1].
As the gain bandwidth product is a measure of both amplification and bandwidth for opamps, the product of the transimpedance (Z) and the bandwidth (BW) should be taken into account in comparison of transimpedance amplifiers. As transimpedance can be exchanged for bandwidth to some extent, a transimpedance-bandwidth-product (ZBW) can be defined for optical receivers.
The transmission of optical data via fiber cables involves electrical-to-optical conversion at the transmission end and optical-to-electrical at the receiving end. These conversion processes are handled by optoelectronic transceiver units that contain electronic devices and semiconductor optical components.

1.4 Transmitting and receiving requirements

In the receiver which is shown in Fig 1.2, the PD converts the received light to a signal current, and the signal swing is amplified to logic levels. Subsequently, the Data Recovery part performs timing and amplitude-level decisions on the incoming signal, which leads to a time- and amplitude-regenerated data stream. The result is then de-multiplexed, thereby reproducing the original channels.
The light-wave traveling through the fiber usually goes under considerable attenuation before reaching the PD. This attenuation requires a subsequent stage to detect and amplify the signal at an acceptable rate. Hence the TIA, the first stage of amplification, should provide wide-band amplification and low input referred noise. To provide the high input sensitivity necessary to receive optical signals weakened by transmitter, the TIA noise must be reduced to a minimum. On the other hand, a high overload tolerance is required to avoid bit errors caused by distortion in the presence of strong optical signals. Furthermore, to ensure stable operation and the required bandwidth, gain can be optimized only within a narrow range. This limitation sometimes causes the output voltage that results from low-power optical signals to be insufficient for further processing. Therefore, the LA often follows to amplify small TIA voltages.

1.5 Technological implementation

In optical communication systems, the front-end of the receiver has a PD and a TIA. Because of the performance requirements for the TIA, the front-end circuit has traditionally used III-V compound semiconductor technologies. On the other hand, their CMOS counterparts, despite having such advantages as low power consumption, high yield that lowers the cost of fabrication, and higher degree of integration, have not performed well enough to survive in such a noisy environment without sacrificing other important attributes. This performance shortcoming is mainly due to the nature of silicon CMOS devices that have limited gain, limited bandwidth. The low voltage headroom in submicron CMOS technologies also is an obstacle to the implementation of broadband amplifiers.

The optical front-end can be realized with monolithic optoelectronic integrated circuits (OEIC) that have all the components in a single chip. In these products, the PDs and circuits are individually optimized, fabricated and packaged in separate processes and connected by external wires. However, the interconnections may cause unwanted parasitic feedback that degrades overall system performance.

1.6 Some important parameters in optical receivers

An optical receiver front-end consists of two major parts, a semiconductor Photo Diode (PD) followed by an electronic signal amplifier. Light traveling through the fiber is attenuated before reaching the PD, thus requiring a highly sensitive receiver to detect the signal. Hence the performance of the receiver is often characterized by the input sensitivity, bandwidth, and gain in the receiver. This sensitivity can be expressed in terms of mean optical input power or root-mean–square (RMS) input-referred noise. Bandwidth is usually determined by the total capacitance contributed by the PD, the preamplifier and other parasitic elements present at the optical front-end.

The fundamental behind the optical to electrical signal conversion is optical absorption. In the operation of the PD, absorbing the incident radiation and in turn generating electron-hole pairs that drift to the metal contacts to generate a current in the external circuit. An equivalent circuit model of the PD is often represented by a current source with a shunt capacitance [2].

Common types of the Photodiode (PD) are p-i-n and avalanche PDs with the types defined based on the photo detection process.

First, the p-i-n consists of a highly resistive middle layer between p and n sections to create a wide depletion region in which a large electric field exists. Most of the incident is absorbed inside i-region thus the drift component of the photocurrent dominates over the slow diffusion component that can distort the temporal response of the PD.

Second, the PD uses an impact ionization mechanism in which an additional multiplication layer is introduced to generate secondary electron-hole pairs that result in an internal current gain. An avalanche PD is often used when the amount of optical power that can come from the receiver is limited, however the avalanche process has major drawbacks in its high noise contribution and in the trade-off between gain and bandwidth.

1.7 Characteristics of transimpedance amplifier

The small photo current generated by the PD must be converted, to a usable voltage signal for further processing. Therefore a preamplifier is used as the first stage and has great

impact on determining the overall data rate and sensitivity that can be achieved in an optical communication system. Typically the preamplifier is required to be able to accommodate wide-band data extending from dc to high frequencies to avoid inter-symbol interference (ISI). These are some parameters which show the performance of the preamplifier and in here we are going to learn about them:

1. Bandwidth
2. Gain
3. Noise
4. Sensitivity
5. BER

As a rule of thumb the amount of BW required for the amplifiers in the receiver side should be 70 percent of the bit rate (BR). For example for an optical receiver to be employed in a 10Gb/s bit-rate system we need to at least have 7GHz bandwidth for the preamplifier.

The Gain required for the preamplifier (TIA) is not defined as a specific value to be mentioned and in the literature, there are a lot of different values achieved for the gain of the TIA but because TIA needs to deliver the voltage to the main amplifier (LA), the input sensitivity of the main amplifier should be satisfied ,therefore normally we need to achieve at least a few mili-volts at the output of the TIA and because we have the amount of the input current as tens or hundreds of micro ampere at the input of the TIA (depend on the optical system) we need to achieve the gain of a few hundreds at least to satisfy the conditions. Normally in the literature the gain of between 40dB-Ohms and 60dB-Ohms has been reported for the recent TIAs.

The sensitivity and noise are related to each other. Since the TIA needs to sense a very small amount of current at the input, the amount of input referred noise should be very low so the amplifier can have a high sensitivity which can sense the very small amount of current.

BER normally in the optical system the amount of BER should be less than 10^{-12} .The definition of BER is the ratio of the number of errors received to the total number of bits. There are some mathematical relations between BER and the BW of the amplifiers in the receiver side which shows if the rule of thumb mentioned above is achieved for the amplifiers in the receiver side the amount of BER will be satisfied.

2. Background and literature review

2.1 Overview

The aim of this chapter is to review some of the previous works which have been done in the TIA area. We aim to discuss the BW extension and review some of the techniques which have been done in the literature to improve the performance of the TIAs.

2.2 BW extension in the TIA design

The general structure for the feedback TIA is shown in the figure below in which we can see that a voltage amplifier with a resistive feedback can be converted to a Transimpedance amplifier [3]. As we can see the light is converted to current using the Photodiode (PD) and then this current is amplified using the TIA and then the voltage signal will be delivered to the main amplifier (Limiting Amplifier).

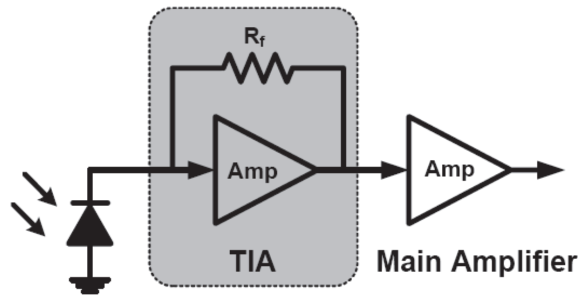


Fig. 2.1 PD, TIA and LA

Now according to the discussion here, there are several obstacles to extend the Bandwidth of a TIA:

1. Photodiode Capacitance (CPD)
2. Inherent parasitic capacitance of the MOS Transistor
3. Loading Capacitance (input capacitance of the main amplifier)

The methods normally we see in the literature on the topic of bandwidth extension are dealing with either of these issues and try to defeat them in some respects and hence extend the Bandwidth of the TIA .There are several bandwidth extension techniques for the TIAs in the literature and in this part we need to discuss these techniques.

For the matter of this discussion we need to define the word bandwidth .The bandwidth is defined as the lowest frequency at which the TIA gain drops by $\sqrt{2}$ or 3dB. Accordingly this bandwidth is often called the 3-dB bandwidth [4].

Some of the techniques which have been done previously in the literature are summarized below.

1. Shunt peaking
2. Series peaking
3. PIP technique
4. Inductor between the stages

2.2.1 Shunt peaking

Shunt peaking is the traditional way to enhance the bandwidth in wideband amplifiers. It uses a resonant peaking at the output of the circuit. It improves the BW by adding an inductor to the output load. It introduces a resonant peaking at the output as the amplitude starts to roll off at high frequencies. Basically what it does is that, it increases the effective load impedance as the capacitive reactance drops at high frequencies [4].

The model for a common source amplifier with shunt peaking is shown in the figure below [5], [16]. As we can see an inductor is added in series with the resistive load and establishes a resonance circuit and reduces the effect of the output capacitance which in this figure consists of all the parasitic capacitances of the drain of the transistor and the loading capacitance of the next stage.

Kromer [7] has used inductive peaking technique in all the 3 stages of the TIA, The main stage is CG but it uses 2 boosting stages in the path of the signal. He could achieve the transresistance gain of 52dB ohms and -3dB BW of 13GHz, although he worked with the technology of 80nm.The amount of Photodiode capacitance he used is 220fF.

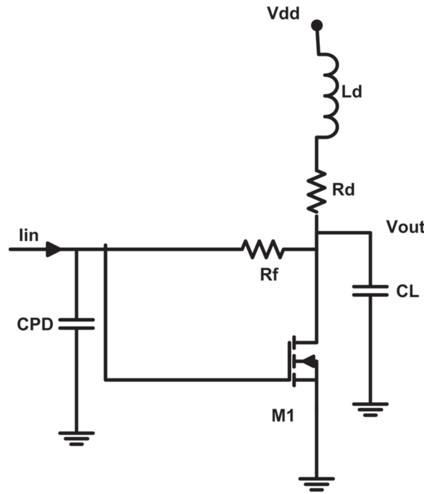


Fig. 2.2 Shunt peaking

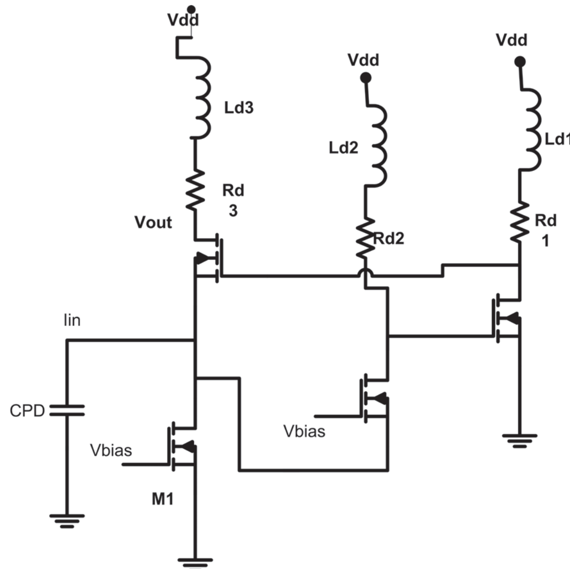


Fig. 2.3 Shunt peaking technique by Kromer

2.2.2 Series peaking

Wu [8] has presented this technique. This technique mitigates the deteriorated parasitic capacitances in CMOS technology. Because the inductor is inserted in series with all the stages in the signal path, it is called series peaking technique. As we can see in the Fig 2.4 the structure of the circuit shows that inductors are used to reduce the effect of the parasitic capacitances in the different stages of the amplifier. As we can see without inductors, amplifier bandwidth is mainly determined by RC time constants of every node.

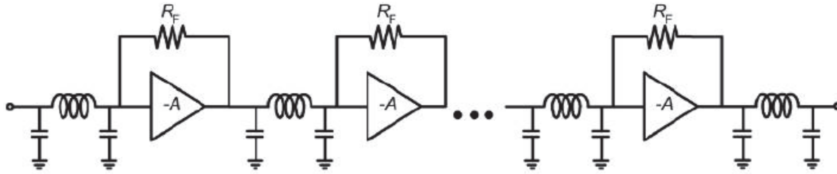


Fig. 2.4 Series peaking technique

This work was done in 0.18um CMOS technology and achieves a gain of 61dB-Ohms and BW of around 7GHz. The amount of PD capacitance in this work is 250fF.

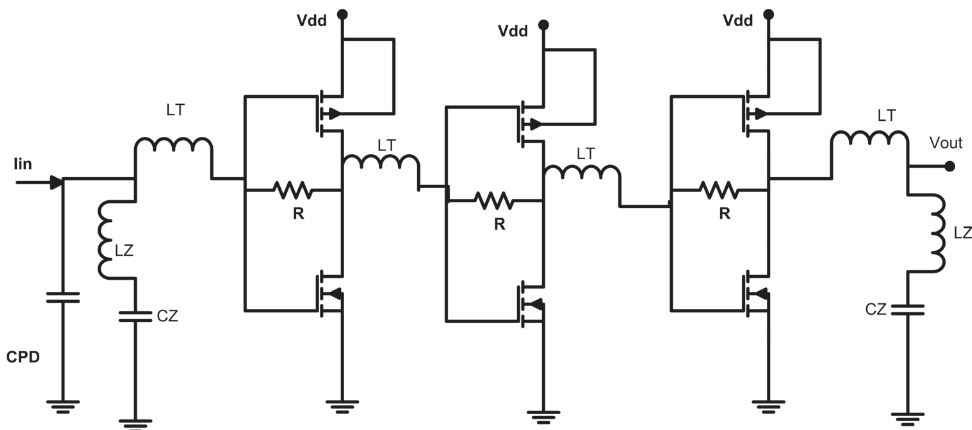


Fig. 2.5 Circuit implemented by Wu

2.2.3 PIP technique

Jin and HSu [9] have proposed this technique to defeat the parasitic capacitances using the combination of several inductors. The combination of the inductors shapes a Π and hence they call it a Pi-type Inductor Peaking (PIP). The Fig 2.6 shows how the combination of 3 inductors in a common source amplifier constructs the PIP technique.

This technique improves the BW of the TIA by resonating with the intrinsic capacitances of the devices. The actual implemented circuit by them is shown in the figure below.

This circuit is done in 0.18 CMOS technology and achieves around 30GHz BW and 51dB-Ohms gain. The amount of PD capacitance in this circuit is the lowest used in the literature and it is 50fF.

2.2.4 Matching inductor between the stages

Analui [10] has mentioned a technique to isolate the effect of parasitic capacitance of different stages to each other. It uses a passive network (inductor) to isolate the effect of capacitors. It has claimed this passive network absorbs the effect of parasitic capacitor of the transistor. This passive network mainly can be an inductor and it can form a ladder filter with the parasitic capacitances of the devices.

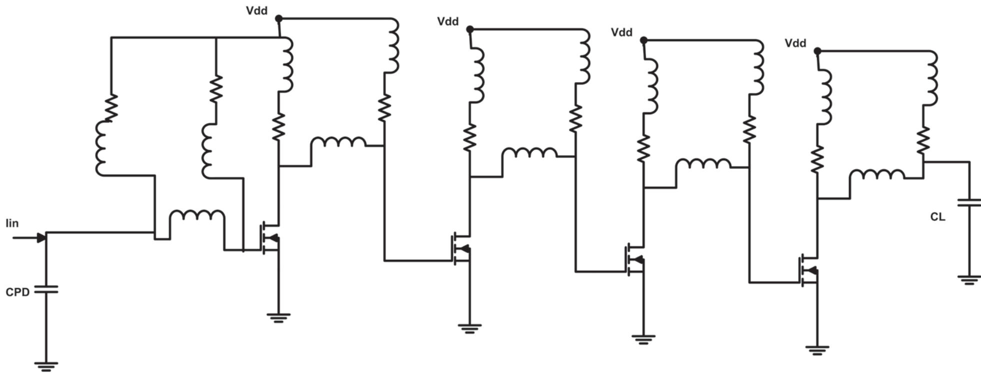


Fig 2.6 Circuit implemented by Jin and HSu

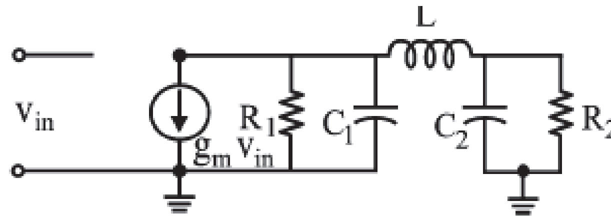


Fig 2.7 Inductor between the stages

The circuit was implemented by Analui. The parasitic capacitances of the devices are shown in the circuit which can form the ladder structure with the deliberately added inductor
He has achieved the gain of 54dB and 3dB BW of 9.2GHz and this work was done in 0.18um BICMOS process using CMOS transistors. The amount of PD in this circuit is 500fF.

2.3 Conclusion

In this chapter we reviewed some of the BW extension techniques available in the literature in the field of TIA design. In general inductive techniques are quite common to extend the BW in the TIAs and researchers have accepted the fact that in order to have wide band circuits. It is worth losing some area in the chip and instead have a better circuit in order to build optical receivers for higher data-rates but still it is a challenge that although it is acceptable to build wideband circuits using spiral inductors, we need to have circuits with fewer number of inductors to have low cost chips.

3. Three stage low power transimpedance amplifier

In this chapter a three-stage Transimpedance Amplifier based on inductive feedback technique and building block of cmos inverter TIA has been proposed. The effects of parasitic capacitances of the MOS transistors and the photodiode capacitance have been mitigated in this circuit [11], [12]. The process of zero-pole cancellation in inductive feedback to extend the BW of the amplifier has been reviewed. To demonstrate the feasibility of the technique the new three stage transimpedance amplifier has been simulated

in a well-known CMOS technology (i.e. 90nm STMicroelectronics). It achieves a 3-dB bandwidth [13] of more than 30GHz in the presence of a 150fF photodiode capacitance and 5fF loading capacitance while only dissipating 6.6mW.

3.1 Introduction

Optical receivers are important in today's high data rate (Gb/s) wireline data communication systems. The requirement for the amplifiers is to be wideband to be able to handle the data. Transimpedance amplifiers (TIAs) at the frontend of the optical receivers do an important job which is the amplification of the current received from the photodiode (PD) to an acceptable level of voltage for the next stage. The bandwidth of CMOS TIAs can be limited by the photodiode (PD) capacitance and parasitic capacitances of the MOS transistors. Bandwidth extension technique essentially is a technique to mitigate the effect of these capacitances in high frequencies when the TIA gain (ratio of the output voltage to input current) starts to roll off. Different circuit techniques for TIAs have been proposed in the past. Shunt peaking is the most well-known technique to enhance the bandwidth of the amplifiers [22]. Multiple inductive series peaking is also a proposed technique for BW extension in the amplifiers [23]. Putting matching networks (inductor) between the stages of the amplifier has been proposed [4]. A Π-type inductor peaking (PIP) technique to enhance the bandwidth of TIAs was recently proposed [24]. Inductive feedback technique [19], [25] has also been applied to extend the BW of TIAs.
The remainder of this chapter is organized as follows: Section 3.2 reviews the inductive feedback technique and the theory of zero pole cancellation for the conventional inverter based TIA [19]. In Section 3.3 the proposed three-stage TIA is introduced. To show the validity of the design simulation results of the circuit and a comparison with other works are shown in Section 3.4. In Section 3.5, conclusions are given.

3.2 Bandwidth extension using inductive feedback technique

This part has been discussed in the previous publication [19] and is reviewed in this paper as the basis for the extension of the work which is discussed in part 3.4 of this paper. The objective of using inductive feedback is to extend the BW of the TIA by deliberately adding a zero to the transfer function of the TIA and hence cancel the dominant pole of the amplifier thereby extending the BW. This can be done by adding an inductor to the feedback path of the TIA. The newly introduced inductor in the feedback path (inductive feedback) adds one zero and one pole to the transfer function of the TIA and by an appropriate design the newly added zero can cancel the dominant pole of the amplifier and hence extend the BW [19]. In order to discuss the technique in detail we consider two TIAs shown in Figures 3.1 and 3.2. In this paper we refer to the circuit in Fig. 3.1 as the TIA with resistive feedback and the circuit in Fig. 3.2 as the TIA with inductive feedback. Fig. 3.3 shows the small signal model of the TIA.
In the small signal model for the TIA we have these definitions:

$$G_m = g_{m1} + g_{m2}\,,\ r_o = (r_{ds1}\,||\,r_{ds2})$$

$$c_i = c_{gs1} + c_{gs2} + c_{PD}\ ,\ c_f = c_{gd1} + c_{gd2}$$

$$c_o = c_{db1} + c_{db2} + c_L$$

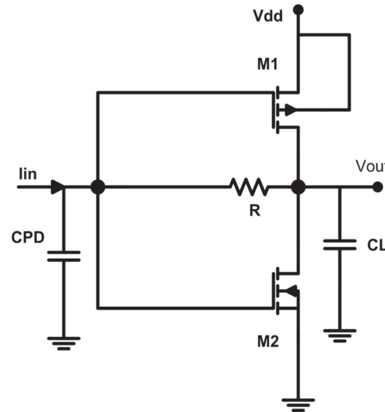


Fig. 3.1 TIA with resistive feedback

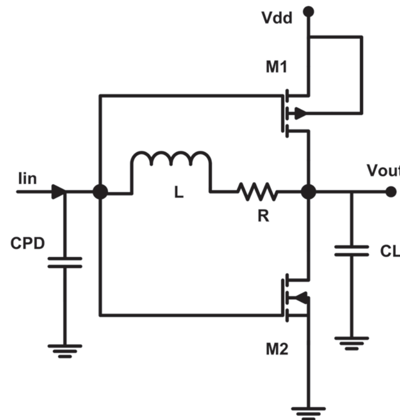


Fig. 3.2 TIA with inductive feedback

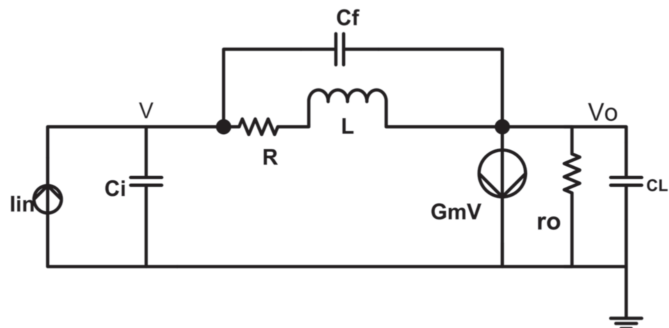


Fig. 3.3 Small signal model of the TIA with inductive feedback

And the transfer function of this circuit is:

$$Z(s) = \frac{as^2 + bs + c}{As^3 + Bs^2 + Cs + D} \tag{1}$$

In which for the case of the Fig. 3.1 (L=0) the coefficients are shown with the index 1 and we have:

$$a_1 = 0\,,\; b_1 = Rc_f\,,\; c_1 = 1 - G_m R$$

$$A_1 = 0\,,\; B_1 = R(c_i c_o + c_f c_o + c_i c_f)$$

$$C_1 = c_i + c_o + R(c_i g_o + c_f g_o + c_f G_m)$$

$$D_1 = g_o + G_m$$

For the case of the circuit in Fig. 3.2 we have the coefficients as (shown with the index 2):

$$a_2 = Lc_f\,, b_2 = R_{c_f} - LG_m\,, c_2 = 1 - G_m R$$

$$A_2 = L(c_i c_o + c_f c_o + c_i c_f)$$

$$B_2 = R(c_i c_o + c_f c_o + c_i c_f) + L(c_i g_o + c_f g_o + c_f G_m)$$

$$C_2 = c_i + c_o + R(c_i g_o + c_f g_o + c_f G_m)$$

$$D_2 = g_o + G_m$$

Now considering the transfer function of the system in Fig. 3.1, the dominant pole of the system (-3db BW) can be approximately calculated as D_1 / C_1.

$$P = \frac{g_o + G_m}{C_i + C_o + R(C_i g_o + C_f g_o + C_f G_m)} \tag{2}$$

In the proposed approach, the dominant pole is cancelled by adding a zero. This can be achieved by adding an inductor in the feedback path of the amplifier giving the circuit in Fig 3.2. As we can see adding an inductor to the feedback path adds one pole and one zero to the transfer function and the newly added zero is approximately:

$$Z = \frac{R}{L} \tag{3}$$

By a judicial choice of the inductance we can cancel the dominant pole of the circuit in Fig. 3.1 which determines the -3db BW and hence extend the BW. An approximate value for the amount of the inductor can be calculated by solving the equation P=Z, giving

$$L = \frac{R(C_i + C_o) + R^2(C_i g_o + C_f G_m)}{g_o + G_m} \tag{4}$$

3.3 Zero-pole cancellation process

The zero-pole analysis in this part has been taken from the previous publication [19] and is reviewed to show the theory for the extension of the work in part 3.4. The circuit has been simulated using a well-known sub-micron CMOS technology (i.e. 90nm CMOS STMicroelectronics). Simulations are done with a single supply (i.e. Vdd=1.2 V) and in the presence of a 150fF photodiode capacitance and 5fF loading capacitance. The pole-zero analysis outlined here was done using the schematic of the circuit with ideal inductor values to show the process of zero-pole cancellation more clearly. Based on the pole-zero analysis for TIA with resistive feedback the circuit has two poles and one zero. The poles are located in the LHP of the s-plane which shows the circuit is stable. The TIA with inductive feedback will have two zeros and three poles. By choosing the inductor according to (4) we can cancel the dominant pole leaving a pair of complex conjugate poles in the circuit. The circuit after having cancelled the single dominant pole will have two complex conjugate poles with a damping factor and natural frequency which can be designed for the desired frequency response. The zero-pole cancellation process has been shown and we can see that by changing the value of the inductor in the circuit the newly added zero is moving towards the dominant pole of the circuit. In the end it reaches to that pole and cancels it and hence this zero can extend the -3dB BW. We can also see that the positions of the complex conjugate poles [14] are changing by sweeping the value of the inductor. The actual values of the poles and zeros extracted from the simulation are shown in Table I.

L(nH)	Zeros (GHz)	Poles (GHz)
0	192.2	-12.7 -22
2	-27.3 223.4	-14.6 -17±17.9j
2.5	-21.8 224.2	-14.9 -13.6±17j
3	-18.1 224.8	-15.2 -11.4±16j
3.5	-15.5 225.2	-15.5 -9.8±15j

Table 3.1 Pole -Zero analysis for the circuit

3.4 Proposed three-stage TIA using the inductive feedback technique

In this part the new proposed TIA is discussed. Cascaded amplifiers are one of the ways to widen the bandwidth of the amplifiers [3], [17] and therefore, we can cascade the previously

discussed single stage transimpedance amplifier to get more Gain*Bandwidth from the amplifier. In this part we introduce the new three stage cascaded TIA using inverter based TIA with inductive feedback. In Figure 3.4 the new transimpedance amplifier has been shown.

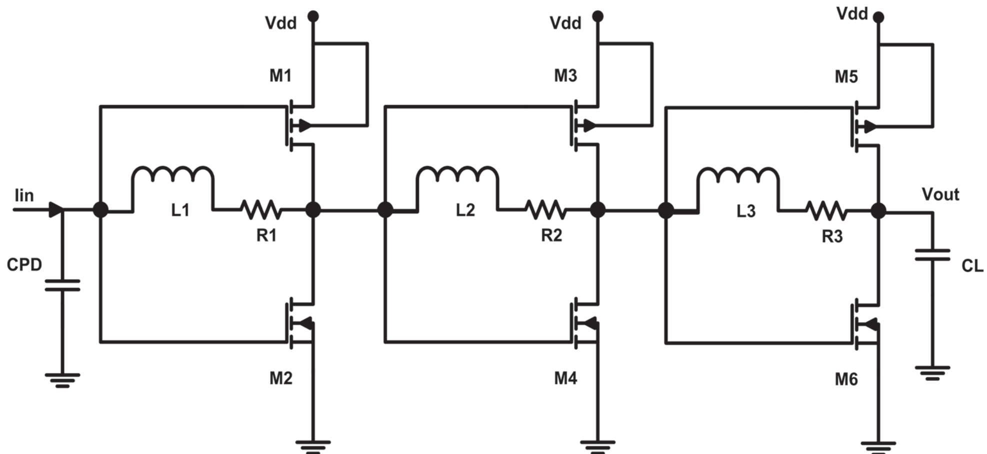


Fig. 3.4 Three stage inverter based TIA with inducitve feedback

In Figure 3.5 the simulation results based on different values of the inductors have been shown. The frequency response of the three-stage TIA has been summarized in table 3.2 as well. In order to fabricate the circuit in sub-micron CMOS spiral inductors are needed [15].In the table the size of the transistors are all 12/0.1(um/um) and the resistor in the feedback path is 400Ohms. The frequency response of the three stage transimpedance circuit for different values of the inductor has been shown in Figure 3.5.
The frequency response of the three-stage transimpedance amplifier has been summarized in table 3.2. For different values of the three inductors for each stage in the table the amounts of the -3dB Bandwidth and gain peaking have been shown. Table 3.3 gives a comparison of this work with other previously published works using other techniques and the new Transimpedance amplifier simulation results together with the other works in the literature has been summarized. As we can see the advantage of this work is to offer high bandwidth consuming very low power consumption in comparison with other previously published works.

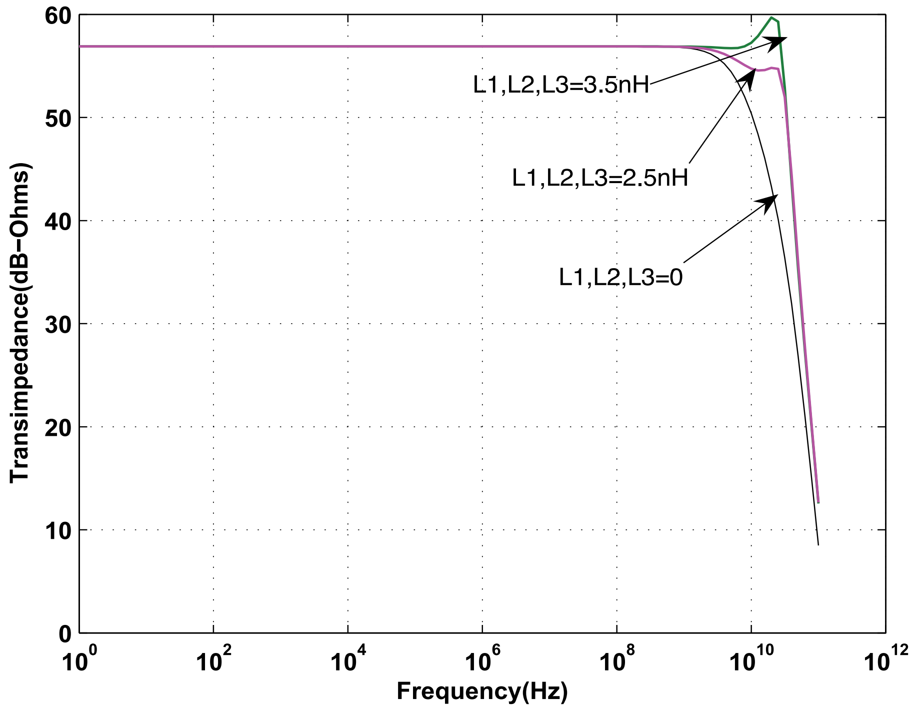


Fig. 3.5 Frequency response of the three stage TIA

Transistor size (um/um)	Resistors(Ohms) R1,R2,R3	Inductors(nH) L1, L2, L3	TIA-Gain (dB-Ohms)	-3dB BW (GHz)	Peaking (dB)
12/0.1	400	0	56.89	5.6	0
12/0.1	400	2.5nH	56.89	27.2	0
12/0.1	400	3.5nH	56.89	30.5	2.4

Table 3.2 Frequency response of the three stage TIA with PD=150fF

	Technology	TIA Gain (dB-Ohm)	-3 dB BW(GHz)	$i_{n,in}$ (pA/√Hz)	Power (mW)	Number of Inductors	PD Cap (fF)
This work	90nm- CMOS	56.8	30.5	36.4	6.6	3	150
Design[5]	90nm- CMOS	50.8	16.7	16.9	2.2	1	150
Design[2]	180nm-CMOS	61	7.2	8.2	70.2	9	250
Design[3]	180nm-BiCMOS	54	9.2	17	137.5	4	500
Design[4]	180nm-CMOS	51	30.5	34.3	60.1	15	50
Design[6]	65nm-CMOS	8	29	N/A	6	1	N/A
Design[7]	80nm-CMOS	52.8	13.4	28	2.2	3	220
Design[8]	180nm-CMOS	62.3	9.0	N/A	108.0	2	150

Table 3.3 Performance of the new TIA and comparison with state of the art

3.5 Conclusion

In this chapter we briefly reviewed bandwidth extension techniques for TIAs and the single stage inverter based transimpedance amplifier using inductive feedback technique has been discussed. The new three stage inverter based TIA using inductive feedback was introduced and the simulation results for the new TIA have been discussed in detail and comparison with the other previously published works has been done.

3.6 Acknowledgements

The tools and design kits were provided by CMC Mircosystems in Concordia University.

4. References

[1] J. Savoj and B. Razavi "High speed CMOS Circuits for Optical Receivers," Kluwer Academic Publishers,Massachusettes 2001

[2] Indal Song "Multi Gb/s CMOS Transimpedance Amplifier with Integrated photodetector for Optical interconnects," Ph.D thesis ,Georigia institute of technology, Nov 2004

[3] Behnam Analui "Signal Integrity Issues in High speed wireline links," Ph.D thesis ,caltech 2005

[4] B. Analui and A. Hajimiri "Bandwidth enhancement for transimpedance amplifier," IEEE J. of Solid-state Circuits, vol.39, pp. 2334-2340, Dec 2003

[5] S. S. Mohan ,M. Hershenson, S. Boyd, T.H.Lee," Bandwidth Extension in CMOS with Optimized On-Chip Inductors" IEEE J. of Solid-State Circuits, vol 35,No 3,pp 346-355 ,Mar2000

[6] S.M. Rezaul Hasan ,"Design of a Low-Power 3.5-GHz Broadband CMOS Transimpedance Amplifier for Optical Transceiver" IEEE Transaction on circuits and systems,Vol.52,No.6,June 2005

[7] C. Kromer et al, "A low-power 20-GHz 52-dBOhms Transimpedance Amplifier in 80-nm CMOS" IEEE J. of Solid-State Circuits, vol 39,No 6,pp 885-894 , June2004
[8] C.-H. Wu, C.-H.Lee, W.-S. Chen, and S.-I. Liu," CMOS wideband amplifiers using multiple inductive-series peaking technique" IEEE J. of Solid-State Circuits, vol 40, pp.548-552, Feb2005
[9] Jun-De Jin and Shawn S.H.Hsu "40-Gb/s Transimpedance Amplifier in 0.18-um CMOS Technology," European solid state circuits conference, 2006 pp.520-523
[10] B. Analui and A Hajimiri "Multi-Pole Bandwidth enhancement technique for Transimpedance amplifiers," Proceeding of the ESSCIRC 2002
[11] Adel Sedra and Kenneth Smith "Microelectronic Circuits" Fifth Edition, Oxford University Press 2004
[12] B.Razavi "Design of Analog CMOS Integrated Circuits" Preliminary Edition Mcgraw-Hill 2000
[13] M. Ingels and M. Steyaert "Integrated CMOS Circuits for Optical Communication" Springer 2004
[14] Ogata Katsuhiko "Modern Control Engineering" Englewood cliffs, N.J Prentice-Hall 1970
[15] Ali Niknejad "Analysis, Design, and Optimization of Spiral Inductors and Transformers for Si RF ICs" Thesis, College of Engineering, University of California at Berkeley
[16] S. Mohan, M. Hershenson, S. Boyd, T. H. Lee "Simple accurate expressions for Planar Inductors,"IEEE journal of Solid state circuits October 1999
[17] The Design of CMOS Radio-Frequency Integrated Circuits T. H. Lee, 2nd edition Cambridge 2004
[18] A.K. Peterson, K. Kiziloglu, T. Yoon, F. Williams, Jr., M.R. Sander, " Front-end CMOS chipset for 10 Gb/s communication," in IEEE RFIC Sym. Dig, June 2003
[19] Omidreza Ghasemi, Rabin Raut, and Glenn Cowan, "A Low Power Transimpedance Amplifier Using Inductive Feedback approach in 90nm CMOS," To be appeared on IEEE International Symposium on Circuits and Systems (ISCAS) 2009, Taipei, Taiwan
[20] Rabin Raut, Omidreza Ghasemi, "A Power Efficient Wide Band Transimpedance Amplifier in sub-micron CMOS Integrated Circuit Technology," IEEE joint NEWCAS/TAISA conference 2008, Montreal, Canada
[21] Yu-Tso Lin, Hsiao-Chin Chen, Tao Wang, Yo-Sheng Lin, and Shey-Shi Lu, "3-10GHz Ultra-Wideband Low-Noise Amplifier Utilizing Miller Effect and Inductive Shunt-Shunt Feedback Technique," IEEE Transactions on Microwave Theory and Techniques, vol. 55, no. 9, Sept. 2007
[22] S. S. Mohan, M. Hershenson, S. Boyd, and T. H. Lee, "Bandwidth Extension in CMOS with Optimized On-Chip Inductors," IEEE J. Solid-State Circuits, vol. 35, no. 3, pp. 346-355, Mar. 2000
[23] C. H. Wu, C .-H. Lee, W.-S. Chen, and S.-I. Liu, "CMOS wideband amplifiers using multiple inductive-series peaking technique," IEEE J. Solid-State Circuits, vol. 40, pp. 548-552, Feb. 2005

[24] Jun-De Jin and Shawn S.H.Hsu, " 40-Gb/s Transimpedance Amplifier in 0.18-um CMOS Technology," European Solid-State Circuits Conference, 2006 pp.520-523

[25] Thoedoros Chalvatzis et al, "Low-Voltage Topologies for 40-Gb/s Circuits in Nanoscale CMOS" IEEE Journal of solid state circuits, VOL. 42, NO.7 , JULY 2007

Part 3

APD and Single Photon Detection

8

Avalanche Photodiodes in High-Speed Receiver Systems

Daniel S. G. Ong and James E. Green
University of Sheffield
United Kingdom

1. Introduction

The avalanche photodiode (APD) is widely used in optical fibre communications (Campbell, 2007) due to its ability to achieve high internal gain at relatively high speeds and low excess noise (Wei et al., 2002), thus improving the system signal-to-noise ratio. Its internal mechanism of gain or avalanche multiplication is a result of successive impact ionisation events. In an optical receiver system, the advantage of internal gain, in the APD, is experienced when the amplifier noise dominates that of a unity-gain photodiode. This increases the signal-to-noise ratio (SNR) and ultimately improves the receiver sensitivity as the gain increases until the APD noise rises to become dominant.

Indium Phosphide (InP) is widely used as the multiplication layer material in commercially available APDs for applications in the 0.9–1.7μm wavelength region with $In_{0.53}Ga_{0.47}As$ grown lattice-matched to it as the absorption layer. It has been predicted that Indium Alluminium Arsenide ($In_{0.52}Al_{0.48}As$) will replace InP, as a more favourable multiplication layer material due to its lower excess noise characteristics (Kinsey et al., 2000). In comparison to InP, tunnelling currents remain lower in InAlAs due to its larger bandgap. While holes ionise more readily than electrons in InP, the opposite holds true for InAlAs and InGaAs, as electrons ionise more readily than holes; thus making the InGaAs/InAlAs combination superior to InGaAs/InP in a SAM APD, in terms of lower excess noise, higher gain-bandwidth product, and improved sensitivity. Studies have also shown that the breakdown voltage of InAlAs APDs is less temperature dependent compared to InP (Tan et al., 2010), which would be useful in temperature sensitive applications, thus making temperature control less critical.

The sensitivity performance criterion for digital receivers is its bit-error rate (BER), which is the probability of an error in the bit-identification by the receiver. The receiver sensitivity is defined as the minimum average optical power to operate at a certain BER; 10^{-12} being a common standard for digital optical receivers. The sensitivity of APD-based high speed optical receivers is governed by three main competing factors, namely the excess noise, avalanche-buildup time and dark current of the APD. Generally, the excess noise and avalanche-buildup time increases with APD gain. Thus, for a fixed multiplication layer thickness, there is a sensitivity-optimised gain that offers a balance between SNR while keeping the degrading contributions from the excess noise factor and intersymbol-interference (ISI) at a minimum. More importantly, changing the thickness of the multiplication layer strongly affects the receiver sensitivity, as the aforementioned three

factors change. Reducing the thickness of the multiplication layer serves to reduce the excess noise factor, due to the dead space effect, (Li et al., 1998) and minimise ISI via reducing carrier transit times across the avalanche region. On the other hand, the increase in the field in thin layers accentuates tunnelling currents at exponential rates (Forrest et al., 1980a). Thus, careful attention is required when determining the multiplication layer thickness for an optimum APD design.

It is, therefore, very useful and interesting to model the sensitivity of an APD-based receiver system accurately. Such models have been developed but none included some form of dark current mechanism, which can significantly affect the receiver's sensitivity. Characterisation of the APD excess noise factor in test structures is also necessary in order to model the BER of an APD-based receiver system. Several efforts have been made to systematically characterise promising detector material systems including InP and InAlAs.

In this chapter, we will describe the model used to investigate the receiver-sensitivity-optimisation of InP and InAlAs APDs, which include dark current contributions from tunnelling current. A comprehensive assessment of the measurement systems reported in the literature is also provided followed by two suggestions for an improved design. The results of the BER calculations on receiver systems using InP APDs will be presented, followed by a discussion on the competing effects of performance-determining factors. A straightforward comparison between InP and InAlAs APDs will then be presented with an analysis on the difference.

2. Impact ionisation

The impact ionisation process occurs when a carrier injected into a high-field region gains enough energy from the applied field and collides with the lattice structure to produce an electron-hole pair. In an electron-initiated process, as depicted schematically in Figure 1(a), an energetic electron at a higher state of the conduction band scatters with an electron at the top of the valence band via Coulombic interaction, and promotes it to the bottom of the conduction band (Singh, 1995). As this process can have a cascading effect, the net result is the creation of many secondary electrons and holes from a single primary electron, generated through absorption of a photon. A similar process occurs in hole-initiated impact ionisation with similar results, as shown in Figure 1(b).

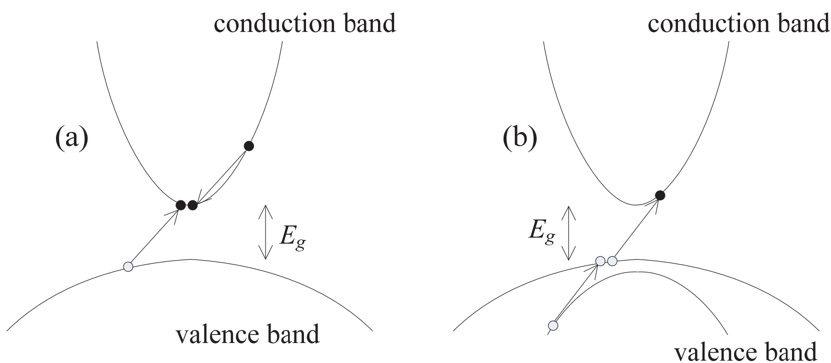


Fig. 1. Schematic wavevector diagrams depicting (a) electron-initiated and (b) hole-initiated impact ionisation events.

Due to conservation of energy and momentum, a threshold energy, E_{th}, prerequisite has to be satisfied by the primary carrier. This energy has to be greater than the band gap, E_g, as the carrier also experiences non-ionising collision processes such as phonon scattering, which involves carriers gaining energy, losing energy or exchanging momentum. On average, carriers will lose energy by phonon scattering because the emissive phonon scattering rate is proportional to n_p+1 whereas the phonon absorption rate is proportional to n_p, where n_p is the phonon occupation number, which depends on the phonon energy, $\hbar\omega$, given by $n_p = \frac{1}{\exp\left(\frac{\hbar\omega}{k_B T}\right) - 1}$, where k_B is Boltzmann's constant and T is the absolute temperature.

The generation rate or mean number of ionisation events per unit distance for a carrier is known as the impact ionisation coefficient. The electron and hole ionisation coefficients, α and β respectively, are functions of electric field, temperature and material.

Carriers with energy less than E_{th} are unable to initiate impact ionisation and have to traverse a distance, within the high electric field region, known as the dead space before they acquire sufficient energy. A carrier that has gained E_{th} is said to be *enabled*, as its ionisation probability is no longer zero.

The mean multiplication factor, M, or gain is the ratio of the total number of carriers generated to the number of carriers injected. In electrical current terms, this is given by $M = I_p/I_i$, where I_p is the generated output photocurrent (where carrier multiplication occurs) and I_i is the initial photocurrent (before carrier multiplication). M can be calculated using the local model (Stillman and Wolfe, 1977) where the multiplication layer width is assumed to be much greater than dead space. Neglecting dead space and solving electron and hole continuity current equations in the multiplication layer, M is given by

$$M(x) = \frac{\exp\left[-\int_0^x \left(\alpha(x') - \beta(x')\right)\, dx'\right]}{1 - \int_0^w \alpha(x') \exp\left[-\int_0^{x'} \left(\alpha(x'') - \beta(x'')\right) dx''\right] dx'} \tag{1}$$

where α and β are position-dependent ionisation coefficients, and electrons are injected from $x = 0$ and holes from $x = w$, i.e. electrons drift in the positive x direction, holes otherwise.

Assuming a uniform electric field, i.e. an ideal *p-i-n* diode, α and β have no spatial dependence and (1) simplifies to

$$M(x) = \frac{(\alpha - \beta)e^{-(\alpha-\beta)x}}{\alpha e^{-(\alpha-\beta)w} - \beta} \tag{2}$$

Hence, pure electron mean multiplication factors, M_e and M_h, are given by

$$M_e = \frac{(\alpha - \beta)}{\alpha e^{-(\alpha-\beta)w} - \beta} \tag{3}$$

and

$$M_h = \frac{(\alpha - \beta)e^{-(\alpha-\beta)w}}{\alpha e^{-(\alpha-\beta)w} - \beta} \quad (4)$$

Rearranging equations #, a and β can be determined by measuring M_e and M_h in ideal *p*-*i*-*n* structures, based on the simplified assumptions outlined above, as

$$\alpha = \frac{1}{w}\left[\frac{M_e - 1}{M_e - M_h}\right]\ln\left(\frac{M_e}{M_h}\right) \quad (5)$$

and

$$\beta = \frac{1}{w}\left[\frac{M_h - 1}{M_h - M_e}\right]\ln\left(\frac{M_h}{M_e}\right) \quad (6)$$

3. Excess noise

The stochastic nature of the impact ionisation process results in fluctuations in the multiplication factor. This noise, introduced by impact ionisation, is caused by the unpredictability in the production position of the secondary carrier.
For an APD under illumination, assuming the incident photons have a Poisson distribution generating a primary photocurrent, i_{pr}, in a circuit of bandwidth, B, the mean number of photogenerated carriers is given by

$$\langle m \rangle = \eta \phi T_c \quad (7)$$

where η is the quantum efficiency, ϕ is the photon flux in photons per second, and T_c is the collection time interval.
For a measurement circuit with bandwidth B, the minimum distinguishable time interval between received current pulses can be defined by the Nyquist criterion as $T_c = 1/(2B)$. Hence, the total current collected in time interval, T_c, and the associated variance, are given by

$$\langle i_{pr} \rangle = \frac{e\langle m \rangle}{T_c} \quad (8)$$

and

$$\sigma_{pr}^2 = \left(\frac{e}{T_c}\right)^2 \sigma_{\langle m \rangle}^2 \quad (9)$$

where σ_{pr}^2 is the variance in photocurrent, $\sigma_{\langle m \rangle}^2$ is the variance in number of photogenerated carriers, e is the unit of electron charge. From (7) and (8), the mean photocurrent is, therefore, given by

$$\langle i_{pr} \rangle = e\eta\phi \quad (10)$$

and from (8) and (9), noting that $\langle m \rangle = \sigma^2_{\langle m \rangle}$ for a Poisson distribution of photons, the variance in photocurrent is given by

$$\sigma^2_{pr} = 2e\left\langle i_{pr} \right\rangle B \tag{11}$$

These simplified derivations show that even without avalanche gain, variance in the photocurrent is expected due to the random nature of the photocurrent generation. Note, also, that (11) is identical to the shot noise formula for variance in a current.
Hence, for an APD considered as an *ideal noiseless multiplier* with multiplication, $\langle M \rangle$, the mean photocurrent, $\left\langle i_{ph} \right\rangle$, is given by

$$\left\langle i_{ph} \right\rangle = \left\langle i_{pr} \right\rangle \langle M \rangle \tag{12}$$

and the mean square noise current is given by

$$N_{ideal} = 2e\left\langle i_{pr} \right\rangle \langle M \rangle^2 B \tag{13}$$

Equation (13) describes the ideal (noiseless) multiplication process, where the stochastic nature of the avalanche multiplication process is excluded. To account for the noise associated with the multiplication process, the excess noise factor, F, is introduced into (13), giving

$$N = 2e\left\langle i_{pr} \right\rangle \langle M \rangle^2 BF \tag{14}$$

where F is expressed as

$$F = \frac{\left\langle M^2 \right\rangle}{\langle M \rangle^2} \tag{15}$$

Equation (15) shows that the average multiplication, $\langle M \rangle$, has statistical fluctuations and F in (14) describes how much the avalanche noise deviates from an ideal multiplier. When there is no multiplication noise, F = 1 and only shot noise exists. Hence, F permits the noise performance of APDs to be considered in the same terms as that of other system components.

4. The Random Path Length model

Unlike the local model described earlier, non-local models account for the dead space and one such model is the Random Path Length (RPL) model (Ong et al., 1998). The RPL model is a simple model that is able to predict multiplication and excess noise characteristics in APDs by modelling the transport of carriers during the impact ionisation process. The model operates by consideration of the ionisation path length probability distribution function, $P(x)$, for each carrier as it passes through the device. For the *hard threshold dead space model*, which is considered here, the probability for an electron to impact ionise for the first time after travelling a distance x in a uniform electric field, E, is given by

$$P_e(x) = \begin{cases} 0 & , x \le d_e^* \\ \alpha^* \exp\left[-\alpha^*(x - d_e^*)\right] & , x > d_e^* \end{cases} \quad (16)$$

where a^* is the enabled electron ionisation coefficient and d_e^* is the electron hard threshold dead space, given by

$$d_e^* = \frac{E_{the}}{q\Im} \quad (17)$$

and E_{the} is the electron ionisation threshold energy, q is the electron charge and $\Im$ is the applied electric field. From (16), the average distance between electron initiated ionising collisions is

$$\int_0^\infty x P_e(x)\, \mathrm{d}x = \frac{1}{\alpha^*} + d_e^* \quad (18)$$

and the mean ionisation coefficient is the reciprocal of this, that is

$$\alpha = 1 \Big/ \left(d_e^* + \frac{1}{\alpha^*} \right) \quad (19)$$

From (16), the probability that a carrier travels a distance x without impact ionising is

$$S_e(x) = \begin{cases} 1 & , x \le d_e^* \\ \exp\left[-\alpha^*(x - d_e^*)\right] & , x > d_e^* \end{cases} \quad (20)$$

Thus, a random electron ionisation path length, l_e, can be expressed by substituting uniformly distributed numbers, r, between 0 and 1 for $S_e(x)$ to give

$$l_e = d_e^* - \frac{\ln(r)}{\alpha^*} \quad (21)$$

Similar expressions for the hole impact ionisation path length can be obtained by substituting $P_e(x)$, $S_e(x)$, α , α^* , d_e^* and l_e with $P_h(x)$, $S_h(x)$, β , β^* , d_h^* and l_h, in (16)–(21).
The RPL simulation is composed of n number of trials, where the choice of n is a trade-off between accuracy and computation time. A trial in the RPL simulation is complete when all the carriers have left the multiplication region. Each injected carrier gives rise to a multiplication value, m, which is a random variable due to the stochastic nature of the impact ionisation process. The mean multiplication and excess noise factor can be calculated using

$$M = \frac{1}{n} \sum_{i=1}^{n} m_i \quad (22)$$

and

$$F = \frac{1}{nM^2} \sum_{i=1}^{n} \left(m_i^{\,2} \right) \quad (23)$$

where m_i is the multiplication resulting from trial i in the RPL simulation.

5. BER model

The current developed in an APD, by chains of impact ionisation events, take time to build up. Materials with disparate ionisation coefficients tend to have longer chains of ionisation events; thus having longer current buildup times compared to currents developed by shorter chains. This buildup time, which is stochastic, has an associated bandwidth limit and thus governs the APD speed and ultimately, the level of ISI in the receiver system.

To understand the stochastic nature of the APD buildup-time-limited bandwidth and its statistical correlation with the gain, Sun *et al.* introduced the *shot-noise equivalent bandwidth* (Sun et al., 2006), defined as $B_{\text{sneq}} = \langle M^2/T_{bu}\rangle / 2\langle M\rangle^2 F$, where T_{bu} is the avalanche buildup time. The quantity B_{sneq} is the bandwidth that, when used in the usual formula for APD-amplified shot noise, $\sigma^2 = 2e\langle M\rangle^2 F B_{\text{sneq}}\eta P/h\nu$, gives the correct value of the shot-noise variance, where η is the APD quantum efficiency, P is the optical power, h is Planck's constant, and ν is the photon's frequency. Due to the stochastic coupling between T_{bu} and M, B_{sneq} is generally greater than the conventional 3dB bandwidth of the APD, $B_{3\text{dB}}$, which is taken as the 3dB-drop point in the Fourier transform of the APD's mean impulse-response function. This discrepancy can be as high as 30%, leading to a similar error in the prediction of the APD-amplified shot-noise variance if $B_{3\text{dB}}$ is used in place of B_{sneq}.

The Gaussian-approximation method was used to calculate the BER and is described as follows. The output of the integrate-and-dump receiver was approximated by a Gaussian random variable with the exact mean and variance, and the BER was computed using (Agrawal, 1997)

$$\text{BER} \approx \frac{1}{4}\left[\text{erfc}\left(\frac{\theta-\mu_0}{\sqrt{2}\sigma_0}\right)+\text{erfc}\left(\frac{\mu_1-\theta}{\sqrt{2}\sigma_1}\right)\right] \tag{24}$$

where μ_0 and σ_0^2 denote the mean and variance for the receiver's output conditional on the present bit (i.e., the information bit corresponding to the receiver's present integration period) being '0,' and μ_1 and σ_1^2 are similar quantities conditional on the present bit being '1.' The decision threshold, θ, is taken as

$$\theta = \frac{\mu_0\sigma_1+\mu_1\sigma_0}{\sigma_0+\sigma_1} \tag{25}$$

which is a convenient approximation to the optimal decision threshold that minimises the BER (Agrawal, 1997). The expressions for the parameters μ_0, σ_0^2, μ_1 and σ_1^2 are derived as (Sun et al., 2006)

$$\mu_0 = \frac{1}{2}\frac{n_0\langle M\rangle}{\kappa\lambda}\left(1-e^{-\kappa\lambda}\right) \tag{26}$$

$$\sigma_0^2 = \frac{1}{4}\frac{n_0^2\langle M\rangle^2}{\kappa\lambda^2}\frac{\left(1-e^{-\kappa\lambda}\right)^4}{\left(1-e^{-2\kappa\lambda}\right)}+\frac{n_0\langle M\rangle^2 F}{2\kappa\lambda}\left(1-e^{-\kappa\lambda}-\kappa\lambda e^{-\kappa\lambda}\right)+\sigma_J^2 \tag{27}$$

$$\mu_1 = \mu_0 + \frac{n_0 \langle M \rangle}{\kappa\lambda}\left(\kappa\lambda - 1 + e^{-\kappa\lambda}\right) \quad (28)$$

$$\sigma_1^2 = \sigma_0^2 + \frac{n_0 \langle M \rangle^2 F}{\kappa\lambda}\left(\kappa\lambda - 2 + 2e^{-\kappa\lambda} + \kappa\lambda e^{-\kappa\lambda}\right) \quad (29)$$

where n_0 is the average number of absorbed photons per '1' bit, $\kappa = 4B_{sneq}/2\pi B_{3dB}$ is the *bandwidth correction factor*, which accounts for the discrepancy between B_{sneq} and B_{3dB}, and λ, is the *detector-speed factor*, which is a measure of the detector's relative speed, defined as $\lambda = 2\pi B_{3dB}/R_b$, where $R_b = 1/T_b$ is the bit transmission speed, and T_b is the bit duration. Finally, the term σ_J^2 represents the variance of Johnson noise accumulated in the integration time. Note that μ_0, σ_0^2, μ_1 and σ_1^2 are quantities that are averaged over all possible past bit patterns.

The expressions in (26)–(29) are generalisations of the traditional expressions for the output statistics of APD-based receivers found in optical communication literature (Agrawal, 1997). Whilst these expressions capture the usual effects of shot noise and the excess noise due to avalanche multiplication, they additionally capture the effects of ISI, relative speed of the detector, as well as the stochastic coupling between the APD's gain and buildup time through the effective use of the parameters κ and λ. For an instantaneous detector, i.e. $B_{3dB} = \infty$, the detector-speed factor λ is infinite, and the expressions shown in (26)–(29) collapse to the traditional expressions for the receiver mean and variance in the absence of ISI (Agrawal, 1997): $\mu_0 = 0$, $\sigma_0^2 = \sigma_J$, $\mu_1 = n_0 \langle M \rangle$, and $\sigma_0^2 = \sigma_J + n_0 \langle M \rangle^2 F$. Moreover, in detectors for which the gain is unity, e.g. a *p-i-n* diode, the bandwidth correction factor κ is unity, resulting in simplified versions of (26)–(29) that continue to capture the effect of ISI.

The term in μ_0 and the first two terms in σ_0^2, as shown in (26) and (27), respectively, are due entirely to contributions from ISI resulting from the random stream of bits (preceding the present bit). In particular, they arise from contributions from photo-generated carriers generated within the bits that precede the present bit. The second term in (28) and second term in (29) are due to contributions from carriers generated during the present bit. The parameters μ_0, σ_0^2, μ_1 and σ_1^2 shown above do not include the contribution from the APD's tunnelling current generated in the multiplication region of the APD.

5.1 Variant of the Gaussian-approximation

In the previous section, it was assumed that the receiver output, conditional on the state of the present bit, is a Gaussian random variable. Ong et al. (Ong et al., 2009) replaced this with the more realistic assumption that the receiver output, conditional on the state of the present bit *and the entire past bit stream*, is a Gaussian random variable. This enables one to compute the BER conditional on the entire past bit stream, and then average the resulting pattern-specific BERs over all possible past bit patterns to obtain the overall average BER. The advantage of this approach is that it relaxes the often unrealistic assumption of a uni-modal probability density function for the receiver output conditional on the state of the present bit. While this variant approximation will yield an improved approximation of the average BER, it comes with a slight increase in computation.

The mean and variance of the ISI contributions in the receiver output from the kth past bit alone are respectively given by

$$\mu_{\mathrm{ISI},k} = \frac{2n_0 \langle M \rangle \mathrm{e}^{-\kappa\lambda k}}{\kappa\lambda} \left(\cosh(\kappa\lambda) - 1\right) \tag{30}$$

and

$$\sigma^2_{\mathrm{ISI},k} = \frac{n_0 \langle M \rangle^2 F}{\kappa\lambda} \mathrm{e}^{-\kappa\lambda k} \left(\mathrm{e}^{-\kappa\lambda} - 1\right)\left(1 - \kappa\lambda \mathrm{e}^{-\kappa\lambda} - \mathrm{e}^{-\kappa\lambda}\right) \tag{31}$$

Considering an arbitrary past bit pattern, I_j, of length L bits, the mean of the receiver output when the current bit is '0' can be calculated by adding up the contributions from each of the ISI terms from the past bits in the pattern I_j; which yields the expression

$$\mu_0\left(I_j\right) = \sum_{k=1}^{L} a_k\left(I_j\right) \mu_{\mathrm{ISI},k} \tag{32}$$

where $a_k(I_j) = 0$ unless the kth bit in the pattern I_j is a '1' bit, in which case $a_k(I_j) = 1$.
Similarly, the variance of the receiver output associated with the pattern I_j when the current bit is '0' can be calculated by adding up the ISI contributions from past bits in the specific pattern as well as contribution from Johnson noise to obtain

$$\sigma_0^2\left(I_j\right) = \sum_{k=1}^{L} a_k\left(I_j\right) \sigma^2_{\mathrm{ISI},k} + \sigma_J^2 \tag{33}$$

5.2 Generalisation of the model to include tunnelling current

The un-multiplied band-to-band tunnelling current, I_{tunn}, is modelled by (Forrest et al., 1980b)

$$I_{tunn} = \frac{\left(2m^*\right)^{0.5} q^3 \Im V A}{h^2 E_g^{\,0.5}} \exp\left(-\frac{2\pi\sigma_T \left(m^*\right)^{0.5} E_g^{\,1.5}}{qh\Im} \right) \tag{34}$$

where m^* is the effective electron mass, V is the applied reverse bias voltage, A is the device area, and σ_T is the tunnelling fitting parameter. The average number of dark carriers generated per bit time interval, T_b, is given by $n_d = I_{tunn}T_b/q$. Since the dark-carrier generation has Poisson statistics, it is plausible to attempt to include the effect of dark carriers on the parameters μ_0, σ_0^2, μ_1 and σ_1^2 by treating dark carriers as photo-generated carriers.
The expression for $\mu_0(I_j)$ including dark carrier generation was obtained as

$$\begin{aligned} \mu_0\left(I_j\right) &= \sum_{k=1}^{L} a_k\left(I_j\right) \mu_{\mathrm{ISI},k} + \frac{n_d \langle M \rangle}{\kappa\lambda}\left(1 - \mathrm{e}^{-\kappa\lambda}\right) + \frac{n_d \langle M \rangle}{\kappa\lambda}\left(\kappa\lambda - 1 + \mathrm{e}^{-\kappa\lambda}\right) \\ &= \sum_{k=1}^{L} a_k\left(I_j\right) \mu_{\mathrm{ISI},k} + n_d \langle M \rangle \end{aligned} \tag{35}$$

while the expression for $\sigma_0^2(I_j)$ was derived as

$$\sigma_0^2\left(I_j\right)=\sum_{k=1}^{L} a_k\left(I_j\right)\sigma_{\mathrm{ISI},k}^2+\frac{n_d\langle M\rangle^2 F}{\kappa\lambda}\left(\kappa\lambda-1+\mathrm{e}^{-\kappa\lambda}\right)+\sigma_J^2 \tag{36}$$

The receiver mean output, $\mu_1(I_j)$, when the present bit is '1' is obtained by adding to $\mu_0(I_j)$ the contributions from the photons in the current bit. Combining this component with the contributions from the ISI terms gives

$$\mu_1\left(I_j\right)=\mu_0\left(I_j\right)+\frac{n_0\langle M\rangle}{\kappa\lambda}\left(\kappa\lambda-1+\mathrm{e}^{-\kappa\lambda}\right) \tag{37}$$

Similarly, adding the contributions from the photons in the current bit to $\sigma_0^2(I_j)$ yields

$$\sigma_1^2\left(I_j\right)=\sigma_0^2\left(I_j\right)+\frac{n_0\langle M\rangle^2 F}{\kappa\lambda}\left(\kappa\lambda-2+2\mathrm{e}^{-\kappa\lambda}+\kappa\lambda\mathrm{e}^{-\kappa\lambda}\right) \tag{38}$$

Finally, for every past bit pattern I_j, $j = 1, \ldots, 2^L$, the *pattern-specific* BER can be calculated as follows

$$\mathrm{BER}\left(I_j\right)\approx\frac{1}{4}\left[\operatorname{erfc}\left(\frac{\theta-\mu_0\left(I_j\right)}{\sqrt{2}\sigma_0\left(I_j\right)}\right)+\operatorname{erfc}\left(\frac{\mu_1\left(I_j\right)-\theta}{\sqrt{2}\sigma_1\left(I_j\right)}\right)\right] \tag{39}$$

where θ is calculated as before from (25). To calculate the overall BER, the ensemble average of the pattern-specific BER is computed over all possible past bit patterns; more precisely,

$$\mathrm{BER}=\frac{1}{2^L}\sum_{j=1}^{2^L}\mathrm{BER}\left(I_j\right) \tag{40}$$

The bit-length parameter, L, can be chosen to be sufficiently large to capture all significant ISI terms; in this work, $L = 10$ was found to be an adequate choice beyond which no significant change in the BER was observed.

6. Measurement of excess noise: A survey

Estimation of the BER, which can be obtained from a certain diode fabricated from a particular material system, relies on the characterisation of the excess noise factor using one or more test structures. The test structure is often a *p-i-n* avalanche photodiode into which light is injected through one of the contacts. To permit this injection, the shape of that contact is often an annulus. Generally many such characterisation diodes are fabricated simultaneously using standard optical lithographic and etching techniques. The excess noise power as a function of multiplication is measured for many values of multiplication such that the value of McIntyre's k_{eff} (McIntyre, 1966) for a particular material system can be established. This value can then be used in conjunction with the methods described herein and noise analyses of the front end receiver circuits to estimate the BER for a specified optical irradiance. In this section we consider each of the noise measurement systems reported in the literature and offer two suggestions for an improved system.

Several excess noise measurement systems have been reported in the literature and comparisons between the circuits described may be drawn. The figures of comparison are,

- The system signal to noise ratio, where the signal is defined as full shot noise exhibited by 1μA.
- The maximum permissible APD junction capacitance. In the case of multi-frequency systems the lowest available frequency is used; this produces the most favourable result. It is assumed that the system input impedance and diode junction capacitance form a first order low pass network.

6.1 Measurement systems

The first report of noise measurements on photodiodes was by Baertsch (1966). Insufficient information is provided to estimate this system's figures of merit so it is excluded from the comparison. Xie et al. (1993) proposed a measurement system that was substantially similar to Toivonen et al. (1992). The Xie et al. (1993) system represents both. Bulman (1983), Ando and Kanbe (1981) and Lau et al. (2006) presented systems based on phase sensitive detection. Xie et al. (1993) and Toivonen et al. (1992) used a DC approach.

6.1.1 Bulman’s system

Figure 2 shows the system reported by Bulman (1983). It is a PSD system in which photocurrent and excess noise are extracted and read out by two lock in amplifiers. The APD is loaded by the AIL13680 through the bias tee. It may be assumed that the APD experiences a 50Ω load. The preamplifier output is fed to a receiver having a calibrated bandwidth. The resulting signal proportional to the noise power contained within the calibrated bandwidth is passed to a lock-in-amplifier.

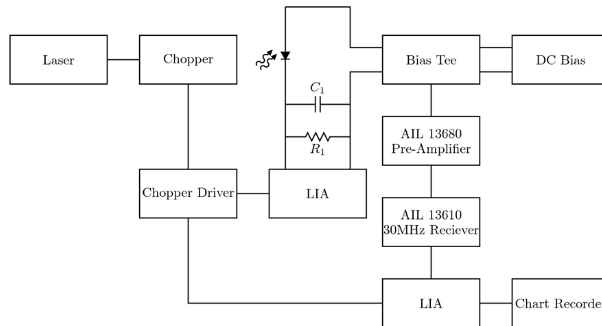


Fig. 2. Bulman's excess noise measurement system

Bulman proposes two methods to quantify the absolute noise power measured. Firstly a *p-i-n* detector is illuminated under unity gain conditions. It is assumed that under these conditions the system will measure full shot noise. A second calibration method is proposed in which a calibrated oscillator is used in place of the APD. This allows the experimenter to set the power which will be measured. Adjusting the power supplied by the oscillator allows the linear range of the system to be estimated. Bulman reports a 30dB range of linear measurement between -140dBm (10^{-17} W) and -110dBm (10^{-14} W). In several III-V semiconductors the impact ionisation coefficients are nearly equal. Under this assumption,

Tager (1965) has shown that excess noise is proportional to M^3. Assuming Bulman's system is used from unity gain with a noise power of -140dBm the maximum multiplication prior to the limit of linearity in noise measurement is ten.

Bulman's report lacks some information regarding the front end amplifier. An Analog Devices AD9618 low noise opamp in non-inverting mode is used as a model. It achieves a gain of 100V/V and a bandwidth of 80MHz with 50Ω input impedance. The equivalent input noise voltage is 1.94 nV/Hz^{½}. Using this model as an approximation the signal-to-noise ratio for Bulman's system is $\mathrm{SNR} = \frac{(100\times 50)^2 \times 2qi_{ph}}{(1.94\times 10^{-9}\times 100)^2} = -36.73\ \mathrm{dB}$. When i_{ph} = 1μA.

The capacitance of the diode which can be tolerated by Bulman's system is given to a first approximation by considering the system input impedance and the junction capacitance as a first order RC network. Assuming that all other parasitic effects (for example diode series resistance) are negligible in comparison, applying the commonly known expression for a first order network yields, $C = \frac{1}{2\pi f R}$, where f = 30MHz and R = 50 ohms. C = 106pF.

6.1.2 The PSD system of Ando and Kanbe

Ando and Kanbe (1981) reported the system shown in Figure 3. It is a PSD system in which the APD is loaded by 50Ω due to the input impedance of the IF amplifier. The measurement system bandwidth is defined by the IF amplifier and is 1MHz centred on 30MHz. The APD is biased using a bias tee and a variable voltage source. The noise power is read out from a lock in amplifier. A power meter and signal generator, with its output passed through a calibrated attenuator, provides a means of relating the absolute signal power to the value measured leaving the IF amplifier. Photocurrent is extracted by DC measurement of the current entering the bias tee. The measurement of devices exhibiting high dark current is therefore difficult with this system. The various connections required to calibrate the system and perform measurements are made using relays.

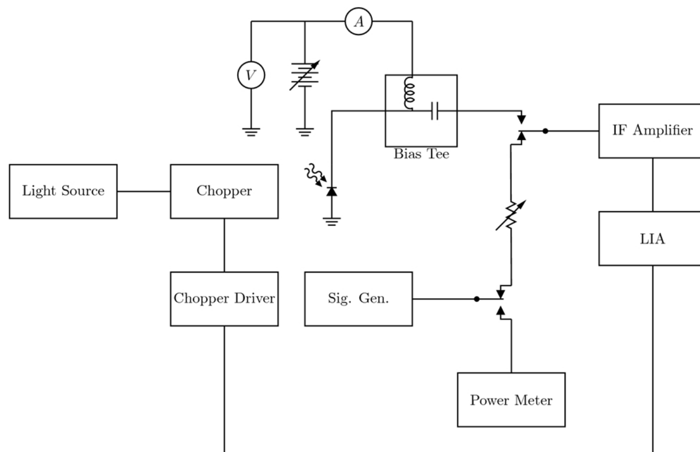


Fig. 3. The PSD noise measurement system due to Ando and Kanbe

Ando and Kanbe did not report any attempts to measure shot noise on their system. They also do not give information regarding the model numbers or manufactures of their system

components. No noise specifications for the instrumentation are given. Assuming that their system adds no noise other than the thermal noise of the 50Ω input impedance within the measurement bandwidth then the signal-to-noise ratio can be computed using $\mathrm{SNR} = \frac{2qi_{ph}BR_{in}}{4k_BT\, B/R_{in}}$, where R_{in}= 50Ω, B = 1MHz, T = 300°K and i_{ph} = 1μA. The junction capacitance which can be tolerated by Ando and Kanbe's system is calculated in a similar way to Bulman's system and produces the same answer C = 106pF.

The authors claim that noise power as low as -130dBm/Hz can be measured with 0.5dB accuracy. This represents a current of 0.125μA developing full shot noise.

6.1.3 A measurement after Xie et al.

The system proposed by Xie et al. (1993) is similar to that proposed by Toivonen et al. (Toivonen et al., 1992). The APD is connected to a micro-strip line and DC voltage is applied via a bias tee.

The measurement is made using a CW light source and a noise figure meter such as the Hewlett Packard 8970A. The system has two significant advantages over PSD systems such as those of Bulman (1983) and Li (Lau et al., 2006). Several measurement frequencies are available up to the limit of the circuits or analyser. Presently Agilent Technologies manufactures noise figure meters capable of measuring 10MHz to 26GHz with variable effective measurement bandwidth. This upper limit can be increased by using heterodyne methods. Xie's system (Xie et al., 1993) was limited to 1.3GHz maximum measurement frequency and 4MHz noise measurement bandwidth. The measurement is, in principle, quicker than a PSD system. The operation of PSD is discussed fully elsewhere (Horowitz and Hill, 1989) but it is sufficient to realise that the time constant of a PSD measurement may be expected to be longer than of a noise figure meter. DC measurements have several disadvantages over PSD however. For example the lowest practically measurable photo-generated noise is higher in CW systems than in some PSD systems. Using a transimpedance amplifier, Li (Li, 1999, Li et al., 1998) has shown that the transimpedance amplifier reported by Lau et al. (2006) can be used as the basis of a noise measuring system with greater (less negative) noise signal to noise ratio than is possible by using a 50Ω measurement system. A further objection to CW systems is that the noise without illumination – the dark noise - should be periodically measured in order to maintain consistency. The dark noise should be stable and sufficiently small, compared to the noise with illumination – combined light and dark noise – that the noise with illumination is dominated by the light noise. If this condition is not met the confidence of the measurement is compromised. Xie et al. (1993) reported measuring noise power as low as -182dbm/Hz without difficulty using the CW system shown in Figure 4. In a 50Ω system -182dbm/Hz is equivalent to full shot noise generated by 8μA of photocurrent. The capacitance which can be tolerated by this measurement system is computed at the lowest useable frequency, as this produces the most favourable result. By the same first order approximation used in Bulman's and Ando and Kanbe's systems Xie's system will exhibit a -3dB (half power) bandwidth of 10MHz when loaded with 636pF.

6.1.4 A PSD system after Li et al.

The system of Li (Lau et al., 2006, Li, 1999) employs phase sensitive detection and a transimpedance amplifier. A schematic diagram is shown in Figure 5.

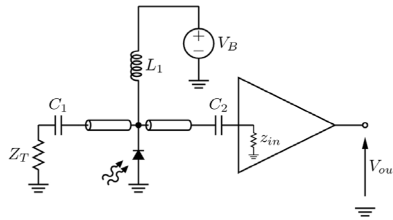


Fig. 4. CW excess noise measurement system after Xie et al.

The laser is chopped by mechanical means at 180Hz and is presented to the diode via a system of optics which is not shown. The TIA is used to convert the diode current into a voltage. This voltage is amplified using a commercial low noise wide band amplifier module (Minicircuits ZFL-500). A precision stepped attenuator (HP355D) is used to vary the system gain permitting measurement of high and low noise devices. The noise signal is separated from the low frequency component of the photocurrent by a Minicircuits SBP-10.7+ LC ladder filter which also defines the noise measurement bandwidth. After filtration, the signal resembles an amplitude modulated noise waveform, where periods of diode illumination produce greater noise amplitude than periods of darkness. Further amplification follows, prior to a wide band squaring and averaging circuit. The output of the squaring and averaging circuit is an approximately square voltage signal, the amplitude of which is proportional to the noise power contained in the measurement bandwidth. The fundamental frequency of the noise power signal is 180Hz. The squaring circuit is based on an Analogue Devices AD835 analogue multiplier. The averaging circuit is a first order RC filter with a time constant of approximately 100µs. The output from the squaring and averaging circuit is measured using a lock-in-amplifier. The photocurrent signal is taken from an auxiliary output of the TIA where the amplitude of the 180Hz square wave is proportional to the photocurrent. The photocurrent signal is measured on a second lock-in-amplifier.

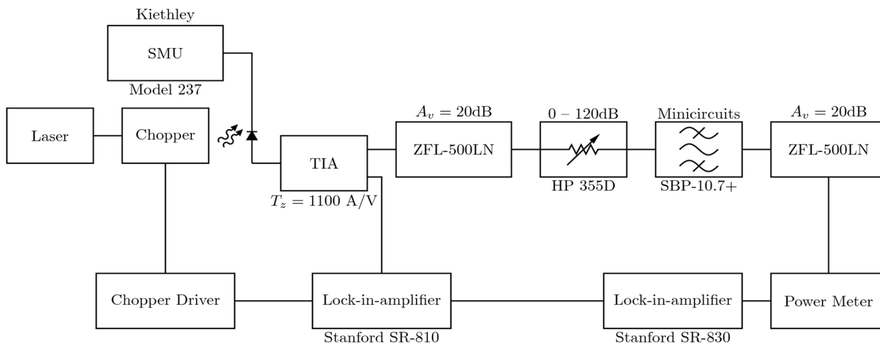


Fig. 5. Schematic diagram of an excess noise measurement system after Li

The system after Li (Lau et al., 2006, Li, 1999) is superior in noise performance to prior reported systems. The transimpedance amplifier provides a signal to noise ratio which is superior to that possible in a 50Ω system. Consider the connection of a photodiode and a 50Ω resistor. Assume that full shot noise generated by i_{ph} = 1µA flows through the resistor

which exhibits thermal noise at T = 300°K. The noise signal to noise ratio is then, $\mathrm{NSNR} = 20\log_{10}\frac{50\sqrt{2qi_{ph}}}{\sqrt{50\times 4k_B T}} = -30.15\ \mathrm{dB}$. The noise signal to noise ratio (also considering 1μA photocurrent) of Li's system is -25.7dB (Li, 1999).

The dynamic range of Li's system is limited at the lower bound by the ability of the lock in amplifier to extract the in-phase excess noise signal from the system's background noise. Practical experimentation by the authors and their colleagues has shown that full shot noise developed by 1μA is approaching the limit and the shot noise from 0.1μA is not reliably measurable. The precise limit is difficult to quantify because it is affected by the prevailing electromagnetic conditions both radiated (passing through the experiment volume) and conducted into the power supply lines. At the upper bound the maximum attenuation of the stepped attenuator provides a limitation however more attenuation could be added without difficulty. The linearity of the transimpedance amplifier at high input current is a second limit. When driven from +/-5V supplies a TIA with a gain of 2200V/A will saturate at approximately 2.25mA input current. Because the relationship between excess noise factor and photo-multiplication varies between material systems it is unwise to speculate the maximum multiplication which can be used. Furthermore if a device is available which can be operated with a very large gain the optical illumination may be reduced in order to reduce the multiplied photocurrent and the excess noise power. In this way higher multiplication values may be measured. In order to measure lower multiplication values a larger primary photocurrent is required. By performing two or more measurements with differing primary photocurrents it is possible, assuming the APD is sufficiently robust, to measure multiplication and excess noise power over any desirable range above the system limit.

The capacitance tolerated by Li's transimpedance amplifier (Lau et al., 2006, Li, 1999) is lower than all of the other systems. The interaction of the APD junction capacitance and the feedback capacitor permits the existence of resonance in the transimpedance amplifier. When the capacitance is sufficiently large oscillation breaks out and the measurement system is saturated. There limit of measureable junction capacitance is however not governed by the presence of oscillation. A result of the interaction of the diode junction capacitance and the feedback capacitance is a dependence of the effective noise power bandwidth of the system on the diode junction capacitance, which is itself dependant on the DC bias voltage applied to the APD. As a result a correction to the measurement bandwidth must be made when processing the measurement data. The limitation of the measurable device capacitance is governed by the quality of the correction which can be achieved and by the presence of oscillation. While it is known that up to 56pF does not cause oscillation, Li placed the limit at 28pF (Li, 1999). This limit was obtained by calibrating the bandwidth of the transimpedance amplifier with several values of capacitance. Having performed the calibration, shot noise due to photo-generated carriers was measured using a unity-gain silicon photodiode. A second data set was gathered in which extra capacitance was placed in parallel with the photodiode to simulate a diode of greater capacitance. The simulated higher capacitance shot noise data was processed using the original calibration. The quality of the fitting of the standard photodiode shot noise and the simulated extra capacitance shot noise data was used as a basis for defining the quality of the correction and hence the maximum capacitance.

6.2 An improved CW noise measurement

We propose two possible improvements to the design proposed by Xie et al. (1993). Both are essentially improvements to the method by which the instrumentation is calibrated. The introduction of a calibrated noise source (HP346B) permits the use of direct noise figure measurement – as opposed to hot/cold measurements, which is a considerable improvement. The noise figure meter (N8973A, or an older model such as the N8970) is designed such that the noise source is connected to the device (for example an LNB) under test. Of course if the device is an electro-optical transducer this is impossible as there is no place to attach the noise source. This leads to the use of a pre-test calibration followed by hot/cold measurements. It would be preferable to use the noise figure analyser (NFA) according to its design principle, i.e. with the noise source in the measurement. The NFA is provided with prior calibration - by the manufacturer - of the noise source's contribution to the system. The system gain is also computable by measuring the effect on the noise output when the noise source is switched on and off - it is pulsed by the NFA. The time average of the change in noise level can provide the gain from the noise input port to the NFA input port. The prior knowledge of the known noise input from the calibrated source (HP346B) allows the NFA to compute the gain and noise figure nearly instantly, a considerable improvement in measurement speed, accuracy and precision. The question is then "How can the noise source be applied to the APD?" It cannot be directly applied. However, a secondary port can be created which permits the connection of an APD and the noise source to the NFA simultaneously. We provide two example designs here, the first uses a 50Ω matched topology similar to that of Xie et al. (1993). The second describes a similar overall structure but using a commercial transimpedance amplifier.
The APD multiplication, excess noise factor and noise power bandwidth can be established simultaneously in one measurement. The limitation of the system bandwidth can be alleviated by two methods. Firstly a higher maximum frequency noise figure meter can be obtained. Agilent Technologies presently manufactures noise figure meters/analysers capable of directly measuring up to 26GHz. The use of heterodyne techniques could extend this considerably. However a relatively inexpensive alternative is to use a lower bandwidth noise figure meter but begin measuring bandwidth once the APD has been biased to achieve a high gain. The high frequency roll off due to a finite gain bandwidth product can be observed at lower frequencies; the unity noise gain bandwidth product can then be inferred. The importance of correct impedance matching cannot be overemphasized.

6.2.1 50Ω system

The system diagram in Figure 6 shows the structure of the measurement setup. A Source-Measure Unit[1] drives a bias tee composed of L_1 and C_1. An example of a suitable tee is the PicoSecond Model 5541A. The APD is connected to a microwave DC block (C_1) and this is in turn connected to a termination (50Ω). The DC block and the termination must be electrically close to the APD even at the highest measurement frequency. It is preferable to fabricate the DC block and the 50Ω termination with the APD as an integrated circuit. From the point of view of the first amplifier the APD is a Norton source coupled to the end of a properly terminated transmission line. Approximately half of the noise power will escape to ground via R_1, the rest will enter the measurement system. It is possible to calibrate the

[1] A precision voltage source and current measuring device, e.g. Keithley models 237, 2400 and 2612

measurement system either manually (i.e. use a 50Ω signal generator to list a table of adjustments for each frequency and post process the measured device data based on these reading) or automatically by using the HP 346B Noise source connected to the first amplifier input instead of the APD. The attenuator setting must be noted down when the calibration is carried out. The first amplifier in the chain must be of the lowest possible noise. Examples include Minicircuits ZFL-1000LN+, ZX60-33LN+ and Pasternack PE1513. The ZFL-1000 has low noise and a reasonably flat gain vs. frequency profile from 100kHz to 1GHz however bandwidth is limited to 1GHz. The ZX60-33LN+ has exceptionally low noise, and reasonable gain vs. frequency characteristics from 50MHz to 3GHz. The PE1513 has relatively poor noise especially as frequency increases, the gain vs. frequency profile is not ideal either; however it is the only device which covers the whole frequency range of the NFA, which is 3 GHz in the case of the N8973A. Unless APDs possessing bandwidths below 50MHz are to be routinely measured the authors preferred choice is the ZX60-33LN.

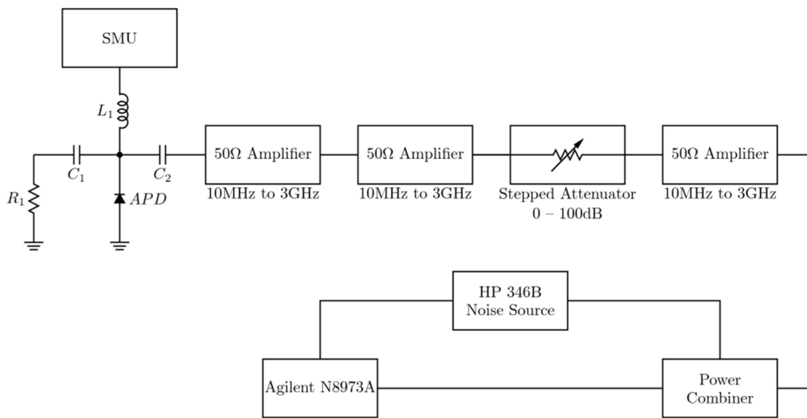


Fig. 6. 50Ω 10MHz to 3GHz excess noise measurement system

The specifications of the second and third amplifiers are considerably less critical than the first. Any microwave device with reasonable noise and gain vs. frequency characteristics will be acceptable. The stepped attenuator should be of the precision type for example the Trilithic RSA35-100 (0dB to 100dB in 10dB steps) would be ideal. The power combiner may be of any type which covers the required bandwidth. A suitable resistive splitter/combiner is the Minicircuits ZX10E-14-S+.

The maximum device capacitance is approximately 2pF to obtain a 3dB point of approximately 3GHz. R_1 must be electrically close to the APD, consequently it is unlikely that the noise contribution of this resistor could be minimised by cooling as was reported by Xie et al. (1993). If the APD was measured at low temperature however it would be plausible to place R_1 and C_1 in the cryostat chamber with the APD, thus obtaining a noise advantage at lower temperatures. A laser is often used to excite electro-optical transducers in characterisation experiments. In this case the laser should be a gas laser possessing a single longitudinal mode, preferably frequency and amplitude stabilised. The authors have met with little success in noise characterisation experiments using semiconductor lasers, the laser relative intensity noise (RIN) is often too great to permit measurement of the detector noise.

6.2.2 TIA CW noise measurement system

The structure of this measurement system is nearly identical to the 50Ω system previously described. The principle difference is the use of a transimpedance amplifier front end instead of a 50Ω system. Figure 7 shows the system diagram.

C_1 provides an AC ground for the APD such that the very great majority of the noise current flows into the TIA. Example TIAs are given in the figure. Commercial TIAs often have input impedance which is not a good approximation to a virtual earth. As a result the maximum permissible device capacitance is often lower than in the 50Ω system case, and is dependent on the particular TIA in use. The MAX3910 provides ~9GHz small signal bandwidth and nearly linear output voltage to input current relationship for photocurrents in the range 0 to 900μ$A_{pk\text{-}pk}$. The small signal gain of this TIA is approximately 1.6kV/A in the linear region.

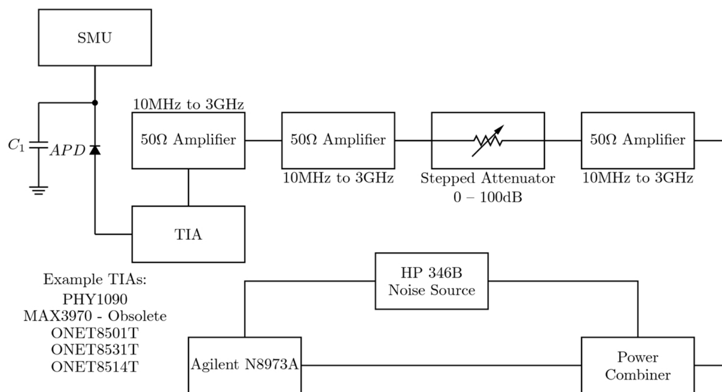


Fig. 7. Transimpedance amplifier excess noise measurement system

Unlike the 50Ω system it is not possible to connect the noise source to the TIA input for calibration purposes. Impedance matching considerations preclude it. This is a major limitation of the TIA measurement compared with the 50Ω measurement. Calibration of the TIA signal path with the noise source is only possible at the TIA output. A plausible method of calibration is to use a unity gain wide band *p-i-n* diode which is known to exhibit shot noise. Any deviation from shot noise can be calibrated out.

7. 10 Gb/s optical communications receiver BER analysis

This section will use the model described in section 3 to analyse the sensitivity of an APD-based receiver system by first investigating the performance of a 10 Gb/s receiver system using InP APDs followed by a discussion on the competing effects of excess noise, APD bandwidth, and tunnelling current on the receiver sensitivity. Similar calculations will then be performed for systems using InAlAs APDs to provide a straightforward and fair comparison with InP.

7.1 Parameters and coefficients

The non-local impact ionisation coefficients and threshold energies of Tan et al. (2008) for InP and Goh et al. (2007a) for InAlAs are used due to the extensive electric field range over which they are valid. The un-multiplied tunnelling current (Forrest et al., 1980b) defined by Equation (34) will use reported experimental InP (Tan et al., 2008) and InAlAs (Goh et al.,

2007b) tunnelling fitting parameters. Since the tunnelling fitting parameters vary with avalanche width, the lowest value, 1.16 for InP and 1.26 for InAlAs, was used for all investigated avalanche widths to assume the worst case scenario. The Johnson noise due to the TIA in the receiver at 10 Gb/s was assumed to be 636 electrons per bit, corresponding to an input noise current density of 10.7 pA/Hz½. Calculations were performed for a series of InP and InAlAs APDs, with active area radius of 15μm and avalanche widths ranging from 0.1 to 0.5μm. A complete list of the parameters used in this section is shown in Table 1.

Parameters	InP	InAlAs
v_e (×10^5 m/s)	0.68	0.68
v_h (×10^5 m/s)	0.7	0.7
E_{the} (eV)	2.8	3.2
E_{thh} (eV)	3.0	3.5
E_g (eV)	1.344	1.45
m^*	$0.08m_o$	$0.07m_o$
σ_T	1.16	1.26

Table 1. Parameters used to simulate the receiver sensitivity performance of InP, InAlAs, and InP and InAlAs APDs.

7.2 InP APD optimisation

Sensitivity versus gain curves were calculated for the InP APDs and the results are shown in Figure 8. The key observation is that for each APD, there exists an optimum mean gain that achieves the lowest sensitivity. In Figure 9, the optimum sensitivity for each device and corresponding mean gain are plotted as functions of the avalanche region width. This allows identification of the optimum avalanche width for a given transmission speed, thereby yielding the optimised sensitivity for a given transmission speed; in this case, 10 Gb/s. The calculations predicted an optimum avalanche width of 0.19 μm for InP APDs, yielding a sensitivity of -28.1 dBm at a gain of 13 for a 10 Gb/s system.

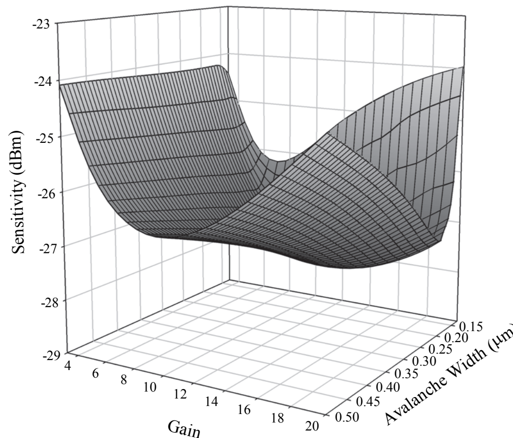


Fig. 8. Receiver sensitivity versus gain for the InP p-i-n APDs, of different avalanche widths, investigated for a 10 Gb/s transmission system.

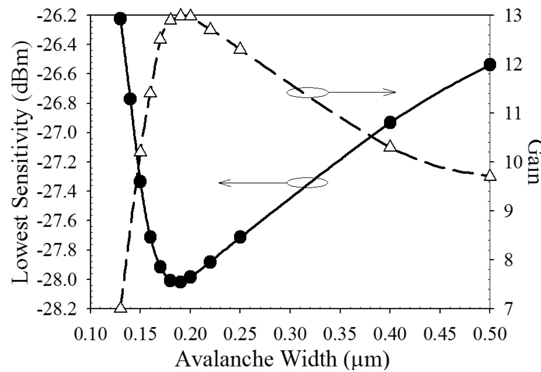


Fig. 9. Lowest sensitivity (solid line, left axis) and its corresponding optimal mean gain (dashed line, right axis) versus InP APD avalanche width for a 10 Gb/s transmission system.

7.3 Competing performance-determining factors

In order to independently assess the significance of (i) ISI, (ii) device bandwidth, and (iii) tunnelling current, three additional sets of calculations were carried out, which shall be referred to as *incomplete* calculations (all at 10 Gb/s). Each set in the incomplete calculations ignores one of the aforementioned three effects. ISI is excluded from the calculations by setting $L = 0$ in (35) and (36). The device bandwidth constraint is removed by setting $\lambda = \infty$, which corresponds to an instantaneous APD. The effect of ISI is also automatically ignored in an instantaneous APD. It is important to note that when ISI is excluded from the model by means of setting $L = 0$, the receiver output is still affected by the bandwidth through the parameter λ in the second terms of (37) and (38), which in turn, represent the attenuation in the receiver output resulting from the APD's bandwidth constraint. This shows the capability of the model to exclude ISI effects alone without the need for assuming an infinite APD bandwidth. Tunnelling current is excluded by setting $n_d = 0$.

Results from each of these three sets of incomplete calculations are compared to those from the *complete* calculation in Figure 10. By observing Figure 9, it is clear that the optimum sensitivity versus width characteristic for a given transmission speed is controlled in a very complex fashion by three device-related factors, namely the tunnelling current, excess noise, and device bandwidth. As the device width decreases, the operating field increases, resulting in increased tunnelling current. The excess noise also decreases with thinner devices confirming, as the dead-space effect becomes more significant (Tan et al., 2008, Forrest et al., 1980a). At the same time, the APD's bandwidth decreases with w; this causes weaker receiver output as well as an increase in the significance of ISI, thereby causing an elevation in the sensitivity.

For the complete calculation results, high sensitivity values for diodes narrower than the optimum avalanche width optimum are due to high tunnelling current. For diodes wider than the optimum avalanche width, sensitivity increases with w, as described above. However, the relative dominance of increasing k_{eff} (resulting in an increase in the excess noise) and decreasing diode bandwidth becomes clear through careful observation of the incomplete calculations. Sensitivity results from the calculations that exclude the bandwidth constraint are only affected by changes in the excess noise when w is increased beyond the

optimum width. Consequently, the sensitivity is observed to increase more slowly with avalanche width compared to that obtained from the complete calculation, suggesting that a decreasing device bandwidth plays a more dominant role than increasing excess noise on sensitivity as w increases. As such, calculations that ignore bandwidth effects will erroneously predict higher optimal device gains compared to those predicted by the complete calculation.

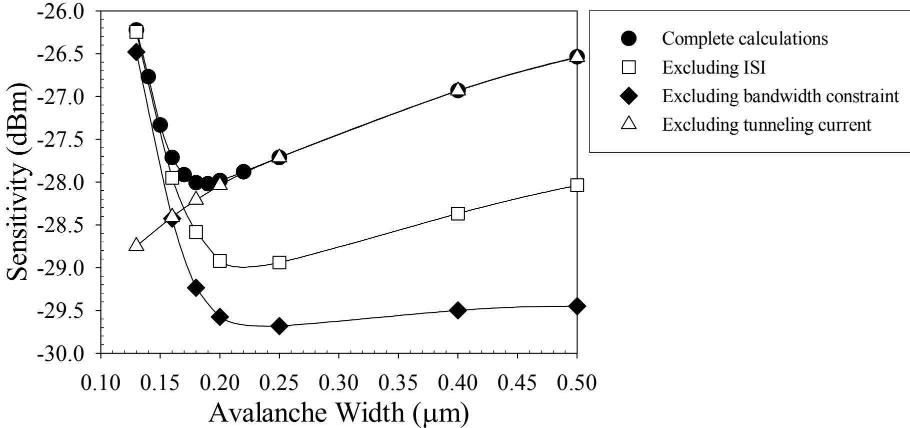


Fig. 10. Sensitivity versus avalanche width for the complete and various incomplete calculation conditions for a 10Gb/s system. Different curves identify the distinct roles of ISI, device bandwidth, avalanche excess noise, and tunneling current.

7.4 Comparison of InP and InAlAs APDs

The optimum sensitivity (optimized over the mean gain) and its corresponding mean gain from the InP and InAlAs calculations are plotted against the avalanche region width, as shown in Figure 11, for a 10 Gb/s system. The calculations predict an optimum w of 0.15μm, with sensitivity of -28.6 dBm and gain of 15, for InAlAs APDs in a 10 Gb/s system.

For any given width, InAlAs provides better sensitivity than InP. However, the improvement is not significant. At their respective optimum avalanche widths, the difference in receiver sensitivities is only 0.5 dBm at both transmission speeds, corresponding to a reduction of 11% in optical signal power at the receiver input. This marginal improvement was also reported by Marshall et al. (2006) albeit with higher sensitivity values, as a result of ignoring the effects of APD bandwidth and ISI. The modesty in this improvement is partly due to a diminishing advantage, as w decreases, in excess-noise characteristics in InAlAs over InP, as shown in Figure 11 in the form of effective ionization coefficient ratio, k_{eff}. At the optimum avalanche widths, the values for k_{eff} are 0.21 and 0.29, for InAlAs (at 0.15μm) and InP (at 0.18μm), respectively. Another factor is the slightly higher gain-bandwidth product in InAlAs compared to InP, 220 and 180 GHz, respectively, at their optimum widths, as shown in Figure 11. The slightly lower tunnelling current in InAlAs APDs compared to those in InP APDs (expected from the slightly larger bandgap of InAlAs), also shown in Figure 11, also contributes slightly to the improvement in receiver sensitivity.

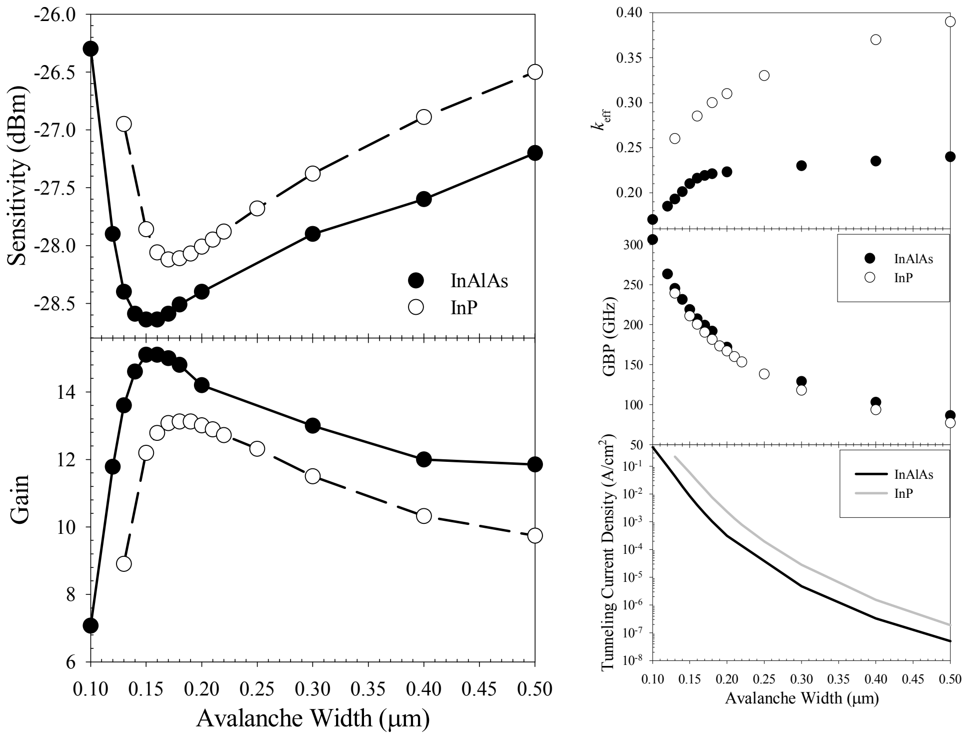


Fig. 11. Optimum sensitivity (left; top) and the corresponding mean gain (left; bottom) versus avalanche width for a 10Gb/s system using InAlAs (closed symbols) and InP (open symbols) APDs. Effective ionization coefficient ratio (right; top), gain-bandwidth product (right; middle), and tunnelling current density (right; bottom), as functions of avalanche width for a 10 Gb/s transmission system using InAlAs and InP. Lines are present to aid visualization.

8. Conclusions

In this chapter the impact ionisation process, from the perspective of APD detector design, has been introduced. The beneficial multiplicative effect on current, and the associated detrimental current fluctuations, excess noise, has been derived. The RPL model has been introduced. This model is routinely used to compute the multiplication and excess noise of thick and thin APD structures. A comprehensive survey of the measurement systems used to characterise the excess noise properties of photodiode structures has been presented, and two improved measurement systems have been suggested. A BER model which includes ISI, excess noise, and tunnelling current has been outlined. The key performance-determining factors which influence the APD and receiver design choices have been analysed. A comparison of InAlAs and InP APDs has been presented and InAlAs offers a marginal sensitivity improvement. An example 10 Gb/s detector and receiver combination has been presented for InAlAs and InP APDs.

9. Acknowledgements

The work reported here was carried out in the Department of Electronic and Electrical Engineering at the University of Sheffield, UK, within the research group of Prof. John David and Dr. Jo Shien Ng, whom the authors thank most sincerely for securing the necessary funding and helping to direct the work.
Daniel S. G. Ong is funded by the University of Sheffield studentship and James E. Green is funded by Engineering and Physical Sciences Research Council (EPSRC).

10. References

Agrawal, G. P. (1997). *Fiber-Optic Communication Systems,* 2nd ed New York, John Wiley & Sons, Inc.

Ando, H. & Kanbe, H. (1981). Ionization coefficient measurement in GaAs by using multiplication noise characteristics. *Solid-State Electronics,* 24**,** 629-634.

Baertsch, R. D. (1966). Low-frequency noise measurements in silicon avalanche photodiodes. *IEEE Transactions on Electron Devices,* ED13**,** 383-385.

Bulman, G. E. (1983). *The experimental determination of impact ionisation coefficients in GaAs and InP.* PhD thesis, University of Illinois.

Campbell, J. C. (2007). Recent advances in telecommunications avalanche photodiodes. *Journal of Lightwave Technology,* 25**,** 109-121.

Forrest, S. R., DiDomenico, M., Jr., Smith, R. G. & Stocker, H. J. (1980a). Evidence for tunneling in reverse-biased III-V photodetector diodes. *Applied Physics Letters,* 36**,** 580-582.

Forrest, S. R., Leheny, R. F., Nahory, R. E. & Pollack, M. A. (1980b). $In_{0.53}Ga_{0.47}As$ photodiodes with dark current limited by generation-recombination and tunneling. *Applied Physics Letters,* 37**,** 322-325.

Goh, Y. L., Marshall, A. R. J., Massey, D. J., Ng, J. S., Tan, C. H., Hopkinson, M., David, J. P. R., Jones, S. K., Button, C. C. & Pinches, S. M. (2007a). Excess avalanche noise in $In_{0.52}Al_{0.48}As$. *IEEE Journal of Quantum Electronics,* 43**,** 503-507.

Goh, Y. L., Massey, D. J., Marshall, A. R. J., Ng, J. S., Tan, C. H., Ng, W. K., Rees, G. J., Hopkinson, M., David, J. P. R. & Jones, S. K. (2007b). Avalanche multiplication in InAlAs. *IEEE Transactions on Electron Devices,* 54**,** 11-16.

Kinsey, G. S., Hansing, C. C., Holmes, A. L., Streetman, B. G., Campbell, J. C. & Dentai, A. G. (2000). Waveguide $In_{0.53}Ga_{0.47}As$-$In_{0.52}Al_{0.48}As$ avalanche photodiode. *IEEE Photonics Technology Letters,* 12**,** 416-418.

Lau, K. S., Tan, C. H., Ng, B. K., Li, K. F., Tozer, R. C., David, J. P. R. & Rees, G. J. (2006). Excess noise measurement in avalanche photodiodes using a transimpedance amplifier front-end. *Measurement Science & Technology,* 17**,** 1941-1946.

Li, K. F. (1999). *Avalanche noise in submicron GaAs and InP structures.* PhD thesis, University of Sheffield.

Li, K. F., Ong, D. S., David, J. P. R., Rees, G. J., Tozer, R. C., Robson, P. N. & Grey, R. (1998). Avalanche multiplication noise characteristics in thin GaAs p^+-i-n^+ diodes. *IEEE Transactions on Electron Devices,* 45**,** 2102-2107.

Marshall, A. R. J., Goh, Y. L., Tan, L. J. J., Tan, C. H., Ng, J. S. & David, J. P. R. (2006). A comparison of the lower limit of multiplication noise in InP and InAlAs based APDs for telecommunications receiver applications. *2006 IEEE LEOS Annual Meeting Conference Proceedings, Vols 1 and 2***,** 789-790.

McIntyre, R. J. (1966). Multiplication noise in uniform avalanche diodes. *IEEE Transactions on Electron Devices,* ED13**,** 164-168.

Ong, D. S., Li, K. F., Rees, G. J., David, J. P. R. & Robson, P. N. (1998). A simple model to determine multiplication and noise in avalanche photodiodes. *Journal of Applied Physics,* 83**,** 3426-3428.

Ong, D. S. G., Ng, J. S., Hayat, M. M., Sun, P. & David, J. P. R. (2009). Optimization of InP APDs for high-speed lightwave systems. *Journal of Lightwave Technology,* 27**,** 3294-3302.

Singh, J. (1995). *Semiconductor Optoelectronics: Physics and Technology*, McGraw-Hill.

Stillman, G. E. & Wolfe, C. M. (1977). Avalanche photodiodes. *In: Semiconductors and semimetals,* R. K. Williardson & A. C. Beer (eds.)**,** 291-393. Academic Press, Inc.

Sun, P., Hayat, M. M., Saleh, B. E. A. & Teich, M. C. (2006). Statistical correlation of gain and buildup time in APDs and its effects on receiver performance. *Journal of Lightwave Technology,* 24**,** 755-768.

Tager, A. S. (1965). Current fluctuations in a semiconductor (dielectric) under the conditions of impact ionization and avalanche breakdown. *Soviet Physics Solid State,* 6**,** 1919-1925.

Tan, L. J. J., Ng, J. S., Tan, C. H. & David, J. P. R. (2008). Avalanche noise characteristics in submicron InP diodes. *IEEE Journal of Quantum Electronics,* 44**,** 378-382.

Tan, L. J. J., Ong, D. S. G., Ng, J. S., Tan, C. H., Jones, S. K., Qian, Y. H. & David, J. P. R. (2010). Temperature dependence of avalanche breakdown in InP and InAlAs. *IEEE Journal of Quantum Electronics,* 46**,** 1153-1157.

Toivonen, M., Salokatve, A., Hovinen, M. & Pessa, M. (1992). GaAs/AlGaAs delta-doped staircase avalanche photodiode with separated absorption layer. *Electronics Letters,* 28**,** 32-34.

Wei, J., Dries, J. C., Wang, H. S., Lange, M. L., Olsen, G. H. & Forrest, S. R. (2002). Optimization of 10-Gb/s long-wavelength floating guard ring InGaAs-InP avalanche photodiodes. *IEEE Photonics Technology Letters,* 14**,** 977-979.

Xie, F. Z., Kuhl, D., Bottcher, E. H., Ren, S. Y. & Bimberg, D. (1993). Wide band frequency response measurements of photodetectors using low-level photocurrent noise detection. *Journal of Applied Physics,* 73**,** 8641-8646.

9

Silicon Photo Multipliers Detectors Operating in Geiger Regime: an Unlimited Device for Future Applications

Giancarlo Barbarino, Riccardo de Asmundis, Gianfranca De Rosa, Carlos Maximiliano Mollo, Stefano Russo and Daniele Vivolo
Università di Napoli "Federico II" - Departement of Physics Sciences and Istituto Nazionale di fisica nucleare - Section of Napoli
Italy

1. Introduction

Photon detectors are indispensable in many areas of fundamental physics research, particularly in the emerging fields of particle astrophysics, nuclear and particle physics, as well as in medical equipment (i.e. PET), in physical check-ups and diagnosis as in-vitro inspection (Radioimmunoassay and Enzyme immunoassay as luminescent, fluorescent, Chemiluminescent Immunoassay), biomedicine, industrial application, in environmental measurement equipment (like dust counters used to detect dust contained in air or liquids, and radiation survey monitors used in nuclear power plants). In astroparticle physics, photons detectors play a crucial role in the detection of fundamental physical processes: in particular, most of the future experiments which aimed at the study of very high-energy (GRB, AGN, SNR) or extremely rare phenomena (dark matter, proton decay, zero neutrinos-double beta decay, neutrinos from astrophysical sources)[3-7] are based on photons detection. The needs of very high sensitivity push the designing of detectors whose sizes should greatly exceed the dimensions of the largest current installations. In the construction of such large-scale detectors no other option remains as using natural media - atmosphere, deep packs of ice, water and liquefied gases at cryogenic temperatures [8-13]. In these (transparent) media, charged particles, originating from interaction or decays of primary particles, emit Cherenkov radiation or fluorescence light, detected by photosensitive devices. Hence, for the improvement in the quality of the experimental results a particular attention should be paid to the improvement of photon detectors performances. In underwater neutrino telescopes (but this is applicable also to other experiments) Cherenkov light, emitted by charged leptons stemming from neutrino interaction, hits photomultipliers (PMT) situated at different distances from the track. This implies, that the response of PMTs should be linear in a very wide range from high illumination to the single photon. Another area of interest is the direct searches of Dark Matter in form of WIMPs: in these experiments it is exploited the scintillation properties of double-phase (liquid-gas) detectors, where primary and secondary scintillation light signals are detected by high-efficiency PMTs, immersed in cryogenic liquids or low temperature gases (89 K for the liquid argon) [14-17]. The next generation of experiments requires further improvement in linearity, gain, and sensitivity (quantum efficiency and single photon counting capability) of PMTs.

To date, the photon detection capabilities of the Vacuum Photomultiplier Tube (VPMT) seem to be unrivalled. Nevertheless standard photomultiplier tubes suffer of the following drawbacks:

- fluctuations in the first dynode gain make single photon counting difficult;
- the linearity is strongly related to the gain and decreases as the latter increases;
- the transit time spreads over large fluctuations;
- the mechanical structure is complex and expensive;
- they are sensitive to the magnetic fields;
- the need of voltage dividers increases failure risks, complexity in the experiments designs and power consumption.

2. Alternatives to the standard photomultipliers tubes

To overcome these limitations, alternatives to VPMT, mainly concentrated on solid-state detectors, are under study. After about one century of standard technology (photocathode and dynode electron multiplication chain), the recent strong developments of modern silicon devices have the potential to boost this technology towards a new generation of photodetectors, based on an innovative and simple inverse p–n junction, PN or PIN photodiodes, avalanche photodiodes—APD and avalanche photodiodes in linear Geiger-mode (GM-APD, SiPM from now on) [18-25]. These solid-state devices present important advantages over the vacuum ones, namely higher quantum efficiency, lower operation voltages, insensitivity to the magnetic fields, robustness and compactness. The step by-step evolution of solid-state photon detectors was mainly determined by their internal gain: a PIN has no gain, an APD can reach a gain of few hundreds, while the GM-APD 10^5–10^6, comparable with that of the vacuum photodetectors; this would allow the GM-APD to achieve single-photon sensitivity and to be used in low-level light applications. This silicon device has become commercially available in the recent years.
We will first discuss the detection of light by silicon devices and then move on to the description of the SiPM and its properties and possible applications.

2.1 Light detection with the photodiode

The basis for detection of light in silicon photodiodes is the p-n junction described in Figure 1, where a depleted region is formed due to carriers diffusion [26].
A junction is formed by diffusing a donor impurity to a shallow depth into silicon which is originally high purity p-type, sometimes called π-type silicon. Thus the layer at the surface is highly doped n-type, often referred as n^+ type with an high concentration of electrons, and the material inside is p type with a relatively low concentration of holes. A schematic view of the structure is shown in Figure 2. The resulting structure, referred to as an n^+-p junction, presents a configuration $n^+p\pi p^+$, where π is a very slight p-type doping. In an analogous fashion a diffused p^+n junction detector can be constructed. Since the density of acceptors in the p-type region is much lower relatively to that of donors in the n^+-type region, the space charge region extends much further into the p region than into the n^+ region. This space-charge region, characterized primarily by acceptor centres in the p-region and filled by donor electrons from the n^+ region, is a charge depleted region of very high resistivity. If electron-hole pairs are produced in this region, the electric field will drive electrons toward the n and holes toward the p side producing a current through the device.

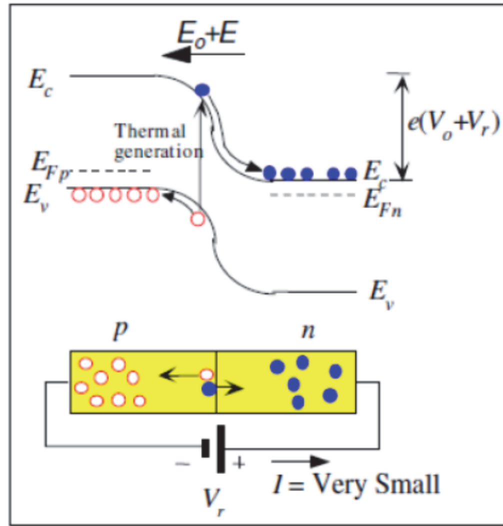


Fig. 1. p-n junction with reversed bias. Energy band diagram is also shown.

2.2 Photon absorption in silicon

Pairs can be produced by light if the energy of the photon is sufficient to bring the electron over the energy band gap.

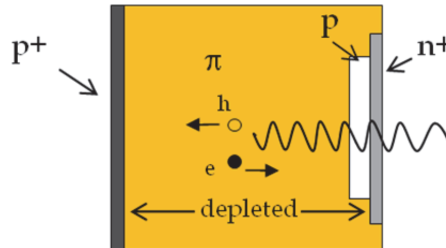


Fig. 2. Schematic view of a p+n junction.

The photon absorption process for photo generation, that is the creation of electron-hole pairs, requires the photon Energy to be at least equal to the band gap energy E_{gap} of the semiconductor material to excite an electron from the valence to the conduction band, namely $h\nu > E_{gap}$, corresponding to $hc/\lambda > E_{gap}$:

$$E_{ph} = h\nu = {}^{hc}\!/_{\lambda} > E_{gap}$$

The upper cut-off wavelength (or the threshold wavelength) λ_{th} is therefore determined by the bandgap energy E_{gap}:

$$\lambda_{th}(\mu m) = \frac{hc}{E_{gap}} = \frac{1.24}{E_{gap}(eV)} \tag{1}$$

In some semiconductors, such as Si and Ge, the photon absorption process for photon energies near E_{gap} requires the absorption and emission of lattice vibrations (vibrations of the Si atoms), namely phonons. The absorption process is said, in these cases, to be indirect as it depends on lattice vibrations which in turn depends on the temperature [27]. Since the interaction of a photon with a valence electron needs a third body, a lattice vibration, the probability of photon absorption is not as high as in a direct transition. As a consequence, the threshold wavelength is not as sharp as for direct band gap semiconductors. During the absorption process, a phonon may be absorbed or emitted. If ξ is the frequency of the lattice vibrations, then the phonon energy is $h\xi$ and the photon energy should be $h\nu > E_{gap} \pm h\xi$. Actually, since $h\xi$ is small (<0.1 eV), the energy needed for absorption is very close to E_{gap}.
In silicon, for which E_{gap}=1.12 eV, the threshold wavelength as given by the Equation 1 is ≈1100 nm.
Incident photons with wavelengths shorter than λ_{th} become absorbed as they travel in the semiconductor and the light intensity, which is proportional to the number of photons, decays exponentially with distance into the semiconductor. The absorption coefficient α determines how far into a material the light of a particular wavelength can penetrate before absorption. In a material with a low absorption coefficient, light is only poorly absorbed, and if the material is thin enough, it will appear transparent to that wavelength.
The absorption coefficient, α, is related to the extinction coefficient, k, by the following formula:

$$\alpha = \frac{4\pi k}{\lambda}$$

where λ is the wavelength. Thus, defining the complex index of refraction as ñ= n – ik, the imaginary component k is related to the absorption, whereas the real component n= c/v_{phase} is related to reflectivity. In Figure 3 the real and imaginary part of the refractive index of silicon is shown [28].

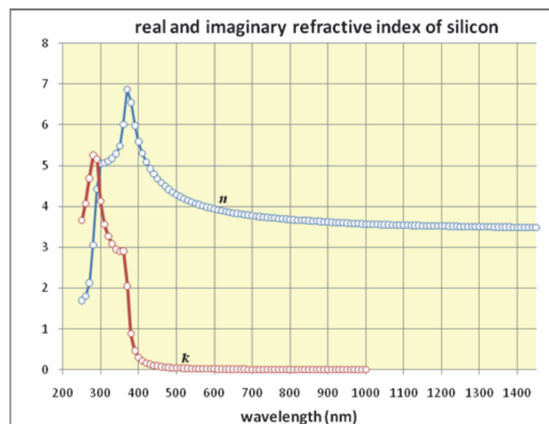


Fig. 3. Real and (negative) imaginary components of the refractive index for silicon at 300 K.

As a consequence of the cut-off wavelength, direct bandgap semiconductor materials (as GaAs, InP) have a sharp edge in their absorption coefficient. Actually, even for those

photons which have an energy above the band gap, the absorption coefficient is not constant, but still depends strongly on the wavelength. The probability of absorbing a photon depends on the probability that a photon and an electron interact in such a way as to move from one energy band to another. For photons which have an energy very close to that of the band gap, the absorption is relatively low since only those electrons directly at the valence band edge can interact with the photon to cause absorption. As the photon energy increases a larger number of electrons can interact with the photon, resulting in a higher absorption probability.

In indirect bandgap semiconductor materials, like silicon, there is a long tail in absorption out to long wavelengths. Figure 4 [27] shows the absorption coefficient α as a function of wavelength λ for various semiconductors: it is clear the different behaviour of α with the wavelength in the case of direct band gap semiconductors (e.g. GaAs, InP) with respect to indirect band gap semiconductors (e.g. Si, Ge). In Figure 5 [29], instead, the absorption length or penetration depth, defined as 1/α, as a function of wavelength for Si is shown.

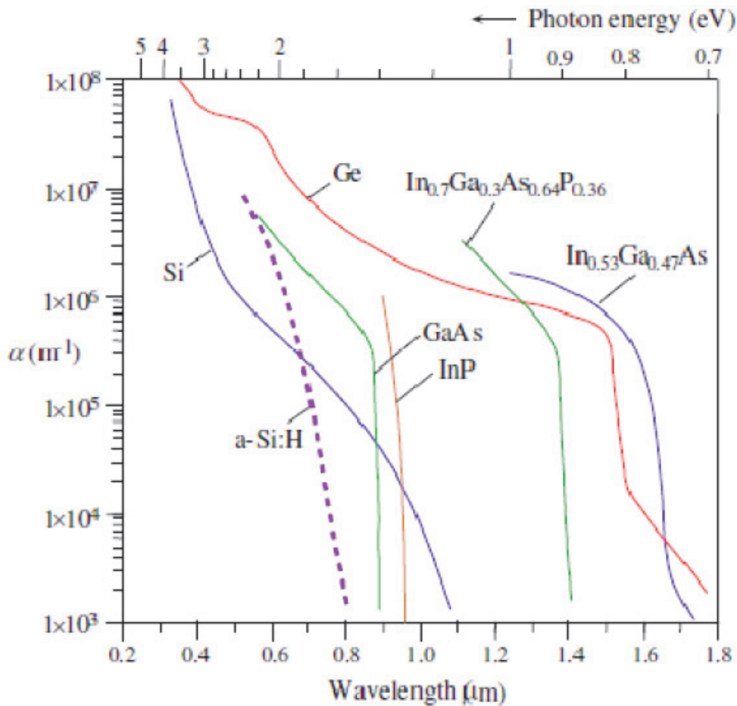


Fig. 4. Absorption coefficient α as a function of wavelength λ for various semiconductors.

To detect light by a photodiode, it first has to enter through the surface and then absorbed in the active volume of the device. Due to the high value of real part of the refractive index of silicon, which is above 3.5 for wavelengths below 1100 nm at 300 K as shown in Figure 3, an antireflective coating is needed to reduce the strong Fresnel reflection of light from the surface of the device.

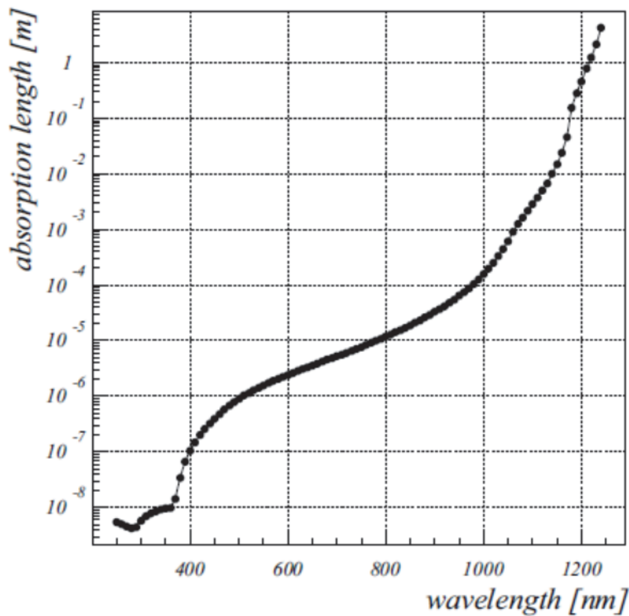


Fig. 5. Absorption length 1/α as a function of wavelength λ for Silicon.

Actually, not all the incident photons are absorbed to create pairs that can be collected and give rise to a photocurrent. The efficiency of the conversion process is measured by the quantum efficiency QE of the detector, defined as

$$QE = \frac{\textit{number of electron} - \textit{hole pairs generated and collected}}{\textit{number of incident photons}}$$

which depends on the wavelength. In the evaluation of QE, the number of electrons collected per seconds is given by I_{ph}/e, where I_{ph} is the measured photocurrent, whereas the number of photons arriving per second is $P_o/h\nu$, with P_o the incident optical power.
Then the QE can be also defined as [27]

$$QE = \frac{I_{ph} / e}{P_o / h\nu}$$

A typical photodiode QE is shown in Figure 6 [30].
If the semiconductor length is comparable with the penetration depth not all the photons will be absorbed, resulting in a low QE.
Therefore, to obtain an high quantum efficiency, the thickness of the depleted layer has to be larger than the absorption length. The absorption length shows strong variations from about 10 nm, for near UV light, to more than 1 mm, in the infrared region. To enhance the sensitivity in the range of blue light, the active region needs to be close to the surface and for the detection of longer wavelengths it has to be thick compared to the absorption length.

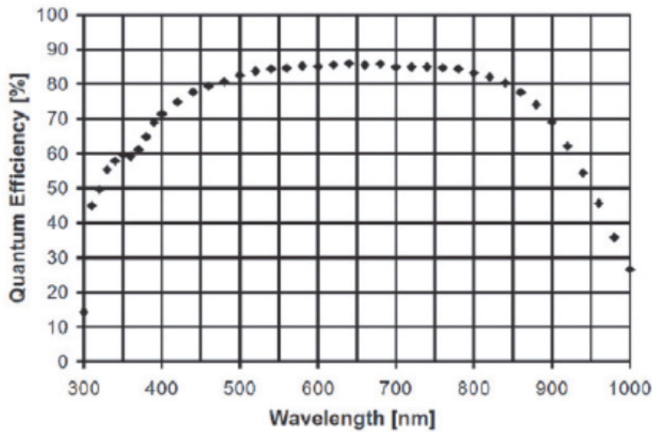


Fig. 6. Typical photodiode QE as a function of wavelength.

2.3 Reverse biased photodiodes

The thickness of the layer can be increased by applying a reverse bias to the diode junction. To obtain a thick depletion layer with low reverse bias, a PIN photodiode is used with an intrinsic layer between the p and n faces of the diode. The photodiode does not present any internal amplification of the signal so the number of charges generated it is equal to the number of detected photons. It can be used for applications in which more than about 10,000 photons are simultaneously detected by the device. Taking into account that the capacitance per unit area C/A, expressed in picofarad per square centimetre, is $C/A=1{,}061/x_0$ where x_0 is the depletion layer thickness expressed in cm, millivolt ranged signal is expected using typical parameters [31]. A typical application in high energy physics for such a device is the calorimetry, in which a large amount of photons has to be detected.

To detect weaker signals, instead, internal amplification is required. This can be obtained, as in gas based devices, by increasing the applied voltage. In fact, if the electric field in the silicon is high enough, primary carriers can produce new pairs by impact ionization. These generated electron-hole pairs are further accelerated by the electric field to a sufficiently high kinetic energy to cause new impact ionization, releasing more electron-hole pairs, which leads to an avalanche of impact ionization processes. Thus, with a single photon absorption, one can generate a large number of electron-hole pairs, all of which contribute to the observed photocurrent, leading an internal gain mechanism. Each absorbed photon creates in average a finite number M of electron–hole pairs exploiting the impact ionization process. The multiplication of carriers in the avalanche region depends on the probability of impact ionization which strongly depends on the reverse bias V_{bias}.

This mode of operation is called linear because the number of the collected carriers is proportional (by a factor M) to the number of absorbed photons. A photodiode with such an amplification region is called the avalanche photodiode (APD). The ionization rate is higher for electrons than for holes, so the amplification process for electrons starts at lower fields and the avalanche grows in the direction of the electrons movement. With the increase in the electric field also holes start to ionize. When the ionization probability is high enough, the amplification can no longer be controlled. This limits the amplification factor in APDs to about ~100. Due to the low amplification, these devices are still not appropriate for detection

of signals of a few photons only. However, signals coming from about 100 photons can be detected.

2.4 Geiger mode APD silicon photomultiplier

To obtain the single photon sensitivity in a silicon device, one needs to operate the APD in the Geiger mode [32]. A diode working in a region near the breakdown voltage can be operated in two different ways depending on whether the bias voltage is below or above the breakdown point. In the first case the device is called avalanche photodiode (APD) described above. In the second case the device is referred to as Geiger-mode APD (GM-APD). In this bias condition, the electric field is so high that a single carrier injected into the depletion region can trigger a self-sustaining avalanche. The carrier initiating the discharge can be either thermally generated (noise source of the device) or photo-generated (useful signal).

In Figure 7 the schematic view of the gain as a function of reverse bias is shown.

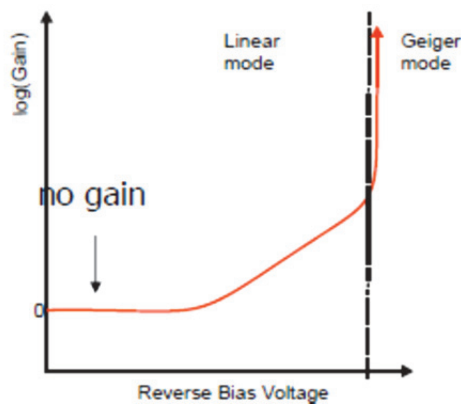


Fig. 7. Schematic view of gain as a function of Vbias.

The main limitation of a single diode working in GM is that the output signal is the same regardless of the number of interacting photons. In order to overcome this limitation, the diode can be segmented in tiny micro-cells (each working in GM) connected in parallel to a single output. Each element, when activated by a photon, gives the same current response, so that the output signal is proportional to the number of cells hit by a photon and the output signal is the sum of the Geiger mode signals of microcells. The dynamic range is limited by the number of elements composing the device, and the probability that two or more photons hit the same micro-cell depends on the size of the micro-cell itself. This structure is called Silicon Photo Multiplier (SiPM) [1].

All the microcells are identical, independent and operate in single photon counting mode. A quenching mechanism is implemented thanks to a specially resistive material technology. Together with the common electrode structure all this gives the possibility to act as a proportional detector for measurements of low intensity photons flux.

The typical density of microcells that can be produced is 1000–5000 per mm^2 and the total number of microcells on our tested photodetectors with sensitive area of 1 mm^2 is of the order of 2000. This defines the dynamic range of the device. The noise conditions of the

SiPM is defined by dark count rate, as in Geiger mode a single thermally generated electron or hole can initiate an avalanche, leading to an electrical pulse that is indistinguishable from the one of a single photon. This gives the main limitation of increasing the sensitive area of SiPM operated in single photon counting mode, but it is not so significant for low photon flux measurement when $N_{phot} > 3$.

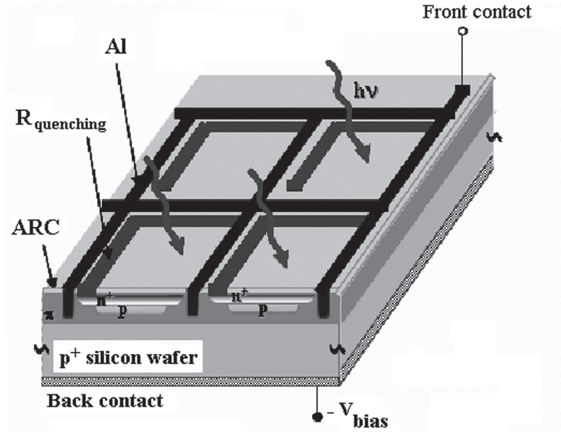


Fig. 8. Structure of the multi cell matrix of a SiPM.

3. Structure of the SiPM

The structure of the silicon photomultiplier is a combination of large number of avalanche microcells on a single substrate and with common quenching mechanism (resistive layer) and common electrodes.

3.1 Structure of avalanche microcell

The schematic structure of the avalanche microcell of a SiPM is shown in Figure 9 and presents a configuration n^+-p-π-p^+, where π represents very slight p-type doping.

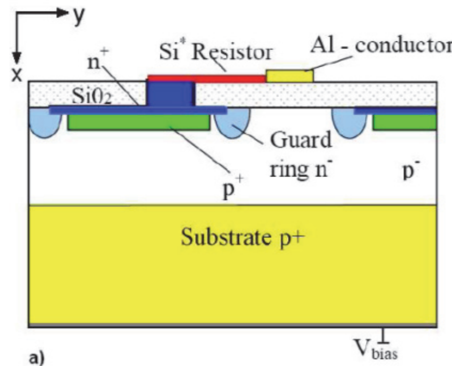


Fig. 9. Schematic structure of avalanche microcell of SiPM.

The n^+ side is thin and is the one which receives light through a window. A thickness of about 1μm of depletion region between the thin n^+ (thickness = 0.1–1.5 μm) and p layers is created thanks to the reverse electric field. There are three p-type layers of different doping levels next to the n^+ layer to suitably modify the field distribution across the structure. The first is a thin p-type layer, the second is a thick lightly p-type doped (almost intrinsic) π-layer of ≈300 μm, and the third is a heavily doped p^+ layer ≈ 3μm thick. On the surface of the avalanche microstructure a thin metal layer is placed (≈ 0.01 μm) with an antireflection coating. Above the n^+ region, a resistive SiO_2 layer (thickness ≈0.15 μm, ρ ≈ 10^5-10^7 Ωcm) limits the Geiger breakdown propagation by a local reduction of the electric field.
The electric field is at a maximum at the n^+p junction, then decreases slowly through the p-layer. The field vanishes at the end of the narrow depletion layer in the p^+ side, as shown in Figure 10 [33].

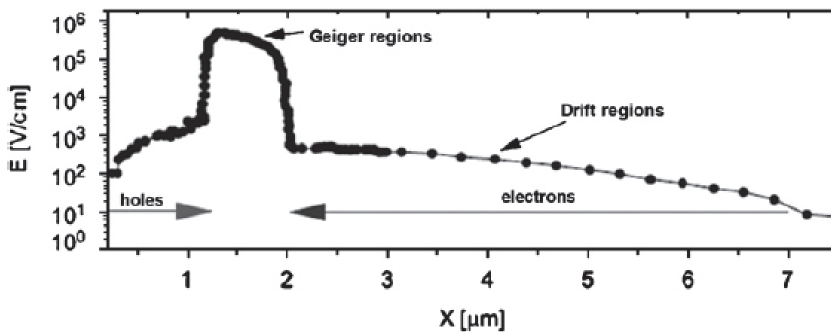


Fig. 10. Configuration of the electric field. The high-field region (E≈5×105 V/cm) is built up in the highly doped n+p.

The absorption of photons of λ≈400 nm takes place mainly in the π-layer. The nearly uniform field here separates the electron–hole pairs and drifts them at velocities near saturation towards the n^+ and p^+ sides, respectively. When the drifting electrons reach the p-layer it may be accelerated by the high fields to sufficiently large kinetic energies to further cause impact ionization and release more electron-hole pairs which leads to an avalanche of impact ionization processes. Thus, from a single electron entering the p-layer, one can generate a large number of electron-hole pairs all of which contribute to the observed photocurrent. In this mode, any electron event in the sensitive area will produce a very large current flow with amplification gain of up to 10^6.

3.2 Operation principle of a SiPM

As mentioned in the previous paragraph the SiPM is a matrix of GM-APDs connected in parallel. A schematic representation of the device is shown in Figure 11. The connection between the cells is made on one side by the low-resistivity substrate and on the other side by a metal layer. The diodes (labelled as D) are asymmetric p–n junctions. Each GM-APD has in series a quenching resistor (R_Q) which is needed to stop the avalanche current and, then, to restore the initial bias condition enabling the detection of a new incoming photon. A reverse bias voltage (V_{bias}) is applied to each junction through the common substrate electrode to deplete the n^+–p junctions.

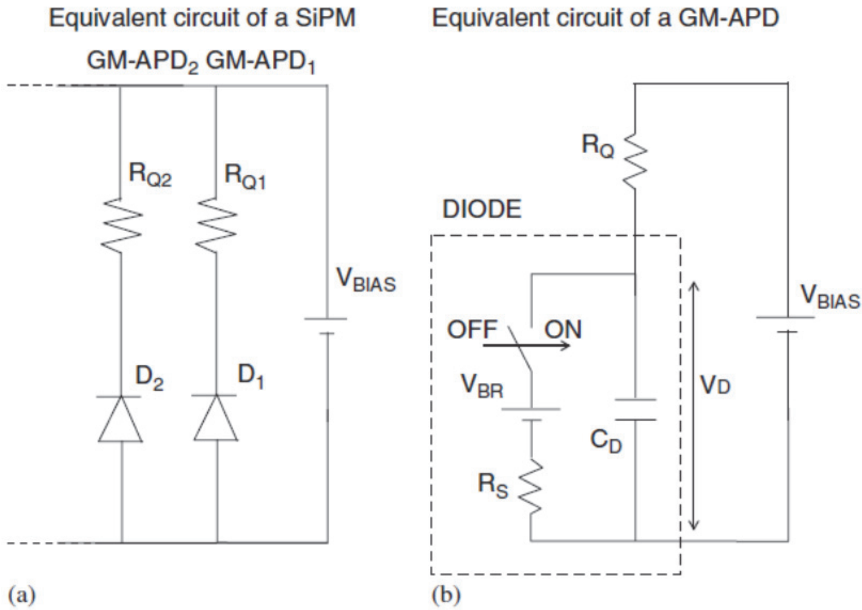


Fig. 11. Equivalent electric scheme of the SiPM

4. Test of static characteristics

The breakdown voltage (V_{break}) and the R_Q values are determined thanks to the reverse and forward current–voltage (I–V) characteristics curves.

The MPPCs used for this work have an active surface of 1 X 1 mm², divided into 1,600 pixels of 25µm x 25µm (Figure 12, Figure 13 and Figure 14), and of 3 X 3 mm², divided in 14,400 pixels of 25µm x 25µm (Figure 15 and Figure 16 : from Hamamatsu data sheet) [34].

Fig. 12. Aspect and external dimensions of the MPPC 1 X 1 mm² under characterization.

Active area 1 × 1 mm type (Typ. unless otherwise noted, Ta=25 °C)

Parameter	Symbol	S10362-11 series			Unit
		-025U, -025C	-050U, -050C	-100U, -100C	
Chip size	-	1.5 × 1.5			mm
Effective active area	-	1 × 1			mm
Number of pixels	-	1600	400	100	-
Pixel size	-	25 × 25	50 × 50	100 × 100	μm
Fill factor *1	-	30.8	61.5	78.5	%
Spectral response range	λ	270 to 900			nm
Peak sensitivity wavelength	λp	400			nm
Quantum efficiency (λ=λp)	QE	70 Min.			%
Photon detection efficiency *2 (λ=λp)	PDE	25	50	65	%
Recommended operating voltage range	-	70 ± 10 *3			V
Dark count	-	300	400	600	kcps
Dark count Max.	-	600	800	1000	kcps
Terminal capacitance	Ct	35			pF
Time resolution (FWHM)	-	200 to 300			ps
Temperature coefficient of reverse voltage	-	50			mV/°C
Gain	M	2.75×10^5	7.5×10^5	2.4×10^6	-

Fig. 13. Operating parameters of the MPPC 1 X 1 mm^2 as delivered by the supplier.

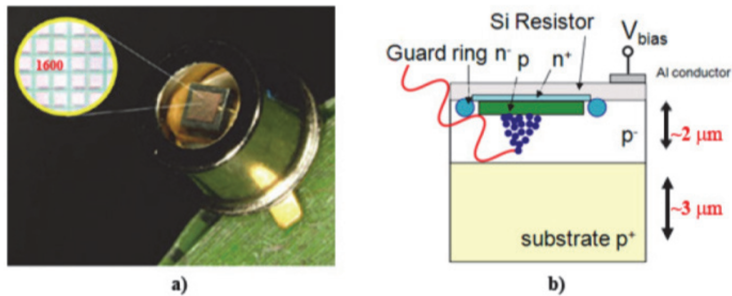


Fig. 14. a) A Photograph of the MPPC S10362-11-025U by Hamamatsu. b) Structure of a MPPC pixel.

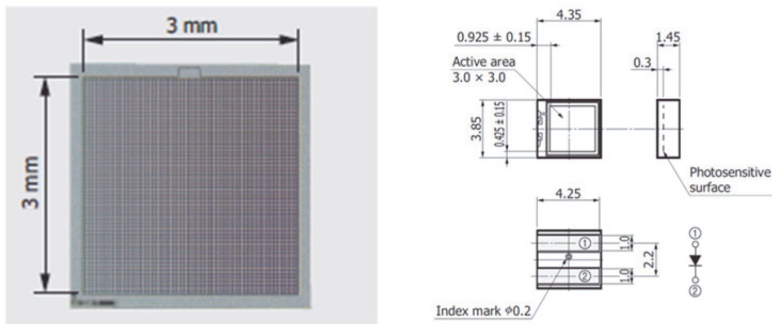


Fig. 15. Aspect and external dimensions of the MPPC 3 X 3 mm^2 under characterization.

Electrical and optical characteristics (Typ. Ta=25 °C, unless otherwise noted)

Parameter	Symbol	S10362-33			S10931			Unit
		-025C	-050C	-100C	-025P	-050P	-100P	
Fill factor *1	-	30.8	61.5	78.5	30.8	61.5	78.5	%
Spectral response range	λ	320 to 900				320 to 900		nm
Peak sensitivity wavelength	λp	440				440		nm
Operating voltage range	-	70 ± 10 *2				70 ± 10 *2		V
Dark count *3	-	4	6	8	4	6	8	Mcps
Dark count Max. *3	-	8	10	12	8	10	12	Mcps
Terminal capacitance	Ct	320				320		pF
Time resolution (FWHM) *4	-	500 to 600				500 to 600		ps
Temperature coefficient of reverse voltage	-	56				56		mV/°C
Gain	M	2.75×10^5	7.5×10^5	2.4×10^6	2.75×10^5	7.5×10^5	2.4×10^6	-

Fig. 16. Operating parameters of the MPPC 3 X 3 mm2 as delivered by the supplier.

A V_{break} of 70.1 V has been obtained for S10362-11-025U 1 X 1 mm² SiPM, thus demonstrating a good uniformity of the V_{break} for different SiPM's. The value of the quenching resistor extracted from the forward characteristics is of ~145 Ω, giving for a single micro-cell a value of 230 kΩ (see Figure 17). In fact the global resistance measured is related as:

$$R_{SiPM} = R_{micro_cell}/N_{micro_cell}$$

where $N_{micro\text{-}cell}$ = 1,600 for the S10362-11-025U model of 1 X 1 mm² SiPM and $N_{micro\text{-}cell}$ = 14,400 for the S10931-025P model of 3 X 3 mm² SiPM.

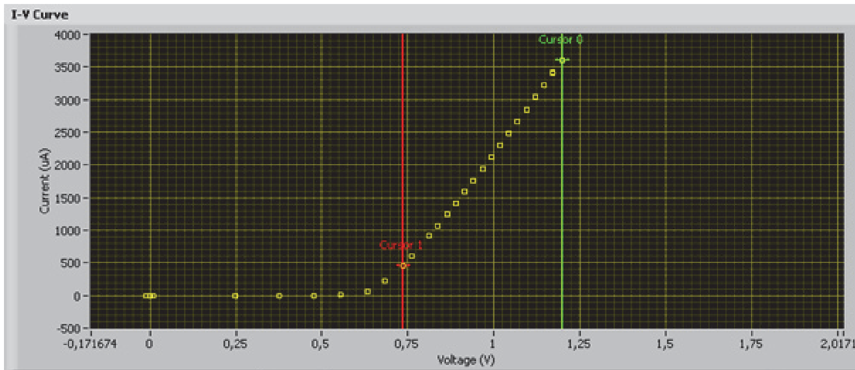


Fig. 17. Forward characteristics of S10362-11-025U SIPM (1 x 1 mm²).

A similar measurement on the 3x3 mm² SiPM led to a global resistance of 26 Ω, giving a single pixel quenching resistance of ~ 380 kΩ.
In the Figure 18, the reverse biased portion of the I-V curve is shown, for the 3x3 mm² SiPM. This side of the curve is used to extrapolate from the fit, the most convenient bias voltage to apply to the device.

5. Dynamic characteristic and basic performances

A circuit model, which emulates the evolution of the signal of a GM-APD was developed in the 1960s to describe the behaviour of micro-plasma instability in silicon [35, 36]. According to this model, the pre-breakdown state can be represented as a capacitance (junction capacitance, C_D) in series with the quenching resistor.

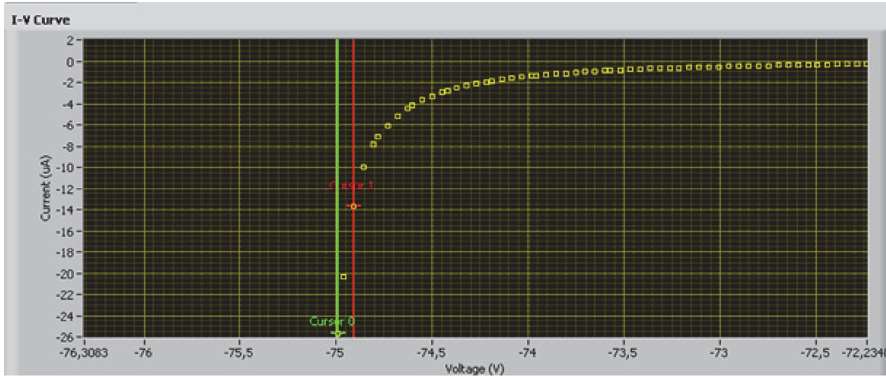


Fig. 18. Reverse side of the Current VS Voltage curve, this time for the 3x3 mm[2] SiPM.

In steady state, the capacitance is charged at $V_{bias}>V_{break}$. When a carrier traverses the high-field region, (switch closed in Figure 11/b) there is a certain probability, to initiate an avalanche discharge. If this happens, the new state of the system can be modelled adding to the circuit a voltage source V_{break} with a series resistor R_S in parallel to the diode capacitance. R_S ($\approx 1k\Omega$) includes both the resistance of the neutral regions inside the silicon as well as the space charge resistance. C_D, originally charged at $V_{bias}>V_{break}$, discharges through the series resistance down to the breakdown voltage with a time constant τ_D given by the product R_SC_D. The discharge current of the avalanche process can take some hundreds of ps. As the voltage on C_D decreases, the current flowing through the quenching resistance tends to the asymptotic value of $(V_{bias}\text{-}V_{break})/(R_Q+R_S)$. In this final phase, if R_Q is high enough (some hundreds of kΩ), the diode current is so low that a statistical fluctuation brings the instantaneous number of carriers flowing through the high-field region to zero, quenching the avalanche.

When the avalanche process is terminated, the switch is again open and the circuit is in its initial configuration with the capacitance charged at V_{break}. At this point it starts recharging to the bias voltage with a time constant $\tau_R=C_DR_Q$ and the device becomes ready to detect the arrival of a new photon (Figure 19). The number of carriers created during an avalanche discharge is given by $n_{carriers}=(V_{bias}-V_{break})C_D/e$, where e is the electron charge.

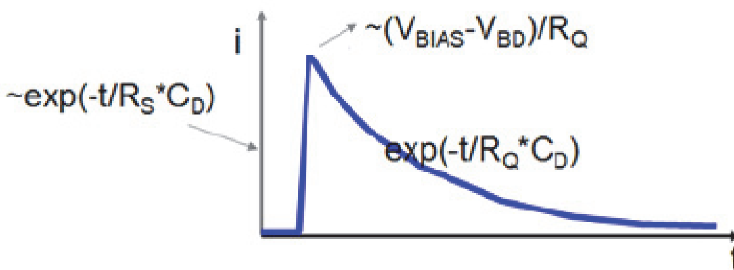


Fig. 19. The time evolution of the current into a cell, when a discharge occurs.

Than the gain of the SiPM (G) is determined by the charge (Q) that can be released from a micro-cell after the breakdown [19]:

$$G = \frac{Q}{e} = \frac{\Delta V\, C_D}{e} = \frac{1}{e}\frac{\Delta V}{R_Q} R_Q C_D = \frac{I_{\max}\tau_R}{e}$$

where $\Delta V = (V_{bias}\text{-}V_{break})$ is the overvoltage and $\Delta V/\ R_Q = I_{max}$.
The time integration of the micro-cell dark pulse allowed the measurement of the gain.
Each diode composing the SiPM reacts independently in the above-described way. Thus, if n cells are activated at the same time, the charge measured at the SiPM output is n times the charge developed by a single GM-APD, giving information on the light intensity.

5.1 Experimental set up

The basic performances of the MPPC's are measured with light from a pulsed laser. All results shown in this section are for a Hamamatsu MPPC of 1x1 mm^2 and 3x3 mm^2.
The general scheme of the experimental setup is shown in the Figure 20. The board accommodating the MPPC is placed inside of a dark chamber; the laser, pulsed at a typical frequency of 100 kHz, is connected via optical fiber with a custom connector as coupling device. Two beam splitters are installed along the fiber way in order to reduce the beam intensity and control, for very low intensities, the number of photons: the 1% outputs are used on both splitters, leading to a 10^{-4} reduction factor.
The optical power of the installation has been measured using a Newport mod. 815 power meter having a sensitivity of 1 nW obtaining a response of about 50 to 2350 photons per pulse.

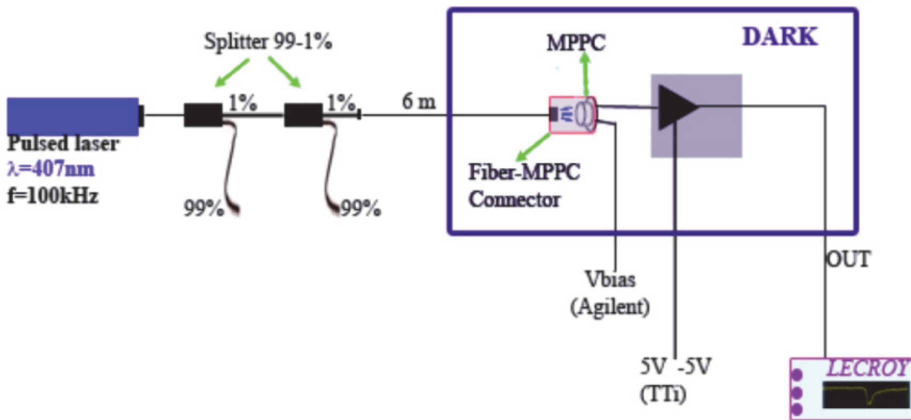


Fig. 20. Scheme of the bench test for the MPPC using laser source and beam splitters.

Figure 21 shows the electric scheme of the board used to amplify the MPPC signals. The detector receives power from a polarization resistor R_P~10 kΩ and its output is connected to the input of a current-to-voltage amplifier. The amplifier is a LMH6624 by National Semiconductor used in inverting configuration and powered by ±5 V. In this configuration

the output of the operational amplifier is directly proportional to the current on the input flowing through the reaction resistance R_f, determining the amplification trans resistance gain and so giving:

$$V_{out} = -i_{in} \cdot R_f \tag{5.1}$$

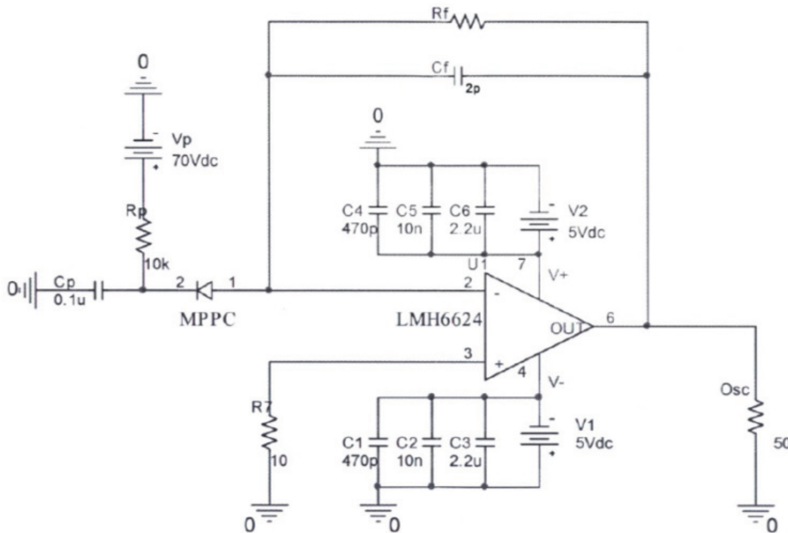


Fig. 21. Schematic diagram of the amplification board designed which accommodates the MPPC.

5.2 Raw signal

Figure 22 shows the raw signal and the output charge spectrum from an MPPC taken with an oscilloscope. The MPPC is illuminated by pulsed light from the laser at low intensity and the oscilloscope is triggered in synch with the laser. The responses for multiple triggers are overlaid in the figure. The charge corresponding to different numbers of fired pixels shows well separated peaks.

This indicates that the gain of each micropixel is uniform, demonstrating the excellent photon counting capability of the MPPC. For bias voltage under 69.8 V on the 1x1 mm^2 device (and 69.4 V on the 3x3 mm^2 device) we did not succeeded in observing any signal, since their amplitude is covered by electronic noise. At 69.8 V (69.4 V for the 3x3 mm^2 device) we observe clear signals of 1 photon equivalent (p.e.) of an amplitude of the order of 2 mV (2.7 mV for the 3x3 mm^2 device). As the bias voltage increases, higher signals are obtained, so 1 p.e. and signals for 2 or 3 p.e. begin to be always observable. The latter are due to an increased probability of thermally generation and crosstalk events. Around 71.5 V signals for 2 or 3 p.e. become very frequent and it is clear that, in these conditions, it becomes difficult to distinguish between any actual signal and dark count.

Some samples of signal as observed at the oscilloscope at different operating voltages are shown in the Figure 23.

We remark that:

- Rise time and fall time of signals are basically independent from the applied voltage. Rise time is around 1-2 ns; fall time is 2-3 ns and is sensitive to the transition capacitance of the pixels and the quenching resistor, since:

$$\tau_{fall} = (R_P + R_Q)C_D \cong R_Q C_D = \tau_R$$

- Optimal working voltage for this MPPC are around 70.5-71.2 V, values that are compatible with the ones suggested by the manufacturer.

Rise and fall time have been measured at 70.8 V reverse bias, by accumulating 10,000 values and obtaining, for the 1x1 mm^2:

$$\tau_{rise} = (1.4 \pm 0.6)\ ns$$

$$\tau_{fall} = (2.0 \pm 0.5)\ ns$$

and 72.1 V reverse bias for the 3x3 mm^2,obtaining:

$$\tau_{rise} = (8.4 \pm 0.6)\ ns$$

$$\tau_{fall} = (9.9 \pm 0.5)\ ns$$

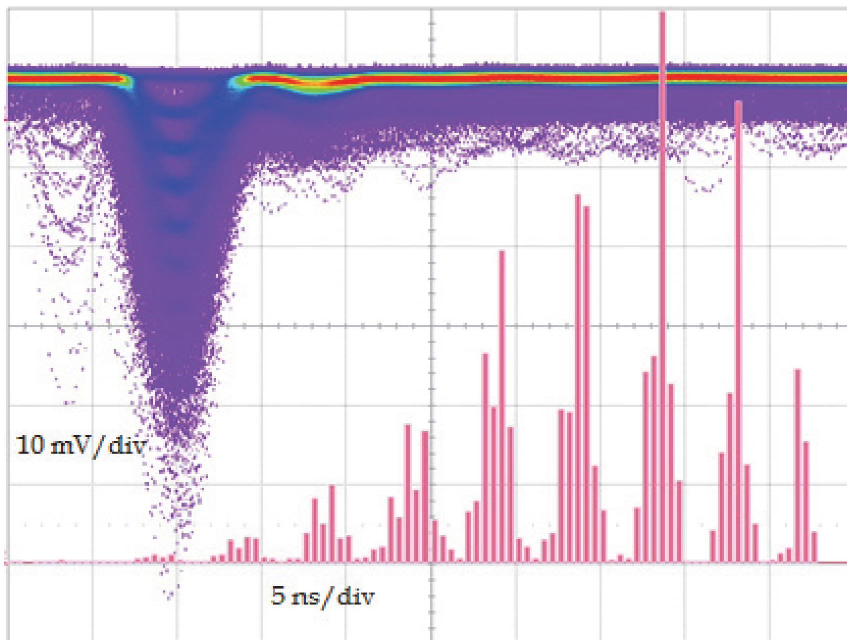


Fig. 22. A collection of pulse signals from MPPC as observed at the oscilloscope for the 3x3 mm^2 MPPC.

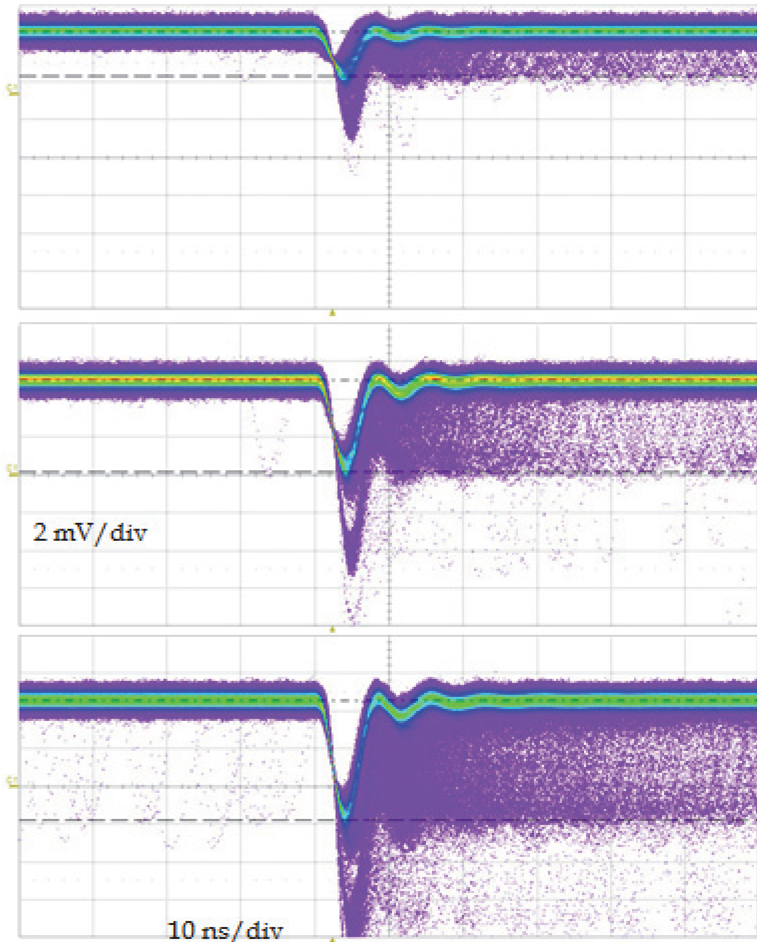


Fig. 23. A collection of pulse signals from MPPC as observed at the oscilloscope for the 1x1 mm^2 MPPC.

5.3 Gain

From the number of ADC counts between a well-separated pedestal and the peak corresponding to a single fired pixel, we derive the charge corresponding to a it, Q. The gain is defined as Q divided by the charge of an electron:

$$G = \frac{Q}{e}$$

Figure 25 shows the gain measured as a function of the applied voltage for the 1x1 mm^2 MPPC. The measurement is performed inside a temperature-controlled chamber and data at 22.5 °C are shown. The measured gain depends linearly on the applied voltage as expected. As a second method, the gain can be calculated by measuring the number of channels between two neighboring peaks. To calculate the gain with the second method we put the

MPPC under very low illuminations conditions, allowing to clearly distinguish between peaks of 1, 2, 3 and 4 p.e.. By changing the bias voltage between 71.5 and 74.1 V in 0.2V steps, we measured the difference in the amplitudes of signals of 2 – 1 p.e., 3 – 2p.e. and 4 – 3 p.e.. Figure 22 shows the measurements obtained for 1x1 mm² MPPC whereas histograms on the left in Figure 24 show the results obtained with a 3x3 mm² MPPC. Alternatively, the gain can also be evaluated by measuring the charge of the signal corresponding to the initial number of photoelectrons. The method is shown in the right histogram in the Figure 24, while in Figure 26 the two methods are compared.

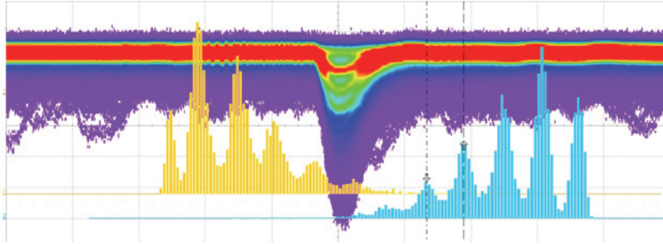


Fig. 24. Pulses from MPPC and gain measurement for the 3x3 mm² MPPC (binning of left histograms is of 5 mV, and ,of right one is 50.0 pVs. Signal shown with 5 mV/div-20ms/div).

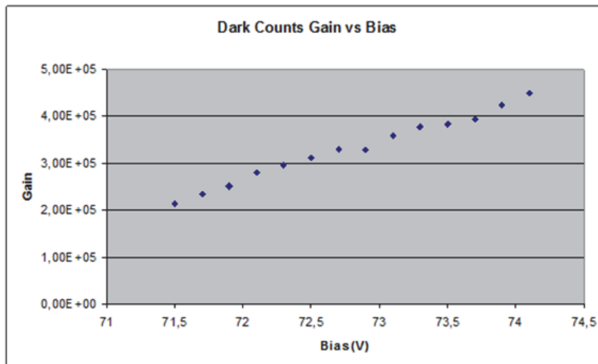


Fig. 25. Measured gain as a function of the applied voltage for the 1x1 mm² MPPC.

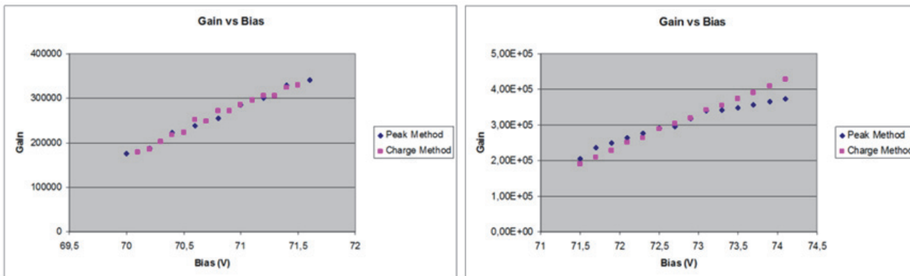


Fig. 26. Comparison between methods for gain evaluation for the 1x1 mm² MPPC (Left) and the for the 3x3 mm² MPPC (Right).

5.4 An estimation of the capacitance

From the gain obtained it is possible to get an estimation of the junction capacitance C_D. In the case of the 3x3 mm^2 MPPC (1x1 mm^2 MPPC), from the linear fit of Figure 26, the slope of the fitting line is

$$b = (906 \pm 9)\ 10^2 V^{-1} \quad ((105 \pm 2)\ 10^3 V^{-1} \text{ in the 1x1 } mm^2 \text{ MPPC})$$

by multiplying this value for the electron charge we get:

$$C_D = (14.51 \pm 0.15)\ fF \quad ((16.74 \pm 0.03)\ fF)$$

From this we can *get an estimation of the value of the quenching resistor:*

$$R_Q = \tau_{fall}/C_D = (680 \pm 40)\ k\Omega \quad ((119 \pm 30)\ k\Omega)$$

Moreover, since

$$G = \frac{(V_{bias} - V_{break})C_D}{e}$$

it is also possible to estimate the breakdown voltage of the device, by extrapolating from the gain line the voltage value corresponding to G=0. In the 3x3 mm^2 MPPC we obtain $V_{break} = (69.4 \pm 0.7)$ V, while $V_{break} = (68.796 \pm 0.005)$ V for the 1x1 mm^2 MPPC.

5.5 Noise considerations

The Geiger-mode micro-cell detection of an event does not give intensity information. The output pulse produced by the detection of a photon is indistinguishable from that produced by the detection of many simultaneously absorbed ones. That means a single thermally generated electron or hole can initiate an avalanche. This gives the main limitation of increasing the sensitive area of Si avalanche structures operated in single photon-counting mode at room temperature. Reduction of the dark counting rate in Si avalanche can be obtained by limiting both the sensitive area 1x1 - 3x3 mm^2) and the thickness of depleted region.

Other improvements can be achieved by minimizing the number of generation-recombination centres, the impurities and crystal defect. In addition, the detector operation at low temperature and a good quality in the fabrication process further improve the single photon detection capability. The main effect to be taken into account is the production of after-pulses by charges from the avalanche process that are temporarily trapped, generating a new avalanche after their release (see Figure 27).

After-pulses with short delay contribute little because the cells are not fully recharged, but have an effect on the recovery time. Operation at low temperatures elongate the delayed release by a factor of 3 when the temperature is reduced by 25 °C [21].

Another effect to be taken into account is the optical cross talk due to photon travelling to a neighbouring cell which trigger an avalanche.

In fact, in an avalanche breakdown, there are 1–3 photons emitted in average per carriers, with a photon energy higher than the band gap of silicon. These photons may travel to another pixel of the matrix and initiate an avalanche breakdown there. A dedicated design, with grooves between the cells acting as an optical isolation, reduces the cross talk till two order of magnitude. Operation at a relatively low gain is advantageous in this case.

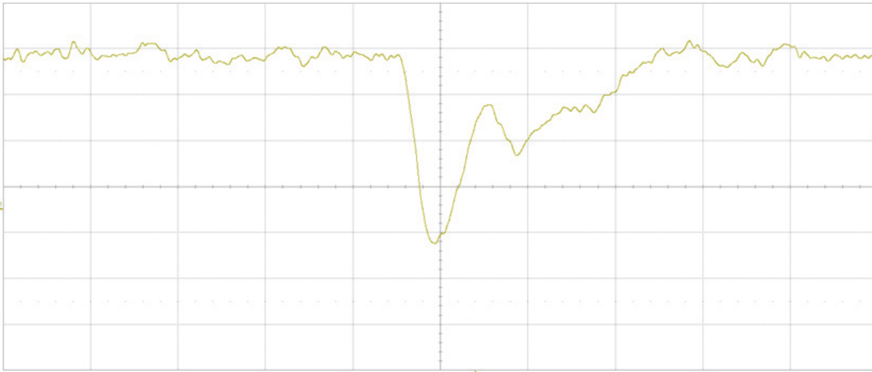


Fig. 27. After pulse event as obtainable at the oscilloscope.

The origin of the cross-talk is presumed to be related to optical photons emitted during avalanche [37] which enter neighboring micro pixels and trigger another Geiger discharge. The probability of causing cross-talk is estimated from the fraction of events with more than one p.e. to that with one p.e. in randomly triggered events without external light. We assume that the events with more than 1 p.e. are caused by the cross-talk from the original Geiger discharge in a single pixel. At low bias voltage, a dark count of 2 p.e. should be related to crosstalk phenomena only because of the low probability that both electrons generate a Geiger discharge.
In order to obtain a complete characterization of the device we have measured the dark counts rate as a function of the supply voltage. For every voltage applied we have performed three measures of rate using three different trigger thresholds: 0.5 p.e., 1.5 p.e. and 2.5 p.e. at 23 °C . Results for these measurements are shown in Figure 28. The noise rate decreases as the temperature becomes lower. The temperature coefficient of noise rate at 0.5 p.e. threshold is −5 %/◦ C. There is a factor 2 reduction of the dark count every 8°C [21, 38]. These observations imply that the dominant component of the noise is due to the discharge of single pixels induced by thermally generated carriers.

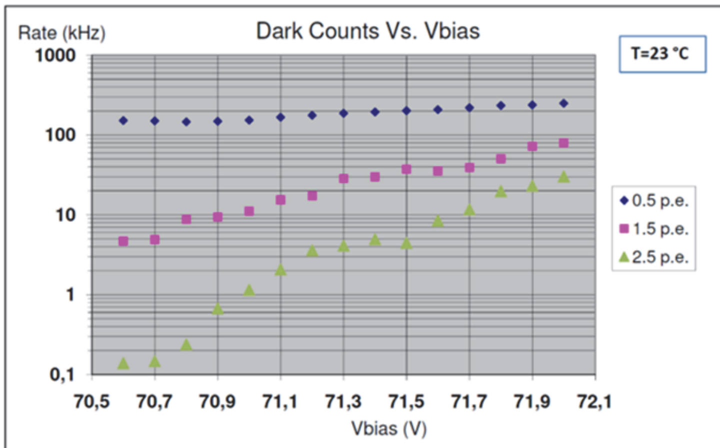


Fig. 28. Dark counts rate generated by the MPPC as a function of the supply voltage.

The measurement of the event rate with 0.5 p.e. trigger gives an estimation of the global noise rate, including the thermal dark counts and the crosstalk events. At 1.5 p.e. of trigger and for low bias voltage, an estimation of the cross talk events only should be possible, since at room temperature we have a low probability that two pixels generate, at the same time, a couple just for thermal excitation.
From the Figure 28 we can remark that the high single rate of the SiPM (if we adopt a low photoelectron threshold) can be easily overcome in those experimental conditions where the time parameter takes a main role. A double coincidence or a gate signal of the right duration can reduce the single rate to acceptable or negligible levels. We have to remind, at this stage of the discussion, that the threshold is of the level of a single or few photoelectrons, a level which would be impossible for classical PMT.
In the following table it is shown the noise rate as a function of the threshold and duration of the coincidence:

gate duration	Treshold 0.5 p.e	Treshold 1,5 p.e	Treshold 2,5 p.e
10 ns	23 Hz	1 Hz	~10^{-10} Hz
20 ns	46 Hz	0.5 Hz	~10^{-10} Hz
50 ns	115 Hz	0.2 Hz	~10^{-10} Hz

These rates are perfectly compatible with the random coincidences rate obtained from the relation $N_1 x N_2 x 2T$. Under these conditions we can see that the dark noise is negligible with respect to the collected events. Moreover, even without artifices like the indicated coincidence technique, with a threshold greater than 3 p.e., the single rate becomes acceptable.

6. Detection efficiency for photons and ionizing particles

The efficiency of an SiPM is the product of several factors and depends on the QE, the geometrical efficiency (ε_{geom}), the Geiger-triggering probability:

$$PDE = QE(\lambda) \times P_{trigger} \times \varepsilon_{geom}$$

The geometrical efficiency ε_{geom} represents the fraction of active area in a micropixel.
Actually, only part of the area, occupied by the micro-cell, is active and the rest is used for the quenching resistor and other connections (see Figure 29). ε_{geom} is defined as the ratio of sensitive to insensitive area, namely the fill factor, and thus depends on the design and layout of the pixels only. It is about 0.3 for a 25 µm pitch sample (as the considered ones) and about 0.7 for a 100 µm pitch sample [34, 39].
The quantum efficiency of the sensitive area is defined by the intrinsic QE of Si (typical QE = 80–90%). The thickness of layers on top of the structure and of the depletion area can be optimized for specific applications. Efficient absorption of photons requires an increase of the thickness in order to maximize photon conversion. On the other side, it is necessary to minimize the depletion area region in order to reduce the dark count rate. Since the QE of the sensitive area is defined by absorption coefficient α in Si, taking into account the probability of reflection of photons on the device surface, photon detection efficiency can be written as:

$$PDE = \varepsilon_{geom}\left(1 - e^{-\alpha x}\right) P_{trigger} (1 - R)$$

where R is the reflection coefficient and x is the position in which the electron-hole pair is generated. The fraction of the light transmitted to the sensitive volume is conditioned by the topmost layers and the resistive one. For short wavelength in the UV region, the situation is more critical. To improve the sensitivity also in this region it is necessary to optimize the top contact technologie, depletion thickness and n-p configurations. The triggering probability $P_{trigger}$ depends on the position where the primary electron–hole pairs are generated and the over-voltage (ΔV). To enhance the triggering probability, we have to take into account that electrons have in silicon a better chance to trigger a breakdown with respect to holes, by about a factor of 2, and their difference decreases with increasing fields, as shown in Figure 30 [40]. If one electron-hole pair is born at position x, then the probability that neither the electron nor the hole causes an avalanche is given by $(1 - P_e) \cdot (1 - P_h)$ where the function P_e is the probability that an electron starting at position x in the depletion layer will trigger an avalanche and the function P_h is the analogous for holes.

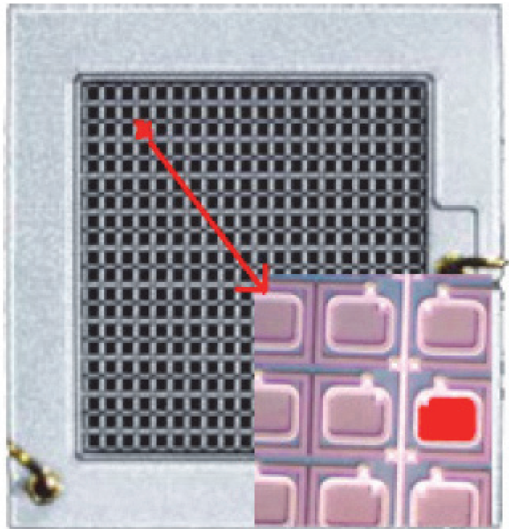


Fig. 29. Matrix of G-APD and evidence of the so called "Fill Factor".

Consequently, the probability $P_{trigger}$ that at x either the electron or the hole initiates an avalanche is given by

$$P_{trigger} = P_e + P_h - P_e P_h$$

Thus, we can write:

$$PDE = \varepsilon_{geom}\left(1 - e^{-\alpha x}\right)(1 - R)\left(P_e + P_h - P_e P_h\right)$$

In case of a photo-generation event, two carriers are created travelling in opposite directions at the absorption point. The contribution to the PDE can be calculated as a function of the generation position by solving two differential equations involving the carrier ionization

rates. If conversion happens in the p depleted region, x is equal to the depleted region thickness (see Figure 2).

In a conventional structure n^+-p-π-p^+, when a pair is generated in the upper side of the high-field region (n^+), the electron is directly collected at the n^+ terminal (see Figure 31); thus, it does not contribute to the triggering. The hole is forced to pass the whole high-field triggering the avalanche. On the contrary, when the pair is generated in the bottom side (p), the situation is symmetrical and only electrons contribute to the triggering probability. So the triggering probability depends on the position where the primary electron–hole pair is generated and on the overvoltage. A high gain operation is favoured.

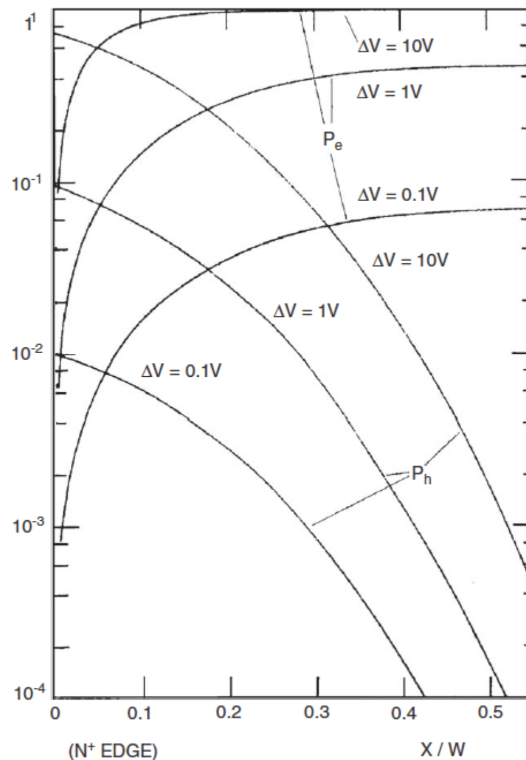


Fig. 30. Avalanche region with width W and the position X which runs from 0 to W starting at the n-edge.

Thus, to maximize the triggering probability, the photon conversion should happen in the p side of the junction, in order to allow the electrons to cross the high-field zone and trigger the avalanche.

As an example for λ>450 nm (green and red light) photons convert deep in p-silicon beyond the high-field region. Electrons drift back into the high-field region, triggering avalanches. Hence in this wavelength range the efficiency is very high. For λ<400 nm photons are absorbed in the first microns of the n^+ layer. Here the holes drift into the high-field region and trigger the avalanche. Under these conditions the QE is reduced in this wavelengths

range. As a reference for λ = 400 nm (corresponding to photon energy = 3.10 eV) the absorption coefficient is 1.2×10^5 cm^{-1} and the thickness required to absorb more than 99% of the light is ~1µm (see Figure 5, where the absorption length as a function of the wavelength is shown) [41-43]. Several solutions exist for increasing the sensitivity at short wavelengths:

- an higher reverse bias voltage would increase the avalanche probability for holes, though the voltage has to be limited due to the increase of cross talk and dark rate
- entrance windows has to be made as thin as possible [44, 45]
- the n^+ layer has to be as shallow as possible (for optimum QE); with standard equipment for detector fabrication, layers with a junction depth of 100 nm can be obtained. The high-field region should be as thin as possible in order to convert photos beyond it.
- Triggering probability can be improved by maintaining the same doping profile configuration but reversing the types, i.e. having a p^+-n-n^--n^+ structure, and making the junction deeper (> 0.4 µm). Hence the roles of electrons and holes are reversed, resulting in avalanches triggered by electrons at short wavelengths (Figure 31).

In conclusion, to maximize the triggering probability: (i) the photo generation should happen in the p side of the junction in order for the electrons to pass the whole high field zone, and (ii) the bias voltage (V_{bias}) should be as high as possible.

A better scenario is obtained when electron bombardment is considered. In Figure 32 a simulation for the range of electrons penetrating into the silicon is shown. The simulation has been computed by using Geant4 Simulation Toolkit [46, 47]. If ionizing particles, like electrons, are detected in a n^+pp^+ junction, the range - i.e. the energy - will determine where the carriers are generated. If the end of range is in the p region beyond the high-field area, both carriers created along the track will be travelling in the opposite directions, contributing to the avalanche-triggering probability. Electrons detection efficiency can be evaluated from the following:

$$EDE = \varepsilon_{geom}(1 - R_{back})P_{trigger} = \varepsilon_{geom}(1 - R_{back})\,(P_e + P_h - P_e P_h)$$

where P_e and P_h are the electron and hole breakdown initiation probabilities and R_{back} is the backscattering probability. When a pair is generated before the high field region, the electron is collected at the n^+ terminal; thus, it does not contribute to the trigger. The hole is forced to pass through the full high-field region and so its triggering probability is given by P_h. For pairs generated beyond the high field region, the situation is reversed and only electrons contribute to the triggering probability P_e. These probabilities depend on the impact ionization rates of holes and electrons, respectively. As pointed out above, the electron has an ionization rate of about a factor 2 higher than the hole.

The reduction of the thickness in n^+ layer allows lowering the detectable electron energy. As an alternative, maintaining the same doping profile configuration but reversing the types, i.e. using a p^+nn^+ structure and making the junction deeper, can improve the triggering probability. In this case the electron range is completely contained inside the p^+ region.

6.1 Dynamic range

SiPMs produce a standard signal when any of the cells goes to breakdown. When many cells are fired at the same time, the output is the sum of the standard pulses. Single photonsproduce a signal of several millivolts on a 50 Ω load. For a matrix of $N_{microcells}$ microcells, the dynamic range is limited by the condition that ($N_{ph} \times PDE/\ N_{microcells}$)<1, where N_{ph} is the number of photons, and PDE the Photon Detection Efficiency of the SiPM.

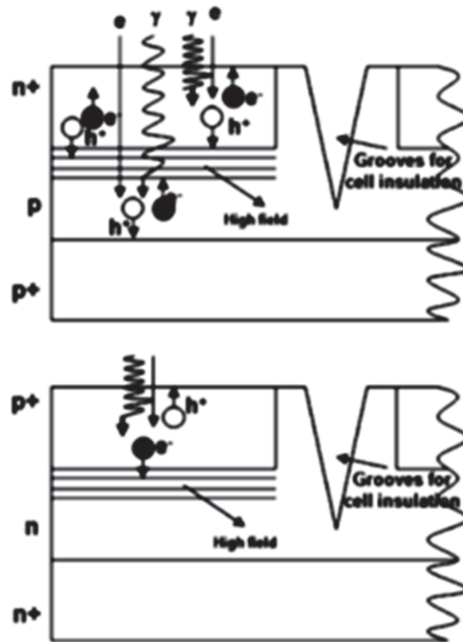


Fig. 31. Photon and electron avalanche induced in the two silicon configurations (p^+nn^+ and n^+pp^+).

In other words, the average number of photons per cell should be less than 1. If the number of detected photons is much smaller than the number of cells, the signal is fairly linear and saturates when the number of photons is about equal to the number of cells. Saturation is well described by:

$$N_{signal} = N_{microcells} \times \left[1 - \exp\left(\frac{-N_{ph} \times PDE}{N_{microcells}} \right) \right]$$

6.2 Timing

The active layers of silicon are very thin (2–4 mm), the avalanche breakdown process is fast and the signal amplitude is big. Therefore, very good timing properties even for single photons can be expected. Fluctuations in the avalanche development are mainly due to a lateral spreading by diffusion and by the photons emitted in the avalanche [48, 49]. As shown in Figure 34 for the case of 1x1 mm^2 MPPC, operation at high overvoltage (high gain) improves the time resolution.

The dependence of the FWHM as a function of the number of photoelectrons as shown in Figure 35 is in fair agreement with Poisson statistics. The resolution with 15 photo-electrons, typical of applications where SiPM are coupled to small volume, high light yield scintillators, is better than 25 ps.

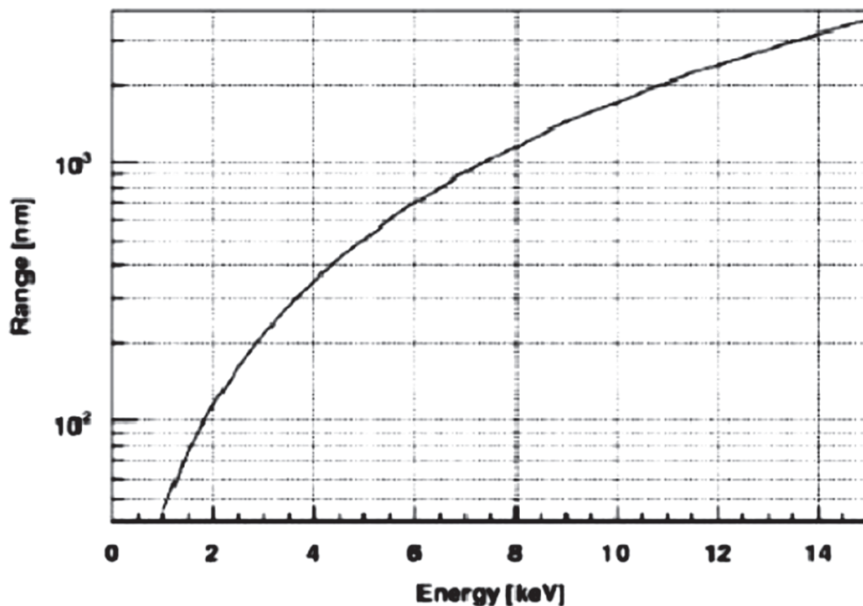


Fig. 32. The range of electrons in Silicon as obtained from a GEANT4-based simulation.

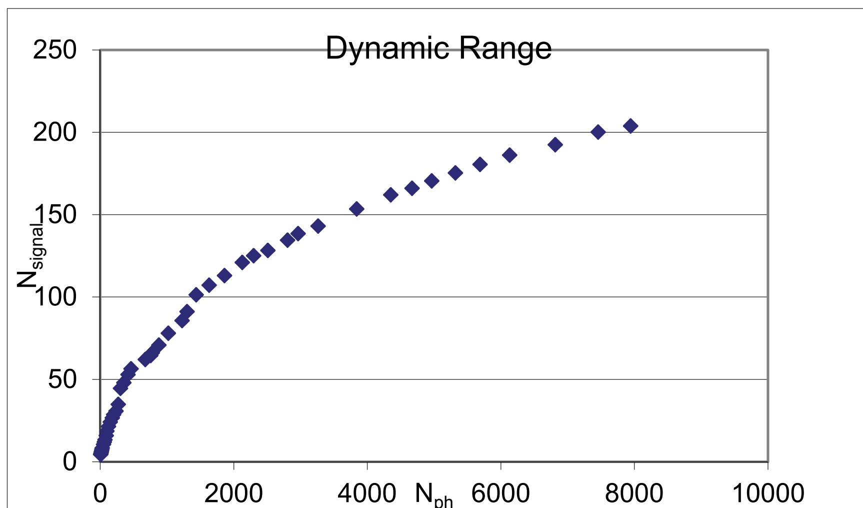


Fig. 33. Dynamic range

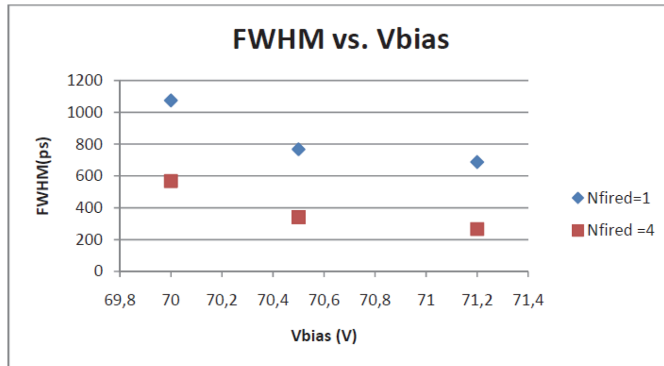


Fig. 34. Time resolution for 1 and 4 photons for the 1x1 mm² MPPC as a function of V_{bias}.

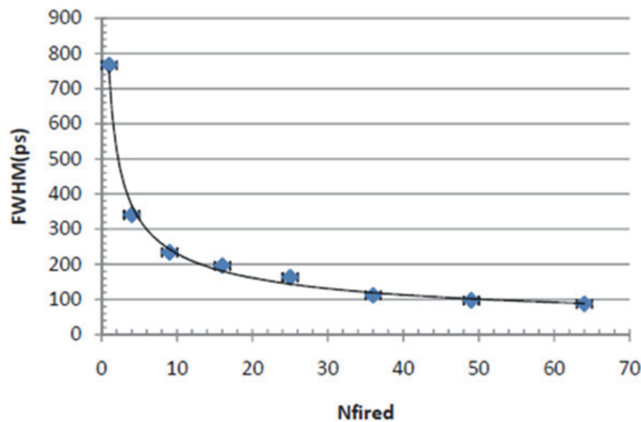


Fig. 35. Time resolution as a function of the number of fired pixels

7. New concepts for semiconductor photomultiplier

The present commercial production of avalanche Geiger-mode photodiodes gives the starting point for a new photomultiplier age, based on p–n semiconductors. As an example, in the Hamamatsu production at least three types of n^+pp^+ Multi-Pixel Photon Counter (MPPC) exist: 1600 (25µmx25µm), 400 (50µmx50µm) and 100 (100µmx100µm) pixels segmented onto a 1x1-mm² total active area. The achieved gain, 10^5–10^6 at 70–72 V reverse bias voltage, makes possible the one photon level detection. The dark count rate is suppressed to a few hundreds kHz level, by setting a threshold at 0.5 p.e.. It decreases to 1 kHz for 1.5 p.e. and it is not significant for 2-3 p.e. Thermally generated free carriers can be further reduced by cooling the device. The temperature coefficient of noise rate at 0.5 p.e. threshold is -5%/°C. With the present structures the most sensitive wavelength region is around 400 nm where the PDE is 25% for the 1600 pixels type, 50% for the 400 pixels type and 65% for the 100 pixels type [34], reflecting the higher geometric factor value.

At present, the silicon wafer cost and the thermal dark current limit the dimensions of the SiPM photodetector at a few mm^2.
Now the question is how to detect photons from large surfaces and/or volumes.
Their transport and/or focusing from surface and volume can be achieved in three different ways:

1. collecting photons and conveying them towards a single SiPM device;
2. enlarging the sensitive detector area by ordering several SiPMs in a pixelated matrix shape or by focusing the light to the sensitive area by Winston cones, pyramidal waveguides or lenses
3. making a photon conversion by a vacuum hemispherical photocathode which focuses photoelectrons on a SiPM (VSiPMT).

7.1 SiPM coupled to WLS fibers

The reduction of geometrical area can be obtained by using wavelength shifter fibres embedded in the plastic scintillator body and connected at the other end to the SiPM.
Light collection from large scintillators or complex geometries can sometimes be aided through the use of optical elements that employ wavelength shifting technique. Many liquid or plastic scintillators incorporate an organic additive whose function is to absorb the primary scintillation light and reradiate the energy at a longer wavelength. It is emitted isotropically uncorrelated respect to the direction of absorbed light.
The same light collection principle can be applied using plastic fibers whose contains a wavelength medium. For best light propagation along the fiber one want a large shift between the optical absorption and the emission band so that minimal self-absorption takes place.
One of the first experience in this technique has been achieved in T2K experiment with the usage of wavelength shifter fibers. In this application [50], the counters are readout via WLS fibers embedded into S-shaped grooves in the scintillator from both ends by multi-pixel avalanche photodiodes operating in a limited Geiger mode. A customized 667-pixel MPPC was developed for T2K by Hamamatsu Photonics [51] with a sensitive area of $1.3 \times 1.3\ mm^2$ and a pixel size of $50 \times 50\ \mu m^2$; the sensitive area is larger than those available previously and relaxes the mechanical tolerances required for coupling to the WLS fibers used extensively in the experiment.

7.2 Compound Parabolic Concentrators (CPC)

Image compression from large-surface detectors can be realized using matrices of single SiPM pixels. Such a device is particularly suitable in experiments detecting the Cherenkov or fluorescence light in the atmosphere. A light concentrator can be used to enhance the number of incident photons on the sensitive surface of the detector.
Some experiment for VHE gamma-ray astronomy (as example VERITAS [52], MAGIC [53] and HESS [54]) already use non-imaging light collectors to concentrate light on photomultiplier tubes, while light concentrators are also widely used in diverse fields, as solar energy production.
Compound Parabolic Concentrators (CPC), also known as Winston cones [55], are light-collection devices intended to concentrate light on a smaller area by maximizing photon density per unit surface. Characteristic parameters for CPCs are: dimensional geometry, compression, acceptance angle and collection efficiency. CPCs are usually produced with amorphous (vitreous) materials like commercially available B270 and BK7.

A three-dimensional compound parabolic concentrator is designed by rotating a parabola around the optical axis. The analytical description of the CPC profile is given by the following equation [55]:

$$(r\cos\theta_{max} + z\sin\theta_{max})^2 + 2a'(1+\sin\theta_{max})^2 r - 2a'\cos\theta_{max}\ (2+\sin\theta_{max})^2 z - \\ -a'^2(1+\sin\theta_{max})(3+\sin\theta_{max}) = 0$$

in which θ_{max} is the acceptance angle and a' the exit aperture radius. r and z are, instead, the reference axes as shown in Figure 36.

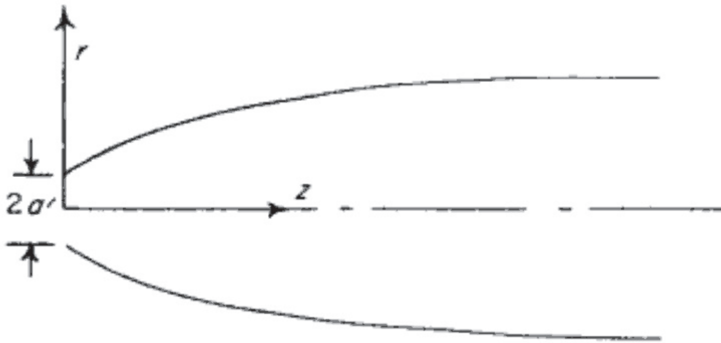


Fig. 36. CPC Profile

As shown in Figure 37 a Winston cone is a double paraboloid built from two off axis parabolas, such that the focal point of one falls to the edge of another. The reflecting surface is obtained by rotating the parabola around the concentrator axis.

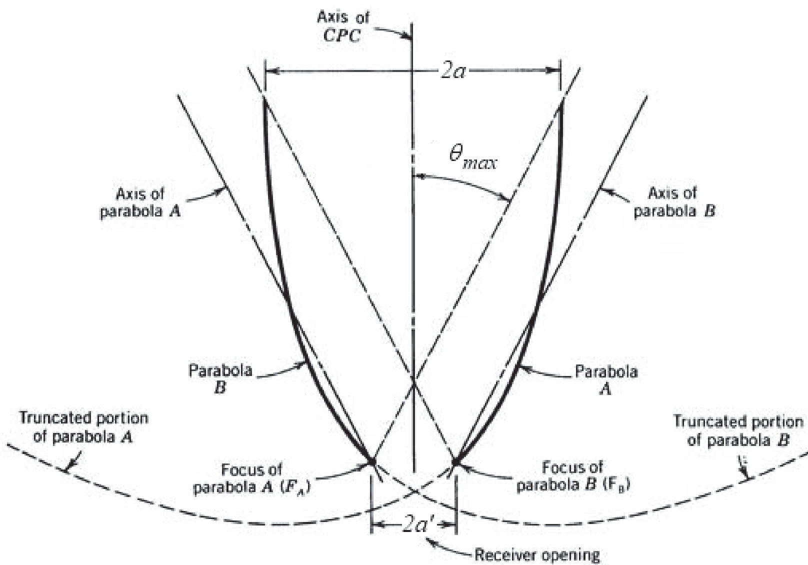


Fig. 37. CPC profile and acceptance angle (**θ_{max}**).

The overall length of the parabolic concentrator is conditioned by the symmetry that must ensure to pass both edging rays and is thus limited by the maximum entrance diameter. The overall length is given by:

$$L = \frac{a' (1+\sin\theta_{max})\cos\theta_{max}}{\sin^2\theta_{max}}$$

Since the diameter of entrance surface is:

$$a = \frac{a'}{\sin\theta_{max}}$$

resulting:

$$L = \left(a + a'\right)\cot g\,\theta_{max}$$

A useful ratio for describing the characteristics of a concentrator is the geometrical concentration ratio or compression [55] defined as:

$$C = entrance\ surface / exit\ surface$$

The theoretical maximum concentration ratio for a three-dimensional design is thus given by:

$$C_{max} = \frac{a^2}{a'^2} = \frac{1}{\sin^2\theta_{max}}$$

where θ_{max} is the acceptance angle. The acceptance angle (or the cut-off angle) is the angle beyond which most of the light entering the concentrator is reflected out of it: the rays inside the collector undergo multiple reflections, and some of the rays that enter at angles smaller than the limit value can be turned back; some rays incident at angles larger than the limit angle are instead transmitted.
The optical concentration ratio, considering losses (the optical efficiency), is the amount of emerging light at the exit aperture compared to the amount of the incident light on the entrance aperture. The attenuation in the concentrator results from reflection losses, scattering and absorption. Light collection efficiency depends on the radiation incident angle (relative to the Winston cone symmetry axis) and on the acceptance angle. In particular, the efficiency drops as large as the acceptance angle.
Thus, defining the transmission efficiency ε_{trans} as:

$$\varepsilon_{trans} = n_d / N_{phot}$$

where n_d is the number of photons reaching the exit aperture and N_{phot} is the total number of photons penetrating the entrance aperture, the collection efficiency (ε_{coll}) can be written as:

$$\varepsilon_{coll} = \varepsilon_{trans} \,.\, C_{max}$$

The collection efficiency is strictly related to the number of multiple reflections before reaching the exit aperture.
Even if the CPC have been designed for solar energy applications, their utilization in low photon detection is attractive to extend the detection surface. In this case, also the impact point

on entrance surface has to be taken into account since this leads to a non homogeneous efficiency. As we will show, better results are obtained with acceptance angles lower than 5°.
As shown in Figure 38/left (related to a cone with an acceptance angle of 10° and 8° incident photons), it's possible to identify areas on the entrance surface for which the number of multiple reflections is almost the same.

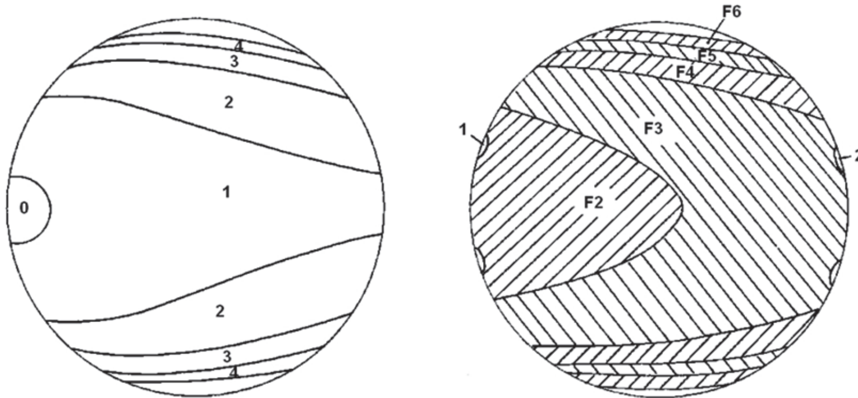


Fig. 38. Left: Transmission zones of rays on the entrance surface of a CPC having θmax = 10° for impinging photons at 8°; Right: Rejections zones of rays on the entrance surface of a CPC having a θmax of 10°, for impinging photons at 11.5°: remark in (1) two little zones where the transmission is preserved.

Likewise, it's possible to identify the areas where the photons exit from the entrance surface after reflections inside the CPC, without reaching the exit surface. These areas are shown in the Figure 38/right, for an 11,5 ° incident photons.
In order to estimate the collection efficiency of the light concentrator and to study its dependence on the length of the funnels and on the angle of incidence, we carried out several Monte Carlo simulations. Photons with given direction were produced at the entrance aperture and their path was followed until they were either absorbed by the funnel walls or left the funnel through one of the apertures. Various types of paraboloids and pyramidal light concentrators were examined in the simulations.
Figure 39 shows a Winston cone simulated with a 0° acceptance angle with an entrance and exit apertures of a radius of 28 mm and 5.5 mm, respectively, corresponding to a concentration ratio of 25.91.
The transmission and collection efficiency for this device as a function of photon incident angle (with an uniform distribution of photon impact point) is shown in Figure 40.
As shown in Figure 40, the transmission efficiency, evaluated as n_d/N_{phot}, is strongly suppressed for non-perpendicular photon incident angles, in the case of devices designed with a 0° acceptance angle. However, the efficiency is about 50% for 10° incident angles.
The collection efficiency takes into account also the compression ratio ($\varepsilon_{coll} = \varepsilon_{trans} \cdot C_{max}$); simulation shows that, at an incidence angle of about 20°, the collection efficiency is 1: the density of photons on the entrance surface is the same of that on the exit surface and the concentrator is useless.

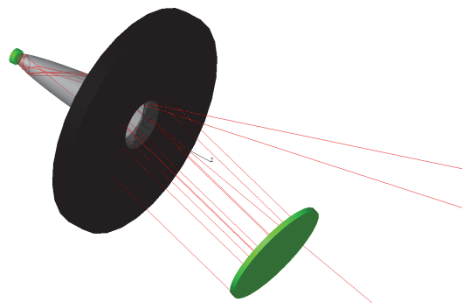


Fig. 39. 3-D model of the CPC with an acceptance angle of 0° used for the simulation.

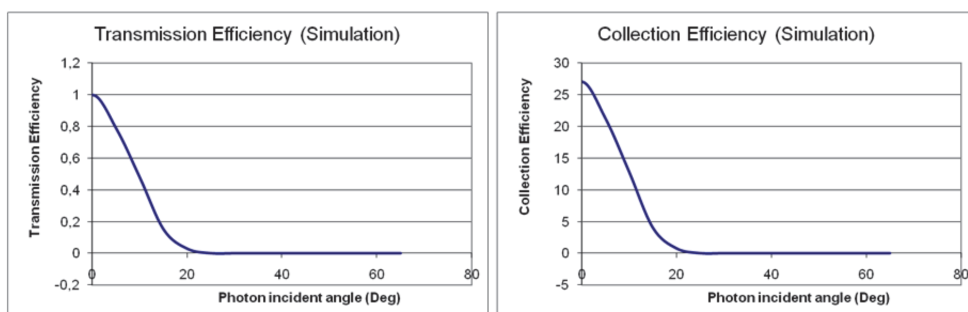


Fig. 40. Simulated transmission (left) and collection efficiency as a function of incident photon angle.

It makes sense the use of a CPC to increase the detection surface of a silicon device only if the devices have a large acceptance angles. To explore this option, a detailed simulation of a CPC with 25° acceptance angle ($CPC_{25°}$) has been performed. $CPC_{25°}$ is an optical glass B270 cone having 9.01 mm entrance diameter, 2.50 mm exit diameter, and is 19.25 mm long, commercially available by Edmund Optics [56]. Figure 41/left shows the tridimensional model used in the simulations while Figure 41/right shows the $CPC_{25°}$ used for the measurements.

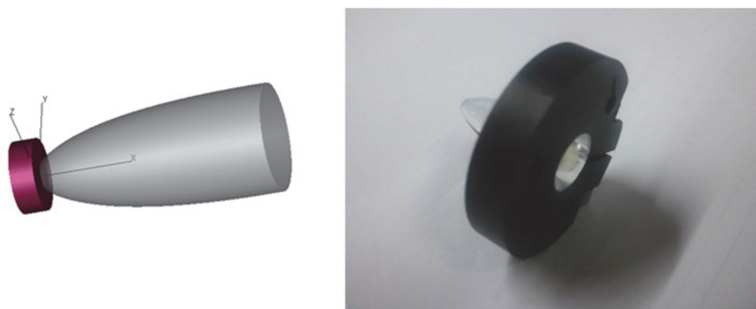


Fig. 41. Left: 3-D model of the $CPC_{25°}$ having an acceptance angle of 25°. Right: A photography of the CPC cone used for measurements mounted on its support.

Simulations show that the concentrator is able to transmit photons with incident angle up to about 25° with a good collection efficiency, ranging between 0.5 and 0.8 depending on the incident angle. A small lack in the transmission efficiency is evidenced for 0° incident photons, with respect to the case of a 0° acceptance angle designed one.
In order to experimentally check for the simulation accuracy, we used two different settings. The first set up is shown in Figure 42.

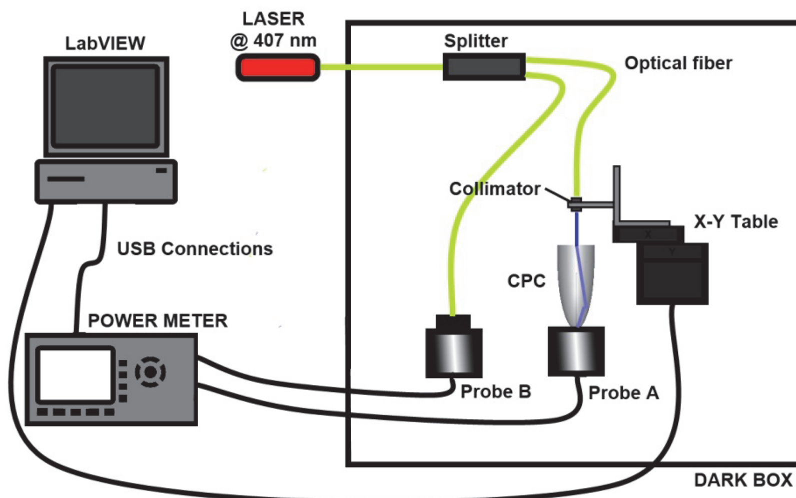


Fig. 42. The experimental setup.

The efficiency of a $CPC_{25°}$ has been measured as a function of the impact position on the concentrator entrance surface by employing a computer controlled x-y movement with a position precision of tens of microns. A λ=407 nm highly collimated pulsed laser beam (spot diameter = 0.9 mm) has been sent on the $CPC_{25°}$ and both beams at the entrance and exit surfaces have been measured by a double channel Power Meter Newport mod. 2936-C.
The intrinsic efficiency measured is strongly dependent on the impact point of the photon on the entrance surface, as shown in Figure 43.
The measurements of the transmission efficiency along the $CPC_{25°}$ diameter, superimposed to the simulation results, are shown in Figure 44 for impinging photons at 0° and at 20°: the simulation correctly reproduces the measurements and confirms the efficiency dependence on the photon impact point on the entrance surface.
Discrepancy in the maximum collection efficiency between data and simulation is mainly due to the optical coupling between the $CPC_{25°}$ and the Power Meter probe, but also to the photon absorption in the B270 glass, to the intrinsic Power Meter probe efficiency and its dependence on the photon incident angle.
The same measurement has been repeated for several angles of incidence of photons .
This study shows as, to increase the detection surface of a SiPM by using parabolic concentrators, devices designed for small acceptance angles are preferred, with the better choice corresponding to 0°, but limiting in this way the field of view. In any case, the detection of radiation produced at fixed angles, as in experiments in which Cherenkov radiation has to be detected, can profit of this solution.

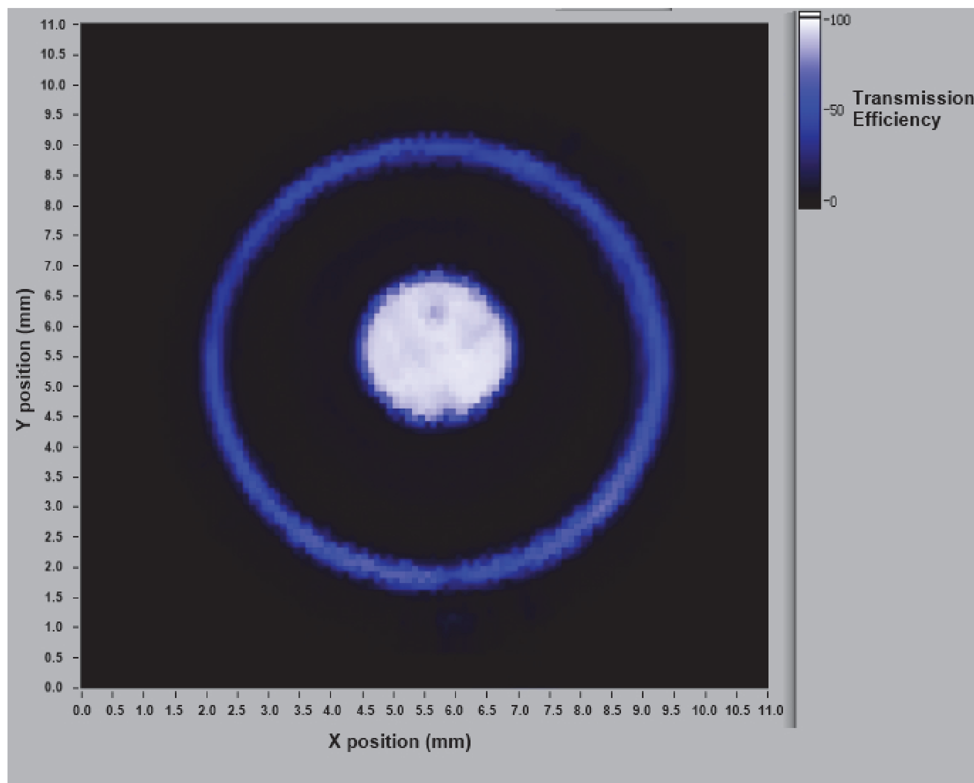


Fig. 43. Transmission efficiency on the entrance surface of the $CPC_{25°}$.

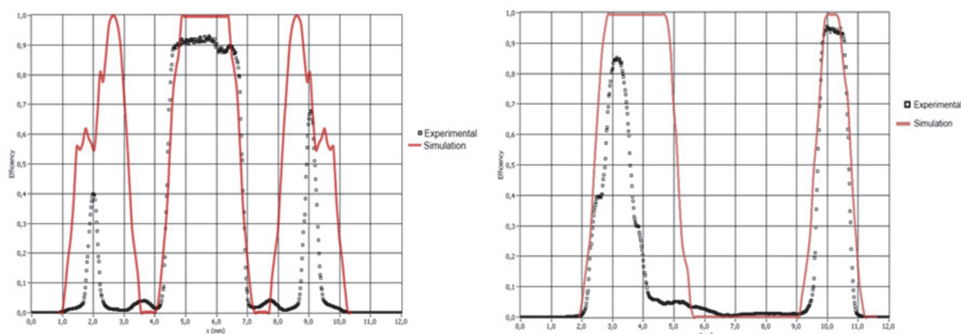


Fig. 44. Comparison between simulation and actual data on the $CPC_{25°}$ (CPC having an acceptance angle of 25°) for impinging photons at 0° (Left) and at 20° (Right).

Differently, a pyramidal device can be considered. Simulation studies have been carried on this geometric structure (Figure 45).

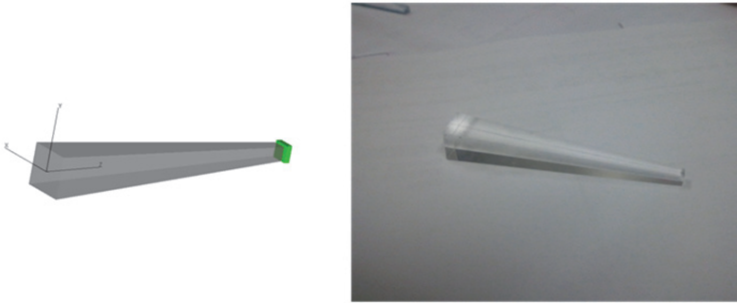


Fig. 45. 3-D model of the pyramidal concentrator used in the simulation and a photography of the pyramidal concentrator used for measurements

The pyramidal concentrator simulated in this work is an optical glass N-BK7 device, with 7.5 x 7.5 mm^2 entrance surface, 2.5 x 2.5 mm^2 exit surface and 50 mm long, commercially available by Edmund Optics [56]. Figure 45/right shows the pyramidal light concentrator used for the measurements.

From simulation, a good transmission efficiency, almost uniform up to 20° is obtained for this geometry, as shown in Figure 46.

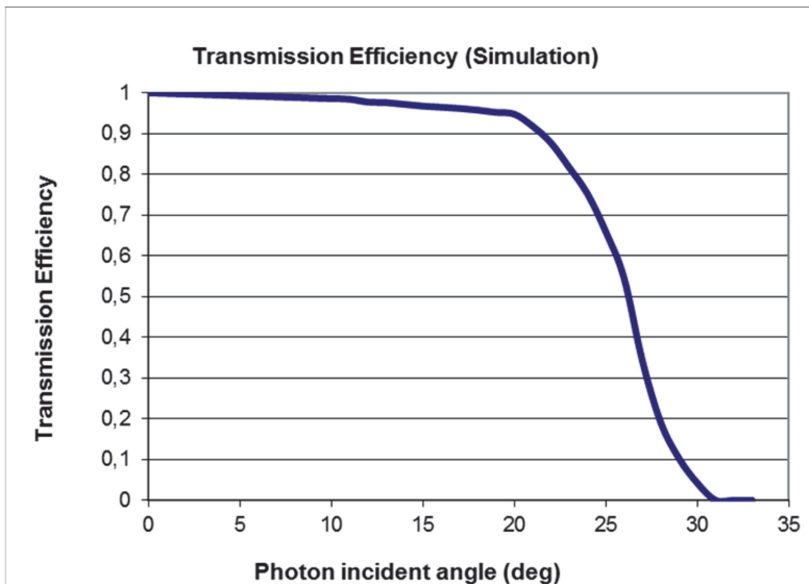


Fig. 46. Efficiency curve of the pyramidal concentrator as a function of the impinging angle of the photons.

Also for this concentrator, the efficiency has been measured as a function of the impact position on the entrance surface, by using the experimental set-up described in Figure 42. The transmission efficiency measured is shown in Figure 47, Figure 48 and Figure 49 for different incident angles.

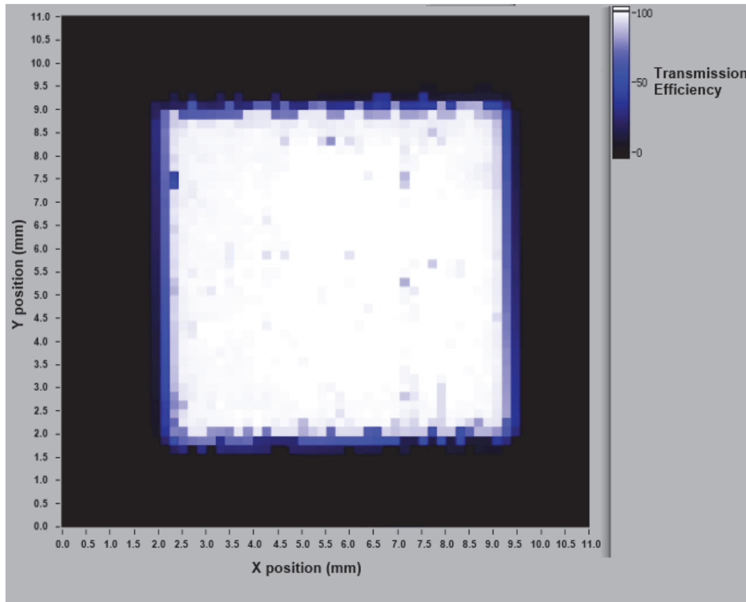


Fig. 47. Measurement of the transmission efficiency on the entrance surface of the pyramidal concentrator for 0° impinging photons.

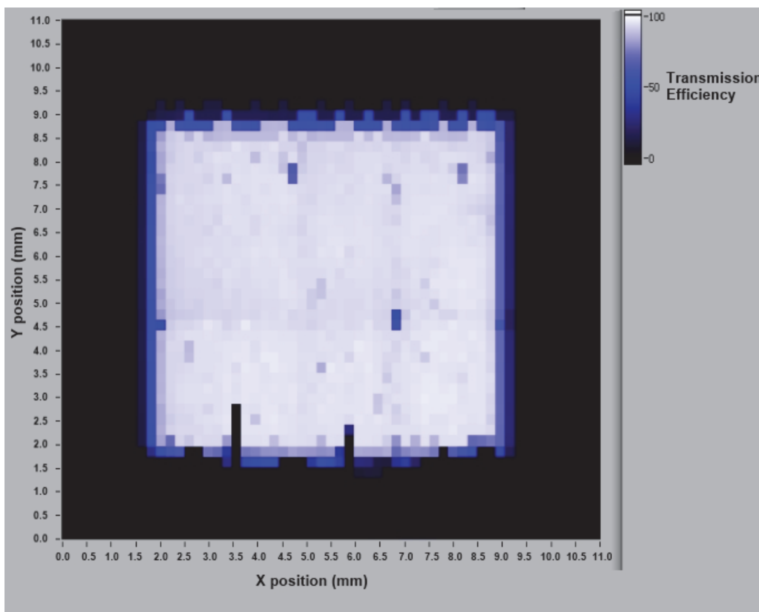


Fig. 48. Measurement of the transmission efficiency on the entrance surface of the pyramidal concentrator for 5° impinging photons.

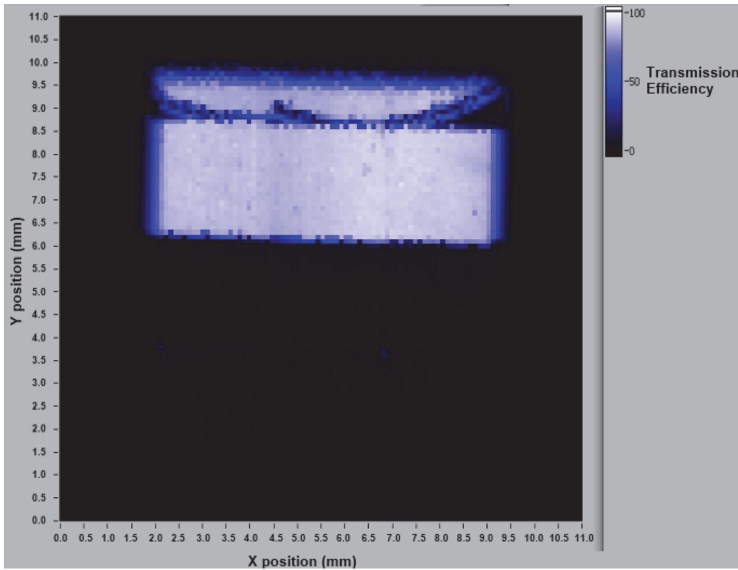


Fig. 49. Measurement of the transmission efficiency on the entrance surface of the pyramidal concentrator for 10° impinging photons.

Measurement results show that transmission efficiency of such a pyramidal light concentrator has a slight dependence on the impact point except for the case of incident angles of 10°. In fact, as shown in Figure 49, in this case large part of the entrance surface (about an half) results in a very low efficiency (about 10%). Actually, this partial inefficiency may be due to the very large angles of exiting photons (>60°), out of the angular acceptance of the Power Meter probe.

A second methods adopted to evaluate the effect of light concentrators is based on the measurement of photons detected with a 3x3 mm² MPPC S10931-025P by Hamamatsu arranged at the exit surface of the concentrator (Figure 50).

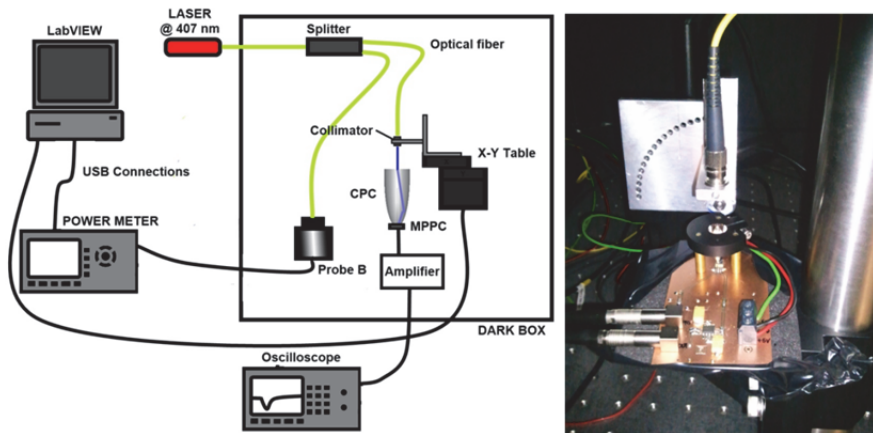


Fig. 50. The experimental set-up: light concentrator on the MPPC

The average number of incident photons on the concentrator, can be determined by measuring the laser power on one of the two outputs of the splitter. The MPPC signal, amplified by the National Semiconductor LMH6624 chip as already described, has been measured on the oscilloscope and an estimation of the overall efficiency of the system (light concentrator + MPPC) has been done by evaluating the number of photons detected by the MPPC. Laser power has been set to 40 photons per pulse on the surface of the $CPC_{25°}$, with an incidence angle of 0°. The maximum total efficiency for several distances of impact point from the center along one diameter (Figure 51) has been measured. Results show that a maximum obtainable efficiency is 0.1, while observing dynamic range it is possible to note that this shape of concentrator is not useful to increase the field of view of a MPPC.

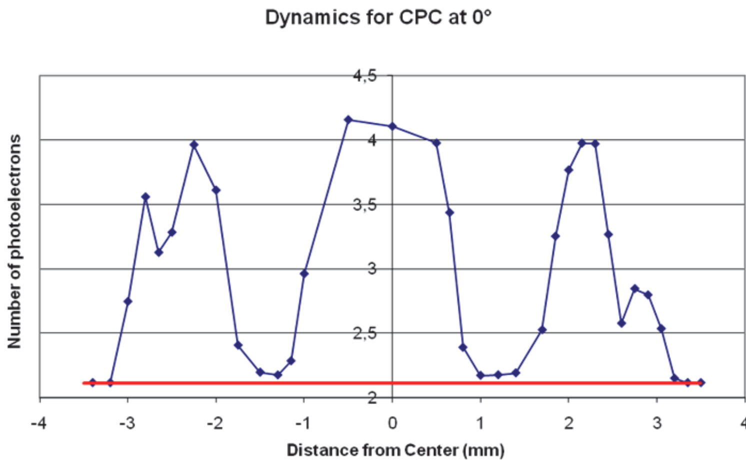


Fig. 51. Number of photoelectrons as a function of the distance of impact point from the centre

The same experimental set-up has been used to measure the efficiency of pyramidal concentrator (see Figure 52).

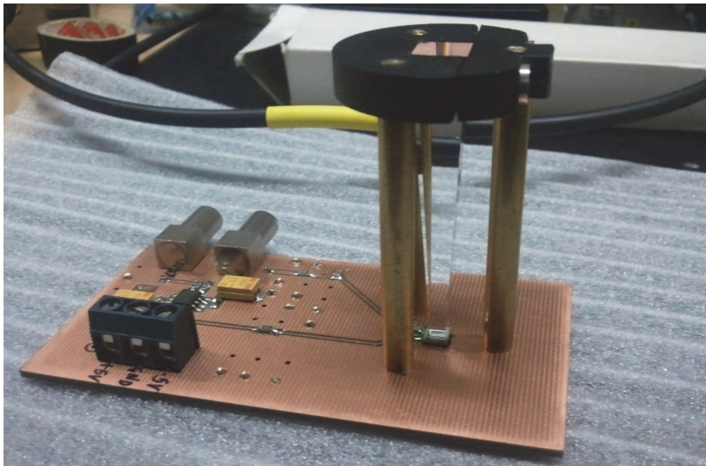


Fig. 52. Experimental set-up for pyramidal concentrator using an MPPC as a sensitive device.

Figure 53 shows the efficiency measurement for several distance of impact point from the center (along one dimension) and for several incident photon angles. Again it can be observed that the maximum total efficiency is 0.1, but in this case the total surface is enhanced. Furthermore, the efficiency results to be very uniform for incident angles in the range 0°-10°.
The efficiency of the concentrator will be surely enhanced by developing SiPMs in which a front-side structure with quenching resistors is integrated into the silicon bulk. In this mode obstacles for light like metal lines or contacts, can be omitted and therefore the fill factor would only be limited by the gaps indispensable for the optical cross-talk suppression and it can reach in principle 100% [57]. Being the fill factor of present SiPM of the order of 30% (see Figure 13 and Figure 16) and taking into account it represent the main contribution to the concentrator efficiency, it will be possible to achieve very interesting overall efficiencies of the order of 35-40%. In this way, the features of a SiPM and concentrator can largely overcome the properties of the classical PMT.

8. Hybrid photo detectors

One option to further improve the angular coverture of a silicon device could be to combine it with vacuum technology. One can replace the dynodes structure of a photomultiplier with a silicon photodiode (HPD – Hybrid Photo Detector or with an avalanche photodiode (HAPD - Hybrid Avalanche Photo Detector). The detection of photon starts as in an ordinary photomultiplier at the photocathode, where a photoelectron is produced.

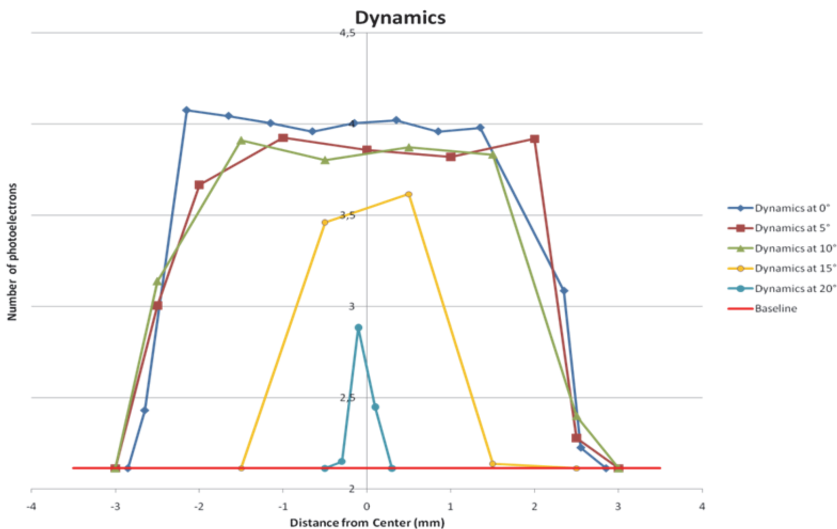


Fig. 53. Efficiency measurement for several distance of impact point from the center (along one dimension) and for several angles of the incident photon.

A pioneering configuration using a gain G = 1 p–i–n diode has been proposed 10 years ago by Hamamatsu and DEP. Using p–i–n diodes the signal amplification is given by the number of electron–hole pairs produced by electrons emitted from the photocathode, accelerated by a high-intensity electric field and injected into the target diode. The gain depends on the energy of incident photoelectrons: at 4 kV operational voltage, about 1100

electron–hole pairs are generated. However, in many applications where an higher sensitivity is required, such a small gain (further reduced by the photocathode QE) limits the low level light detection capability.
More recently, a gain improvement was obtained using diodes in avalanche regime (APD), with a resulting additional gain of ~20–50. The overall photomultiplier gain increases, yet not enough for low light level detection. In addition, the fluctuations in the avalanche multiplication, limit the useful gain range.
The idea, which starts from these experience and from the consolidation of the Geiger-APDs, is to combine them into a sort of classical vacuum tube: electrons emitted by a photocathode can be collected and focused on an array of G-APDs, which acts as the amplifier. The junction works as an electron multiplier with a gain of 10^5–10^6, equivalent to the dynode chain of a classical VPMT. Thus the Vacuum Silicon Photomultiplier Tube (VSiPMT) would consist of the following:

- a photocathode for photon–electron conversion,
- an electric field to accelerate and focus the photoelectrons on a small area covered by the G-APD array,

With 10 kV between photocathode and SiPM, the range of 10 keV photoelectrons impinging on the silicon is 1.5 μm (as shown in Figure 32). In a conventional front illuminated n^+pp^+ structure, the n^+ layer has to be enough shallow (≤ 0.5 μm) in order to have an efficient ionization in the p region. With standard equipment for detector fabrication, layers with a junction depth of 100 nm can be obtained [32]. In addition the high-field region should be as thin as possible in order to have more ionization beyond it, maximizing the electron trigger probability. Otherwise, at lower voltages, a junction structure p^+nn^+ is needed. The ionization should happen in the p^+ side allowing electrons to pass the whole high-field zone with high avalanche efficiency. In both cases photoelectrons spend most of the ionization within the p layer thickness. In contrast with the photon detection, the precise range of the photoelectrons permits the optimization of the thickness of the junction n^+pp^+ or p^+nn^+ layers, minimizing the depletion region with great advantage for lowering the dark current, increasing at same time the efficiency and the time resolution.
Recently C. Joram and al. at CERN [58] performed a very interesting study concerning the response of SiPM devices to the electrons impinging. Results they found have been so encouraging and interesting to influence also the title they give to the article: "Proof of principle of G-APD based hybrid photo detectors". In practice they demonstrate the feasibility of such a solution, already preannounced by our group in [2]. The CERN group, thanks to their usual activities, had the possibility of making a detailed test just dedicated to this aspect. We consider their result as a precious confirmation of our previous theories concerning this arrangement.
Exploiting the full fill factor of a front illuminated SiPM described in a previous paragraph, an ultimate design for a new semiconductor hybrid photomultiplier is possible. It consists of a hemispherical vacuum tube with a deposited photocathode and a special SiPM in which quenching resistor and electric contacts are integrated in the bulk. The admittance of such a component on the sensitive surface allows the full geometrical efficiency of a SiPM used as amplifying element, making it very attractive also in this application. So in this way this hybrid PMT results equivalent to those, already existing, manufactured with APD (gain 10^2), but with a gain comparable to the standard PMT (10^6-10^7).
In conclusion, photoelectrons emitted by the photocathode are accelerated, focused and then amplified by Geiger junctions (Vacuum Silicon PhotoMulTiplier,VSiPMT,). Such an

amplifier, which would substitute the classical dynode chain, presents several attractive features such as small size, low cost, high gain, high efficiency, absence of an external voltage divider, no power consumption, weakened dependence on magnetic fields.
These developments will offer an attractive response to the necessity of increasing active surfaces with high sensitivity.

8.1 SiPM in cryogenics

As a last point we wish to underline the possible applications of SiPMs in cryogenic environments. Both SiPM matrices and VSiPMTs can be considered as a promising alternative to classical photomultipliers for VUV scintillation detection in liquid noble gas experiments requiring a high sensitivity to very low energies to detect neutralino Dark Matter signals (see for instance [59, 60]). The drop of thermal noise at low temperatures, three orders of magnitude at -95° for liquid Xenon, enhances the linearity and the single photon detection capability. The above considerations regarding the front and back illuminated SiPM induce to envisage a quantum efficiency of 25% in VUV region [61]. These figures encourage the use of these new devices in cryogenic environment [62].

9. Conclusions

The so-called silicon photomultipliers (SiPMs, MPPCs by Hamamatsu, etc.) are already replacing photomultiplier tubes in many applications. Recently this new photon detector is operative on different R&D lines in various applications. High performances characterize such a device: reduced dimensions, high gain, low power supply, single photon sensitivity, magnetic field operation, very good time resolution, low cost. In particular low power supply and negligible power consumption together with a single photon counting, make it very interesting also in hostile environments (space, astronomy) and suitable for a wide variety of applications (including medicine and biology). The drawback of this detector are the reduced dimensions limited by the dark count and silicon cost. To collect photons both from great surfaces or/and volumes and to increase the field of view of such a photo detector, three solutions have been presented in this paper. Apart the classical and proven technique of detecting scintillation light from large volume using WLS fibers, which diameter dimensions are compatible with the active part of the present commercial SiPM, two innovative solutions are presented to improve the field of view and/or focus the light: cones and pyramidal light concentrators assembled in single SiPM or organized in a matrix shape. Such a device are applied both to read Cherenkov detector on beam experiments (RICH) and astronomical telescopes. A second solution simplifies and improves the classical PMT performances by replacing the dynode amplification chain with a SiPM in a Hybrid configuration. A similar photo detector using a lower APD gain has been implemented and it is already operative. There is space to implement both solutions. To do this at the best, advanced SiPM have to be realized. Two are the critical features that have to be solved in the next years: the dark noise reduction and the elimination of the dead region around the microcell (fill factor) where the quenching resistor is positioned. Already today it has been proposed a front-side illuminated detector structure with quenching resistors integrated into the silicon bulk. In this concept, the fill factor is only limited by the gaps necessary for optical cross-talk suppression. Compared to existing devices the proposed detector has the potential of a very high photon detection efficiency.

10. References

[1] J. Ninkovic, Nucl. Instr. and Meth. A 580 (2007) 1020.
[2] G. Barbarino et al., A new high-gain vacuum photomultiplier based upon the amplification of a Geiger-mode p-n junction., Nucl.Instrum.Meth.A594:326-331,2008.
[3] Pamela: A payload for Antimatter Matter Exploration and Light-nuclei Astrophysics. P. Picozza et al. Published in Astropart.Phys.27:296-315,2007.
[4] An anomalous positron abundance in cosmic rays with energies 1.5-100 GeV., O. Adriani et al., Nature 458:607-609,2009.
[5] Moon and sun shadowing effect in the MACRO detector. MACRO Collaboration (M. Ambrosio et al.) Astropart.Phys.20:145-156,2003.
[6] Final results of magnetic monopole searches with the MACRO experiment. MACRO Collaboration (M. Ambrosio et al.), Eur.Phys.J.C25:511-522,2002.
[7] Limits on dark matter WIMPS using upward going muons in the MACRO detector. MACRO Collaboration (M. Ambrosio et al.). Phys.Rev.D60:082002,1999.
[8] NEMO Collaboration, Sensitivity of an underwater Cerenkov km**3 telescope to TeV neutrinos from Galactic Microquasar., Astropart.Phys.28:1-9,2007.
[9] First results from a Dark Matter search with liquid Argon at 87 K in the Gran Sasso Underground Laboratory, P. Benetti et al., Astropart.Phys.28:495-507,2008, e-Print: astro-ph/0701286
[10] WARP liquid argon detector for dark matter survey, Presented by C. Rubbia at the 6th UCLA Symposium on Sources and Detection of Dark Matter and Dark Energy in the Universe. Published in New Astronomy Reviews Volume 49, Issues 2-6, May 2005, Pages 265-269
[11] The MAGIC Telescope and the observation of gamma ray bursts, MAGIC Coll. (Denis Bastieri et al.), Nuovo Cim.28C:711-714,2005, e-Print: astro-ph/0504310
[12] Very high energy photons from distant extragalactic sources observed by MAGIC: AGNs and GRBs., MAGIC Coll. AIP Conf.Proc.1018:157-166,2008.
[13] Warp liquid argon detector for dark matter survey, WARP Coll. New Astron.Rev.49:265-269,2005.
[14] Characterization of ETL 9357FLA photomultiplier tubes for cryogenic temperature applications., A. Ankowski et al., Nucl.Instrum.Meth.A556:146-157,2006.
[15] Performance and behavior of photomultiplier tubes at cryogenic temperature, M. Prata et al., NIMA567(2006) 222.
[16] Detection of VUV light at high quantum efficiency with large area avalanche photodiodes (LAAPDs) R. Chandrasekharan et al., Nucl.Instrum.Meth.A567:45-47,2006.
[17] Detection of the VUV liquid argon scintillation light by means of glass-window photomultiplier tubes, P. Benetti, et al., Nucl.Instrum.Meth.A505:89-92,2003.
[18] The recent development and study of silicon photomultiplier, V. Savelev, Nucl.Instrum.Meth.A535:528-532,2004.
[19] Development of the First Prototypes of a Silicon PhotoMultiplier (SiPM) at ITC-irst., N. Dinu et al., Nucl.Instrum.Meth.A572:422-426,2007
[20] Photon counting techniques with silicon avalanche photodiodes -Appl. Opt., H.Dautet et al., 32 (1993) 3894-3900.
[21] D.Renker Nuclear Instruments and Methods in Physics Research A 567 (2006) 48–56.
[22] Limited Geiger-mode microcell silicon photodiode: New results, G. Bondarenko et al. Nucl.Instrum.Meth.A442:187-192,2000.
[23] Silicon photomultiplier and its possible applications, P. Buzhan et al., Nucl.Instrum.Meth.A504:48-52,2003.

[24] Novel type of avalanche photodetector with Geiger mode operation, V. Golovin et al., Nucl.Instrum.Meth.A518:560-564, 2004.
[25] A.N. Otte et al., Nucl.Instrum.Meth.A567:360-363, 2006.
[26] Glenn F. Knoll: Radiation Detection and Measurement John Wiley & Sons, Inc.
[27] S.O.Kasap, Optoelectronics and Photonics, Principles and Practices, Prentice-Hall International (2001)
[28] http://www.pveducation.org/pvcdrom
[29] M.A.Greenand, M.J.Keevers, Progress in Photovoltaics: Research and Applications, Vol. 3, Issue3 313 (1995) 189.
[30] D.Renker,Nucl.Instrum.Meth.A468 (2002) 164.
[31] William J. Price, Nuclear Radiation Detection McGraw-Hill Book Company.
[32] C. Piemonte, Nucl. Instr. and Meth. A 568 (2006) 224.
[33] P. Buzhan, B. Dolgoshein, et al., An advanced study of silicon photomultiplier, ICFA Instrumentation Bulletin, 2001.
[34] MPPC Data Sheet, Hamamatsu Photonics K. K., http:// www.hamamatsu.com.
[35] R.J. McIntrye, J. Appl. Phys. 32 (6) (1961) 983.
[36] R.H. Haitz, J. Appl. Phys. 35 (5) (1964) 1370.
[37] M. Yokoyama et al., Development of Multi-Pixel Photon Counters, SNIC Symposium, Stanford, California, April 2006.
[38] S. Gomi et al., 2006 IEEE Nuclear Science Symposium Conference Record
[39] J. Haba, Status and Perspectives of Pixelated Photon Detector (PPD), Nucl. Instrum. Meth. Phys. Res A, 595, 154-160 (2008).
[40] W.G.Oldham IEEE Transactions On Electron Devices, Vol. Ed-19, No. 9, September 1972.
[41] S. Adachi, Phys. Rev. B 38 (1988) 12966.
[42] J. Humlı´cek, et al., J. Appl. Phys. 65 (7) (1989) 2827.
[43] H.R. Philipp, E.A. Taft, Phys. Rev. 120 (1960) 37.
[44] R. Hartmann, et al., Nucl. Instr. and Meth. A 387 (1997) 250.
[45] G. Collazuol, et al., Nucl. Instr. and Meth. A 581 (2007) 461.
[46] K. Amako, et al., KEK-PREPRINT-2005-28, May 2005, 8pp. (published in IEEE Trans. Nucl. Sci. NS-52 (2005) 910.).
[47] /www.cern.ch/geant4
[48] D. Ferenc et al,. Nucl. Instr. and Meth. A 567 (2006) 166.
[49] A. Lacaita et al., Photon-assisted avalanche spreading in reach-through photodiodes, Appl. Phys.Lett. 62 (1993) 606.
[50] A. Vacheret et al, arXiv1101.1996
[51] M. Yokoyama et al., Nucl. Instr. and Meth. A 610 (2009) 128
[52] A.A. Radu et al., Nuclear Instruments and Methods in Physics Research A 446 (2000) 497-505
[53] F. Lucarelli et al., Nuclear Instruments and Methods in Physics Research A 589 (2008) 415–424
[54] K. Bernlöhr et al., Astroparticle Physics 20 (2003) 111–128
[55] R. Winston, W.T. Welford, High Collection Nonimaging Optics, Academic Press, New York, 1989.
[56] www.edmundoptics.com
[57] J.Ninkovic et al., Nuclear Instruments and Methods in Physics Research A 617 (2010) 407–410.
[58]: C. Joram, A.Rudge, J. Séguinot: Nucl. Instr. and Meth. A 621 (2010) 171–176.
[59] WARP Collaboration, Astropart. Phys. 28 (2008) 495.
[60] XENON Collaboration, Phys. Rev. Lett. 100 (2008) 021303.
[61] J. Ninkovic, et al., Nucl. Instr. and Meth. A 580 (2007) 1013.
[62] E. Aprile, Nucl. Instr. and Meth. A 556 (2006) 215.

10

Near-Infrared Single-Photon Detection

Guang Wu, E Wu, Xiuliang Chen,
Haifeng Pan and Heping Zeng
*State Key Laboratory of Precision Spectroscopy,
East China Normal University
China*

1. Introduction

With the rapid increase of research interest in quantum information (Bennett&Brassard, 1984; Gisin et al., 2002; Knill et al., 2001), the near-infrared single-photon detection received a great boost not only in inventing (or improving) basic devices, but also in improving operation techniques on the conventional devices. Especially, in the application of quantum key distribution (Bennett&Brassard, 1984; Gisin et al., 2002), practical single-photon detectors (SPDs) with small size, operating at room-temperature are in great need.
Avalanche photodiodes (APDs) are usually used to build SPDs. Avalanche photodiodes (APDs) have internal gain due to a process of impact ionization that leads to multiple electron-hole pairs per input photon. Applying a large reverse voltage to the APD will result in a large multiplication gain, until the breakdown voltage (V_{br}) is reached. Usually, the output photocurrent of the APDs is linearly proportional to the intensity of the optical input when the bias voltage is below V_{br}, and this mode is called as "linear mode". When the bias voltage is larger than V_{br}, the electron-hole generation process can become self-sustaining and result in a runaway avalanche, then a single photoexicited carrier can induce a runaway avalanche that gives rise to a detectable macroscopic current, and this mode is called as "Geiger mode". Si-APD SPD exhibits excellent performance with the spectral range from 400 to 1000 nm. Si APD is typically operated in free-running mode. Its detection efficiency is as high as 70% around 700 nm with the dark count rate (DCR) of 10-100 counts per second (Stipčević et al., 2010). InGaAs/InP APD has a spectrum response range from 1200 to 1700 nm, covering the fiber optical communication window at 1310 and 1550 nm. However, InGaAs/InP-APD SPD has a large dark count (e.g. DCR~10^5 per second) and afterpulsing effect. Especially, the serious afterpulsing effect of the InGaAs/InP APD limits the application of high-speed detection. In order to improve its performance, an InGaAs/InP APD is usually operated in gated Geiger mode to suppress dark counts and afterpulsing effect. Recently, (Yuan et al., 2007; Namekata et al., 2006) reported self-cancellation and sine-wave techniques, which exhibited great improvements on the InGaAs/InP-APD SPD. The single-photon detection speed was increased significantly from megahertz to gigahertz.
Previously, it was thought that APDs are unable to resolve the number of photons in a short time interval. (Kardynal et al., 2008) first found that by suppressing the capacitive response down to a sufficiently low level, the weak avalanche current of an InGaAs/InP APD can be discriminated in its early development before saturated. In this mode, the variation in

multiplication gain of the current shows the capability of resolving photon number, called as "non-saturated" Geiger mode (Wu et al., 2009; Yuan et al., 2010). To date, some individual photon detectors have been demonstrated to exhibit interesting photon-number-resolving capability, such as visible light photon counters (Takeuchi et al., 1999), superconducting optical detectors (Miller et al., 2003; Schuster et al., 2007), and field effect transistors with quantum dots (Shields et al., 2000). All those developments have already stimulated vast promising applications although their performances are still limited by the requisite cryogenic operation.

On the other hand, several optical techniques have been developed to obtain high-performance near-infrared single-photon detection. A straightforward way to overcome the incapability of the current infrared detectors is to up-convert the infrared single photons with complete quantum state transfer into their replicas in the visual-near infrared region so that a Si-APD SPD can be used for the single-photon detection. Frequency up-conversion of 1550 nm single photons has been demonstrated by sum frequency mixing (Albota&Won, 2004; Pan et al., 2006). In particular, frequency up-converted single-photon detection facilitates a successful realization of a fiber-based QKD at gigahertz (Thew et al., 2006). Another optical technique for single-photon detection is just beginning (Han et al., 2008; Wu et al., 2010), where a high-gain optical amplifier is invented to amplify single photon to an intense light pulse. The spontaneous fluorescence is the major obstacle of the optical amplification technique to detect the single photon. (Han et al., 2008) used an optical parametric amplification, where the pulse width of the pump laser pulse was only 130 femtosecond. The influence of the spontaneous fluorescence was deeply suppressed in this ultra short period. (Wu et al., 2010) used a short optical bandpass filter to suppress the spontaneous fluorescence that few-photon pulse at 1550 nm could be detected.

In this chapter, we introduce some practical techniques of near-infrared single-photon detection, containing four sections as following: i) InGaAs/InP APD SPD; ii) Photon-number-resolving detector based on a InGaAs/InP APD; iii) Near-infrared single-photon detection with frequency up-conversion; iv) Few-photon detection with linear external optical gain photodetector.

2. InGaAs/InP-APD single-photon detector

The performance of an InGaAs/InP APD SPD is characterized by detection efficiency, dark count, afterpulsing effect, and time jitter, etc. The detection efficiency is mainly determined by two factors (Itzler et al., 2007): i) the quantum efficiency of the APD, which is the probability that a photon excited a carrier to reach the multiplication region, and ii) the avalanche breakdown probability, which figures the probability that a carrier in the multiplication region triggers an avalanche process successfully. The dark counts arise when carriers are created by processes other than photoexcitation, including the thermal excitation and field-mediated creation of free carriers (i.e. tunneling, trapping processes). Typically, the dark count rate of a Si-APD SPD is about 100 counts per second. However, the InGaAs/InP APDs have much more random bulk leakage carriers, leading to the dark count rate in the order of 10^5 counts per second. The afterpulsing effect is a dark count induced by the release of a carrier trapped by a defect in the multiplication region during an earlier avalanche event. In high-speed operation, the afterpulsing effect becomes the major problem that enhances dark counts. The time jitter is the variation in the temporal correspondence between the arrival of the photon and the detection of a resulting avalanche. It originates in

the stochastic nature of the carrier dynamics involved with avalanche breakdown, typically in the order of 100 ps for Si-APD APD.
In order to decrease dark counts and afterpulsing effect, InGaAs/InP APDs are usually operated in gated Geiger mode, by applying a reverse bias above the breakdown voltage with a short gating pulse over a DC bias. However, strong capacitive response of the electric gating pulse will be produced by the junction capacitance of the APD, which buries the weak avalanche pulses when a short gating pulse is used. It is quite important to produce an electric signal equal to the capacitive response of the APD, before cancelling the capacitive response. Here we present three different noise-cancellation techniques to operate the InGaAs/InP APD in gated Geiger mode with the gating width ≤ 1ns.

2.1 Balanced capacitance cancellation technique

An equal capacitance can make a similar electric signal when it is applied on a same gate pulse, as most of the capacitive response is induced by the junction capacitance of the InGaAs/InP APD. Figure 1 is the frequency responses of a InGaAs/InP APD and a variable capacitance, where the InGaAs/InP APD is tested with a DC bias of 50 V. It seems similar between the InGaAs/InP APD and the variable capacitance below 500 MHz. So, capacitance is a simple and economical method to produce a similar capacitive response in a large frequency range. As shown in Fig. 2, a same electric pulse is both applied on the fiber pigtailed InGaAs/InP APD and the variable capacitance. The responses of the InGaAs/InP APD and the variable capacitance are sent to two differential inputs of a differential amplifier, respectively. The common signal such as the capacitive response is deeply suppressed in the differential amplifier, while the avalanche pulse is amplified. There were also several other cancellation techniques which suppressed the capacitive response effectively (e.g. (Bethune&Risk, 2000) produced a capacitive response by a coaxial cable reflection line; (Tomita& Nakamura, 2002) used two APDs to produce two capacitive responses).
The InGaAs/InP APD is cooled to -35 ^{o}C. The width of the electric gating pulse is 1 ns with the gating rate of 25 MHz. A distribution feedback (DFB) laser diode is triggered at 1 MHz with the pulse duration < 300 ps. The laser is synchronized at 1/25 of the electric gating pulse. And the laser pulses are attenuated to 0.1 photon per pulse before sent to the InGaAs/InP APD. The detection efficiency decreases with the bias voltage, seeing Fig. 3, and it becomes saturated at about 30%. However, the dark count and afterpulsing effect will degrade the detector when increase the bias voltage. As shown in Fig. 4, the dark count rate is less than 8×10^{-6} counts per gate at the detection efficiency of 10%. It exponential rises with the DC bias voltage until the detection efficiency of about 30%. Meanwhile, the afterpulsing effect is also measured with the detection efficiency. Figure 4 shows the afterpulsing probability at 40-ns delay after the first avalanche. It is less than 0.1% at the detection efficiency of 10%, and increases more slowly than the dark count rate below the detection efficiency of 25%. Then the afterpulsing effect dominates the dark count of the detector when the detection efficiency is close to 30%. The afterpulsing effect comes from the trapping effect in the multiplication region of the InGaAs/InP APD. The trapped carriers are exponential decay with time (Wu et al., 2006 ; Itzler et al., 2007). So the afterpulsing effect is the major obstacle for high speed operation, especially the adjacent two gating pulses should be less than 3 ns (Liang et al., 2011). Increasing the temperature of the InGaAs/InP APD can speed up the decay of the trapped carriers. However, it will obviously increase the dark count due to the thermal effect.

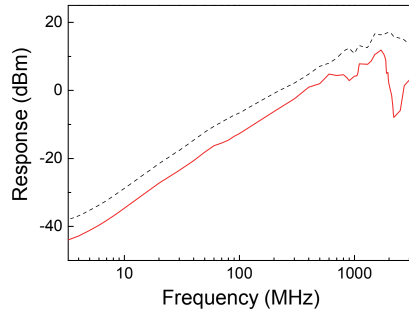


Fig. 1. Frequency response of an APD and a variable capacitance, where dash line: the APD, solid line: the variable capacitance

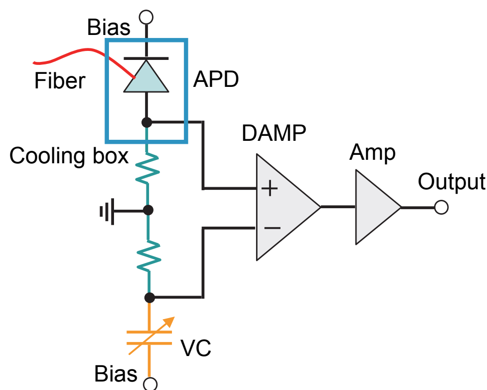


Fig. 2. Schematic of the balanced capacitance cancellation circuit, where Bias: DC bias + gating pulse, VC: a variable capacitance, DAMP: a differential amplifier, and AMP: a broadband amplifier

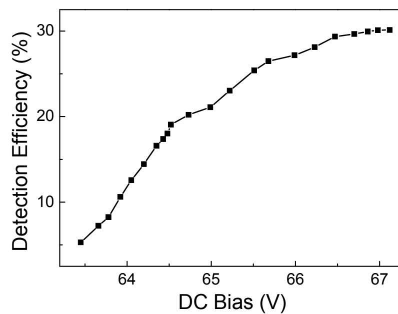


Fig. 3. Detection efficiency as a function of DC bias voltage

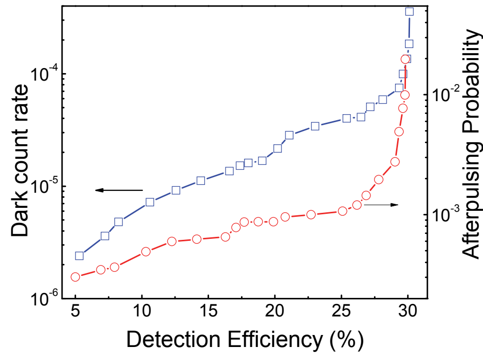


Fig. 4. Dark count rate and afterpulsing probability as a function of detection efficiency

Except rising the temperature, the other method to suppress the afterpulsing effect relies on decreasing the carriers passing through the APD, where shortening the gating width can decrease carriers. Figure 5 presents afterpulsing probability as a function of the gating width. It proves that short gating width is effective to suppress the afterpulsing effect for high speed operation.

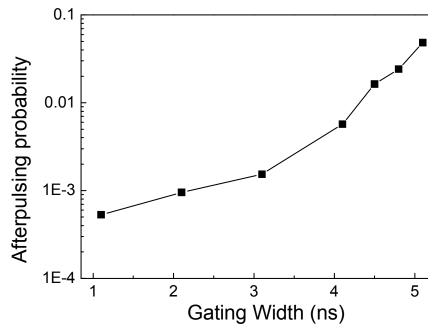


Fig. 5. Afterpulsing probability with the electric gating width

2.2 Self-cancellation technique

As mentioned above, shortening the gating width can decrease the number of bulk carriers passing through the InGaAs/InP APD, so it can weaken the afterpulsing effect for high speed operation. However, the avalanche current becomes weaker. It requires higher sensitivity, as well as better capacitive-response cancellation, to catch the avalanche pulse in the capacitive response. As shown in Fig. 1, the frequency responses between the InGaAs/InP APD and the variable capacitance are quite different when the frequency > 500 MHz. So, the variable capacitance cannot produce absolutely same capacitive response with the InGaAs/InP APD, although it has a same value of the capacitance. The self-cancellation technique solves the problem nicely. Figure 6 is the schematic of this technique. The electric signal on the cathode of the InGaAs/InP APD is sent to a 50/50 power splitter to produce

two equal components. Then the two identical components are combined by a differencer, where one of the components is delayed by one gating period. The output of the differencer is the difference of the two components. Actually, they are the signals of two adjacent gating periods. The capacitive response is cancelled by itself. As a result, weak avalanche pulse can be discriminated at high-speed gating rate. (Yuan et al., 2010) promoted the gating rate as high as 2 GHz with the gating width of only 250 ps.

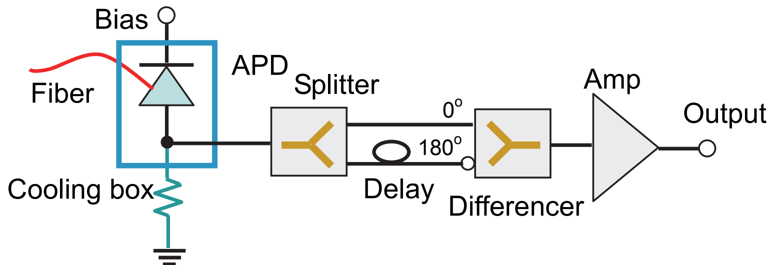


Fig. 6. Schematic of the self-cancellation circuit

2.3 Optical self-cancellation technique

In self-cancellation technique, the electric signal transmits through two coaxial cables. Due to the large transmission loss of the coaxial cable, the delay of one component cannot be too long; resulting in the gating rate should be high (e.g. > 200 MHz). Moreover, the electric circuit of the self-cancellation has a very wide bandwidth > 2 GHz. It should take more attention on designing and manufacturing for high cancellation ratio of the capacitive response. The optical self-cancellation technique gives a simple method to realize self-cancellation in wide bandwidth, including the operation at low gating rate.

Figure 7 is the schematic of the optical self-cancellation. The InGaAs/InP APD response is magnified by a low-noise broadband amplifier to trigger a DFB laser diode at 1550 nm. The response bandwidth of the laser diode is 2.5 GHz, fast enough to transfer the electronic signal to light pulse while keeping the same shape. In this way, the AC electronic signal is transformed to optical signal, preserving the original information from the InGaAs/InP APD including the capacitive response and the avalanche pulse. The fiber connecting the splitter and the detectors has different lengths to introduce a delay of one gating period between the two components. A fiber stretcher is employed to precisely control the delay

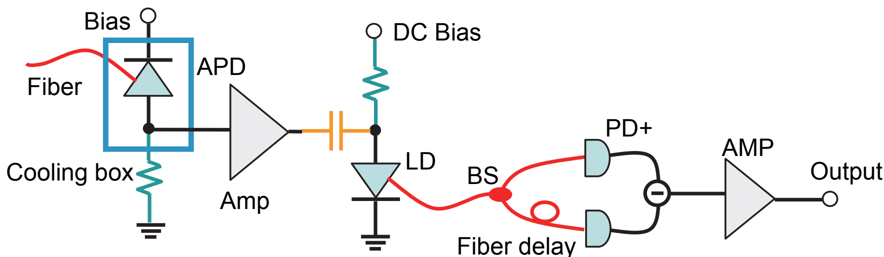


Fig. 7. Schematic of the optical self-cancellation circuit, where LD is a 1550-nm DFB laser diode, BS is a 50/50 fiber splitter, PD+ and PD- are the balanced optical detector

between the two components with 0.17-ps resolution. Two conventional photodiodes are used to detect the optical signals from each fiber. The response of the photodiodes exactly replayed the detection signal of the APD. At the output of the balancer, the identical capacitive response is subtracted. With this optical self-differential photodetector, the weak avalanche current can be measured (Wu et al., 2009).

3. Photon-number-resolving detection based on a InGaAs/InP APD

It was thought that a single APD cannot resolve the incident photon number without time or space multiplexing techniques since the gain on the APD is saturated in Geiger mode. However, recent research result reveals that the avalanche current is proportional to the photon number of the input light pulse when the APD is operated in non-saturated Geiger mode. Figure 8 gives a typical avalanche trace. It is recorded by a 6-GHz digital oscilloscope with the gating width of 5 ns. The current grows gradually first within area (a), and then it becomes saturated in area (b). Area (a) is the non-saturated Geiger mode period that the current is proportional to the input photon number. However, the saturation inhibits all the variation in the early avalanche development in area (b). The avalanche is just beginning in area (a), which the current is much weaker than the current in area (b). Through the optical self-cancellation technique, the non-saturated avalanche pulse is observed successfully.

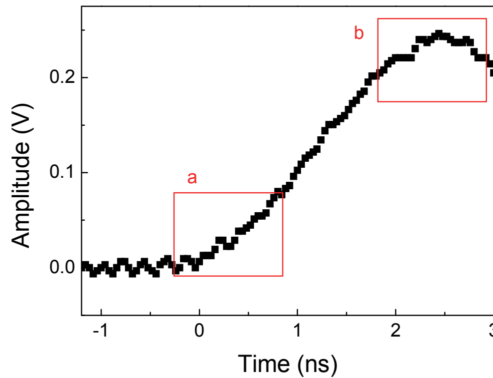


Fig. 8. An avalanche trace in 5-ns electric gate

Figure 9 is a typical histogram of the output peaks of the avalanche pulses. The distribution of the peak output of the avalanche pulses shows 3 peaks. Obviously, these distribution peaks are induced by different input photon number. The input light is from a DFB laser. This coherent light source obeys the Poissonian distribution, where the photon number *(n)* is determined by the probability:

$$p(\mu, n) = \frac{\mu^n}{n!} e^{-\mu} \tag{1}$$

where μ the is the mean photon number per pulse. The probabilities of the peak output of the avalanche pulses are calculated according to the Poissonian distribution, which is given by:

$$P(V) = \sum_{n=0}^{\infty} p(\mu, n) \cdot \rho(n, V) \quad (2)$$

where $\rho(n, V)$ is the distribution of the peak output of the avalanche pulses when they are induced by n-photon. It shows a Gaussian-like distribution. The calculated data fits well with the measured data as shown in Fig. 9, proving that the avalanche current in non-saturated Geiger mode is proportional to the input photon number. The width of n-photon peak is $\sqrt{n}$ $(n > 1)$ scaled to the 1-photon peak, which is caused by the statistical fluctuation. The width of 1-photon peaks is determined by the avalanche multiplication, and the excess noise derived from the statistical nature of the avalanche multiplication of the InGaAs/InP APD.

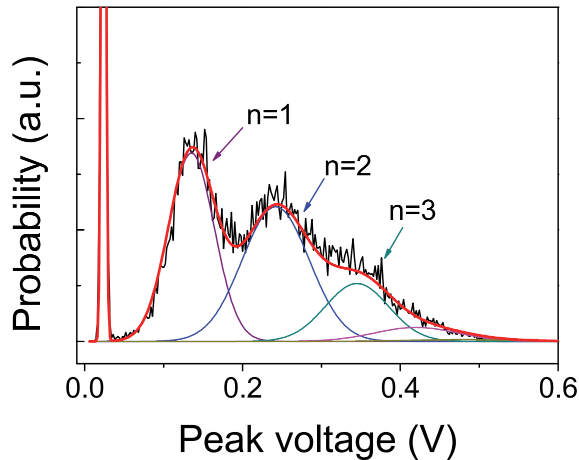


Fig. 9. Distribution of the peak output of the avalanche pulses, where the black line is the measured data, the red line is the calculated data. The detected mean photon number is 1.9 per pulse at the detection efficiency of 10%

Figure 10 is the color-grading waveforms of the avalanche pulses in non-saturated Geiger mode. It is recorded by a 6-GHz digital oscilloscope with the integration time of 0.1 second. Three peaks of the distribution of the avalanche pulses clearly appear in the waveforms. They are induced by 1-, 2-, and 3-photon, respectively.

Figure 8 shows that the non-saturated Geiger mode exits in a short period of the early avalanche development. As a result, in order to observe the capability of the photon-number-resolving (PNR) of the InGaAs/InP APD, the gating width should < 2 ns. In order to figure out the relation between the PNR performance and the avalanche multiplication, the distributions of the peak output of the avalanche pulses at different detection efficiency are measured as shown in Fig. 11. It is hard to resolve the photon number at low detection efficiency. And the optimal period of the detection efficiency for PNR is from 10% to 20%.

When the detection efficiency increases to 36%, all the peak output of the avalanche pulses reach the maximum amplitude of 960 mV, the saturation effect appears obviously and the peak voltage is independent of the incident photon number more than 2. This sets the upper boundary for the InGaAs/InP APD to resolve photon numbers.

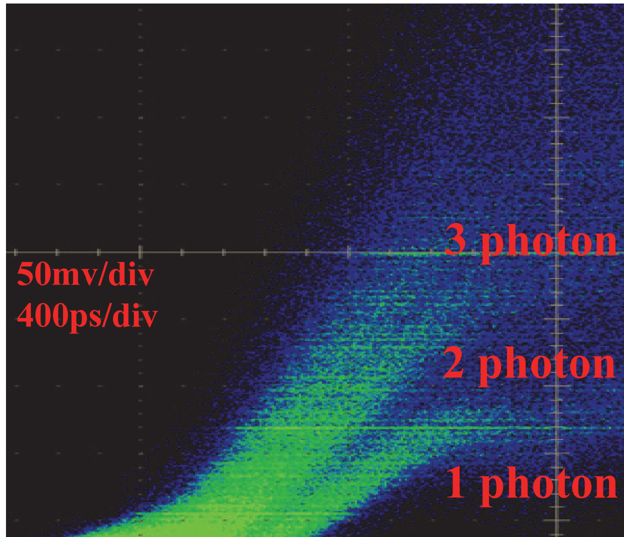


Fig. 10. Color-grading waveforms of the avalanche pulses

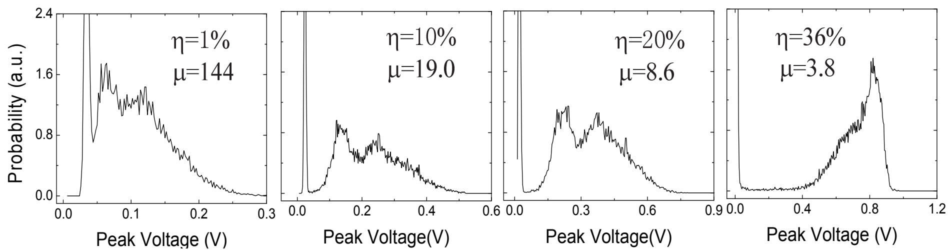


Fig. 11. Distribution of peak output of the avalanche pulses at different detection efficiencies

The PNR performance is time resolved with the input laser delay, since the non-saturated Geiger mode is observed in a short gated mode. Figure 12 shows the photon count rate varies with the laser pulse delay. The electric gating width is about 1.2 ns, while the effective detection gating width is about 300 ps. Three delays of the input laser are observed, they are signed as (a), (b), and (c) in Fig. 12. Figure 13 is the distribution of the peak output of the avalanche pulses at these three points. Obviously, the PNR performance is similar good at points (a) and (b). And the PNR performance degrades at point (c). As shown in Fig. 8, the avalanche multiplication gain increases for about 2 ns until saturated. So, the avalanche current obtains a larger gain when the photon arrives at the rising edge of the electric gate than that at the falling edge. Therefore, a large multiplication gain is good for PNR performance before the InGaAs/InP APD is saturated.

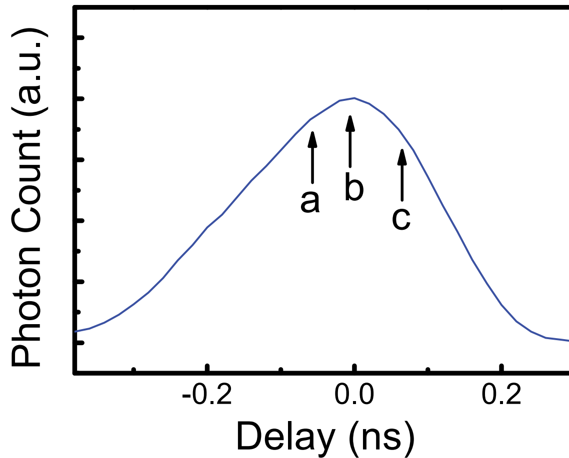


Fig. 12. Photon count rate as a function of the laser pulse delay

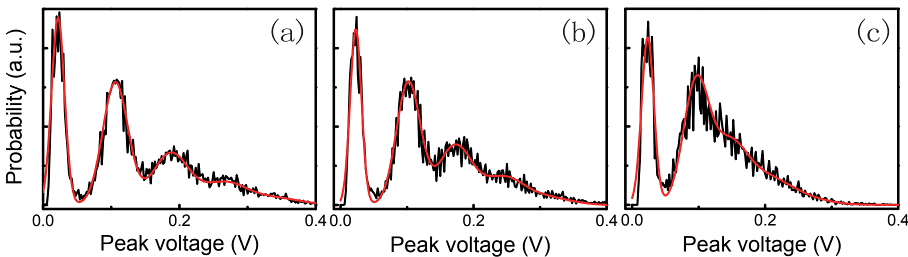


Fig. 13. Distribution of the peak output of the avalanche pulses at points (a), (b), and (c), respectively, where the detected mean photon numbers are 1.33, 1.35, and 1.32, respectively

4. Near-infrared single-photon detection with frequency up-conversion

The single-photon frequency up-conversion can be considered as the sum-frequency generation (SFG) process as shown in Fig. 14. Suppose that the pump laser is in the single longitudinal mode. The solution to the coupled-mode equations for the phase-matched interaction is given by (Kumar, 1990):

$$\begin{aligned}\hat{a}_1(L) &= \hat{a}_1(0)\cos(|gE_p|L) - \hat{a}_2(0)\sin(|gE_p|L),\\ \hat{a}_2(L) &= \hat{a}_2(0)\cos(|gE_p|L) + \hat{a}_1(0)\sin(|gE_p|L),\end{aligned} \tag{3}$$

where $\hat{a}_1$ and $\hat{a}_2$ are annihilation operator for the signal and upconverted fields, respectively, g denotes the nonlinear coupling coefficient, and L is the length of the nonlinear medium. As indicated in Eq. (3), a complete quantum conversion occurs from $\hat{a}_1$ to $\hat{a}_2$ when $|gEp|L = \pi/2$ is satisfied. The single-photon conversion efficiency can be written as:

$$\eta = \sin^2(0.5\pi \cdot \sqrt{P_p / P_c}) \tag{4}$$

where P_p is the effective pump power, and P_c is the pump power at unity conversion efficiency.
The complete quantum conversion demands a large nonlinearity of the nonlinear media together with a strong pump field. Thus, periodically poled lithium niobate (PPLN) is usually employed in the single-photon frequency up-conversion since it has a large nonlinear coefficient (d_{eff} =14 pm/V) and provides a large quasi-phase-matching (QPM) interaction length in the order of 10 mm. The single-photon frequency up-conversion has been demonstrated in a PPLN waveguide or bulk PPLN. With a PPLN waveguide, the requirement on the pump field power can be lowered since the power of the optical field can be confined to a small volume in the waveguide to have a very high intensity. The PPLN waveguide scheme requires subtle processes to prepare a monolithic fiber pigtailed PPLN waveguide, which will induce an avoidless big insertion loss (Tanzilli et al., 2002; Langrock et al., 2004&2005). And the bulk PPLN scheme requires a high pump power, e.g. using a resonant pump cavity with a stable cavity lock to enhance the circulating pump power (Albota et al., 2004); or enhancing single-photon frequency up-conversion by intracavity laser pump (Pan et al., 2006).

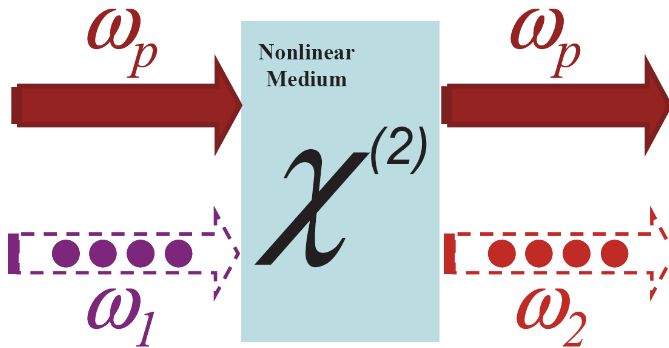


Fig. 14. Schematic of single-photon frequency up-conversion

For single-photon frequency up-conversion, one of the key parameters is the signal to noise ratio. If the noise is much larger than the signal photons, it will be meaningless to take the trouble to do the up-conversion. Therefore, suppressing the noise will much improve the performance of the single-photon frequency up-conversion in the applications. Figure 15 shows the possible noise sources in the intracavity enhanced up-conversion system discussed in the section above. The dark counts from the Si-APD SPD (10~200 counts per second depending on the device) could be neglected since the dark counts from the background photons are much larger. The main contribution to the background photons comes from the strong pump field. The background photons at 808 and 1064 nm comes from the solid-state laser itself. Besides the up-conversion process with the incident single photons, other nonlinear effects also takes place in the nonlinear media, such as second harmonic generation (SHG) of the pump laser at 532 nm and the optical parametric generation (OPG) fluorescence. These background photons could be removed by the filter system since they are at different wavelengths from the signal photons. However, among the background photons, there are some of the same wavelengths with the signal photons at 631 nm. They are caused by up-conversion of the parametric fluorescence caused by the

strong pump field. At first, spontaneous down-conversion of the strong pump took place in the nonlinear media as $\omega_{1064nm}=\omega_{1550nm}+\omega_{3400nm}$. In this process, the parametric fluorescence photons at 1550 nm are of the same wavelength with the incident signal photons. And since the temperature of the nonlinear media is tuned for the phase matching of SFG for $\omega_{1064nm} + \omega_{1550nm} = \omega_{631nm}$, these noise photons are up-converted together with the incident signal photons with high efficiency. Therefore, some of the output photons at 631 nm are not the replica of the incident signal photons but the noise from the up-converted parametric fluorescence. Unfortunately, these background photons can not be removed spectrally by the filters and contributed a lot to the dark counts on the Si-APD SPD.

Several groups have proposed the long-wavelength pump scheme to overcome the troublesome up-converted parametric fluorescence (Langrock et al., 2004; Dong et al., 2008; Kamada et al., 2008). By choosing a comparatively long-wavelength pump, which means the energy of the pump photons is lower than that of the signal photons, the parametric fluorescence from the down conversion will not fall in the incident infrared signal photon spectral regime. As a result, the pump induced parametric fluorescence can be efficiently suppressed and the dark counts will be greatly lowered. We have demonstrated an efficient single-photon frequency up-conversion system for the infrared photons at 1064 nm with ultralow dark counts (Dong et al., 2008). The pump source was provided by a mode-locked erbium-fiber laser. The repetition rate of the pulse train was 15.8 MHz and the pulse duration was measured to be 1.4 ps. The average output power of the amplifier was measured to be 27 mW. The peak power of the pulsed laser was ~220 W, high enough to achieve unity conversion efficiency in the system. With such a pulsed pump source, no cavity enhancement was required, much simplifying the whole system. A long-pass filter with 1000 nm cutting off was placed in front of the PPLN crystal to block the stray light from the erbium doped fiber amplifier (EDFA), such as the pump for the EDFA from the LD at 980 nm and the green and red up-conversion emission of the EDFA. In this long-wavelength pump system, the relatively lower energy pump photon would not induce undesired parametric fluorescence at the signal wavelength 1550 nm, and the dark counts at SFG wavelength from followed up-conversion of the parametric fluorescence was eliminated. Moreover, besides that the Si-APD SPD did not respond to pump light at 1550 nm, the up-conversion fluorescence by the second harmonic of the strong pump was not phase matched at this working temperature, thus the noise from that process could also be ignored. Thanks to sufficient suppression of the intrinsic background photons, the narrow bandpass filter was even not necessary in the filtering system, increasing the transmittance of the filtering system. After the filtering system, we measured the dark counts of the whole detection system and got a count rate of ~150 counts per second, when there were neither signal nor pump photons feeding. Moreover, when there was pump feeding, the dark count rate was still around 150 counts per second, indicating that the dark counts were not from the nonlinear parametric processes caused by the strong pump but mainly due to dark counts of Si-APD SPD and ambient background light. With this system, we achieved so far the lowest noise to efficiency ratio of ~160 for a near unity conversion efficiency (93%) as shown in Fig. 16.

The single-photon frequency up-conversion has not only shown a solution to the sensitive detection of the infrared weak signals but also provided a technique to manipulate quantum states of the photons. Novel ideas on the techniques for single-photon frequency up-

conversion come forth from time to time, highlighting its applications in the quantum information processing.

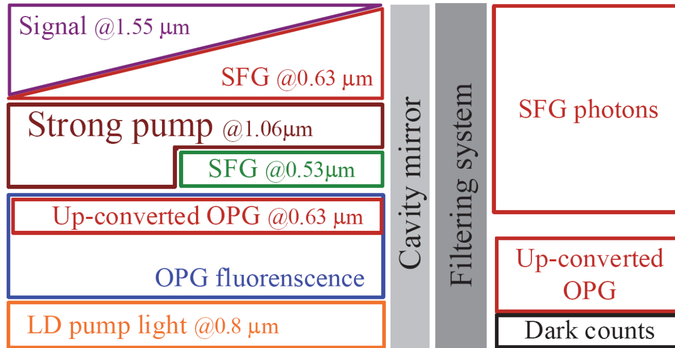


Fig. 15. Noise of the intracavity single-photon up-conversion

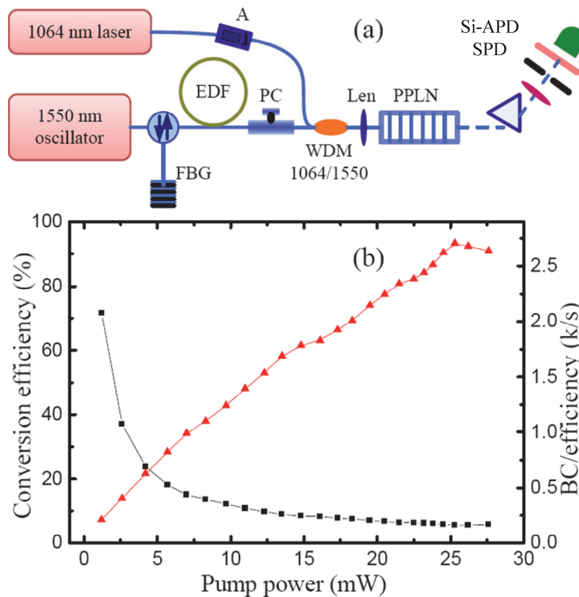


Fig. 16. Schematic of the long-wavelength pumped frequency up-conversion

5. Few-photon detection with linear external optical gain photodetector

Different from the most methods to amplify the photo-excited carrier with a large internal electric multiplication gain by electronic devices, we employed the optical devices to amplify the few-photon before detecting by a conventional PIN photodiode. Interestingly, the photodiode response showed a linear dependence on the incident photon signals, promising a novel few-photon detection technique.

Single-photon amplification by stimulated emission becomes the focus of research interest in recent years due to its application in quantum cloning (Simon et al., 2000; Fasel et al., 2002). In order to detect the amplified photon signals with conventional PIN photodiodes, the amplifier should be chosen under the constraint of a high gain. In addition, the amplifier noise due to the spontaneous emission should be suppressed enough to allow the identification of photons due to the stimulated emission. Er-doped optical fibers are commonly used in the optical fiber communication as amplifiers due to their large gain up to 40 dB around 1550 nm. But the spontaneous emission always accompanies the stimulated emission and will be amplified as well, which would be the big barrier to identify the signal photons from the noise. In order to suppress the amplified spontaneous emission (ASE), we separated the amplification into two steps. Figure 17(a) shows the setup of the external-gain photodetector based on the single-photon amplification. The light source is a laser diode modulated by an intensity modulator at 25.0 MHz with pulse duration of 325 ps. The output spectrum of the laser is shown by the green line in Fig. 18(b). The central wavelength is at 1550.20 nm and the full width at half maximum (FWHM) is 0.02 nm. The output of the laser is attenuated to contain only a few photons per pulse. Then, the photons are sent to the first EDFA for amplification. In order to detect the stimulated emission photons, spectral filtering is necessary because the ASE spectrum of the EDFA covered a broad range from 1527.36 to 1563.84 nm. Firstly, an inline bandpass filter (IF_1) centered at 1550 nm with the FWHM of 3 nm is inserted to roughly extract the amplified signal photons from the broadband fluorescence.

Secondly, the combination of the two fiber Brag gratings ($FBG_{1,\ 2}$) with the FWHM of 0.18 nm form another bandpass filter. By tuning the temperature to combine the rising edge of FBG_1 and the falling edge of FBG_2, a final bandwidth of the bandpass filter was determined to be 0.06 nm. Finally, a fiber polarization controller (PC) together with a polarization beam splitter (PBS) helps to remove the ASE noise of the orthogonal polarization. Then, the optical signal is sent to another EDFA for the amplification again. Since the incident photons are pre-amplified while most of the ASE noise is removed before the second amplification, the ASE of the second EDFA itself is much suppressed and instead the stimulated amplification is enhanced. Spectral filtering is not as strict as in the first step. The filtering system for the second amplification is composed of a bandpass inline filter (IF_2) with the FWHM of 3 nm and a fiber Brag grating FBG_3 with the FWHM of 0.18 nm. The PBS is not even necessary in the second step because the ASE of the orthogonal polarization in the second EDFA is so weak that it could be ignored. The black line in Fig. 17(b) shows the ASE spectrum after the two-step amplification. The spectral width is mainly constrained by the combined FBG filters in the first step. When the signal photons are sent in, the peak at 1550.20 nm raises on the top of the ASE spectrum as shown by the red line in Fig. 17(b), indicating the stimulated amplification of the incident photons. The total gain of the two EDFAs is measured separately to be about 42.7 dB, indicating that an incident photon could be amplified to $\sim 10^4$ photons per pulse (about 1 mW of the peak power) after the two-step amplification. The optical pulse signal is detected by a PIN photodiode. The variance of the ASE noise is measured and plotted as a function of the ASE output power as shown in Fig. 17(c). Since the main voltage noise is derived from the ASE beat on the PIN photodiode, the variance of the noise increased nonlinearly with the average output power, indicating that the ASE noise could be considered as a classical noise. The voltage noise amplitude is in Gaussian distribution with an FWHM of ~140 mV (Fig. 17(d)).

Figure 18 plots the color-grading waveforms of the output voltage measured by the DPX acquisition mode of a 2.5-GHz oscilloscope with an average incident photon number of μ = 4

and 16. From the oscilloscope traces, it is observed that the peak output signal amplitude changes with the incident photon numbers, showing the evidence of the photon number resolving ability of the detector. The amplitude of the peak output signal shows a linear dependence on the input average photon number as shown in Fig. 20, indicating that the EDFAs and the photodiode are far from saturation under such milliwatt optical input power and capable of registering more than 16 individual photons. By taking into account the optical amplification, the photodiode response and the electronic amplification, the sensitivity of the whole setup is obtained by fitting the curve in Fig. 19 to be 15.39 mV/photon.

The photon statistics is studied by analyzing the histograms of the voltage signal acquired by the oscilloscope. The temporal resolution of oscilloscope is set to 100 ps, and the voltage amplitude resolution is set to 4 mV. A 500-ps sampling window is used in the analysis. Figure 20 plots the histograms of the peak output signal voltage recorded for different incident photon numbers of μ = 4 and 16 per pulse. The red lines in Fig. 21 show a simulation of the experimental data assuming a Poisson distribution for the incident photons. Due to the ASE noise variance, the distribution histograms are broadened, dimming the boundary for different photon numbers. By taking into account the Poisson distribution of the incident photons, the single photon response of the system is obtained to be 68 mV by fitting the curves in Fig. 20, and the quantum efficiency of the EDFAs is calculated to be 22.7%. Due to the linearity of the external-gain photodetector, the curves of the peak voltage kept the shapes of the ideal Poisson distribution of the input photons. The probability statistics of the peak output voltage could be also observed in Fig. 18 directly by its color-grading.

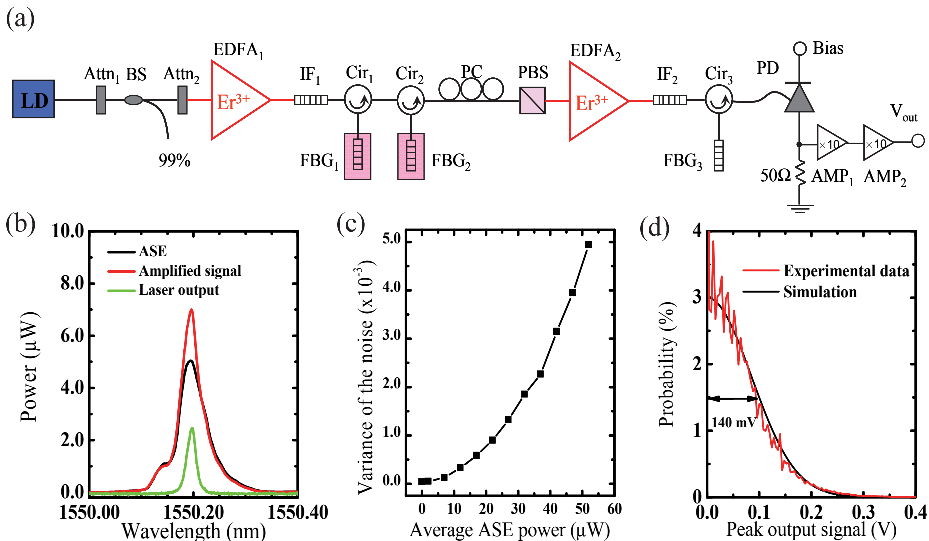


Fig. 17. (a) Schematic of the external-gain detector, where IF1,2: inline filters, FBG1-3: fiber Bragg gratings, Cir1-3: fiber circulators, PC: polarization controller, PBS: polarization beam splitter, PD: conventional pin photodiode, AMP: RF amplifiers. (b) Spectra of the laser (blue), ASE of the EDFA (green) and amplified single-photon signal (red). (c) variance of the ASE noise with the ASE power; (d) Distribution of the peak output of one photon.

Due to the lager spontaneous emission of the EDFA, the detector cannot discriminate single-photon pulses. The FWHM of the bandpass filter is 0.06 nm as well as about 7.4 GHz. The laser pulse width is ~325 ps, corresponding to the laser bandwidth in the order of 10 GHz. The filers fits well with the laser pulse, but the insertion loss of the filters is about 15dB, most of them comes from FBG_1 and FBG_2. So, more efforts are needed to suppress the spontaneous emission to decrease the insertion loss. However, the detector can be used as a sensitive power meter at single-photon lever, as the integral output has a good linearity to the input photon number, while the noise is averaged.

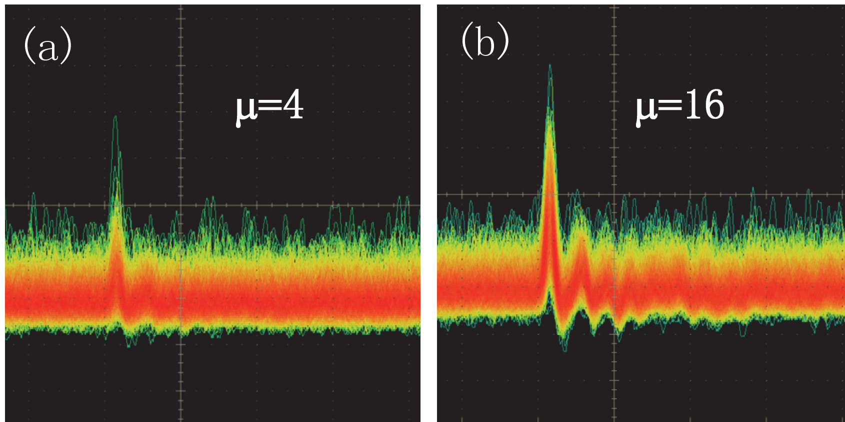


Fig. 18. Waveforms of the voltage output recorded by the oscilloscope with incident photon number of (a) 4 and (b) 16.

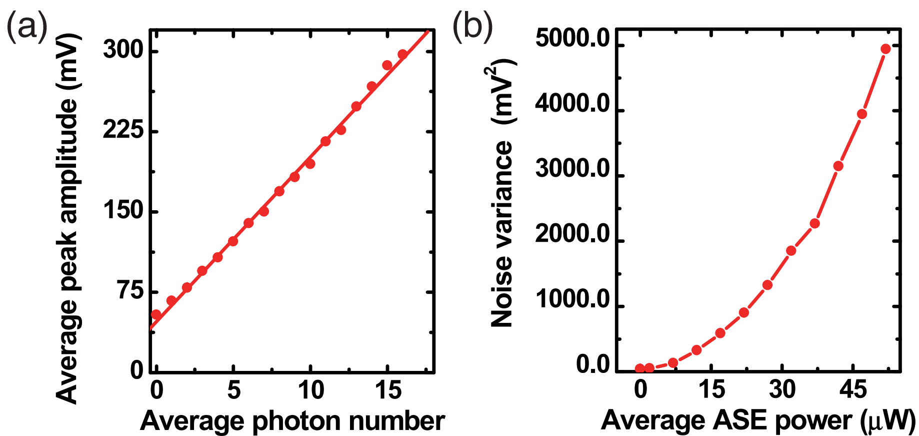


Fig. 19. Peak output voltage of the amplified photon signals and ASE noise variance. (a) Average peak amplitude as a function of the incident photon number. (b) ASE noise variance dependent on the output power.

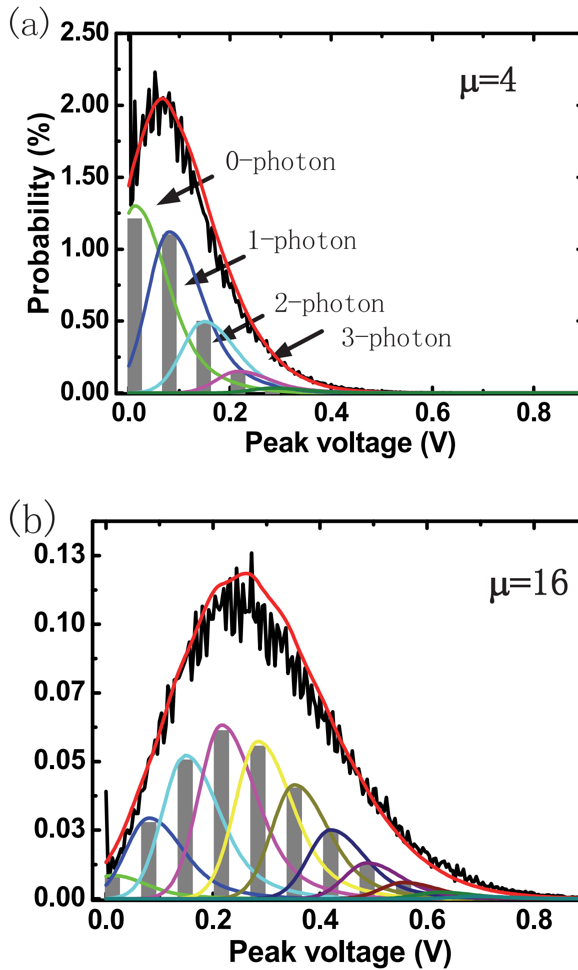


Fig. 20. Distributions of peak output signal for incident pulses of different average photon numbers, together with the calculated distributions

6. Conclusion

InGaAs/InP APDs are typically operated in gated Geiger mode for near-infrared single-photon detection. The afterpulsing effect becomes the major obstacle for high speed detection with the gating rate > 200 MHz, which should use short gating width (e.g. ~1 ns). The capacitive response of the InGaAs/InP APD must be cancelled to distil the weak avalanche pulse. We introduce three kinds of capacitive-response cancellation techniques, where the balanced capacitance cancellation technique is robust for operation at the gating rate < 200-MHz; and the self-cancellation and optical self-cancellation techniques are effective in higher gating rate up to 2 GHz. On the other hand, InGaAs/InP APDs are operated in non-saturated Geiger mode when the gating width < 2 ns. In this mode, the

output of the InGaAs/InP APD is proportional to the input photon number. And we prove that the PNR performance is determined by the multiplication gain of the InGaAs/InP APD and input time of the photons.

Optical techniques are potential to realize high performance near-infrared single-photon detection. One of them is the single-photon frequency up-conversion. The major problem of up-conversion is the background photons induced by the optical nonlinear process, which could be resolved by using long-wavelength pump laser, and the background photons are suppressed at a negligible level. The other optical technique is just starting that detects few-photon pulse with a conventional linear photodiode after amplified by an external optical amplifier. Up to now, it still need efforts to realize an ultra-low noise optical amplifier for few-photon detection.

7. Acknowledgment

This work was funded, in part, by the National Natural Science Fund of China (10904039, 10525416, 10990101, and 91021014), Key Project Sponsored by the National Education Ministry of China (108058), Research Fund for the Doctoral Program of Higher Education of China (200802691032), and Shanghai Rising-Star Program (10QA1402100).

8. References

Bennett, C. H. & Brassard, G. (1984). Quantum cryptography: Public key distribution and coin tossing. *Proc. IEEE Int. Conf. Comput. Syst. Signal Process.*, pp. 175–179

Gisin, N.; Ribordy, G.; Tittel, W. & Zbinden, H. (2002). Quantum cryptography. *Rev. Mod. Phys.*, Vol.74, pp. 145–195

Knill, E.; Laflamme, R. & Milburn, G. J. (2001). A scheme for efficient quantum computation with linear optics. *Nature* Vol.409, pp. 46-52

Stipčević, M.; Skenderović, H. & Gracin, D. (2010). Characterization of a novel avalanche photodiode for single photon detection in VIS-NIR range. *Opt. Express*, Vol.18, pp. 17448-17459

Yuan, Z. L.; Kardynal, B. E.; Sharpe, A. W. & A. J. Shields (2007). High speed single photon detection in the near infrared. *Appl. Phys. Lett.* Vol.91, 041114

Namekata, N.; Sasamori, S. & Inoue, S. (2006). 800 MHz single-photon detection at 1550-nm using an InGaAs/InP avalanche photodiode operated with a sine wave gating. *Opt. Express.* Vol.14, pp. 10043–10049

Kardynal, B. E.; Yuan, Z. L. & Shields, A. J. (2008). An avalanche-photodiode-based photon-number-resolving detector. *Nature Photonics* Vol.2, pp. 425-428

Wu, G.; Jian, Y.; Wu, E; Zeng & H. P. (2009). Photon-number-resolving detection based on InGaAs/InP avalanche photodiode in the sub-saturated mode. *Opt. Express* Vol.17, 18782

Yuan, Z. L.; Dynes, J. F.; Sharpe, A. W. & Shields, A. J. (2010). Evolution of locally excited avalanches in semiconductors. *Appl. Phys. Lett.* Vol.96, 191107

Takeuchi, S.; Kim, J.; Yamamoto, Y. & Hogue, H. H. (1999). Development of a high-quantum-efficiency single-photon counting system. *Appl. Phys. Lett.* Vol.74, pp. 1063-1065

Miller, A. J.; Nam, S. W.; Martinis, J. M. & Sergienko, A. V. (2003). Demonstration of a low-noise near-infrared photon counter with multiphoton discrimination. *Appl. Phys. Lett.* Vol.83, pp. 791-793

Schuster, D. I.; Houck, A.; Schreier, J.; Wallraff, A.; Gambetta, J. M.; Blais, A.; Frunzio, L.; Majer, J.; Johnson, B.; Devoret, M. H.; Girvin, S. M. & Schoelkopf, R. J. (2007). Resolving photon number states in a superconducting circuit. *Nature* Vol.445, pp. 515-518

Shields, A. J.; O'Sullivan, M. P.; Farrer, I.; Ritchie, D. A.; Hogg, R. A.; Leadbeater, M. L.; Norman, C. E. & Pepper, M. (2000). Detection of single photons using a field-effect transistor gated by a layer of quantum dots. *Appl. Phys. Lett.* Vol.76, pp. 3673-3675

Albota, M. A. & Wong, F. N. C. (2004). Efficient single-photon counting at 1.55 μm by means of frequency upconversion. *Opt. Lett.* Vol.29, 1449

Pan, H. F.; Dong, H. F. & Zeng, H. P. (2006). Efficient single-photon counting at 1.55 mm by intracavity frequency upconversion in a unidirectional ring laser. Appl. Phys. Lett. Vol.89, 191108

Thew, R. T.; Tanzilli, S.; Krainer, L.; Zeller, S. C.; Rochas, A.; Rech, I.; Cova, S.; Zbinden, H. & Gisin, N. (2006). Low jitter up-conversion detectors for telecom wavelength GHz QKD. *New J. Phys.* Vol.8, 32

Han, X. F.; Weng, Y. X.; Wang, R. Chen, X. H.; Luo, K. H.; Wu, L. A. & Zhao, J. M. (2008). Single-photon level ultrafast all-optical switching. Appl. Phys. Lett. Vol.92, 151109

Wu, G.; Jian, Y.; Wu, E & Zeng, H. P. (2010). Linear external optical gain photodetector at few-photon level. *IEEE Photon. Tech. Lett.* Vol.22, pp. 688-690

Wu, G.; Zhou, C. Y.; Chen, X. L. & Zeng, H. P. (2006). High performance of gated-mode single-photon detector at 1.55 μm. *Opt. Comm.* Vol.265, pp. 126-131

Itzler, M. A.; Ben-Michael, R.; Hsu, C. F.; Slomkowski, K.; Tosi, A.; Cova, S.; Zappa, F. & Ispasoiu, R. (2007). Single photon avalanche diodes (SPADs) for 1.5 μm photon counting applications. *J. Mod. Opt.* Vol.54, pp. 283–304

Bethune, D. S. & Risk, W. P. (2000). An autocompensating fiber-optic quantum cryptography system based on polarization splitting of light. IEEE J. Quantum Eletron. Vol.36, pp. 340-347

Tomita, A. & Nakamura, K. (2002). Balanced, gated-mode photon detector for quantum-bit discrimination at 1550 nm. *Opt. Lett.* Vol.27, pp. 1827-1829

Liang, Y.; Jian, Y.; Chen, X. L.; Wu, G.; Wu, E & Zeng, H. P. (2011). Room-temperature single-photon detector based on InGaAs/InP avalanche photodiode with multichannel counting ability. *IEEE Photon. Tech. Lett.* Vol.23, pp. 115 – 117

Yuan, Z. L.; Sharpe, A. W.; Dynes, J. F.; Dixon, A. R. & Shields. A. J. (2010). Multi-gigahertz operation of photon counting InGaAs avalanche photodiodes. Appl. Phys. Lett. Vol.96, 071101

Kumar P. (1990). Quantum frequency conversion. *Opt. Lett.* Vol.15, pp. 1476-1478

Tanzilli, S.; Tittel, W.; De Riedmatten, H.; Zbinden, H.; Baldi, P.; De Micheli, M.; Ostrowsky, D. B. & Gisin, N. (2002). PPLN waveguide for quantum communication. *Eur. Phys. J. D* Vol.18, pp. 155-160

Langrock, C.; Diamanti, E.; Roussev, R. V.; Yamamoto, Y.; Fejer, M. M. & Takesue, H. (2004). Periodically poled lithium niobate waveguide sum-frequency generator for efficient single-photon detection at communication wavelengths. Opt. Lett. Vol.29, pp. 1518-1520

Langrock, C.; Diamanti, E.; Roussev, R. V.; Yamamoto, Y.; Fejer, M. M. & Takesue, H. (2005). Highly efficient single-photon detection at communication wavelengths by use of upconversion in reverse-proton-exchanged periodically poled LiNbO3 waveguides. *Opt. Lett.* Vol.30, pp. 1725-1727

Dong, H. F.; Pan, H. F.; Li, Y.; Wu E & Zeng, H. P. (2008). Efficient single-photon frequency upconversion at 1.06 μm with ultralow background counts. *Appl. Phys. Lett.* Vol.93, 071101

Kamada, H.; Asobe, M.; Honjo, T.; Takesue, H.; Tokura, Y.; Nishida, Y.; Tadanaga, O. & Miyazawa H. (2008). Efficient and low-noise single-photon detection in 1550 nm communication band by frequency upconversion in periodically poled LiNbO3 waveguides. *Opt. Lett.* Vol.33, pp. 639-641

Simon, C.; Weihs, G. & Zeilinger, A. (2000). Optimal quantum cloning via stimulated emission. *Phys. Rev. Lett.* Vol.84, pp. 2993-2996

Fasel, S.; Gisin, N.; Ribordy, G.; Scarani, V. & Zbinden, H. (2002). Quantum cloning with an optical fiber amplifier. *Phys. Rev. Lett.* Vol.89, 107901 (2002)

11

Geiger Avalanche Photodiodes (G-APDs) and Their Characterization

Giovanni Bonanno, Massimiliano Belluso,
Sergio Billotta, Paolo Finocchiaro
and Alfio Pappalardo
INAF - Osservatorio Astrofisico di Catania,
INFN – Laboratori Nazionali del Sud Catania
Italy

1. Introduction

In many fields and in particular in astrophysical observations, a chronic problem is the photon-starving condition, which becomes severe when images are to be obtained in short acquisition times (from micro to milliseconds), as happens in hot areas of astrophysics: optical counterparts of high-energy gamma-ray bursts, study and interpretation of Supernovae bursts. CCDs are inherently unable to provide accurate measurements of such fast low-intensity transients at high frame rates. To respond to single photons, suitable detectors must provide output signals that are sufficiently high to be individually processed by electronic circuits. Therefore, only detectors with an internal mechanism that provides a high multiplication of charge carriers are suitable, namely vacuum tube photomultipliers (PMTs), solid-state avalanche photodiodes (APDs) and electron-multiplying CCDs (EM-CCDs). In PMTs, the photocathodes available for the visible spectral range provide fairly good quantum efficiency and low noise, whereas cathodes for the red and near-infrared range have lower quantum efficiency and must be cooled to reduce the dark-count rate. PMTs are bulky, and so not suitable for assembly in large arrays, fragile, sensitive to electromagnetic disturbances and mechanical vibrations, require high supply voltages (1–2 kV) and are costly devices, particularly the high-performance models. EM-CCDs exploit an internal multiplication process to achieve sub-electron readout noise, thus being able to detect single photons. Their quantum efficiency is very high, and they are inherently suited to imaging applications. However, due to their readout technique, they cannot provide frame rates higher than a few kilo-frames per second, and cannot be used in extreme time-resolved measurements. Single photons can be detected efficiently by avalanche diodes operating in Geiger mode, known as Single-Photon Avalanche Diodes (SPADs). Avalanche photodiodes have the typical advantages of solid state devices (small size, low bias voltage, low power consumption, ruggedness and reliability, suited to building integrated systems). In the last few years, a new kind of planar semiconductor device has slowly but steadily come out, namely the silicon photomultiplier (SiPM), with promising features that, in some respect, could even replace traditional photomultiplier tubes (Kovaltchouk et al, 2005). Based on a Geiger mode avalanche photodiode elementary cell, it consists of an array of *n*

independent identical microcells whose outputs are connected together. The final output is thus the analog superposition of *n* ideally binary signals (Buzhan et al., 2003). This scheme, along with the sensitivity of each individual cell to single photons, appears to result, in principle, in the perfect photo-sensor capable of detecting and counting single photons in a light pulse. Unfortunately, this is not the case, considering that this kind of device has several drawbacks and all of them are mainly derived from its noise features; due to lattice defects and impurities in the basic material, the dark counts cannot be reduced below a given rate, and as these mainly have a thermal origin, one could be tempted to solve the problem by cooling the device itself. This works to a given extent; however, another problem sets in, namely afterpulsing, due to charge carriers trapped within the semiconductor during the avalanche signal and later exponentially released. Cooling the device results in an increase of the exponential decay constant, and therefore, the lowest operating temperature becomes a tradeoff between random thermal counts and long-lasting afterpulse counts (Ghioni et al., 1996). This could represent an intrinsic limitation to the implementation of large-area G-APD detectors, if one actually needs the single photon sensitivity. Nonetheless, the suitable use of G-APDs depends strongly on a particular application; although dark counts are a problem for low-level light applications, if there is ample light, one can set the threshold at several photoelectrons and thus suppress them. Such a tradeoff can be useful to optimize the energy resolution. Therefore, although not capable of totally replacing the traditional photomultiplier tubes, the SiPM already promises to fulfill a wide set of requirements coming from numerous applications. Thanks to its properties, the multi-element G-APD is currently promising to find widespread use in nuclear medical imaging applications like Positron Emission Tomography (PET). In this application the G-APD is usually coupled to Lutetium Orthosilicate (LSO) or Lutetium-Yttrium Orthosilicate (LYSO) scintillators which convert gamma-rays into optical photons in the blue and in the near ultraviolet wavelength ranges (Melcher, 1992).

In this chapter, we describe the single and multi-element avalanche photodiode operating in Geiger mode. Their characterization in terms of noise and Photon-Detection Efficiency (PDE) is treated in great detail together with the adopted experimental setups, partly based on optical systems i.e. light sources, filters, monochromator and integrating spheres, and partly based on typical particle counting equipments. The developed technique to obtain very accurate PDE measurements based on single photon counting with subtraction of dark noise, and avoiding as much as possible the noise contribution due to cross-talk and afterpulses is here detailed, as well as the apparatus used for charge signal measurements based on a pulsed laser system, a charge-to-digital converter (QDC) and a time-to-digital converter (TDC). Some measurements and results on various single element G-APDs and multi-element G-APDs, manufactured by different companies, are also discussed. Finally the two most known methods (photocurrent and photon counting) to evaluate the PDE are compared and a discussion on how the noise due to afterpulse and optical cross-talk may influence the measurements is also given.

2. Single element G-APD

A single element G-APD, also called Single Photon Avalanche Diode (SPAD) is essentially a p–n junction biased at a voltage above the breakdown. At this bias, the electric field is so high (higher than 3×10^5 V/cm) that a single charge carrier injected in the depletion layer

can trigger a self-sustaining avalanche and the output current rises (sub-nanosecond rise-time) to a level of milliamps. If the primary carrier is photo-generated, the leading edge of the avalanche pulse marks (with picoseconds jitter) the arrival time of the detected photon. The current continues to flow until the avalanche is quenched by lowering the bias voltage to or below the breakdown level: the lower electric field is no longer able to accelerate the carriers to impact-ionize with lattice atoms. Then the bias voltage must be restored, in order to be able to detect another photon. The circuit that performs such operations is usually referred to as a quenching circuit (Cova et al., 1996). This operation requires suitable electronics, able to:

1. sense the leading edge of the avalanche current;
2. generate a standard output pulse, synchronous with the onset of current;
3. quench the avalanche by lowering the bias below the breakdown voltage;
4. restore the photodiode voltage to the operative level.

The most commonly used circuit in studies on Geiger-mode avalanche photodiodes is the passive-quenching circuit: the avalanche current quenches itself simply by developing a voltage drop across a high-impedance load (RL > 100 kΩ). Such a circuit is very simple and can easily be employed, but sets severe limitations on the maximum admissible photon counting rate and on detector performance in general (Cova et al., 1996). In fact, it was the introduction of the active-quenching circuit (AQC) concept by S. Cova (Cova et al., 1981) that opened the way to practical application of SPADs. Many AQC types have since been reported, with circuit structure and mounting that evolved from standard NIM cards (Brown et al., 1987) to small SMT boards suitable for compact detector modules (Ghioni et al., 1996). But, if some limitation on photon counting rate and on timing response is accepted then, by using quenching resistors, many SPADs can be easily integrated in one chip and a so called multi-element G-APD can be manufactured with sensitive areas comparable to those of small photomultiplier tubes.

A schematic diagram of the AQC developed at INAF Catania Astrophysical Observatory Laboratory for Detectors (Billotta et al., 2009) is shown in Fig. 1. The G-APD is reverse biased through the cathode at $V_{break} + V_{ex}$, where V_{break} is a voltage slightly lower than the

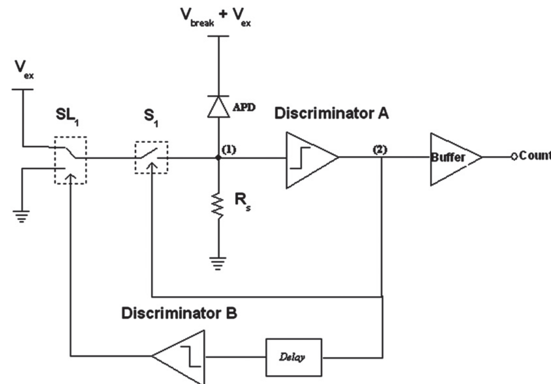


Fig. 1. The single element G-APD is biased and driven by an active quenching circuit (AQC), designed and realized at the INAF Catania Astrophysical Observatory Laboratory for Detectors (COLD), that provides for extinguishing the avalanche, bringing the SPAD to its waiting conditions and after a changeable hold-off time making the SPAD ready to detect another photon.

breakdown and V_{ex} brings the total reverse bias over breakdown. When an avalanche is triggered, the current flowing on Rs activates the discriminator A, which varies the state of node 2, giving a pulse synchronized with the avalanche. A buffer provides for the output of the pulse. Two feedback loops are used, one to quench immediately the diode to reduce the charge trapping and then avoiding afterpulses, and the other to delay the system reset, keeping quenched the diode for a dead time T known as the hold-off time. The first feedback loop acts on S_1 switch forcing the diode anode at V_{ex} voltage, giving as the total voltage V_{break} and thus leaving the diode quenched. The hold-off time is user selectable. After the time T the discriminator B by means of the switch SL_1 forces the node 1 at ground, making the SPAD ready for a new detection. At the same time the discriminators A and B open S_1 and switch SL_1 to V_{ex}.

A sketch of a single SPAD with integrated passive quenching resistor, manufactured by ST Microelectronics and tested at our laboratory, is shown in Fig. 2. The voltage can be applied through the pads 1 and 3 or through the pads 2 and 3 in order to, respectively, include or exclude the quenching resistor R_L.

In Fig. 3 is shown the reverse characteristic at room temperature obtained biasing the device including (curve a) and excluding (curve b) the quenching resistor. Measured current above the BV in curve (a) is lower due to the quenching of the dark events.

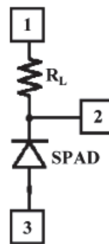


Fig. 2. Sketch of a single SPAD with integrated quenching resistor. The voltage is applied through pads 1–3 to include, or 2–3 to exclude, the quenching resistor R_L.

The plots of Fig. 3 show that the breakdown voltage is about 29.5 V, while the leakage current is few picoamps. The low value of the leakage current indicates a low generation of electrical carriers and then can be considered the first evidence of low defects of this particular kind of G-APD (Mazzillo et al.).

In order to work as a photodetector, a diode must be able to remain biased above the breakdown voltage for a sufficient time, of the order of a few milliseconds. This means that the generation–recombination phenomenon, which would trigger the avalanche, must be kept very low. Since thermally generated carriers can trigger an avalanche, it is possible to observe output current pulses also when a SPAD is kept in the dark: such an average counting rate is called dark-count rate and is one of the key parameters in defining detector noise (Ghioni et al., 1991). This aspect, as can be seen subsequently, becomes critical in multi-element G-APDs. The relevant characteristics of SPAD detectors are:

- dimensions ranging from 20 μm to 200 μm;
- dark count rates that, depending on working temperature, overvoltage and element dimensions, may vary from 20 counts/s to 1000 counts/s;

- photon detection efficiency (PDE), determined by the quantum efficiency (QE) and the trigger probability (TP), that can reach values around 60 % in the visible band.

To understand how the single element G-APD characteristics strongly depend on the bias voltage, or better, on the over voltage OV, the dark noise and the PDE are here considered. In Fig. 4 the dark noise rate of a typical device is represented as a function of OV values.

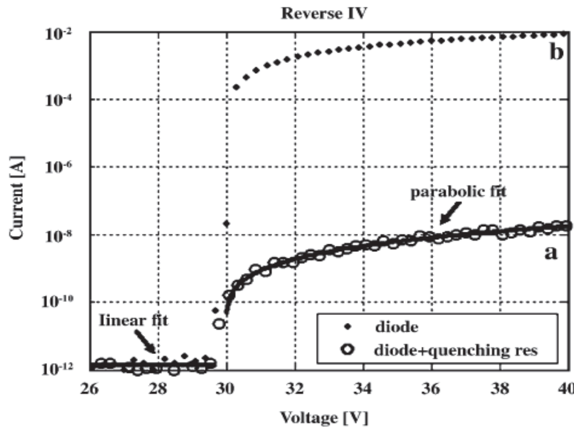


Fig. 3. SPAD reverse I–V characteristics around the breakdown voltage. (a) The quenching resistor was included. (b) The quenching resistor was excluded.

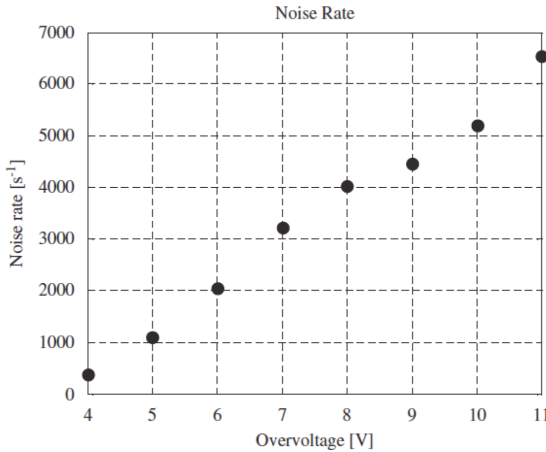


Fig. 4. SPAD dark noise rate at room temperature and increasing overvoltage.

These measurements were done at room temperature. As can be seen in Fig. 4, the dark noise rate increases almost linearly with OV. This evidence suggests that afterpulsing effects are not quite relevant also when high OV is applied to the device. In fact if afterpulsing were dominant the dark noise rate would grow steeper than linearly. Fig. 5 shows QDE (PDE with 100% of trigger probability) in the 350 ÷ 900 nm spectral range for a typical device as a function of the OV.

Both QDE curves show the same shape and reach the maximum at wavelengths between 400 and 500 nm. At 450 nm and 10 % of overvoltage we found a QDE of 30% and a QDE of 17% at 5% of overvoltage. This behavior is strictly related with the triggering probability that increases with the OV. Both QDE curves show also that biasing the device with these values, the Geiger efficiency is far from being near its maximum value, in fact the QDE scales almost linearly with the overvoltage. This means that the best operating condition will be a tradeoff between the acceptable dark count rate (that can be lowered by cooling the device) and the PDE, that of course, is preferred as high as possible. In Fig. 6 are plotted PDE values obtained from characterization of this kind of devices operated at the best condition of OV and temperature, carried out at the INAF - Catania Astrophysical Observatory (Billotta et al., 2009).

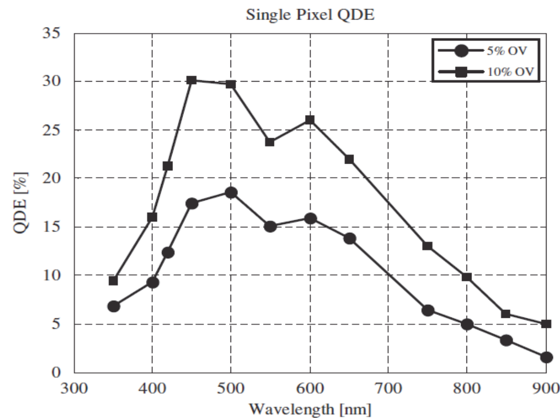


Fig. 5. SPAD quantum detection efficiency at room temperature and increasing overvoltage.

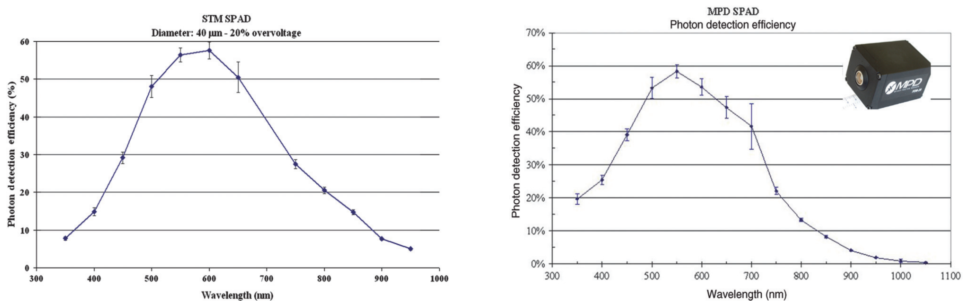


Fig. 6. (Left) PDE of a STMicroelectronics 40 µm device biased at 20% overvoltage; we can note the peak of about 60% at 600 nm. (Right) PDE of a MPD 100 µm SPADs. This device shows a PDE peak of about 60% at 550 nm.

3. Array of single element G-APD

There are applications that may require arrays of single element G-APDs. For instance, one of the toughest problems affecting ground-based telescopes is the presence of the

atmosphere, which distorts the spherical wavefront, creating phase errors in the image-forming ray paths. Even at the best sites, ground-based telescopes observing at visible wavelengths cannot achieve an angular resolution in the visible better than telescopes of 10 to 20 cm diameter, because of atmospheric turbulence alone. Adaptive optics is the answer to this problem: a deformable mirror is inserted in the light path of the telescope, and its control signal is based on measurement of the incoming wavefront, performed by a suitable high-sensitivity detector. Single element G-APD arrays can be used as curvature wavefront sensors (CWFS). By using a pulsed laser system, the array performs better than a CCD, thanks to its gating function and parallel readout, which allow faster loop cycles (Zappa et al., 2007). In Fig. 7 is shown a drawing of a planar array fabricated by STMicroelectronics (Mazzillo et al., 2007). This is manufactured by the integration of 25 pixels with a square geometry of 5 × 5.

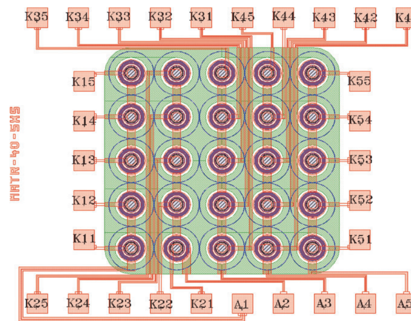


Fig. 7. Plan view of an array fabricated by STMicroelectronics. This device is manufactured by the integration of 25 pixels with a square geometry of 5 × 5.

STMicroelectronics has designed arrays with three different pixel diameters: 20, 40 and 60 mm. Separation distances between adjacent pixels are in the range of 160 and 240 mm according to different diameters. Anode contacts are in common for each row, while each cathode is separately contacted and available from outside by different pads. The typical breakdown voltage is about 30 V.

4. Multi-element G-APD

Multi-element G-APDs also known as silicon photomultipliers (SiPMs) or multi-pixel photon counters (MPPCs), have been developed during recent years as a possible alternative to vacuum photomultiplier tubes (PMT) and avalanche photodiodes (APD). A multi-element G-APD, is a photodetector constituted by hundreds to thousands of single G-APD, and the discharge is quenched by a small transparent polysilicon resistor (passive quenching) integrated on each cell's cathode. The independently operating cells are connected to the same readout line and therefore the combined output signal corresponds to the sum of all fired pixels. A schematic diagram is shown in Fig. 8.

A typical G-APD reaches an intrinsic gain for a single photoelectron of 10^6, comparable to that of vacuum phototubes (PMTs). Fig. 9 shows a G-APD of a 10 × 10 array manufactured by STMicroelectronics.

A snapshot of persistence plots taken on a digital scope is shown in Fig. 10, where the (upper plot) low-light-level pulses, generated by a laser, produce the typical equally spaced

electrical signals corresponding to discrete numbers of photons detected. The signal rise time is below 2 ns; its duration is 10 ns. The dark-count signals, shown on the lower plot, basically show up as one-cell pulses.

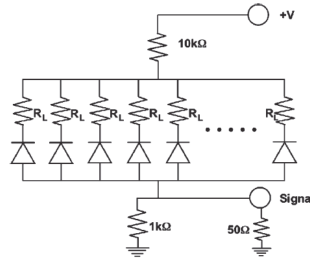


Fig. 8. Electrical schematic of the SiPM, its biasing circuit, and output signal extraction.

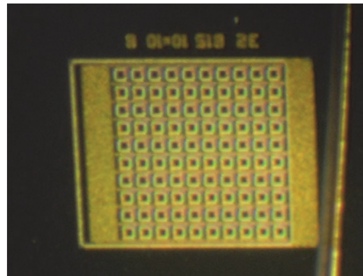


Fig. 9. Microphotograph of the 10 × 10 SiPM with 50 µm pitch. Each cell's active area, 30 µm wide, appears in the picture as a light transparent polysilicon resistor frame surrounding a darker central spot.

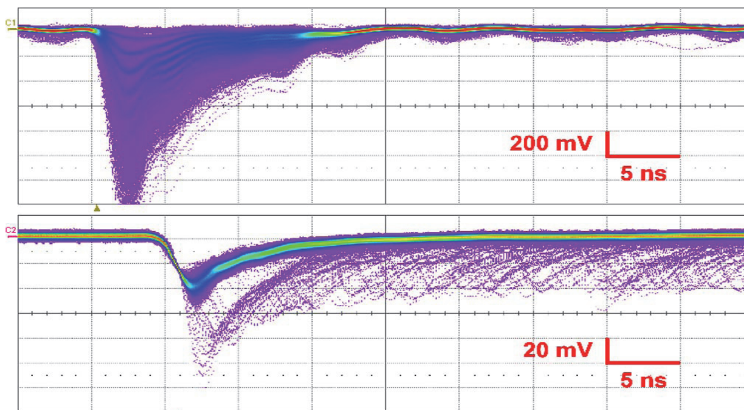


Fig. 10. Persistence plots on a digital scope of a 10 × 10 cells G-APD output signal. The upper plots is obtained by illuminating the detector with low-light-level pulses generated by a laser. The equally spaced signals correspond to discrete numbers of photons detected. The lower plot basically shows as the dark counts are due to one-cell pulse.

In multi-element G-APDs the dead area due to the quenching resistor introduces another parameter that is named "fill factor". This parameter is responsible of the lower PDE as compared to that of the single element. Indeed, as it is better explained in section 9, the PDE is the product of three factors: the quantum efficiency (QE), the trigger probability (TP) and the fill factor (FF). The electro-optical characteristics of these devices are detailed in the following sections.

5. Characterised detectors

The characterisation activity described here regards two kinds of G-APD operating in photon counting regime: SiPMs manufactured by STMicroelectronics and MPPCs manufactured by Hamamatsu. These G–APDs are biased slightly above the breakdown by an overvoltage (around 10% for the STMicroelectronics and few percents for the Hamamatsu). The passive quenching resistor is integrated on the cathode for the STMicroelectronics and on the anode for the Hamamatsu. In particular here we present the characterisation results for three G-APDs:

1. a 10 × 10 single elements (100 cells) STMicroelectronics SiPM;
2. a 10 × 10 single elements (100 cells) Hamamatsu MPPC (S10362-11-100C);
3. a 20 × 20 single elements (400 cells) Hamamatsu MPPC (S10362-11-050C).

The detectors have the following geometric characteristics:
the 100-cells SiPM manufactured by STMicroelectronics has dimensions of 0.5 × 0.5 mm^2, a pitch of 50μm and a fill factor of 36 %;
the 100-cells MPPC Hamamatsu has dimensions of 1 × 1 mm^2, a pitch of 100μm and a fill factor of 78.5%;
the 400-cells MPPC Hamamatsu has dimensions of 1 × 1 mm^2, a pitch of 50 μm and a fill factor of 61.5 %;
and the following electrical characteristics:
the STMicroelectronics SiPM has a breakdown voltage around 29.5 Volts at room temperature, with a variation coefficient of 35 mV/°C;
both Hamamatsu MPPCs have a breakdown voltage around 68.6 V at room temperature. The 100-cells MPPC has a gain $G=2.4 \times 10^6$ (at a Bias Voltage = 69.7 V), the 400-cells has a $G=7.1\times 10^5$ (at a Bias Voltage = 69.8 V).
Furthermore we have to note that, on the contrary of Hamamatsu MPPC, each STMicroelectronics SiPM cell is surrounded by a suitable trench filled with opaque material to avoid that photons produced during the Geiger discharge may be detected by neighboring cells. This effect is commonly named as "optical cross-talk" (Dolgoshein, 2006).

6. Experimental setups for multi-element G-APDs electro-optical parameters measurements

As written in the previous section, in this chapter we report on a complete characterisation of three different G-APDs. The purpose is to explain how accurate measurements of the relevant electro-optical parameters can be carried out to better qualify the detector performances. We measured the following characteristics:

- dark counts, afterpulse and optical cross-talk that represent the noise sources;
- charge response and gain;
- photon detection efficiency (PDE).

To evaluate the above characteristics four different setups are used:

1. A counting system (for simple dark-count measurements) sketched in Fig. 11. The detector output is connected to an amplifier (a FTA810B, with gain 200 and rise time below 1ns) that produces a voltage signal, and forms the input signal of the discriminator (a Lecroy 4608) that, depending on the threshold levels generates the pulses to be counted.
2. A self-correlated timing apparatus (for afterpulse measurements) sketched in Fig. 12. The SiPM signal is acquired by the discriminator that generates two delayed signals, one to start and the other to stop the time-to-amplitude converter TAC (an Ortec 457). The TAC allows a tunable range between 50 ns and 5 µs.
3. A charge signal apparatus (for SiPM response characterization) sketched in Fig. 13. A laser (a 671 nm pulsed laser with FWHM pulse width of 40 ps), through an optical fiber, illuminates the G-APD. The G-APD output signal is connected to a discriminator (Lecroy 4608). The TTL output from the laser is used to generate the gate for the QDC.
4. An optical system (block diagram in Fig. 14) and an electronic equipment (sketched in Fig. 15) for PDE measurements. In the optical system a Xenon lamp is used as radiation source, the wavelength selection is performed by a Czerny-Turner monochromator (FWHM ≤ 1 nm in the 130–1100 nm spectral range) and a beam splitter directs the monochromatic radiation towards an integrating sphere that guarantees a spatial integration of the radiant flux on a 1 cm^2 reference photodiode (NIST traced) and on the detector to be characterized. Furthermore, we designed the detector housings, in such a way to have same aperture and distance from the centre of the sphere. The calibrated photodiode allows to evaluate the number of photons per unit area, and then, after proper rescaling, the number of photons on the detectors. The optical apparatus used for PDE measurements is one of the available facilities at "INAF-Catania" laboratory (Bonanno et al., 1996). The electronic equipment is essentially constituted by a counting system and an electrometer (Keithley 6154) that measure the photo-current from the NIST calibrated photodiode. The electrometer is connected to a PC through an IEEE 488 interface.

The actual configuration of the electronics was slightly modified and fine-tuned with respect to what is shown in the figures based on specific needs.

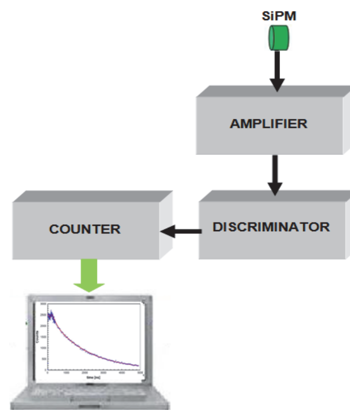


Fig. 11. Sketch of the setup utilized for dark counts measurement.

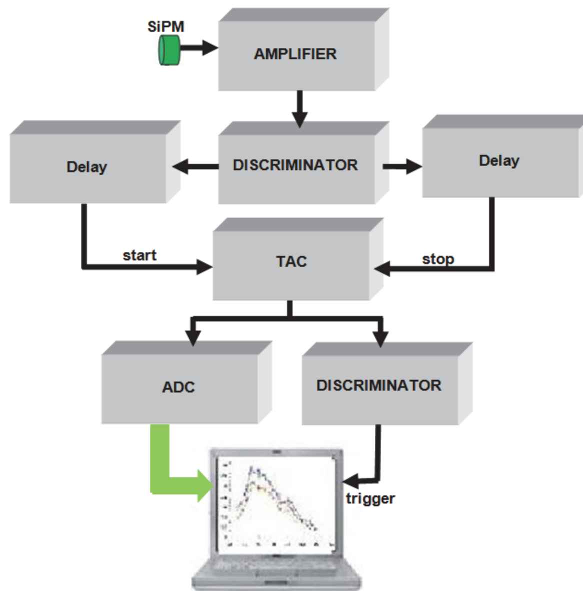


Fig. 12. Sketch of the apparatus for the self-correlated timing, used for afterpulse measurements.

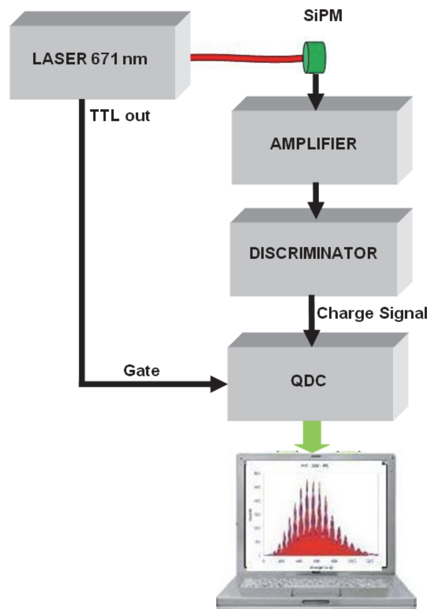


Fig. 13. Sketch of the electronics for the charge measurements used for SiPM response characterization.

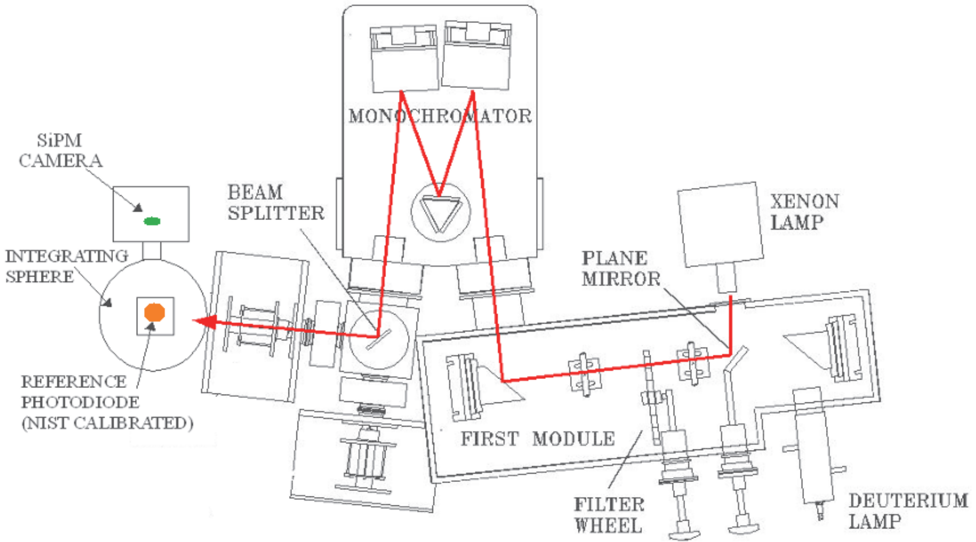


Fig. 14. Schematic of the optical apparatus used for PDE measurements. The light path is also shown.

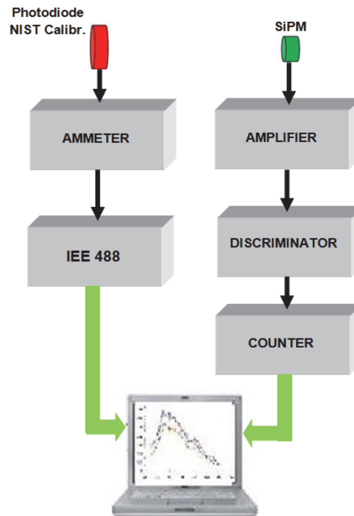


Fig. 15. Sketch of the electronics implemented for the PDE measurement.

7. Dark counts and afterpulse measurements

In order to evaluate the dark-count rate of the characterized devices, each detector is placed into a thermally stabilized light-tight box. By using the electronic setup shown in Fig. 11, the

number of noise pulses generated per unit time as a function of the discriminator threshold is measured. Fig. 16 shows the resulting plots for the G-APDs. The STMicroelectronics device is biased with a voltage of 32.5 V while the Hamamatsu device is biased with 69,7 volts. The threshold is normalized to the one-photon signal amplitude.

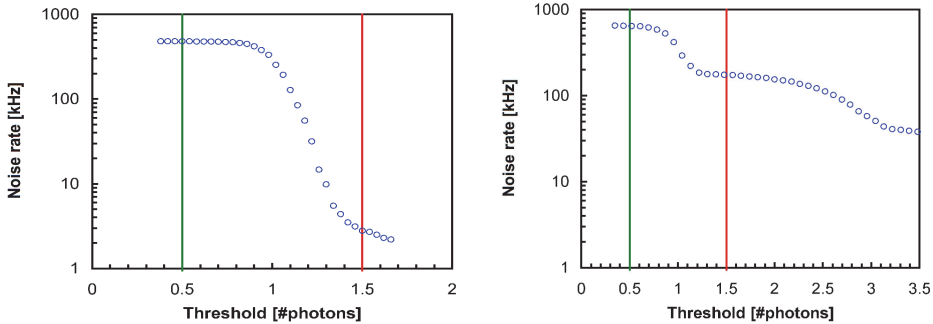


Fig. 16. Measured noise rate as a function of the discriminator threshold for the STMicroelectronics device (left), and for the Hamamatsu device (right). The noise level at 1.5 pe- is quite low for the STMicroelectronics G-APD, indicating a low level of correlated noise, while for the Hamamatsu MPPC the noise level is still considerable (also at 2.5 pe- and 3.5 pe-), confirming a non-negligible correlated noise level.

From Fig. 16 is clearly evident that even though the two G-APDs have dark noise of the same level at 0.5 pe- threshold, the noise of the STMicroelectronics device is strongly reduced at 1.5 pe- threshold. The same amount of reduction is not seen for the Hamamatsu sensor even at 3.5 pe- threshold, indicating the presence of correlated noise. A rough estimate of such correlated noise can be obtained by computing the ratio between the counting rate at 1.5 pe- and 0.5 pe- threshold. This ratio is about 0.5% for the STMicroelectronics G-APD and 27% for the Hamamatsu device. The cause of this difference is surely due to the optical trench technology adopted by STMicroelectronics that reduces considerably the optical cross-talk contribution. From these measurements rises the necessity to measure the correlated noise. An evaluation of the contribution of this noise to the real signals, can be given, for example by measuring the distribution of the time intervals between two consecutive dark pulses. For this purpose the setup sketched in Fig.12 is used. The amplified signal is passed to the discriminator, that generates two fast logic outputs that, in turn, are appropriately delayed, with one used as start and the other as stop of the TAC (Fig. 12). After a precision calibration of the time scale (Finocchiaro et al., 2008) a delay configuration is chosen to have a self-coincidence peak below the overall TAC threshold, and thus the system was only triggered whenever, following the main signal, there was another pulse between 50 ns and the full time range. In these conditions, the distribution of time intervals between two consecutive signals is measured. More details about the afterpulse noise measurement can be found in (Finocchiaro et al., 2008). In Fig. 17 can be seen the result of measurements for both G-APDs.

From Fig. 17 is evident that while the overall behavior for the STMicroelectronics G-APD is exponential, according to Poisson's law (with a small bump due to afterpulsing around 200 ns), for the Hamamatsu device we found a decreasing function with at least three exponential slopes indicating a non-random behavior.

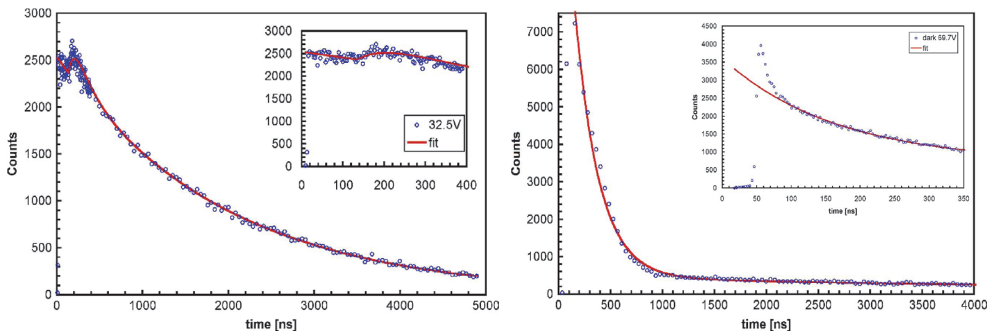


Fig. 17. Distribution of the time interval between two consecutive dark pulses for the STMicroelectronics SiPM biased at 32.5 V (left) and the Hamamatsu MPPC biased at 69.7 V (right).

8. Gain measurements

The gain measurements are of fundamental importance in computing the Photon Detection Efficiency (PDE) considered as ratio between the photo-current of the tested detector and that of the calibrated one. Uncertainties on the gain measurement directly affect the PDE values. As written in sectio n 2, each single element of the devices operates in Geiger mode, and then the interaction of one photon produces an electron-hole pair followed by an avalanche multiplication. The avalanche multiplication factor is the gain that we indicate with G, and depends on the bias voltage. For the G measurements we used the setup sketched in Fig. 13. By setting the laser intensity at various levels, the charge spectrum for each detector has been acquired. The G has been obtained by computing the average spacing between two consecutive peaks in terms of QDC channels. Values in the 10^4 to 10^5 range have been found. As an example in Fig. 18 the STM 100-cells SiPM and the Hamamatsu 100-cells MPPC charge spectra are shown.

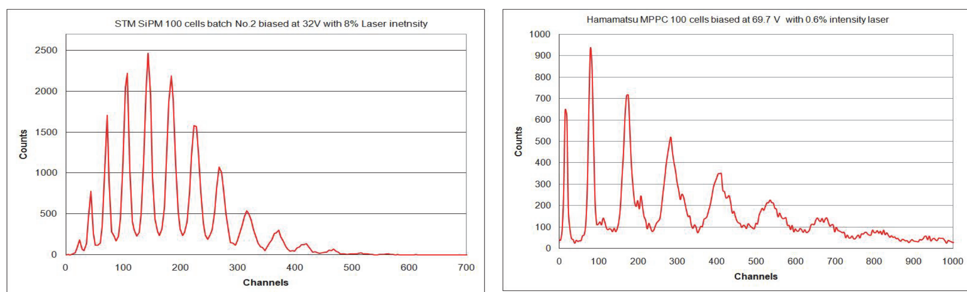


Fig. 18. On the left is plotted the charge spectrum from the STM 100-cells SiPM. On the right is the charge spectrum from Hamamatsu 100-cells MPPC.

Surprisingly for both Hamamatsu devices we have found values of G about one order of magnitude smaller than those reported on the data sheets provided with the detectors. To locate the error sources, we also checked the amplifier by using a calibrated source.

Different configurations have been investigated and some other measurements were carried out with the same results.

9. PDE measurements

Only a fraction of the photons impinging on the sensor will actually trigger an avalanche to produce a detectable signal (Piemonte 2006). Essentially three effects influence a G-APD response efficiency:

1. physical (reflection/absorption by passive layers, material), that is the so called net quantum efficiency (QE);
2. electrical (photon arrival in regions where the electric field is not suitable for triggering the avalanche), that represents in practice the probability that an event occurs and generally is named Trigger probability (TP).
3. geometrical (dead areas between cells), and is generally known as fill factor (FF);

The overall efficiency of the sensor, as for the single element, is generally named Photo Detection Efficiency (PDE), and it relates the real number of impinging photons to the measured effect (photo-electrons) and is the product of the three above mentioned effects:

$$PDE = QE \times TP \times FF \tag{1}$$

In the following sections the reader will be introduced into an important aspect to be considered when the detector PDE has to be evaluated with high accuracy. The requirement to have a well defined methodology, taking care, not only on the precision of all involved instruments, but also on the implemented procedure, is crucial to obtain precise measurements. Here we will demonstrate how the extra noise sources, optical cross-talk and afterpulse, may influence the PDE measurements. In fact, to measure the detector PDE essentially two approaches can be used:

1. one consisting in measuring the generated charges considered as current, that we name: “Photocurrent” method,
2. and another consisting in counting each produced event, that we name “Photon counting” method.

The PDE measurements for both methods have been carried out by using the optical setup sketched in Fig. 14 and the electronic setup sketched in Fig. 15.

The first consideration, to obtain accurate measurements, is addressed to the different dimensions of both detectors, the G-APD and the reference photodiode. In fact, while the tested devices have dimensions of squared millimeter, the reference detector have a sensitive area of 1 cm^2 (leakage current less than 1pA), thus in the “Photon counting” case, we have to adjust the photon flux level (from about 10^5 to about 10^7 phs $\cdot mm^{-2} \cdot s^{-1}$) in such a way that the reference detector was still sensitive and the detectors were safely in the single photon regime with negligible pile-up.

9.1 Photocurrent method

The “Photocurrent” method consists in comparing the photocurrent of the characterized detectors with respect to that of the NIST calibrated reference photodiode. In this case the setup apparatus of Fig. 15 is simplified by substituting the amplifier, the discriminator and the counter with an ammeter. In practice we have two identical systems, one for the tested and one for the reference detector, and simply we have to do measurements of the photo-generated current in both sensors. The following formula explains how the method works:

$$PDE = [(I_{Det} - I_{DarkDet}) / (I_{PhD} - I_{DarkPhD})] \times G^{-1} \times PDE_{PhD} \times (A_{PhD} / A_{Det}) \quad (2)$$

Where I_{Det}-$I_{DarkDet}$ is the current measured in the tested detector, I_{PhD}-$I_{DarkPhD}$ is the current measured in the calibrated photodiode, G is the gain (N_{el}/pe-), PDE_{PhD} is the PDE of the calibrated photodiode and A_{PhD}/A_{Det} is the detectors area ratio.

We operated the detectors at room temperature and measured the PDE of the STM SiPM biased at 32.5V (10% OV) and that of the 100 and 400 cells MPPC biased respectively at 69.8V (~2% OV) and at 69.4V (~2% OV). Using the G values obtained with our measurements, we found unreasonable PDE values (higher than expected). Thus, the sole alternative we had was using the G values given by the manufacturers. Despite a sort of uncertainty of the method, due to the fact that we have to rely on manufacturer's measurements accuracy, we decide to compute the PDE. We made the PDE computation only on the two Hamamatsu MPPCs. The obtained values are plotted in Fig. 19.

As expected the PDE of the 100 cells MPPC at 450 nm has a peak of about 50%, while the 400 cells MPPC has a peak of 30% because of the different fill factor. Now we have to investigate if these results are realistic or the noise contribution has to be taken into account and avoided as much as possible. It is clear that a technique, based on photocurrent measurements, is unable to discriminate from extra-generated pulses, i.e. afterpulses and optical cross-talk pulses, and thus two questions rise:

- Can we include in each PDE value an amount of pulses that is considered "noise"?
- Can we say that the obtained PDE values are accurate?

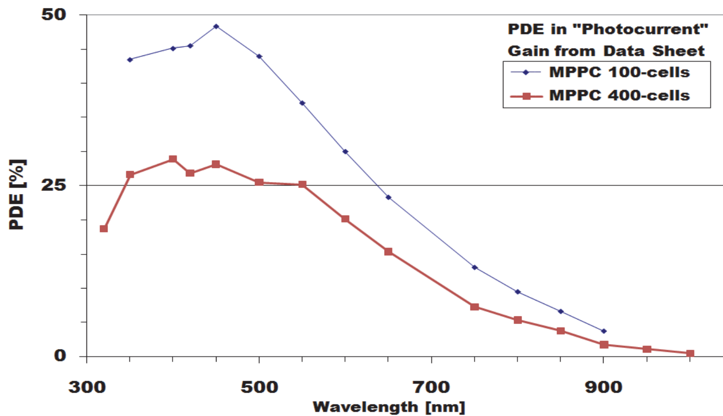


Fig. 19. PDE plots of the two Hamamatsu G-APDs: the 100-cells MPPC and 400-cells MPPC by using the "Photocurrent" method.

If it is impossible to discriminate the extra pulses with respect to the real signal, probably the photocurrent method may lead to overestimated PDE values, and will be better to use another method that can discriminate the real photo-events from extra pulses.

9.2 Photon counting method

The "Photon counting" method is based on measuring the G-APDs count rate due to the real photo-events and comparing it to the photocurrent measured by the ammeter converted into number of electrons per second. The formula of this method is:

$$PDE = \left[\left(CR_{Det} - CR_{DarkDet} \right) / \left(I_{PhD} - I_{DarkPhD} \right) \times PDE_{PhD} \right] \times e^{-} \times \left(A_{PhD} / A_{Det} \right) \quad (3)$$

Where CR_{Det}-$CR_{DarkDet}$ is the measured count rate, e^- is the electron charge and I_{PhD}-$I_{DarkPhD}$, PDE_{PhD}, A_{PhD}/A_{Det} are the same as on formula (2).

By using this method the afterpulse and the cross-talk can be characterized and taken into account in the right way, in fact we can set the threshold at a convenient value and can acquire the signal at a selected time (by varying the time length of the digital output pulse from the discriminator) away from the eventual afterpulse contribution. The first step to carry out the PDE measurement is to analyse the count rates as a function of the threshold. As seen in Fig.16 of section 7, a threshold equivalent to 0.5 photons can be selected as this value is in a safe plateau region. In the tested devices we found that the afterpulse probability is not appreciable after ≈100ns and thus we settled the output logic signal duration from the discriminator longer than this value. We counted the number of pulses per unit time both in dark conditions (~ 600 KCnts/s for the 100-cells MPPC, ~ 500 KCnts/s for the 400-cells MPPC, ~ 500 KCnts/s for the 100-cells STMicroelectronics device) and with monochromatic light conditions (photon signal ranging from ~ 100 KCnts/s to ~ 500 KCnts/s), recording at the same time the light level seen by the reference detector, for several wavelengths. We also carefully tuned the light intensity to keep at negligible levels the pile-up probability. As an example here the analysis made on both the STMicroelectronics and Hamamatsu 100-cells G-APDs is presented. For both devices we evaluated the PDE by measuring all the contributing signals, noise and photons with two gate logic signal durations and accounted for the dead time. For the STMicroelectronics we selected the duration of 50 ns and 500 ns and the resulting PDE plots are shown in Fig. 20, while for the Hamamatsu device we selected the duration of 100 ns and 1000 ns and the resulting plots are shown in Fig. 21. The unappreciable difference between the two sets of

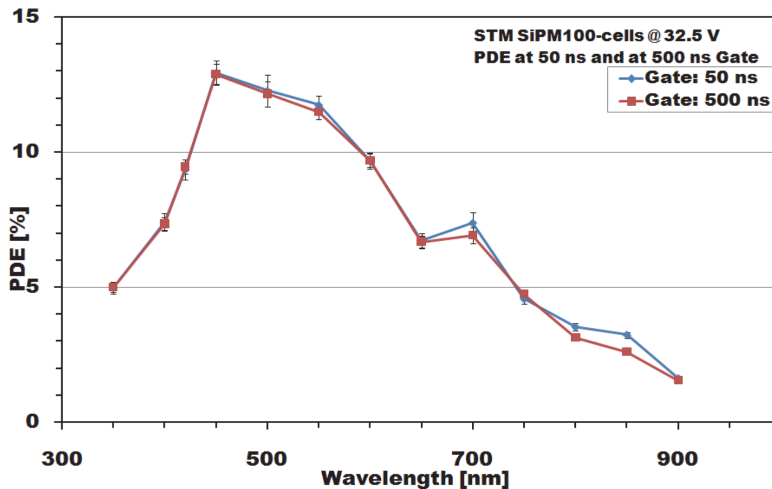


Fig. 20. PDE of the 100-cells SiPM STMicroelectronics device biased at 32.5 V, measured and reconstructed with our method using logic signal durations of 50ns and 500ns respectively. As can be noted the difference between the two sets of measurements is unappreciable, meaning that the afterpulse effect not influence each measure.

measurements, for both G-APDs, demonstrates that the afterpulses are not influent on each measure and strongly supports the correctness of this method.

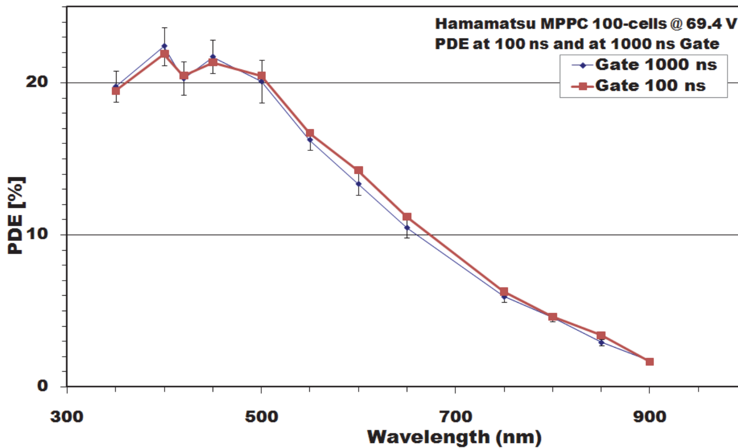


Fig. 21. PDE measured for the Hamamatsu 100 cells biased at 69.4 V using gate signals of 100ns and 1000ns. As can be noted the difference between the two sets of measurements, also in this case, is within the error-bar, meaning that in these measurements the afterpulses are not a problem.

As can be noted from Figs. 20 and 21 the PDE plots of the two G-APDs are quite different specially in the 350 ÷ 450 nm spectral region. This is essentially due to the different technology adopted by the two manufacturers. In the case of Hamamatsu device (that uses the so called p-on-n junction technology) the photons impinging in the first layers of material are absorbed more efficiently than those arriving in the same region of the STMicroelectronics device (that uses the so called n-on-p junction technology).

10. Comparison between “photocurrent” and “photon counting” methods

In order to compare the photocurrent method with the photon counting one, we have plotted in Fig. 22 the PDEs obtained with the two methods for the Hamamatsu MPPC 100-cells.
As can be seen from Fig. 22, the PDE obtained with the photocurrent method is systematically higher than that measured with the photon-counting mode in all the spectral range. Moreover the error-bars associated to the PDE values are very low (not exciding the point itself) demonstrating the high accuracy of measurements and the real difference between the two PDE curves. Unequivocally, Fig. 22 shows that each PDE value obtained using the photocurrent method doubles that of the photon counting operating mode. We, thus have to conclude that the extra noise pulses heavily influence the detector PDE evaluation. A different way that allows us to better clarify the real difference between the two methods, is to represent the two PDE plots as in Fig. 23 where the left axis is used to represent the PDE values obtained with the photocurrent method and the right axis refers to the PDE values obtained in photon counting mode. In order to better understand this figure, it is extremely important to note that the right axis scale (that refers to the photon counting mode) is exactly half of that of the principal axis (that refers to the photocurrent mode).

From the Fig. 23 we can observe that even if the two PDE plots came from different methods, there's an amazing over-position between the two plots. This demonstrates that at each wavelength the PDE values obtained with the two different methods can be related between themselves, and by noting the scale of the left axis respect to right axis, the relation is that each value almost doubles the corresponding. And then, definitively, we can conclude that the PDE of this device in photon counting mode is half of that in which we can't avoid the extra pulses contribute.

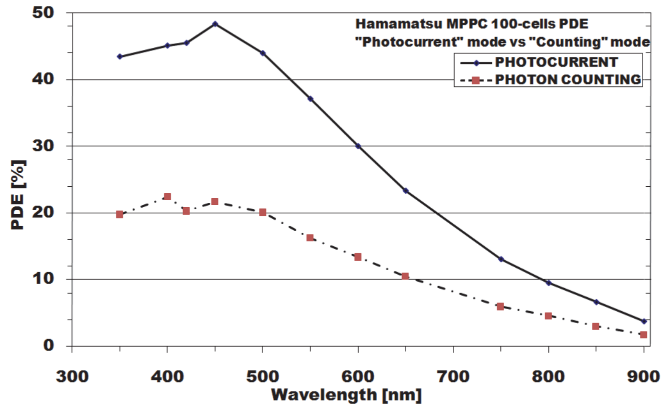


Fig. 22. PDE measurements for a 100-cells Hamamatsu MPPC. The solid line refers to the PDE obtained with the photocurrent method, while the dashed line refers to the PDE obtained with the photon counting technique. Unequivocally the PDE values obtained using the photocurrent method doubles that of the photon counting.

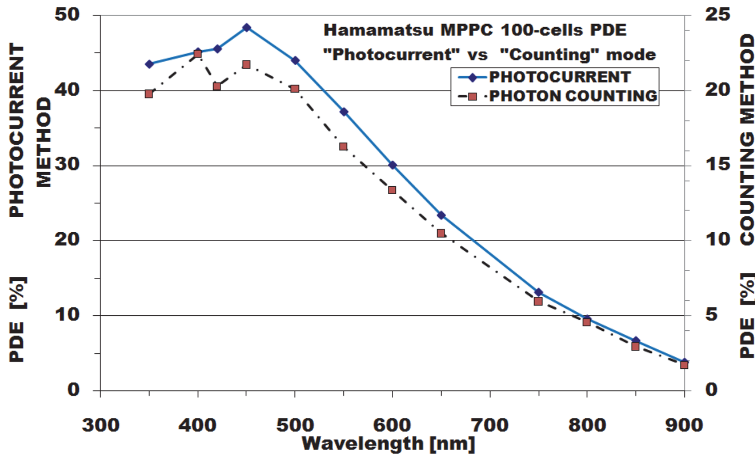


Fig. 23. "Photocurrent" method versus "Photon counting" method. The solid line refers to the PDE (values on the left axis) obtained with the photocurrent method, while the dashed line refers to the PDE (values on the right axis) obtained in photon counting regime. The right axis scale is half of that that refers to the PDE obtained with the photocurrent method.

11. Conclusion

In this chapter, a detailed description of a particular kind of photodiodes able to work in Geiger avalanche mode recently named G-APDs has been described. Starting from a description of the relevant characteristics of the single G-APD we extended to describing the multi-element G-APD as a photodetector constituted by hundreds/thousands of single elements. By discussing in detail the manufacturing technology and the relevant electro-optical characteristics of these devices, we tried to give an idea of the real achievable performance in application such as Nuclear Physics or Astrophysics. The characterisation in terms of noise, and Photon-Detection Efficiency (PDE) has been treated in great detail for both kind of devices together with the adopted experimental setups. Some measurements and results on various single element G-APDs and multi-element G-APDs, manufactured by various companies have been also presented. Finally, emphasis has been given to the developed technique to obtain very accurate PDE measurements based on single photon counting with subtraction of dark noise, and avoiding as much as possible cross-talk and afterpulses. We discussed and compared the two commonly used techniques to measure the PDE, the photocurrent consisting in measuring the photo-generated current in the detector, and the photon counting consisting in measuring the signal considered as number of photons. The comparison between the two methods has pointed out the vulnerability of the photocurrent method that gives PDE values overestimated with respect to those from photon counting. We demonstrated unequivocally that this is essentially due to the fact that the photocurrent technique cannot discriminate the afterpulse and the cross-talk effects. On the contrary, the photon counting method allows to characterize and accurately discriminate the two noise effects providing PDE values quite close to the real ones, but needs to operate in appropriate signal conditions, in fact very fast events can be lost and the total counted events can be lower than those expected. Then we can conclude that the photon counting is a method well suited for PDE measurements because it definitely deals with true photons, reducing as much as possible the contribution of extra pulses.

12. References

S. Billotta et al., JMO, Vol. 56, 273–283 (2009).
G. Bonanno et al., SPIE Proceedings, 2808, p.242 (1996).
R.G. Brown et al., Appl. Opt. 26, 2383 (1987).
P. Buzhan et al., Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip., vol. 504, no. 1–3, 48–52, (2003).
S. Cova et al., Appl. Opt. 35, 1956 (1996).
S. Cova et al., Rev. Sci. Instrum. 52, 408 (1981).
B. Dolgoshein et al., Nuclear Instruments and Methods in Physics Research A 563, 368–376 (2006)
P. Finocchiaro et al., IEEE Trans. on Electron Devices, Vol. 55, no. 10, 2757-2764 (2008).
P. Finocchiaro et al., IEEE Trans. on Nucl. Sci., Vol. 56 no. 3, 1033-1041 (2009).
M. Ghioni et al., Rev. Sci. Instrum., vol. 67, no. 10, 3440–3448, (1996).
M. Ghioni and G. Ripamonti, Rev. Sci. Instrum. 62 163 (1991).
V. D. Kovaltchouk et al., Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip., vol. 538, no. 1–3, 408–415 (2005).
M. Mazzillo et al., Nucl. Instrum. Methods A Vol. 591, 367–373 (2008).
M. Mazzillo et al., Sens. Actuators A, Vol.138, 306–312 (2007).
C. L. Melcher and J. S. Schweitzer, IEEE Trans. Nucl. Sci., vol. 39, no. 4, 502–505 (1992).
C. Piemonte, Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip., vol. 568, no. 1, 224–232 (2006).
F Zappa et al., JMO Vol. 54, 163-189 (2007).

12

Design of High Quantum Efficiency and High Resolution, Si/SiGe Avalanche Photodiode Focal Plane Arrays Using Novel, Back-Illuminated, Silicon-on-Sapphire Substrates

Alvin G. Stern
AG STERN, LLC, Newton, MA 02467
USA

1. Introduction

The design and development of large scale, high quantum efficiency and high resolution silicon and silicon-germanium (Si/SiGe) avalanche photodiode (APD) focal plane arrays (FPAs) is an active topic of research due to the wide range of scientific, medical and industrial applications for such high sensitivity imagers. Avalanche photodiodes can attain single photon sensitive operation due the large internal device gain that compensates and can fully eliminate the electronic readout noise normally limiting the sensitivity of solid-state detector devices, hence their importance in electronic imaging. Large, wafer scale arrays of ultra sensitive, high resolution silicon and silicon-germanium avalanche photodiodes have not been developed yet, primarily due to the increased fabrication complexity of such detector devices and arrays compared to the more common, non-avalanching detectors such as CCDs and CMOS-APS devices. One major fabrication challenge for avalanche type detectors is the requirement of providing effective optical isolation between adjacent detectors in an array since the avalanche gain process produces photons that could create false detection events in neighboring pixels and thereby increase the noise. Providing effective optical crosstalk isolation becomes more difficult for higher resolution arrays. While it is common for CCD arrays to have a pixel pitch between 12-30 μm and for CMOS-APS devices to have pixel pitch below 10 μm, it becomes more challenging to architect arrays of avalanche photodiodes for example, having such a small pitch due to optical crosstalk. The second major fabrication challenge for linear mode avalanche type detectors, especially critical in arrays is the detector gain uniformity. Detector gain uniformity is a critical performance parameter since an increase in gain excess noise will make the detector arrays unsuitable for precision metrology applications. As solid-state avalanche detectors are made smaller, it becomes more difficult to control the gain excess noise due to smaller area multiplication regions where the effects from slight variations in doping profiles and electric fields produce greater gain variability compared to larger area detectors.

In this chapter, design aspects of a novel, back-illuminated silicon-on-sapphire material system are presented and compared to present substrate technologies to illustrate the capability of the novel substrates in solving optical crosstalk and detector gain uniformity fabrication challenges for producing high quantum efficiency and high resolution wafer scale arrays of Si/SiGe APD detector arrays. The novel substrate design incorporates a single crystal, epitaxially grown aluminum nitride antireflective layer between sapphire and silicon to improve optical transmittance into the silicon from sapphire. A $\lambda/4$-MgF_2 antireflective layer deposited on the backside of the sapphire improves optical transmittance from the ambient into the sapphire. The high transmittance, back-illuminated silicon-(AlN)-sapphire substrates represent an enabling technology for producing radiation tolerant, high resolution, wafer scale arrays of solid-state light detectors. (Stern & Cole, 2008) The Si and SiGe solid-state avalanche photodiodes for example, could be produced in highly uniform wafer scale arrays by liquid crystallographic etching of mesa pixels due to sapphire acting as a natural etch stopping layer. Mesa detectors and arrays would retain high quantum efficiency and sensitive-area-fill-factor respectively, due to light focusing monolithic sapphire microlenses beneath each pixel. The space between mesa detectors could be filled with metal to form a low-resistance contact across the array and also block direct pixel-to-pixel optical crosstalk. The closely integrated monolithic sapphire microlenses also help to address detector gain uniformity by focusing optical k-vectors directly into the active multiplication region of the avalanche photodiodes, thereby helping to improve the gain uniformity of the detectors and arrays. Coupled with recent advances in dual linear and Geiger-mode avalanche detector design, the novel substrates will enable wide dynamic range focal plane arrays operating near room temperature, capable of imaging over the full range of natural illumination conditions from AM 0 in space to a cloudy moonless night. (Stern & Cole, 2010)

The novel, back-illuminated silicon-on-(AlN)-sapphire substrates offer the possibility of solving the fabrication challenges currently limiting the low cost availability of highly sensitive, wide dynamic range Si and SiGe avalanche photodiode arrays, including direct pixel-to-pixel optical crosstalk and detector gain uniformity. There still exists however, the phenomenon of indirect optical crosstalk by multiple reflections in the finite thickness, 50 μm thick sapphire substrate. It will be shown through detailed calculations and analysis means that indirect optical crosstalk through the 50 μm thick sapphire substrate although present, will not prevent high resolution, 27 μm pixel pitch Si/SiGe APD detector arrays operating in the highest (Geiger-mode) gain regimes with low noise across the full 1024x1024 pixel FPA for a *f/#* = 5.6 optical system. This significant result confirms that the novel substrates will enable a new class of highly sensitive, solid-state, wide dynamic range, Si/SiGe detector arrays.

2. Technology of silicon avalanche photodiode focal plane arrays

The present approaches to fabricating solid-state Si/SiGe avalanche photodiode (APD) arrays have been constrained by the less than optimal substrates available for fabricating such specialized light detector arrays. Two prevailing approaches have been used in fabricating such APD detector device arrays and both approaches borrow heavily from the fabrication and substrate technology used in more common CCD and CMOS-APS sensor arrays. The first approach shown in Fig. 1 is the simplest and uses conventional CMOS foundry processing for electronic circuits that is also ordinarily used to fabricate low cost,

front-illuminated CMOS-APS sensor arrays, to fabricate front-illuminated avalanche photodiode arrays. The silicon APD focal plane array design approach in Fig. 1 is known as planar CMOS technology because the detector array is fabricated in the same silicon substrate as the integrated pixel control readout electronics. The planar CMOS approach is cost effective because new substrate technology is not needed and existing silicon IC fabrication technology can be leveraged. Planar CMOS technology has been adapted in novel ways for silicon APD arrays by researchers in Italy and Switzerland. (Charbon, 2008; Guerrieri et al., 2009; Niclass et al., 2005) The usual limitations for solid-state detector arrays apply in using the planar silicon CMOS approach including reduced quantum efficiency inherent for front-illuminated devices and less than optimal array sensitive-area-fill-factor due to the space taken up by the pixel electronics.

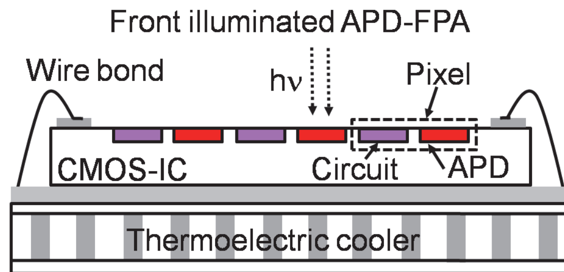


Fig. 1. Planar CMOS technology approach for fabricating cost effective silicon APD focal plane arrays.

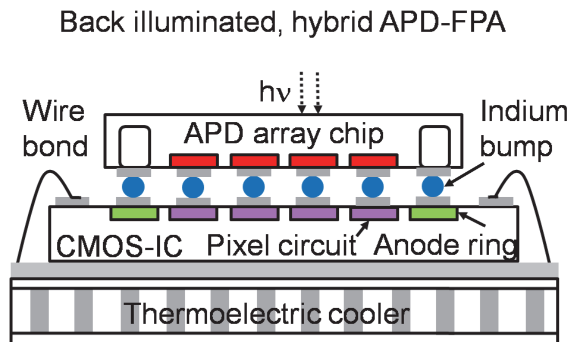


Fig. 2. Hybrid approach for fabricating high performance Si/SiGe APD focal plane arrays.

The second approach shown in Fig. 2, uses a hybridized focal plane array that consists of a back-illuminated detector array chip which is flip-chip bump-bonded or otherwise electrically mated to CMOS readout electronics. (Stern et al., 2003) The hybrid approach offers greater flexibility than the planar CMOS approach because the detector array can be designed in a different substrate material system from the CMOS control electronics. For example, the APD detector array could be fabricated from silicon, silicon-germanium, indium phosphide, indium gallium arsenide or mercury cadmium telluride. Moreover, back-illumination inherently supports higher detector quantum efficiency and array sensitive-area-fill-factor compared to

front-illuminated planar arrays. The planar CMOS APD-FPA approach in Fig. 1 and the hybrid approach in Fig. 2 can both support integration of light focusing microlens arrays to increase the effective sensitive-area-fill-factor of the APD-FPAs, however, the planar CMOS approach is less amenable to microlens integration for the APDs since they would need to be epoxied to the CMOS chip and it is difficult to control epoxy thickness uniformity and refractive index matching. The hybrid APD-FPA approach however, supports microlenses to be monolithically integrated to the detectors without epoxy. The hybrid fabrication approach for silicon APD arrays has been implemented in the United States and is the preferred fabrication method resulting in higher performance arrays, albeit at increased cost. The hybrid approach shown in Fig. 2, has been used to fabricate focal plane arrays of silicon APD detectors using conventional silicon substrates that are back-thinned and either epoxied or oxide bonded to optically transparent quartz substrates followed by flip-chip bump-bonding to silicon CMOS readout ICs as shown in Figs. 3-4 respectively.

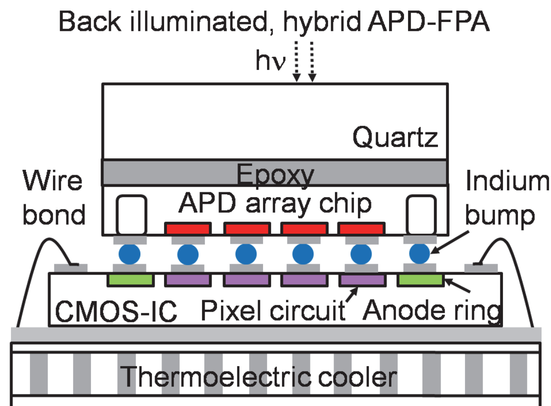


Fig. 3. Back-illuminated APD detector array silicon is thinned and epoxied to a quartz support wafer.

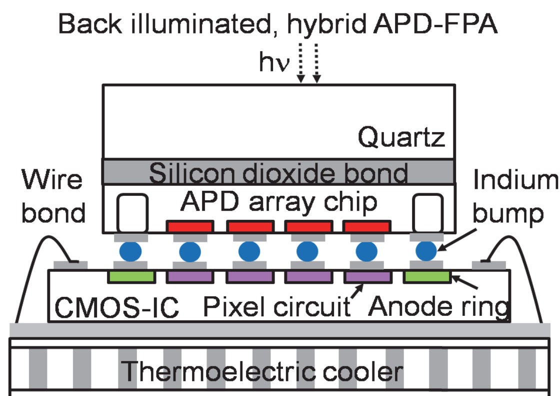


Fig. 4. Back-illuminated APD detector array silicon is thinned and oxide bonded to a quartz support wafer.

The approaches for manufacturing hybrid Si/SiGe APD-FPAs shown in Figs. 3-4 are still non-optimal because the quartz substrate does not provide optimal light transmittance into the device silicon and also because quartz is not resistant to the common hydrofluoric acid (HF) etchant, used in silicon device processing. This may constrain silicon detector devices to be processed in the bulk silicon wafer prior to silicon thinning and subsequent oxide bonding or epoxying to the quartz substrate. As a result, the ultra sensitive detector devices might become damaged during the epoxying or oxide bonding process.

2.1 Back-illuminated, silicon-on-sapphire substrates with improved antireflective layers for Si/SiGe APD-FPAs

The silicon-on-sapphire material system is particularly well adapted for fabricating back-illuminated, hybrid Si/SiGe APD-FPAs. Silicon-on-sapphire was discovered in 1963 by researchers working at the Boeing Corporation. Workers experimented with thermal decomposition of silane gas on a sapphire crystal polished into the shape of a sphere, thereby exposing all possible crystal planes, and discovered that (100) Si resulted from epitaxial growth on the R-plane surface of sapphire. (Manasevit & Simpson, 1964) The advantages of (100) silicon-on-(R-plane)-sapphire (SOS) substrates soon became apparent in fabricating high speed, radiation resistant SOS-CMOS circuits for space electronics including the microprocessor of the Voyager I spacecraft launched in 1977. The problem of high defect densities due to lattice mismatch in the silicon close to the sapphire interface where FETs are fabricated, caused device reliability problems and kept integrated circuit production yields low. The resulting increased cost of production prevented the technology from gaining a wide market share for consumer electronics. In 1979, Lau discovered a method to improve the epitaxial growth of (100) silicon on R-plane sapphire, resulting in lower defect densities in the silicon near the sapphire interface. (Lau et al., 1979) In 1991, Imthurn developed a method of directly bonding a silicon wafer to the sapphire R-plane followed by thinning the silicon using chemical mechanical polishing to proper device thickness. He subsequently fabricated silicon test diodes that exhibited reverse dark currents one order of magnitude lower than similar devices fabricated in heteroepitaxially grown SOS. (Imthurn et al., 1992)

Although silicon-on-sapphire was originally developed for integrated circuit applications, it also has many ideal attributes for use as a substrate material, supporting back-illuminated, solid-state, Si/SiGe detector arrays. Sapphire is an anisotropic, dielectric crystal of the negative uniaxial type that is weakly birefringent (n_o-n_e= 0.008) and possesses broadband optical transmittance ranging from the deep ultraviolet (λ_0 = 200 nm) to the midwave IR (λ_0 = 5500 nm). Sapphire is extremely resilient, supporting thinning below 100 µm which is an important requirement for high resolution, back-illuminated detector arrays. Sapphire can be optically polished to better than an 80-50 scratch and dig surface finish and can be etched using inductively coupled plasma (ICP) to fabricate light focusing microlenses beneath the silicon detectors. (Park et al., 2000) Sapphire is chemically resistant to most liquid etchants at room temperature and therefore functions as an ideal etchstop material during liquid crystallographic etching with tetramethyl ammonium hydroxide (TMAH) solution to define the silicon pixel mesa arrays. To enable high quantum efficiency, back-illuminated silicon detector arrays, the refractive index mismatch between air, sapphire and silicon has to be corrected however. The wide bandgap semiconductor material aluminum nitride (AlN), is closely lattice matched and refractive index matched to both sapphire and silicon and offers

the prospect of enabling fabrication of high transmittance (100) silicon-on-(AlN)-sapphire substrates for back-illuminated silicon imagers. In 2008, Stern proposed introducing an epitaxially grown lattice matched and refractive index matched, single crystal aluminum nitride antireflective layer between the silicon and sapphire. (Stern & Cole, 2008) The $\lambda/4$-AlN antireflective layer helps to improve the back-illuminated optical transmittance from sapphire into the device silicon. A $\lambda/4$-MgF_2 antireflective layer can be deposited on the back surface of the thinned sapphire substrate to improve optical transmittance from the ambient into the sapphire. Figure 5 illustrates the back-illuminated silicon-on-sapphire substrate with $\lambda/4$-AlN and $\lambda/4$-MgF_2 antireflective layers. Research shows that further improvement on the Si-(AlN)-sapphire substrate design from Fig. 5 can be achieved by incorporating an antireflective bilayer between sapphire and silicon, consisting of single crystal AlN and amorphous silicon nitride (a-SiN_X) as shown in Fig. 6. The design of the novel AlN/a-SiN_X antireflective bilayer is analyzed in detail in Sec. 2.2.

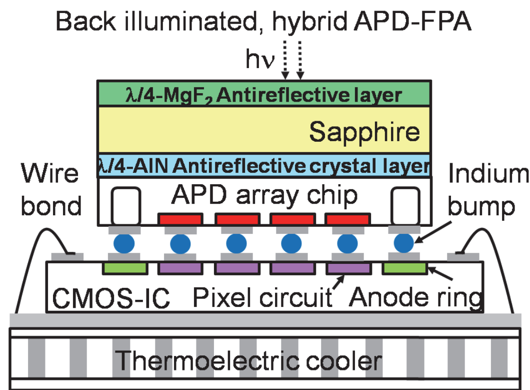


Fig. 5. Back-illuminated, hybrid, silicon-on-sapphire APD-FPA with $\lambda/4$-AlN and $\lambda/4$-MgF_2 antireflective layers.

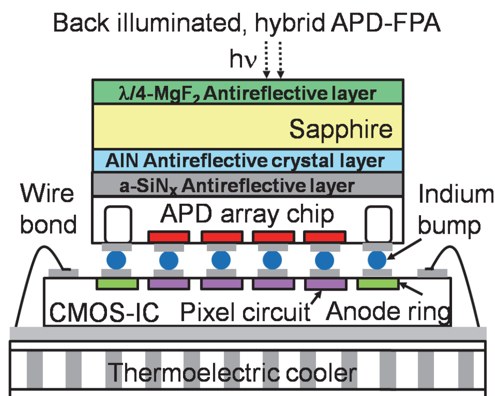


Fig. 6. Back-illuminated, hybrid, silicon-on-sapphire APD-FPA with AlN, SiN_X and $\lambda/4$-MgF_2 antireflective layers.

In contrast to silicon on quartz shown in Fig. 4, the Si-(AlN)-sapphire and Si-(AlN/a-SiN_X)-sapphire substrates shown in Figs. 5-6 can be prepared prior to detector device fabrication because the sapphire, AlN and a-SiN_X material layers are not affected by hydrofluoric acid (HF) or other etchants used in silicon device processing. Moreover, Si-(AlN/a-SiN_X)-sapphire substrates with $\lambda/4$-MgF_2 provide nearly optimal back-illuminated light transmittance into silicon as will be shown in Sec. 2.2, and in addition, microlenses can be directly fabricated in sapphire. (Park et al., 2000) Figure 7 shows a back-illuminated, crystallographically etched silicon mesa APD pixel with monolithically integrated sapphire microlens. Fig. 8 shows a crystallographically etched silicon mesa APD-FPA with monolithic, light focusing sapphire microlenses.

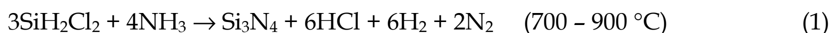
Fig. 7. Back-illuminated, silicon-on-sapphire mesa APD detector pixel with monolithic sapphire microlens.

Fig. 8. Back-illuminated, hybrid, silicon-on-sapphire APD-FPA with monolithic sapphire microlenses.

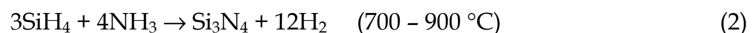
The sapphire substrate shown in Fig. 7 incorporates an antireflective bilayer between sapphire and silicon consisting of single crystal AlN and amorphous or a-SiN_X to improve optical transmittance into the device silicon. The space between mesa APD detector pixels is filled by a low resistance Al or Cu metal anode grid that provides low resistance anode contact at the base of each device mesa and also functions to block direct pixel-to-pixel optical crosstalk by line of sight light propagation. The monolithic sapphire microlens aligned beneath the mesa APD focuses light under the full height of the silicon mesa and away from the reduced height sidewalls. (Stern & Cole, 2008)

2.2 Advanced, very high transmittance silicon-on-sapphire substrate design for Si/SiGe APD-FPAs

A variation on the back-illuminated Si-(AlN)-sapphire substrate described in Sec. 2.1, provides improved optical transmittance into the device silicon by using an advanced antireflective bilayer design between sapphire and silicon consisting of single crystal AlN and non-stoichiometric, silicon rich, amorphous (a-SiN_X) with $x < 1.33$ as shown in Fig. 6.
Stoichiometric, fully dense, silicon nitride (Si_3N_4 or $SiN_{1.33}$) is an amorphous dielectric having a high optical bandgap, E_g = 5.3 eV and low optical absorption coefficient from UV to infrared. (Sze, 1981) Amorphous silicon nitride or a-SiN_X thin films have many applications in silicon processing and device fabrication including surface and bulk passivation of silicon, antireflective layers for silicon solar cells, barrier layers against Na and K ion diffusion and CMOS transistor device isolation using the LOCOS method. (Plummer et al., 2000) In addition, silicon rich a-$SiN_{X<1.33}$ that has a higher refractive index and lower tensile strain than stoichiometric a-$SiN_{1.33}$ has important applications for high speed optical interconnects in silicon nanophotonics and for silicon micromachining in MEMS and MOEMS applications. (Gardeniers at al., 1996)
Due to the ubiquity and importance of a-SiN_X thin films, much effort has been expended in developing optimized, application specific deposition methods for such films. Deposition of a-SiN_X is most readily achieved using low pressure (< 1 Atm.) gaseous precursors reacting either at low or high temperatures. High temperature, stoichiometric a-$SiN_{1.33}$ films are most commonly deposited on substrates in a low pressure chemical vapor deposition (LPCVD) reactor according to the chemical reaction in Eq. (1). (Rosler, 1977)



$$3SiH_2Cl_2 + 4NH_3 \rightarrow Si_3N_4 + 6HCl + 6H_2 + 2N_2 \quad (700 – 900\ °C) \quad (1)$$



$$3SiH_4 + 4NH_3 \rightarrow Si_3N_4 + 12H_2 \quad (700 – 900\ °C) \quad (2)$$

In Eq. (1), dichlorosilane (DCS) is shown as the silicon containing reactant species however, silane (SiH_4) can also be used as shown in Eq. (2). The advantage of DCS over silane is that the HCl byproduct can help remove metallic impurities from substrate surfaces by reacting to form volatile metal halides. Recently, much effort has been placed in developing low substrate temperature a-SiN_X deposition methods using plasma enhanced chemical vapor deposition (PECVD) and hot filament chemical vapor deposition (HFCVD), as such methods can be used to conserve valuable thermal budget during silicon device processing. In PECVD, a plasma reactor is used to enhance the chemical deposition while allowing substrate temperatures to remain in the low 200 – 450 °C temperature range. (Lowe et al., 1986) In HFCVD, an energized tungsten or tantalum filament heats the reactant gases while allowing low substrate temperatures to be used. (Verlaan et al., 2007)

For the present application however, conservation of thermal budget is not a concern because the Si-(AlN/a-SiN_X)-sapphire substrate can be fabricated before the mesa APD detector device. The a-SiN_X antireflective layer shown in Fig. 6, can be fabricated by direct deposition using LPCVD at elevated temperature, on a full thickness (100) silicon wafer according to the chemical reaction in Eq. (1). The sought after a-SiN_X antireflective layer characteristics listed in order from greatest to least in importance include, (1) refractive index, (2) optical bandgap E_g, (3) tensile strain in the layer and (4) surface and bulk passivation properties for silicon. Each of the four characteristics of the a-SiN_X antireflective layer will be analyzed and/or discussed in order of importance for the present application.
The primary role of the a-SiN_X is to function as an antireflective layer in conjunction with AlN as shown in Fig. 6, therefore, it is critical to design the layer to have a refractive index meeting the condition, $n_{a\text{-}SiN} = (n_{AlN} \cdot n_{Si})^{0.5}$ over most of the wavelength range of interest, to yield maximum optical transmittance from sapphire into the device silicon. The Sellmeier dispersion relation for stoichiometric a-$SiN_{1.33}$ is given in Eq. (3), with constants for the equation listed in Table 1.

$$n^2(\lambda) = 1 + \frac{A_1 \lambda^2}{\lambda^2 - \lambda_1^2} \quad \rightarrow (\lambda \text{ in nm}) \tag{3}$$

Parameter	Value
A_1	2.8939
λ_1	139.67 x 10^{-3}

Table 1. Sellmeier dispersion relation constants for stoichiometric a-$SiN_{1.33}$ from Eq. (3).

The real refractive index of a-$SiN_{1.33}$ is plotted in Fig. 9 according to Eq. (3), using the parameters in Table 1. The real refractive indices for MgF_2, sapphire, AlN and Si are also plotted for reference.

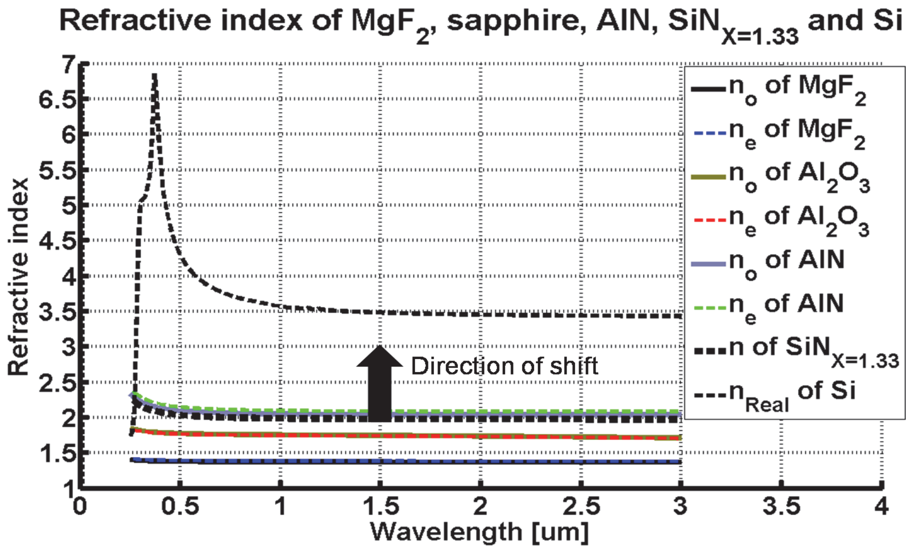


Fig. 9. Real refractive indices for MgF_2, sapphire, AlN, stoichiometric a-$SiN_{1.33}$ and Si are shown as a function of optical wavelength.

The Sellmeier relation given by Eq. (3) and plotted in Fig. 9, shows that stoichiometric a-$SiN_{1.33}$ has a refractive index that is too low to provide refractive index matching between sapphire and silicon in conjunction with AlN as shown in Fig. 6, and the arrow in Fig. 9, represents the direction of vertical shift of the refractive index as a function of wavelength curve that would be required to provide refractive index matching. The Si content in a-$SiN_{1.33}$ therefore should be raised ($x<1.33$), to increase the refractive index of the layer, by increasing the DCS:NH_3 gas flow ratio in Eq. (1). (Gardeniers et al., 1996)

The statistical experiments performed by Gardeniers, studied the properties of silicon rich a-$SiN_{X<1.33}$ films deposited according to Eq. (1), by varying the primary process parameters including (1) temperature, (2) total pressure, (3) total gas flow, (4) DCS:NH_3 gas flow ratio. Although their goal was to optimize the a-SiN_X thin films for micromechanical or MEMS applications requiring low tensile strain, their results also confirmed an important theoretical model described by Makino, predicting the a-SiN_X thin film refractive index as a function of the nitrogen to silicon ratio (x = N:Si) in the film. The model assumes that the refractive index of a-$SiN_{X\leq 1.33}$ films is a "bond-density-weighted linear combination" of a-Si and a-$SiN_{1.33}$ reference refractive indices and is given by Eq. (4). (Makino, 1983)

$$n = \frac{\left[(4/x)-3\right]n_0 + 6n_{1.33}}{(4/x)+3} \tag{4}$$

In Eq. (4), the quantity n_0 represents the refractive index of a-Si and $n_{1.33}$ represents the refractive index of stoichiometric a-$SiN_{1.33}$. Although the refractive index model in Eq. (4) does not consider the presence of residual hydrogen in the a-SiN_X thin film in the form of Si-H and N-H bonds, this only becomes a problem when applying the model to low temperature deposited films that usually contain larger amounts of residual hydrogen than films deposited at high temperature. The a-$SiN_{X\leq 1.33}$ thin films deposited according to Eq. (1) contain negligible amounts of hydrogen due to high temperature deposition and therefore, the model given by Eq. (4) is valid for the proposed fabrication approach. The experiments by Gardeniers, verified Eq. (4) using a value for the refractive index of a-Si equal to that of crystalline silicon or $n_0 = 3.9$ (at $\lambda_0 = 633$ nm), by measuring the values of N:Si = x, in the films they grew, measuring the refractive indices of those films and correlating the measured refractive indices to the calculated ones from Eq. (4) using the measured values of x as input to Eq. (4). Using the results from Gardeniers together with Eqs. (3-4), it becomes possible to calculate the N:Si ratio value x, in the silicon rich a-$SiN_{X<1.33}$, that yields the nearly optimal refractive index as a function of wavelength curve shown in Fig. 10, for achieving refractive index matching with AlN between the sapphire substrate and device silicon.

The N:Si ratio value in the nearly optimal thin film that yields the curve in Fig. 10 is given as N:Si = 0.62, corresponding to a-$SiN_{0.62}$. The nearly optimal refractive index as a function of wavelength needed for refractive index matching between sapphire and silicon will be provided by a-$SiN_{0.62}$, however, it is also necessary to consider the extinction coefficient of the antireflective layer now having a reduced optical bandgap compared to stoichiometric a-

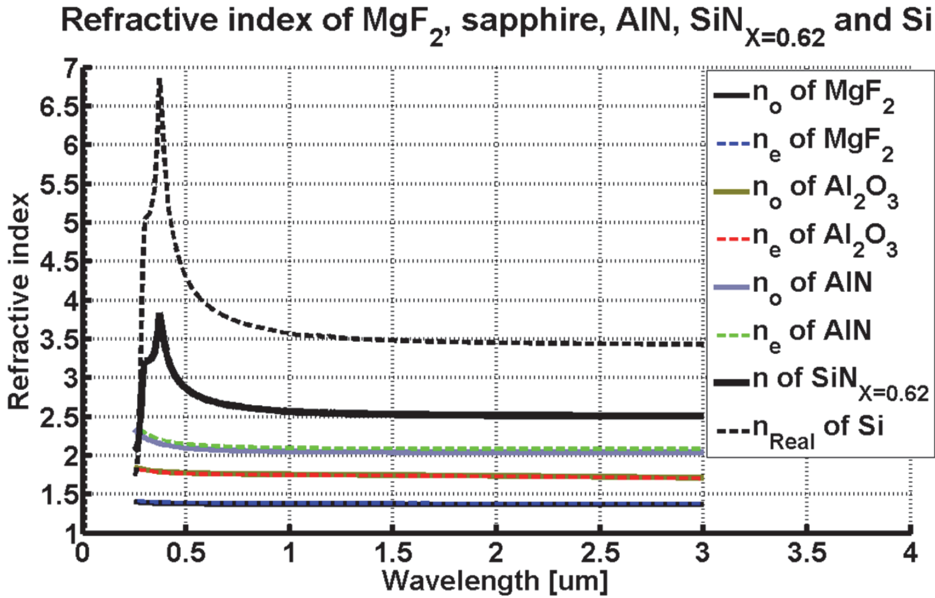


Fig. 10. Real refractive indices for MgF_2, sapphire, AlN, a-$SiN_{0.62}$ and silicon are shown as a function of the optical wavelength.

$SiN_{1.33}$, before calculating the back-illuminated optical transmittance of the novel Si-(AlN/a-$SiN_{0.62}$)-sapphire substrate.

Data for the extinction coefficient as a function of wavelength was not collected in the a-SiN_X samples deposited by high temperature LPCVD according to Eq. (1) by Gardeniers, however, it is still possible to infer the absorbance of the a-$SiN_{0.62}$ antireflective layer from Fig. 10, using data collected by Verlaan, who used HFCVD to deposit a-$SiN_{0.62}$ with identical stoichiometry to the nearly optimal antireflective layer in Fig. 10, and measured the extinction coefficient of the sample over the visible wavelength range from 400–650 nm. (Verlaan et al., 2007) Although HFCVD used by Verlaan maintains the substrate at a lower temperature of 230 °C during deposition compared to high temperature LPCVD used by Gardeniers, the resulting a-SiN_X from HFCVD has a density approaching the density of material deposited by high temperature LPCVD while retaining more hydrogen. Despite these differences between LPCVD and HFCVD deposited thin films, the a-$SiN_{0.62}$ sample data from Verlaan may be used to infer the expected absorbance as a function of wavelength of the LPCVD deposited a-$SiN_{0.62}$ antireflective layer from Fig. 10. To calculate the expected absorbance as a function of wavelength for the a-$SiN_{0.62}$ antireflective layer in the Tauc absorption region from 250 nm to λ_{Eg}, from Verlaan's data, the optical bandgap $E_{g\text{-}opt}$, of the HFCVD deposited a-$SiN_{0.62}$ must first be calculated using the Tauc equation given in Eq. (5). (Tauc, 1974)

$$\sqrt{\hbar\omega\alpha(\omega)} = \sqrt{B}\left(\hbar\omega - E_{g-opt}\right) \tag{5}$$

In Eq. (5), $\hbar$ is the reduced Planck constant, $\omega = 2\pi\nu$ is the angular frequency, $E_{g\text{-}opt}$ is the optical bandgap in eV and B is a slope parameter with units of [$cm^{-1}eV^{-1}$]. Table 2 lists the measured extinction coefficient as a function of the optical wavelength from 400-650 nm for HFCVD deposited a-$SiN_{0.62}$ and substrate temperature of 230 °C.

Wavelength (nm)	Extinction coefficient	Wavelength (nm)	Extinction coefficient
400	0.12	550	0.012
425	0.09	575	0.007
450	0.064	600	0.005
475	0.043	625	0.003
500	0.030	650	0.002
525	0.018		

Table 2. Extinction coefficient of HFCVD deposited a-$SiN_{0.62}$ (Verlaan et al., 2007)

Using the measured data for a-$SiN_{0.62}$ from Table 2 and knowing that the extinction coefficient of a material is related to its absorption coefficient as $0.5(\alpha/k_0)$, it is possible to plot the left side of Eq. (5) as a function of energy as shown in Fig. 11. Fitting a straight line to the linear part of the scatter plot of points in Fig. 11 and extrapolating to the x-axis yields the optical bandgap $E_{g\text{-}opt}$ = 2.1 eV for the HFCVD deposited a-$SiN_{0.62}$ from Verlaan. Using the optical bandgap value $E_{g\text{-}opt}$ = 2.1 eV for a-$SiN_{0.62}$ and corresponding wavelength λ_{Eg}= 590 nm, the absorption coefficient as a function of wavelength for the Tauc absorption region is calculated using Eq. (5) and plotted from 250 nm to λ_{Eg}= 590 nm in Fig. 12 by fitting to the measured data from Table 2.

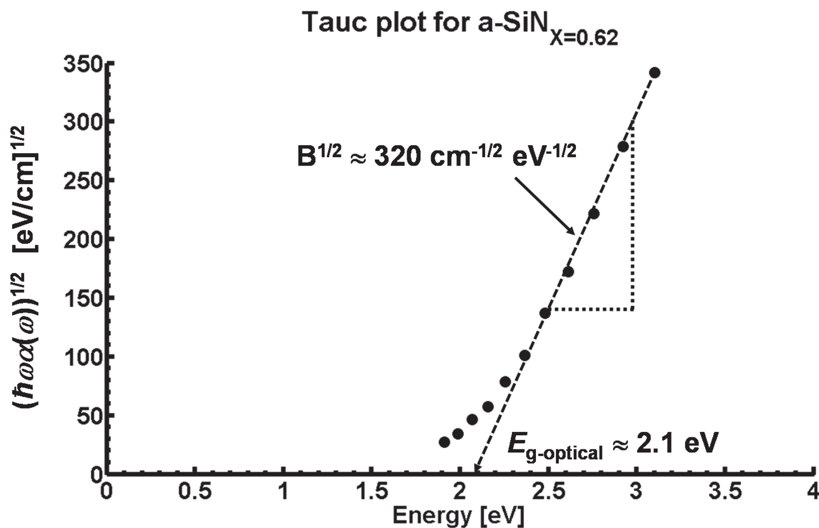


Fig. 11. Determination of the optical bandgap $E_{g\text{-}opt}$ using Tauc plot for a-$SiN_{0.62}$ deposited by HFCVD.

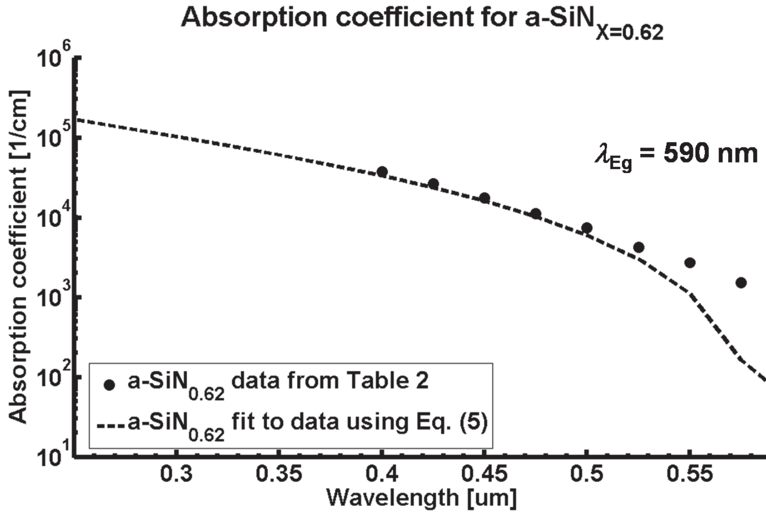


Fig. 12. Tauc absorbance of a-$SiN_{0.62}$ calculated using the optical bandgap $E_{g\text{-}opt}$ = 2.1 eV from Fig. 11.

It will be assumed that the high temperature LPCVD deposited a-$SiN_{0.62}$ antireflective layer from Fig. 10 is characterized by a similar absorption coefficient over the Tauc absorption region from 250 nm to λ_{Eg}= 590 nm as calculated in Fig. 12 for low temperature HFCVD deposited material, where $\alpha \approx 1\times10^5$ cm^{-1} at 250 nm. In practice, the absorption coefficient in the Tauc absorption region for high temperature LPCVD deposited a-$SiN_{0.62}$ should be lower than the calculation in Fig. 12, which represents a worst case scenario. Assuming the worst case of high absorption as calculated in Fig. 12, entails that for an a-$SiN_{0.62}$ antireflective layer thickness below 50 nm, the absorbance may still be considered negligible for wavelengths between 250 nm to λ_{Eg}= 590 nm and the layer can therefore be modeled as a lossless dielectric. Equation (6) expresses the impedance of a material as a function of the real refractive index $n(\lambda)$, and the absorption coefficient $\alpha(\lambda)$.

$$\eta(\lambda) = \sqrt{\frac{\mu_0}{\varepsilon_0\left(n(\lambda) - j\frac{1}{2}\frac{\alpha(\lambda)}{k_0}\right)^2}} \tag{6}$$

To calculate the optical power transmittance of TE and TM waves into silicon for the back-illuminated (MgF_2)-sapphire-(AlN/a-$SiN_{0.62}$)-Si substrate, the full wave transfer matrix M_{STACK} for the material layers in the substrate has to be obtained. This result needs to be put into a scattering matrix form that yields the reflection coefficients for the incident waves which in turn allow the reflected and transmitted optical power to be calculated. The wave transfer-scattering matrix theory is described in the text by Saleh & Teich. (Saleh & Teich, 2007) The matrix M_{STACK} for air-(MgF_2)-sapphire-(AlN/a-$SiN_{0.62}$)-Si results from multiplying together nine wave transfer matrices including four for propagation through MgF_2, sapphire, AlN, SiN_X and five matrices for the material interfaces as shown in Eq. (7).

$$M_{\eta-STACK} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} = M_9 M_8 M_7 M_6 M_5 M_4 M_3 M_2 M_1 \quad (7)$$

The matrices M_1, M_3, M_5, M_7 and M_9 represent wave transfer matrices at the air-MgF_2, MgF_2-sapphire, sapphire-AlN, AlN-a-$SiN_{0.62}$ and a-$SiN_{0.62}$-silicon interfaces while matrices M_2, M_4, M_6 and M_8 are propagation matrices through MgF_2, sapphire, AlN and a-$SiN_{0.62}$. All nine matrices are expressed in terms of the complex impedances of the materials given by Eq. (6). Using a Monte Carlo integration approach, it is possible to calculate the back-illuminated optical transmittance into the APD device silicon as a function of wavelength, for TE waves normally incident to the sapphire substrate plane of the mesa APD pixel from Fig. 7, as shown in Fig. 13.

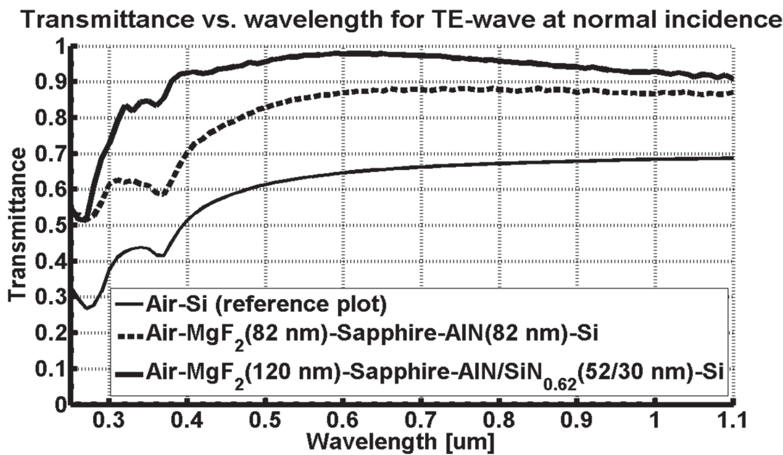


Fig. 13. Optical power transmittance into silicon of a TE wave normally incident from air to the back-illuminated APD substrate.

From the calculation in Fig. 13, it is evident that using silicon rich a-$SiN_{0.62}$ prepared by high temperature LPCVD as an antireflective layer together with AlN, provides the required refractive index matching to enable very high transmittance, back-illuminated silicon-on-sapphire wafer substrates, while retaining a sufficiently high optical bandgap to be treated as a lossless dielectric in this application, as calculated in Figs. 11-12. The silicon-on-sapphire substrate represented by the thick solid curve in Fig. 13, having an AlN/a-$SiN_{0.62}$ antireflective bilayer of 52/30 nm thickness respectively between sapphire and silicon and 120 nm thick MgF_2 antireflective layer between air and sapphire, provides significantly improved back-illuminated transmittance into silicon as compared to the silicon-on-sapphire substrate represented by the dashed curved in Fig. 13, having only an 82 nm thick $\lambda/4$-AlN antireflective layer between sapphire and silicon and an 82 nm thick $\lambda/4$-MgF_2 antireflective layer between air and sapphire. Other advantages of using silicon rich a-$SiN_{X<1.33}$ films include lower tensile strain than stoichiometric films which is important for reducing the injection of vacancy and interstitial defects in silicon. High temperature LPCVD a-SiN_X films unfortunately contain only a trace amount of hydrogen compared to PECVD and HFCVD films, and hydrogen is very useful for bulk and surface

passivation of silicon defects and the lengthening of carrier lifetimes in silicon. It is particularly challenging to optimize a-SiN_X films for high transmittance antireflective layers, that are better prepared using high temperature LPCVD and also for bulk and surface silicon passivation with hydrogen, which require low temperature deposition using PECVD or HFCVD. Such novel substrates however, represent an enabling technology for the next generation of high performance Si/SiGe APD focal plane array imagers.

3. Optical crosstalk in silicon-on-sapphire APD-FPAs

The phenomenon of optical crosstalk in high resolution avalanche photodiode arrays is well known and has been the subject of extensive theoretical and experimental study. (Akil et al., 1998, 1999; Lahbabi et al., 2000; Rech et al., 2008) Optical crosstalk in APD arrays results primarily from photon emission that occurs during impact ionization in the avalanche carrier multiplication process, erroneously triggering neighboring APD pixels and thereby producing noise. The increased APD-FPA noise from optical crosstalk might prevent high sensitivity imaging. Past scientific literature has mainly addressed the phenomenon of direct pixel-to-pixel optical crosstalk which occurs between immediately adjacent detector pixels by line of sight light propagation, however, indirect optical crosstalk might also be present. Indirect optical crosstalk is more difficult to block and occurs from light undergoing multiple reflections in the planar sapphire substrate waveguide. Two forms of indirect optical crosstalk exist in the silicon-on-sapphire APD-FPA due to multiple reflections and they include reflected light generated in the multiplication region of the APD during a detection event as well as reflected incident ambient illumination as shown in Figs. 14-15 respectively.

The back-illuminated silicon-on-sapphire material system readily enables the design of Si/SiGe APD focal plane arrays with zero direct pixel-to-pixel optical crosstalk by line of sight propagation as understood from Fig. 14, where the Al or Cu metal anode grid across the mesa APD detector array performs the important secondary function of blocking direct

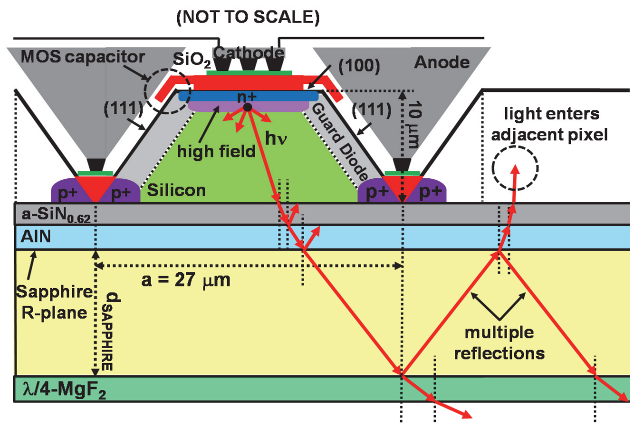


Fig. 14. Indirect optical crosstalk from APD emitted light coupled into the sapphire substrate waveguide.

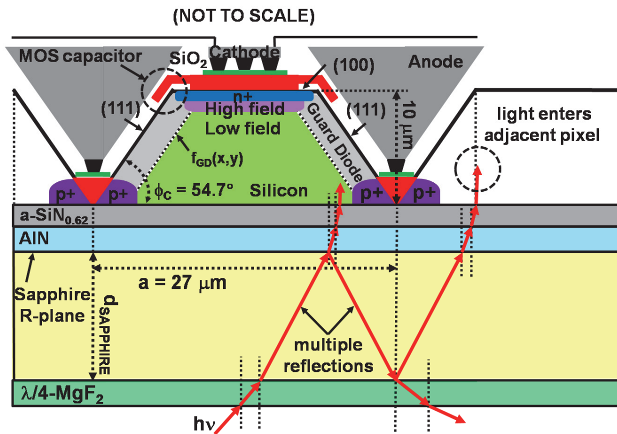


Fig. 15. Indirect optical crosstalk from ambient incident illumination coupled into the sapphire substrate waveguide.

line of sight light propagation between adjacent detectors. In Fig. 14, the mesa APD is shown by approximation to emit light isotropically from a central point in the high electric field avalanche multiplication region. A fraction of the emitted photons are coupled into the planar sapphire substrate waveguide where they might undergo multiple reflections and subsequent transmission into the neighboring silicon mesa APD pixels. In Sec. 3.1-3.2, the fraction of APD emitted light during linear or Geiger-mode detection events transmitted into neighboring 27 µm mesa APD pixels is calculated for planar Si-(AlN)-sapphire and Si-(AlN/a-$SiN_{0.62}$)-sapphire substrates with back-side λ/4-MgF_2 antireflective layer. In Sec. 4, the fraction of incident light on a 27 µm mesa APD from an isotropic point source at infinity, transmitted into neighboring detectors as shown in Fig. 15, is calculated for planar Si-(AlN)-sapphire and Si-(AlN/$SiN_{0.62}$)-sapphire substrates with back-side λ/4-MgF_2 antireflective layer.

Indirect optical crosstalk through the planar sapphire substrate can be reduced by thinning the sapphire substrate and fabricating microlenses beneath each APD detector pixel as shown in Fig. 7. Although microlenses increase APD-FPA performance as described in Sec. 2.1, they also increase the cost and fabrication complexity of the detector arrays. (Stern & Cole, 2008) It will be shown in detailed analysis and calculations in Sec. 5, that indirect optical crosstalk through the planar sapphire substrate waveguide from light generated by impact ionization or from ambient illumination will not increase detector noise to levels that prevent high sensitivity imaging and therefore, fabricating sapphire microlenses is not an imperative for achieving low noise, high resolution Si/SiGe APD-FPAs using back-illuminated silicon-on-sapphire.

3.1 Indirect optical crosstalk from APD emitted light, coupled into the sapphire waveguide; Si-(AlN)-sapphire

The fraction of light emitted by an APD during detection events, coupled into the sapphire substrate waveguide and transmitted to neighboring detectors thereby contributing to indirect optical crosstalk, can be calculated by modeling and simulation of the mesa APD detector pixel. Using the Monte Carlo method, it is possible to calculate how photons are emitted from

the APD multiplication region and transmitted into adjacent mesa APD detector pixels as shown in Fig. 16, for Si-(AlN)-sapphire substrates with back-side $\lambda/4$-MgF_2 antireflective layer. The thickness for both the AlN and MgF_2 antireflective layers is taken to be 82 nm as indicated in Fig. 16. The high electric field mesa APD multiplication region shown in Fig. 16 is 300-500 nm thick and consists of more highly doped p-type silicon formed by boron impurity diffusion. (Stern & Cole, 2008) Although impact ionization resulting in avalanche multiplication of charge carriers as well as photon emission, occurs throughout the volume of the high electric field region, for modeling purposes, it will be assumed that photons are only emitted isotropically from a single point in the center of the multiplication region located at a height h_M = 9 µm, above the silicon pixel base plane as shown in Fig. 16.
Using the Monte Carlo simulation approach, points are randomly generated in the silicon base plane area of the 27 µm mesa APD. Trajectories of light propagation or optical k-vectors are created by connecting straight lines from the isotropic point source in the multiplication region to the randomly generated points in the silicon base plane area of the 27 µm mesa APD as shown in Fig. 17.

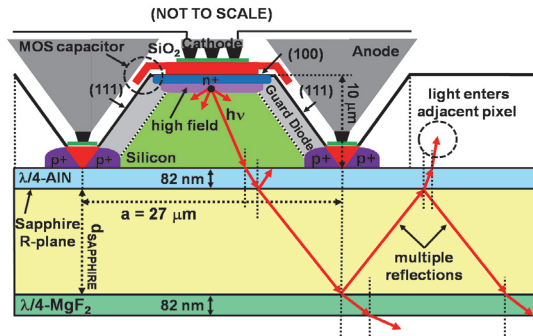


Fig. 16. Indirect optical crosstalk from APD emitted light coupled into the (AlN)-sapphire substrate waveguide.

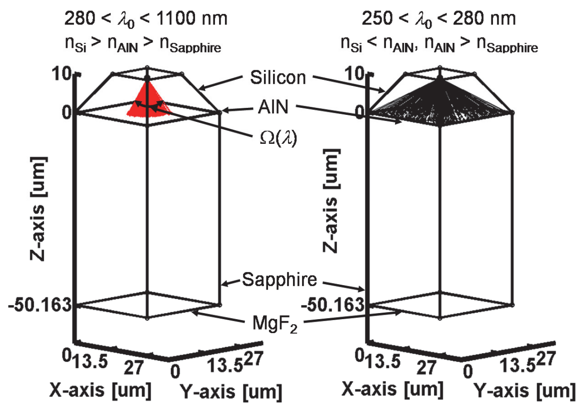


Fig. 17. 3-D ray tracing shows the optical k-vector transmission cone for 280<λ_0<1100 nm light into the sapphire substrate.

In Fig. 17, 3-D ray tracing is used to calculate paths of light propagation for the randomly generated optical k-vectors over a 250 < λ_0 < 1100 nm wavelength range, emitted from the APD multiplication region, transmitted into the sapphire substrate and undergoing multiple reflections. The calculation in Fig. 17 shows that light between the wavelengths of 280-1100 nm (left pixel in Fig. 17) emitted from the isotropic point source in the multiplication region, can only exit the mesa APD through the sapphire substrate waveguide if the optical k-vector from the point source is emitted into a cone characterized by a wavelength dependent solid angle $\Omega(\lambda)$, subtended by a corresponding circular base area and having height h_M = 9 μm of the isotropic point source. The calculation also shows that for 250 < λ_0 < 280 nm wavelengths (right pixel in Fig. 17) light can be transmitted into the sapphire substrate through the major part of the 27 μm mesa APD base area because the refractive index of Si is smaller than AlN as shown in Fig. 10, and the incidence angle at the AlN-sapphire interface does not exceed the critical angle θ_C except when optical k-vectors with 270 < λ_0 < 280 nm, are incident near the corners of the mesa APD base area. For 280 < λ_0 < 1100 nm wavelengths, the refractive index of silicon is larger than AlN as well as sapphire and total internal reflection (TIR) occurs if the incidence angle of the optical k-vector at the silicon-AlN or AlN-sapphire interface exceeds the critical angle θ_C. The critical angles θ_C for light emission out of the silicon mesa into the AlN and from AlN into the sapphire substrate, depend on the refractive indices of the silicon, AlN and sapphire material layers, that in turn depend on the optical wavelength as shown in Eq. (8).

$$\theta_C = a\sin\left(\frac{n_2(\lambda)}{n_1(\lambda)}\right); \quad n_1 > n_2 \tag{8}$$

In Eq. (8), the light propagates from a material with refractive index n_1 to a material with refractive index n_2 where $n_1 > n_2$. Using Eq. (8), the critical angle $\theta_C(\lambda)$ is calculated in Fig. 18 as a function of wavelength at the Si-AlN (dashed line) and AlN-sapphire (thin solid line) material interfaces. The effective critical angle $\theta_{C\text{-eff}}(\lambda)$, (thick solid line) shown in Fig. 18, for light transmission from Si into sapphire through AlN, is equal to the critical angle for light transmission directly from Si into sapphire. The corresponding effective, fractional solid angle Ω_F for light transmission from Si into sapphire through AlN, as a function of wavelength is calculated as the effective light transmission cone solid angle $\Omega_{eff} = \pi(\theta_{C\text{-eff}}(\lambda))^2$ divided by 4π sr solid angle of the sphere into which the isotropic point source emits, as shown in Fig. 19.

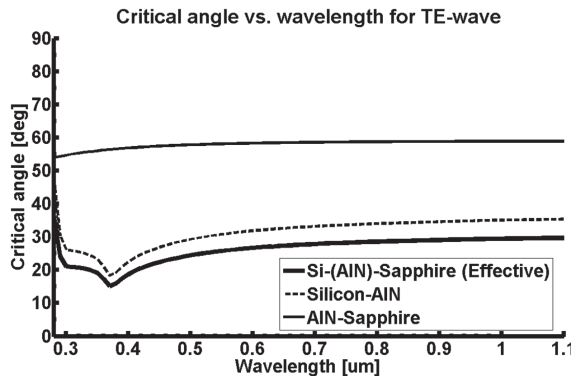


Fig. 18. Critical angle for Si-AlN, AlN-Sapphire, Si-Sapphire.

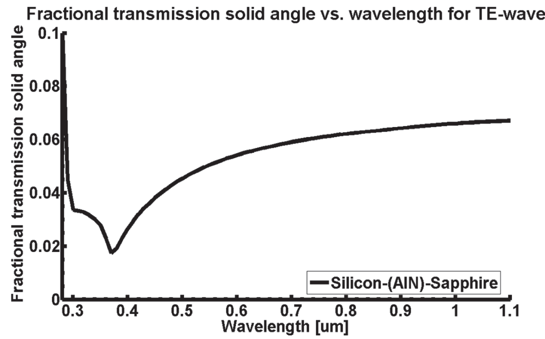


Fig. 19. Fractional solid angle of the light transmission cone.

Reducing the fractional solid angle of the light transmission cone calculated in Fig. 19, would help to prevent optical k-vectors with large incidence angle at the Si-AlN and AlN-sapphire interfaces from propagating into the sapphire substrate where they can undergo multiple reflections and transmission into distant APD detectors in the array to produce optical crosstalk at a distance. Reducing the effective fractional solid angle of the light transmission cone requires a large refractive index contrast ratio between the Si semiconductor device layer and the optically transparent supporting substrate and does not depend on the thin antireflective layers such as AlN between the Si and sapphire where $n_{Si} > n_{AlN} > n_{SAPPHIRE}$.

It will be assumed that any optical k-vectors reflected back into the silicon APD by TIR will not have a second pass, or opportunity to escape the mesa pixel by transmission into the sapphire substrate waveguide and even if such TIR optical k-vectors might be transmitted through the (111) sidewalls of the mesa, the light will subsequently be blocked by the anode metal grid and will not contribute to optical crosstalk. Thus, only the optical k-vectors emanating from the isotropic point source in the APD multiplication region and contained by the light transmission cone calculated in Fig. 19 for 280 < λ_0 < 1100 nm wavelengths or contained by the solid angle subtended by most of the silicon mesa base area for 250 < λ_0 < 280 nm wavelengths, will couple into the sapphire substrate and therefore contribute to the indirect optical crosstalk. Using the result from Fig. 19, it is possible to calculate the fraction of light emitted by the isotropic point source in the mesa APD multiplication region that is transmitted through the sapphire substrate to other APD detectors in the array as a function of wavelength. Multiple reflections may occur in the sapphire substrate for the APD emitted light, and such reflections might not necessarily be bounded by the areas of the eight numbered and immediately adjacent 27 μm mesa APD detector pixels shown in Fig. 20.

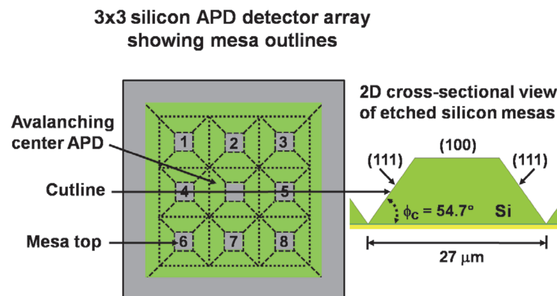


Fig. 20. 3x3 array showing eight immediately adjacent APDs.

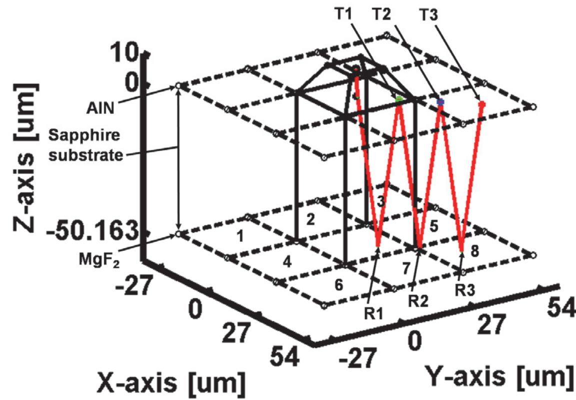


Fig. 21. 3-D ray tracing shows simulated multiple reflections.

The optical transmittance into adjacent detectors numbered 1-8 as well as other detectors outside of the immediately adjacent numbered pixels shown in Fig. 20, is obtained by calculating the fraction of light transmitted into silicon after each successive reflection cycle in the sapphire substrate for an optical k-vector as shown in Fig. 21. The first reflection cycle in the sapphire substrate is indexed as T_1 followed by the second and third cycles with index T_2, T_3 …. T_N where T_N is the highest calculated reflection in the substrate. The results from Fig. 19 and Fig. 21, are used to calculate the fraction of light emitted by the isotropic point source in the mesa APD multiplication region, that will be transmitted through the sapphire substrate to other APD detectors in the array as a function of wavelength as shown in Figs. 22-23.

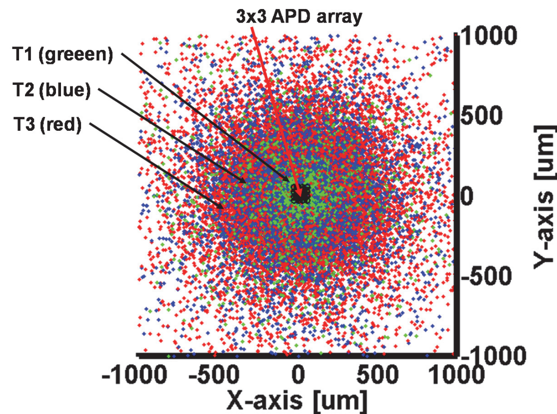


Fig. 22. Average crosstalk distance for 50 µm thick sapphire.

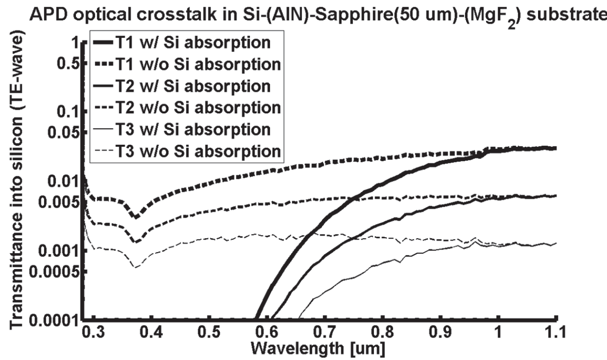


Fig. 23. Indirect APD optical crosstalk in 50 µm thick sapphire.

The average distance of light transmittance points *T*1, *T*2 and *T*3 into the neighboring APD pixels, from the avalanching center mesa APD (shown in Fig. 20) is calculated in Fig. 22 for a 50 µm thick sapphire substrate. In Fig. 23, the fraction of light emitted by the isotropic point source in the mesa APD multiplication region and transmitted to neighboring APD pixels is calculated for a maximum of three reflection cycles, *T*1, *T*2 and *T*3, with and without light absorption in the silicon. (Lahbabi et al., 2000) On the first reflection cycle represented by *T*1 (shown in Figs. 21-23), between 1-5% of the isotropically emitted light from the APD multiplication region having wavelength 280-1100 nm, is transmitted into neighboring pixels while the second reflection cycle *T*2, transmits 0.1-0.5% and the third reflection cycle *T*3, transmits 0.05-0.1% of emitted light into the neighboring pixels. The results in Fig. 24 show that the average distance of *T*1 for a 10 µm thick sapphire substrate corresponds to a radius of a circle contained by the eight adjacent pixels of the avalanching center APD shown in Fig. 20.

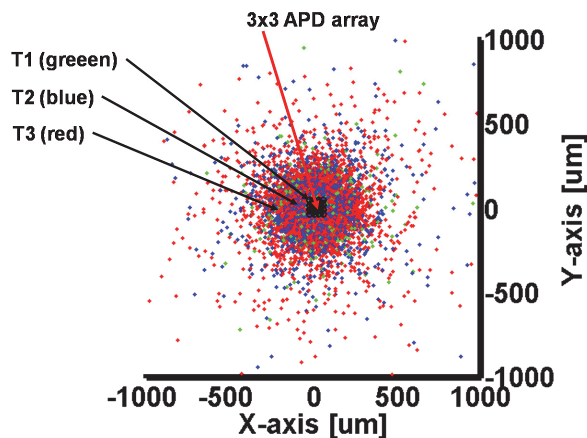


Fig. 24. Average crosstalk distance for 10 µm thick sapphire.

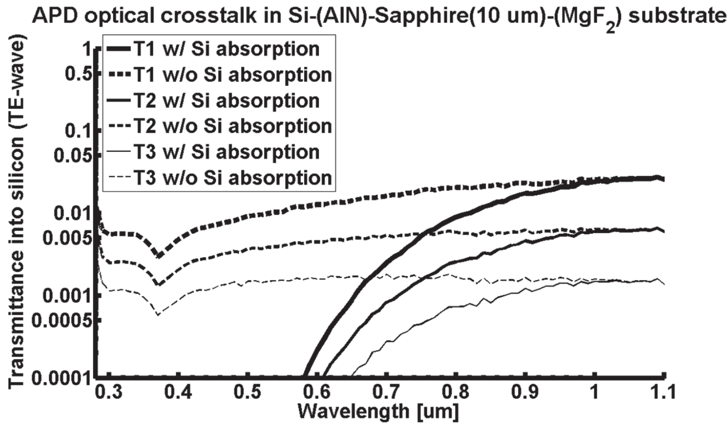


Fig. 25. Indirect APD optical crosstalk in 10 µm thick sapphire.

The results in Figs. 22-25 will be analyzed in Sec. 5 for their effect on the signal-to-noise ratio of APD detectors in an array.

3.2 Indirect optical crosstalk from APD emitted light, coupled into the sapphire waveguide; Si-(AlN/$SiN_{0.62}$)-sapphire

The fraction of light emitted by a 27 µm mesa APD during detection events, coupled into the sapphire substrate waveguide and transmitted to neighboring detectors thereby contributing to indirect optical crosstalk, can be calculated for very high transmittance Si-(AlN/a-$SiN_{0.62}$)-sapphire substrates with back-side $\lambda/4$-MgF_2 antireflective layer described in Sec. 2.2, by following an approach similar to that in Sec. 3.1 for Si-(AlN)-sapphire substrates. Using the Monte Carlo method, it is possible to calculate how photons are emitted from the APD multiplication region and transmitted into adjacent mesa APD detector pixels as shown in Fig. 26, for silicon-(AlN/a-$SiN_{0.62}$)-sapphire substrates with back-side $\lambda/4$-MgF_2 antireflective layer. The thickness of the AlN, a-$SiN_{0.62}$ and MgF_2 layers is d_{AlN} = 52 nm, $d_{\mathrm{a\text{-}SiN_0.62}}$ = 30 nm and d_{MgF2} = 120 nm as indicated in Fig. 26. The high electric field mesa APD multiplication region shown in Fig. 26 is 300-500 nm thick and consists of more highly doped p-type silicon formed by boron impurity diffusion. (Stern & Cole, 2008) Although impact ionization resulting in avalanche multiplication of charge carriers as well as photon emission, occurs throughout the volume of the high electric field region, for modeling purposes it will be assumed that photons are only emitted isotropically from a single point in the center of the multiplication region located at a height h_M = 9 µm, above the silicon pixel base plane as shown in Fig. 26.

Using the Monte Carlo simulation approach, points are randomly generated in the silicon base plane area of the 27 µm mesa APD. Trajectories of light propagation or optical k-vectors are created by connecting straight lines from the isotropic point source in the multiplication region to the randomly generated points in the silicon base plane area of the 27 µm mesa APD as shown in Fig. 27.

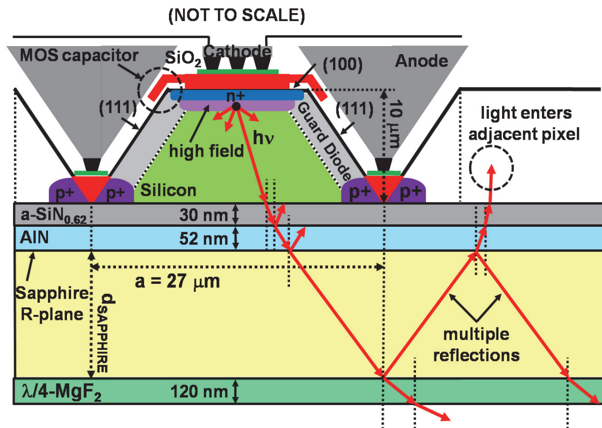


Fig. 26. Indirect optical crosstalk from APD emitted light coupled into the (AlN/a-SiN$_{0.62}$)-sapphire substrate waveguide.

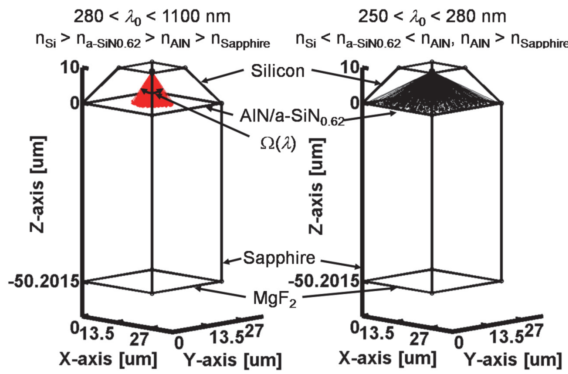


Fig. 27. 3-D ray tracing shows the optical k-vector transmission cone for 280<λ_0<1100 nm light into the sapphire substrate.

As in Sec. 3.1, 3-D ray tracing is used in Fig. 27 to calculate paths of light propagation for the randomly generated optical k-vectors over a 250 < λ_0 < 1100 nm wavelength range, emitted from the APD multiplication region, transmitted into the sapphire substrate and undergoing multiple reflections. The calculation in Fig. 27 shows that light between the wavelengths of 280-1100 nm (left pixel in Fig. 27) emitted from the isotropic point source in the multiplication region, can only exit the mesa APD through the sapphire substrate waveguide if the optical k-vector from the point source is emitted into a cone characterized by a wavelength dependent solid angle $\Omega(\lambda)$, subtended by a corresponding circular base area and having height h_M = 9 µm of the isotropic point source. The calculation also shows that for 250 < λ_0 < 280 nm wavelengths (right pixel in Fig. 27) light can be transmitted into the sapphire substrate through the major part of the 27 µm mesa APD base area because the refractive index of Si is smaller than a-SiN$_{0.62}$ and AlN as shown in Fig. 10, and the incidence angle at the AlN-sapphire interface does not exceed the critical angle θ_C except when optical

k-vectors with 270 < λ_0 < 280 nm, are incident near the corners of the mesa APD base area. For 280 < λ_0 < 1100 nm wavelengths, the refractive index of silicon is larger than a-$SiN_{0.62}$, AlN as well as sapphire and total internal reflection (TIR) occurs if the incidence angle of the optical k-vector at the silicon-a-$SiN_{0.62}$, silicon-AlN or AlN-sapphire interface exceeds the critical angle θ_C. The critical angles θ_C for light emission out of the silicon mesa into the a-$SiN_{0.62}$, from a-$SiN_{0.62}$ into AlN and from AlN into the sapphire substrate, depend on the refractive indices of the silicon, a-$SiN_{0.62}$, AlN and sapphire material layers, that in turn depend on the optical wavelength as given by Eq. (8). In Eq. (8) the light propagates from a material with refractive index n_1 to a material with refractive index n_2 with $n_1 > n_2$.
Using Eq. (8), the critical angle $\theta_C(\lambda)$ is calculated in Fig. 28 as a function of wavelength at the different material interfaces. The effective critical angle $\theta_{C\text{-eff}}(\lambda)$, (thick solid line) shown in Fig. 28, for light transmission from Si into sapphire through a-$SiN_{0.62}$ and AlN, is equal to the critical angle for light transmission directly from Si into sapphire. The corresponding effective, fractional solid angle Ω_F for light transmission from Si into sapphire through AlN, as a function of wavelength is calculated as the effective light transmission cone solid angle $\Omega_{eff} = \pi(\theta_{C\text{-eff}}(\lambda))^2$ divided by 4π sr solid angle of the sphere into which the isotropic point source emits, as shown in Fig. 29.

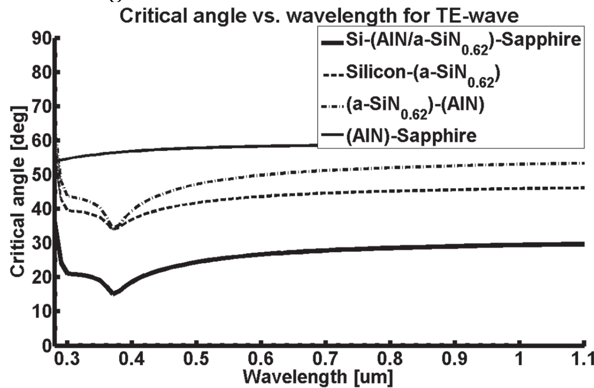


Fig. 28. Critical angles for material interfaces in the substrate.

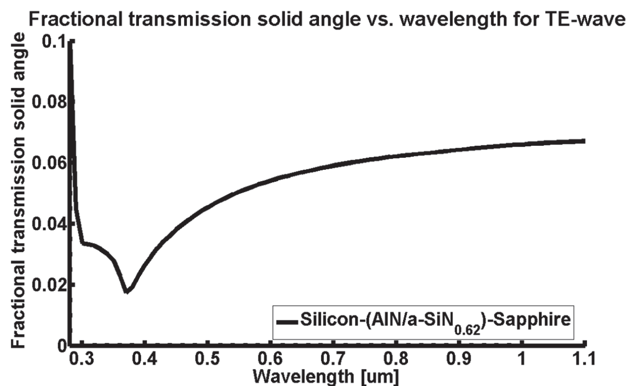


Fig. 29. Fractional solid angle of the light transmission cone.

Reducing the fractional solid angle of the light transmission cone calculated in Fig. 29, would help to prevent optical k-vectors with large incidence angle at the silicon-a-$SiN_{0.62}$, a-$SiN_{0.62}$-AlN and AlN-sapphire interfaces from propagating into the sapphire substrate where they can undergo multiple reflections and transmission into distant APD detectors in the array to produce optical crosstalk at a distance. Reducing the effective fractional solid angle of the light transmission cone requires a large refractive index contrast ratio between the Si semiconductor device layer and the optically transparent supporting substrate and does not depend on the thin antireflective layers such as a-$SiN_{0.62}$ and AlN between the Si and sapphire where $n_{Si} > n_{a\text{-}SiN_0.62} > n_{AlN} > n_{SAPPHIRE}$.

It will be assumed as in Sec 3.1 that any optical k-vectors reflected back into the silicon APD by TIR will not have a second pass, or opportunity to escape the mesa pixel by transmission into the sapphire substrate waveguide and even if such TIR optical k-vectors might be transmitted through the (111) sidewalls of the mesa, the light will subsequently be blocked by the anode metal grid and will not contribute to optical crosstalk. Thus, only the optical k-vectors emanating from the isotropic point source in the APD multiplication region and contained by the light transmission cone calculated in Fig. 29 for $280 < \lambda_0 < 1100$ nm wavelengths or contained by the solid angle subtended by most of the silicon mesa base area for $250 < \lambda_0 < 280$ nm wavelengths, will couple into the sapphire substrate and therefore contribute to the indirect optical crosstalk. Using the result from Fig. 29, it is possible to calculate the fraction of light emitted by the isotropic point source in the mesa APD multiplication region that is transmitted through the sapphire substrate to other APD detectors in the array as a function of wavelength. Multiple reflections may occur in the sapphire substrate for the APD emitted light, and such reflections might not necessarily be bounded by the areas of the eight numbered and immediately adjacent 27 µm mesa APD detector pixels shown in Fig. 30.

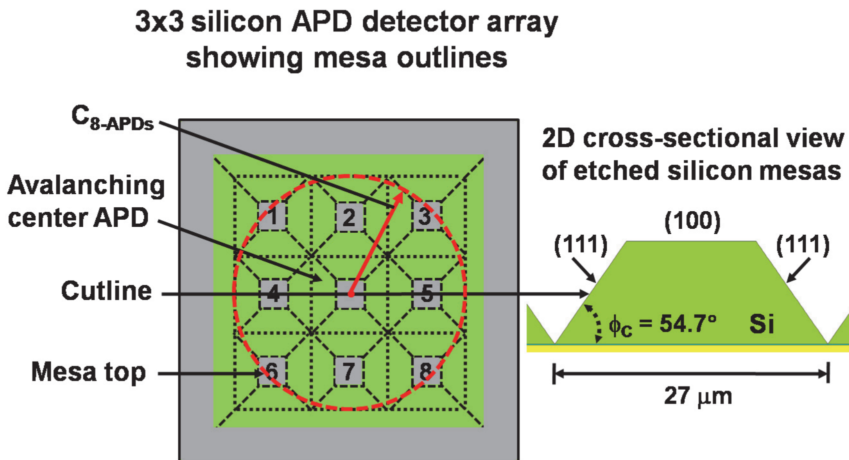


Fig. 30. 3x3 array showing eight immediately adjacent APDs.

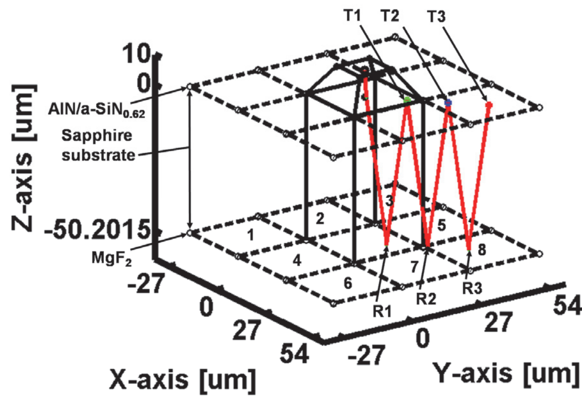


Fig. 31. 3-D ray tracing shows simulated multiple reflections.

The optical transmittance into adjacent detectors numbered 1-8 as well as other detectors outside of the immediately adjacent numbered pixels shown in Fig. 30, is obtained by calculating the fraction of light transmitted into silicon after each successive reflection cycle in the sapphire substrate for an optical k-vector as shown in Fig. 31, using the wave transfer matrix-scattering matrix method discussed in Sec. 2.2. The first reflection cycle in the sapphire substrate is indexed as T_1 followed by the second and third cycles with index T_2, T_3 …. T_N where T_N is the highest calculated reflection in the substrate. The results from Fig. 29 and Fig. 31, are used to calculate the fraction of light emitted by the isotropic point source in the mesa APD multiplication region, that will be transmitted through the sapphire substrate to other APD detectors in the array as a function of wavelength as shown in Figs. 32-33.

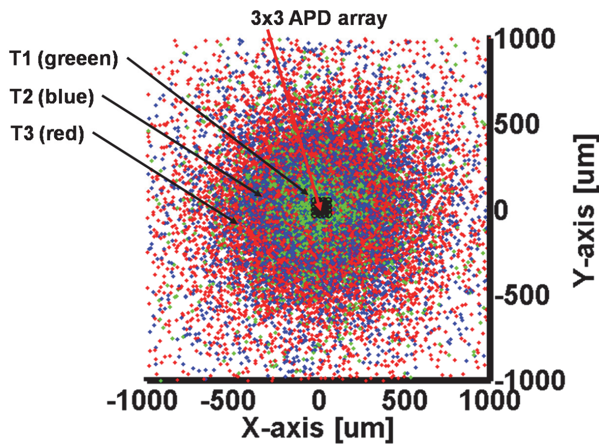


Fig. 32. Average crosstalk distance for 50 μm thick sapphire.

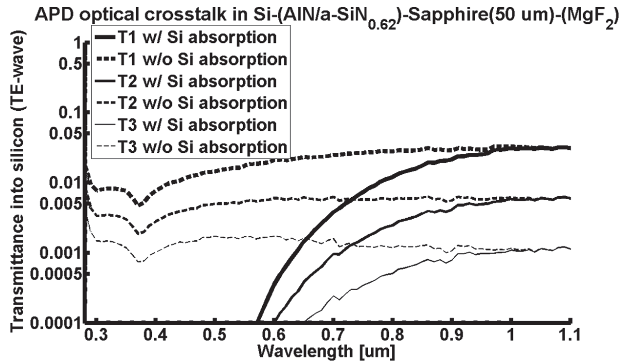


Fig. 33. Indirect APD optical crosstalk in 50 μm thick sapphire.

The average distance of light transmittance points *T*1, *T*2 and *T*3 into the neighboring APD pixels, from the avalanching center mesa APD (shown in Fig. 30) is calculated in Fig. 32 for a 50 μm thick sapphire substrate. In Fig. 33, the fraction of light emitted by the isotropic point source in the mesa APD multiplication region and transmitted to neighboring APD pixels is calculated for a maximum of three reflection cycles, *T*1, *T*2 and *T*3, with and without light self-absorption in the silicon. (Lahbabi et al., 2000) On the first reflection cycle represented by *T*1 (shown in Figs. 31-33), between 1-5% of the isotropically emitted light from the APD multiplication region having wavelength 280-1100 nm, is transmitted into neighboring pixels while the second reflection cycle *T*2, transmits 0.1-0.5% and the third reflection cycle *T*3, transmits 0.05-0.1% of emitted light into the neighboring pixels. The results in Fig. 34 show that the average distance of *T*1 for a 10 μm thick sapphire substrate corresponds to a radius of a circle contained by the eight adjacent pixels of the avalanching center APD shown in Fig. 30.

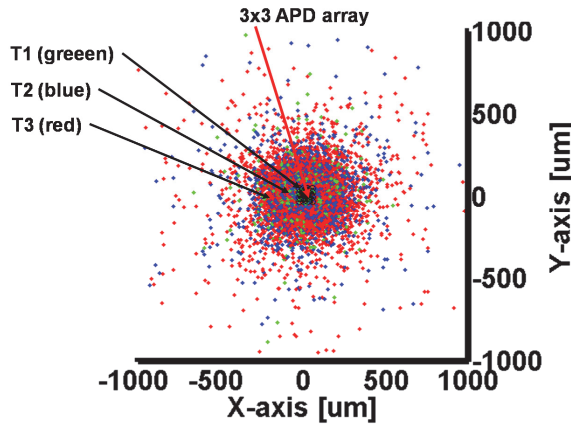


Fig. 34. Average crosstalk distance for 10 μm thick sapphire.

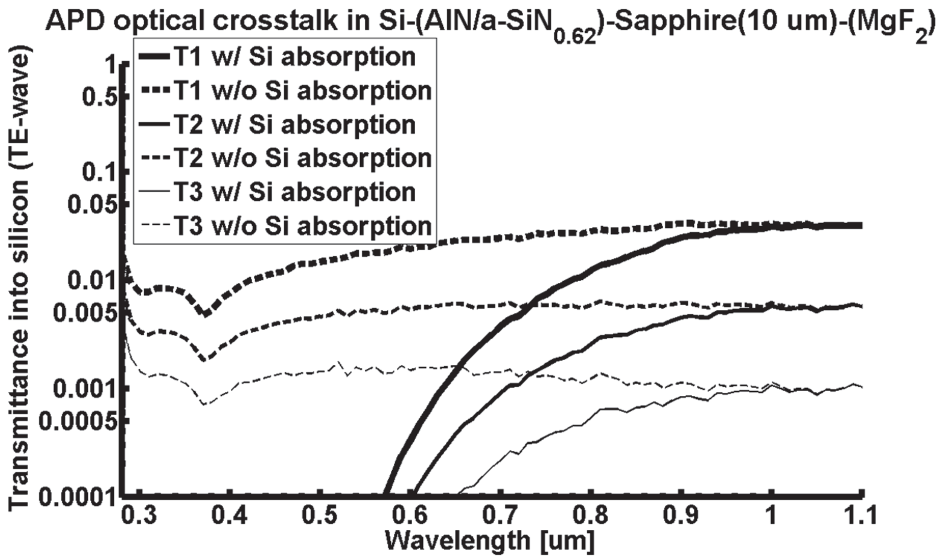


Fig. 35. Indirect APD optical crosstalk in 10 μm thick sapphire.

The results in Fig. 34 show that the average distance of $T1$ for a 10 μm thick sapphire substrate corresponds to a crosstalk radius $C1_{CT} \approx 40$ μm of a circle fully inscribed into the square area formed by the eight adjacent pixels of the avalanching center APD shown in Fig. 30, where $C1_{CT} < C_{8\text{-APDs}} = 40.5$ μm. Comparing the calculated results obtained in Sec. 3.1-3.2 for indirect optical crosstalk resulting from light emission during impact ionization in 27 μm mesa APDs, respectively in Si-(AlN)-sapphire and Si-(AlN/a-$SiN_{0.62}$)-sapphire substrates with back-side $\lambda/4$-MgF_2 antireflective layer, it is evident that the higher transmittance substrate with (AlN/a-$SiN_{0.62}$) antireflective bilayer, also exhibits higher levels of indirect optical crosstalk. This result is expected since a larger fraction of light at points $T1$, $T2$ and $T3$ will be transmitted from sapphire into neighboring silicon mesa APDs due to the more efficient antireflective (AlN/a-$SiN_{0.62}$) bilayer between sapphire and silicon compared to the $\lambda/4$-AlN monolayer. In Sec. 3.3, a figure of merit is introduced for comparing the performance of the two different silicon-on-sapphire substrates analyzed in Sec. 3.1-3.2, based on the level of noise increase in the APD detector array resulting from indirect optical crosstalk from light emitted by the avalanche process. The results in Sec. 3.1-3.2 will be analyzed in Sec. 5 to assess their effect on the signal-to-noise ratio of the APD detectors in an array.

3.3 Figure of merit for the noise performance of silicon-on-sapphire substrates due to the APD emitted light

The results from the analysis of indirect optical crosstalk for 27 μm mesa APDs fabricated in Si-(AlN)-sapphire and Si-(AlN/a-$SiN_{0.62}$)-sapphire substrates with $\lambda/4$-MgF_2 back-side

antireflective layer in Sec. 3.1 and 3.2 respectively, show that the latter substrate with more efficient (AlN/a-$SiN_{0.62}$) antireflective bilayer between sapphire and silicon also produces greater levels of indirect optical crosstalk due to light emitted by the avalanche process. It is useful to be able to describe the levels of indirect optical crosstalk in 27 μm mesa APD arrays using silicon-on-sapphire substrates from light emitted by the avalanche process, in terms of a figure of merit that allows comparison of the detector noise performance for the different back-illuminated substrates including Si-(AlN)-sapphire and Si-(AlN/a-$SiN_{0.62}$)-sapphire.
Fundamentally, optical crosstalk between closely spaced APD detectors in a high resolution array due to light emitted by the avalanche process, produces an increase in the detector noise in the array above the noise level of a standalone detector. To understand how the enhancement or increase in detector noise in an array occurs due to indirect optical crosstalk, it is helpful to consider the examples presented in Figs. 24 and 34, where indirect optical crosstalk from APD emitted light occurs primarily between an APD detector and its eight nearest neighbors, resulting from thinning of the sapphire substrate to $d_{SAPPHIRE}$ = 10 μm. Assuming that the APDs are operating either in linear mode with gain or in non-linear Geiger-mode so that impact ionization and avalanche multiplication of charge carriers can occur, then the APD emitted photon flux resulting from impact ionization and avalanche gain will be given by Eq. (9). (Stern & Cole, 2010)

$$\Phi_{APD0} = \eta_p \langle G \rangle \left(T\eta_{abs}\beta\Phi + \eta_a \Phi_e \right) \tag{9}$$

In Eq. (9), Φ_e describes the average number of thermally generated dark electrons per second and $T\eta_{abs}\Phi$ describes the average number of photogenerated electrons per second where T (shown in Fig. 13) represents the optical power transmittance into the device, η_{abs} represents the absorption efficiency of light in the silicon and Φ represents the incident photon flux. In Eq. (9) it is assumed that both photogenerated and thermally generated electrons traversing the multiplication region of the APD produce secondary electrons through avalanche multiplication with an efficiency β and η_a respectively. (Stern & Cole, 2010) The electrons traversing the multiplication region of the APD produce photons with an efficiency η_P for each traversing electron. A higher average APD gain $<G>$ produces more photons since greater numbers of electrons traverse the multiplication region and the light generating efficiency $\eta_P(E)$, depends on the electric field E, in the multiplication region which is greater at higher detector gain. The APD emitted photon flux in Eq. (9) has a wavelength range of $350 < \lambda_0 < 1100$ nm and therefore can be written as $\Phi_{APD}(\lambda)$. (Akil et al., 1998, 1999)
In the 27 μm mesa APD arrays analyzed in Secs. 3.1-3.2, the photons generated in the multiplication region and emitted isotropically, can only be transmitted to the eight nearest neighboring pixels through the wafer substrate. An increase in APD detector noise in an array occurs when a fraction of the APD emitted photon flux Φ_{APD0} from Eq. (9) is transmitted to the neighboring pixels, thereby increasing the multiplied electron flux ($T\eta_{abs}\beta\Phi + \eta_a\Phi_e$), in those devices that in turn increases their emitted photon flux Φ_{APD}, creating a positive feedback effect. The crosstalk generated multiplied electron flux is defined according to Eq. (10).

$$\Phi_{CT0} = T1\eta_{abs}\beta\Phi_{APD0} \tag{10}$$

In Eq. (10), Φ_{CT0} represents the multiplied electron flux generated in neighboring APD detectors as a result of the APD emitted photon flux Φ_{APD0} given by Eq. (9). The quantity $T1 >> T2 >> T3$ was calculated in Sec. 3.1-3.2 for Si-(AlN)-sapphire and Si-(AlN/a-$SiN_{0.62}$)-sapphire substrates with $\lambda/4$-MgF_2 back-side antireflective layer and represents the fraction of the isotropically emitted APD light that is transmitted into neighboring APD detectors as shown in Figs. 23 and 33. Since the sapphire substrate $d_{SAPPHIRE}$ = 10 μm, the multiplied electron flux Φ_{CT0} from Eq. (10) is produced in the eight adjacent detectors as shown in Figs. 20 and 30. The eight adjacent APD detectors however, each produce the same multiplied electron flux Φ_{CT0}, in their respective eight adjacent pixels and therefore, the total multiplied electron flux in the APD will increase in a first approximation to ($T\eta_{abs}\beta\Phi + \eta_a\Phi_e + \Phi_{CT0}$). Positive feedback will further increase Φ_{CT} and to calculate the increase, an indirect crosstalk parameter D is defined according to Eq. (11).

$$D = \frac{\Phi_{CT0}}{T\eta_{abs}\beta\Phi + \eta_a\Phi_e} = \eta_p \langle G \rangle T1\eta_{abs}\beta \tag{11}$$

The indirect optical crosstalk parameter D in Eq. (11) represents the ratio between the multiplied electron flux generated in neighboring APD detectors as given by Eq. (10), with respect to the multiplied electron flux in the APD ($T\eta_{abs}\beta\Phi + \eta_a\Phi_e$), due to dark electrons and non-crosstalk, photogenerated electrons shown in Eq. (9). The indirect optical crosstalk parameter D, represents a useful figure of merit for the APD array design, describing the degree of indirect optical crosstalk that occurs through the substrate for different mean gain $\langle G \rangle$ in the APD. The normal range of values for D should be $0.0 < D < 1.0$. A lower D value for a given mean gain $\langle G \rangle$, represents a higher performing substrate characterized by lower levels of indirect optical crosstalk. The total multiplied electron flux Φ_{CT-TOT} in the APD due to indirect optical crosstalk can be calculated as shown in Eq. (12), using the indirect optical crosstalk parameter D.

$$\Phi_{CT-TOT} = \sum_{k=0}^{\infty} \Phi_{CT0} D_{k+1}^{k} = T1\eta_{abs}\beta\Phi_{APD0}\left(1 + D + D^2 + ... + D^k\right) \tag{12}$$

In Eq. (12), k takes on integer values from 0 to ∞. It is evident from Eq. (12) that if the value of the indirect optical crosstalk parameter D, is between $0.0 < D < 1.0$, then Φ_{CT-TOT} converges, however, if $D > 1$, then the noise current in the array will increase without bound. In practice, APD quench times in the Geiger-mode will limit the noise current growth, however, the imaging array will become dominated by noise and effectively rendered unusable. The total electron flux in an APD due to indirect optical crosstalk as given by Eq. (12), represents a mean value and should be independent of the distance of indirect optical crosstalk in the sapphire substrate, remaining valid whether the substrate has a thickness $d_{SAPPHIRE}$ = 10 or 50 μm.

The optical crosstalk parameter D can be calculated using Eqs. (9-11), as a function of the mean detector gain $\langle G \rangle$, and different illumination conditions, for imaging arrays comprised of 27 μm mesa APDs fabricated using Si-(AlN)-sapphire and Si-(AlN/a-$SiN_{0.62}$)-sapphire substrates with $\lambda/4$-MgF_2 back-side antireflective layer. The values of parameters used to calculate D are given in Table 3.

Parameter	Value
Pixel size	27 µm
Pixel area, A_{PIXEL}	729×10^{-8} cm^2
Pixel height	10 µm
Focal plane array size	1024 x 1024
FPA side length	2.7648 cm
FPA area, A_{FPA}	7.644 cm^2
Camera lens focal length	21 cm
Focal ratio setting, $f/\#$	5.6
Camera entrance aperture area, $A_{APERTURE}$	11.04 cm^2
Area of the sun's image projected onto the FPA, $A_{SUN\text{-}FPA}$	0.0309 cm^2
Total number of pixels that record the sun's projected image	4238 pixels
APD focal plane array temperature	T = 243 K
Photon generation efficiency in APD multiplication region	$\eta_p = 2.9 \times 10^{-5}$

Table 3. Indirect optical crosstalk calculation parameters.

The total unmultiplied electron flux due to photogenerated and dark electrons is calculated for the 27 µm mesa APD in Figs. 36-37.

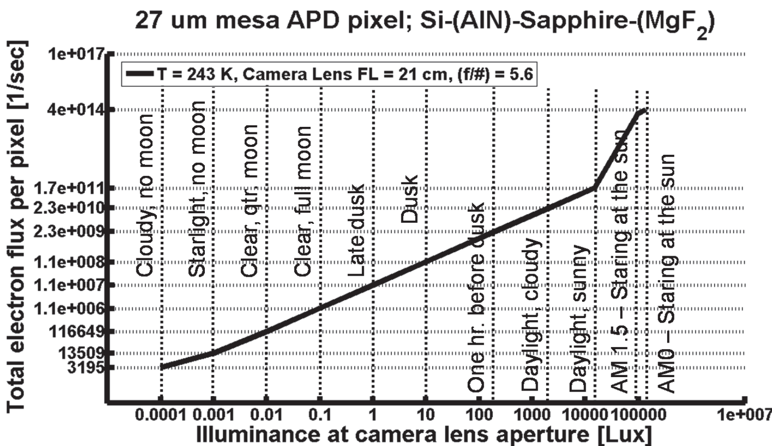


Fig. 36. Total unmultiplied electron flux ($T\eta_{abs}\Phi + \Phi_e$) in APD.

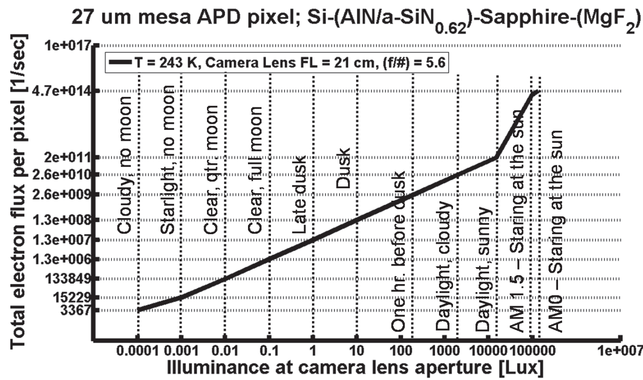


Fig. 37. Total unmultiplied electron flux ($T\eta_{abs}\Phi + \Phi_e$) in APD.

In Figs. 36 and 37, the unmultiplied total electron flux in the 27 μm mesa APD detector fabricated on Si-(AlN)-sapphire or Si-(AlN/a-$SiN_{0.62}$)-sapphire with $\lambda/4$-MgF_2 back-side antireflective layer, is shown to increase as the illumination level at the camera lens increases. The camera lens has focal length F = 21 cm and an aperture stop setting $f/\#$ = 5.6 as indicated in Table 3. The APD detector array operating temperature is set to T = 243 K as provided by a two stage thermoelectric cooler. Using the results from Figs. 36-37 with Eqs. (9-11), the indirect optical crosstalk parameter D for APD emitted light is calculated as a function of the average APD gain $<G>$ in Fig. 38, for the lowest illumination condition occurring on a cloudy moonless night.

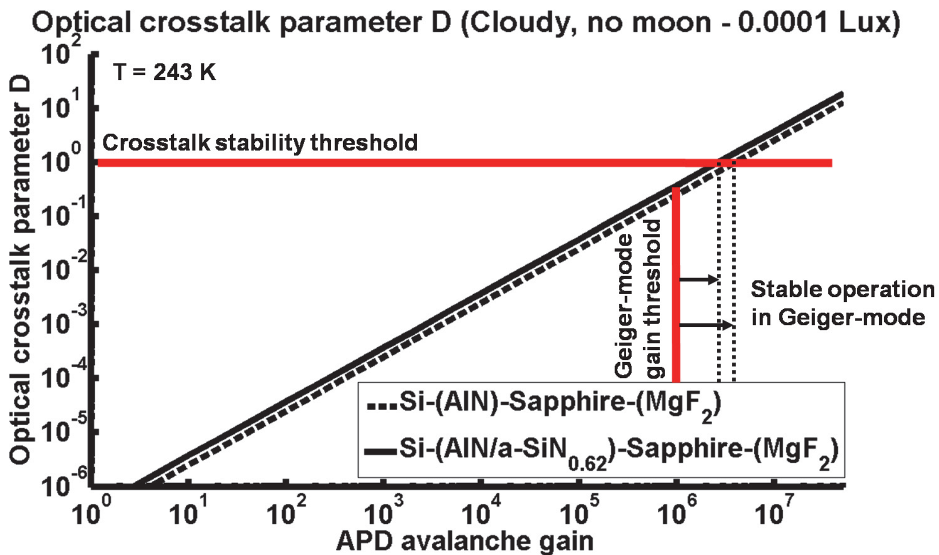


Fig. 38. Optical crosstalk parameter D as a function of APD detector gain for the lowest natural illumination condition of 0.0001 lux at the camera lens, having focal length F = 21 cm and $f/\#$ = 5.6.

The calculation in Fig. 38 considers a worst case example of crosstalk in the FPA, without silicon self-absorption of APD emitted light and approximates the spectral characteristic of the APD emitted photon flux Φ_{APD0} given by Eq. (9), as having a sharp emission peak at 2 eV corresponding to λ_0 = 620 nm, rather than a broad emission spectrum of 350 < λ_0 < 1100 nm described by Akil. The theory of Akil assumes that light emission below 2 eV occurs due to indirect interband transitions, while bremsstrahlung generates the emission from 2.0-2.3 eV and above 2.3 eV, direct interband transitions dominate, however, the theory does not consider light self-absorption in silicon. The theory of Lahbabi assumes an indirect interband recombination model and considers self-absorption of light in the silicon which for a multiplication region located at a height h = 9 μm above the silicon-sapphire interface in the 27 μm mesa APD, will absorb most of the UV and visible light as shown in Figs. 23 and 33, hence the transmission of mainly red light and NIR radiation into the sapphire substrate. Therefore, approximating that Φ_{APD0} given by Eq. (9) occurs at a monochromatic wavelength λ_0 = 620 nm corresponding to a photon energy of 2 eV, is consistent with the results of Akil, Lahbabi and Rech, for the 27 μm mesa APD design presented here. (Akil et al., 1998, 1999; Lahbabi et al., 2000; Rech et al., 2008)
The important result from Fig. 38 confirms that both Si-(AlN)-sapphire and Si-(AlN/a-$SiN_{0.62}$)-sapphire wafer substrates with $\lambda/4$-MgF_2 back-side antireflective layer will support stable APD detector array operation at T = 243 K in both the linear mode and Geiger-mode gain regimes, for the lowest levels of natural illumination of 0.0001 lux at the camera lens. The 27 μm mesa APD detector must have an average gain $\langle G\rangle \leq 4 \times 10^6$ or $\langle G\rangle \leq 3 \times 10^6$ for Si-(AlN)-sapphire and Si-(AlN/a-$SiN_{0.62}$)-sapphire wafer substrates respectively, to preserve an optical crosstalk parameter D < 1, necessary for stable array operation. Such a value of the gain is three times in magnitude above the commonly recognized $\langle G\rangle = 1 \times 10^6$ gain threshold for Geiger-mode operation. The APD detector must therefore be designed and operated in a manner as to prevent the average gain from exceeding the limits for stable array operation. The result from Fig. 38 shows that the planar, high transmittance, back-illuminated, silicon-on-sapphire wafer substrates described, will indeed support stable operation of high quantum efficiency and high resolution 27 μm mesa APD detector arrays operating at the lowest level of natural illumination of 0.0001 lux at the camera lens in dual linear and Geiger-mode. Calculations in fact, confirm stable, wide dynamic range operation of the APD array over the full range of natural illumination conditions (shown in Figs. 36-37) from AM 0 in space to the example in Fig. 38 of a cloudy moonless night. In Sec. 4, the indirect optical crosstalk from ambient incident illumination is calculated for the planar, back-illuminated, silicon-on-sapphire wafer substrates supporting high resolution, 27 μm mesa APD detector arrays. The contribution of indirect optical crosstalk to the APD detector signal-to-noise ratio (SNR) will be analyzed in Sec. 5.

4. Optical crosstalk from ambient light coupled into the sapphire waveguide

It has been demonstrated in Sec. 3 that only a relatively small fraction of the photons generated by impact ionization in a 27 μm mesa APD and emitted isotropically, are coupled into the planar sapphire substrate waveguide and transmitted to neighboring detectors, thereby contributing to an overall increase in noise levels in the array. In this section, a similar analysis considers indirect detector array optical crosstalk due to ambient light from

a point source at infinity, incident on the back-illuminated, sapphire substrate waveguide undergoing multiple reflections and transmission into adjacent mesa APD detectors as shown in Fig. 15 and Fig. 39.

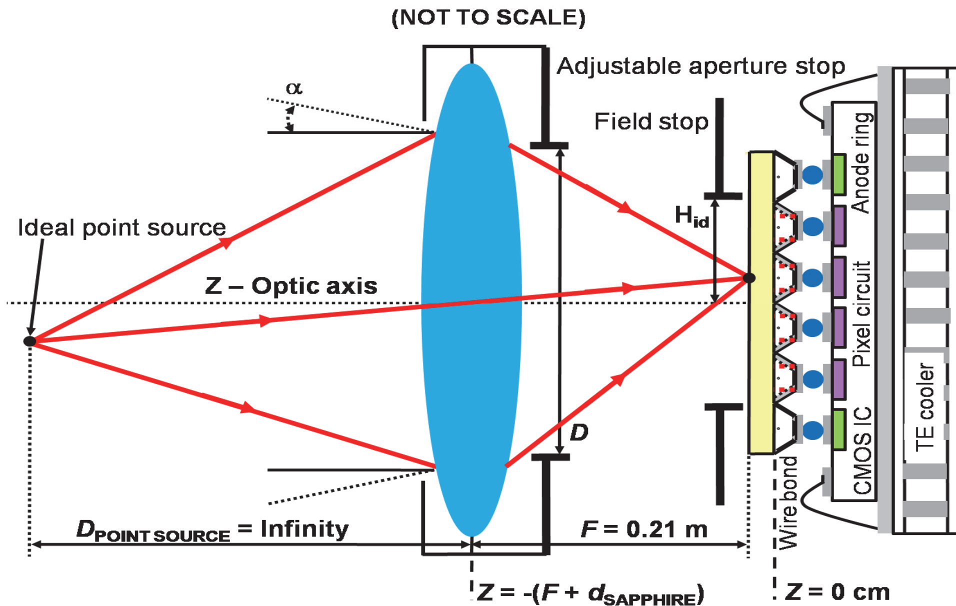


Fig. 39. Isotropic point source at infinity illuminates 27 µm mesa APD in 1024x1024 FPA with $f/\#$ = 5.6 camera lens.

In Fig. 39, an ideal, isotropic point source of light is assumed to be located at an infinity distance, illuminating a 27 µm mesa APD detector in a 1024x1024 pixel FPA through a camera lens with focal ratio setting $f/\#$ = 5.6. The camera lens and aperture stop or iris are circular, therefore, the Airy formula predicts a central disk or spot radius in the image plane for the ideal point source given approximately by Eq. (13).

$$r_{\mathrm{SPOT}} \approx 1.22 \frac{F\lambda_0}{D} \tag{13}$$

In Eq. (13), F and D are the camera lens focal length and diameter respectively and λ_0 is the optical wavelength given in micrometers. The diameter of the central Airy disk will therefore be approximately 5.6 µm as calculated from Eq. (13) with λ_0 = 0.41 µm and $f/\#$ = 5.6, which is significantly smaller than the mesa APD detector pixel size of 27 µm. The subsequent analysis and calculation of indirect optical crosstalk will therefore assume that the point source of light at infinity is focused to an infinitesimal rather than a finite diameter point in the image plane, located directly at the center of the 27 µm mesa APD base area as shown in Fig. 39. The optical k-vectors from the infinite distance point source of light arrive at various incidence angles at the image plane after focusing by the camera lens and are transmitted into the sapphire waveguide where they can undergo multiple reflections.

A numerical or Monte Carlo simulation approach is used to calculate the fraction of light incident on a pixel which is transmitted to neighboring detectors by multiple reflections in the sapphire substrate. The simulation is performed on a 27 μm mesa APD pixel located in the center of the 1024x1024 FPA, aligned with the optical axis of the imaging system shown in Fig. 39. Selecting the center pixel in the FPA for analysis as opposed to selecting a corner pixel, simplifies the indirect optical crosstalk calculation by ensuring that all of the optical k-vectors emanating from the camera lens toward the center point of the pixel in the image plane are meridional rays contained in the same plane as the optical axis, hence there are no skew rays present. In a 1024x1024 APD-FPA with 27 μm pixels and camera lens focal ratio $f/\# = 5.6$ as shown in Fig. 39, the optical crosstalk from light incident at pixels near the corners of the FPA will be greater than for pixels near the optic axis, due to larger optical k-vector incidence angles at corner pixels. The following Sec. 4.1-4.2, analyze and calculate the indirect optical crosstalk from ambient light incident on 27 μm mesa APDs fabricated using respectively, Si-(AlN)-sapphire and Si-(AlN/a-$SiN_{0.62}$)-sapphire substrates with $\lambda/4$-MgF_2 back-side antireflective layer.

4.1 Indirect optical crosstalk from light incident on back-illuminated, sapphire waveguide; Si-(AlN)-sapphire

To study the nature of indirect optical crosstalk in APD-FPAs fabricated on planar, back-illuminated, Si-(AlN)-sapphire substrates with $\lambda/4$-MgF_2 back-side antireflective layer without microlenses, due to incident illumination from a point source located at infinity as shown in Fig. 39, a Monte Carlo modeling and simulation approach is used. A 3-D Cartesian coordinate system can be defined where the z-axis represents the optic axis of the camera system as shown Fig. 39, and the camera lens with focal length $F = 21$ cm is located in the x-y plane at $z = -(F + d_{SAPPHIRE} + d_{AlN} + d_{MgF2})$. The 1024x1024 APD-FPA with 27 μm mesa pixels is located at $z = 0$ cm. Figure 40 shows a 3x3 array of 27 μm mesa APD detectors and Fig. 41 shows the points of light transmittance $T1$, $T2$ and $T3$ in the sapphire substrate due to multiple reflections, for an optical k-vector incident to the $F = 21$ cm camera lens with focal ratio setting $f/\# = 5.6$, from a point source located at infinity.

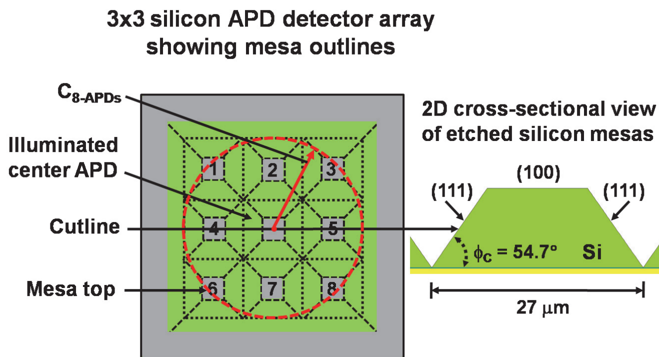


Fig. 40. 3x3 array showing eight immediately adjacent APDs.

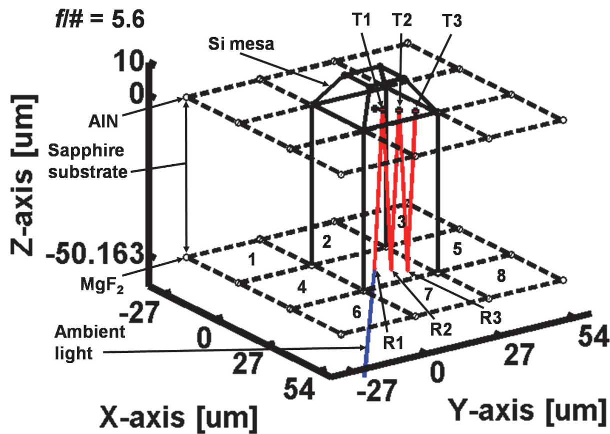


Fig. 41. 3-D ray tracing shows multiple reflections for *f*/# = 5.6.

In Fig. 42, 3-D ray tracing is used to calculate paths of light propagation for the randomly generated optical k-vectors from a point source at infinity over a 250 < λ_0 < 1100 nm wavelength range, transmitted into the sapphire substrate and undergoing multiple reflections for a camera focal ratio *f*/# = 5.6. In Fig. 42, even after three reflection cycles, the points of transmittance at *T*3 occur inside the 27 μm mesa pixel base area for the APD aligned with the camera optic axis and located in the center of the 1024x1024 FPA. Therefore, a camera focal ratio setting *f*/# = 5.6 produces negligible indirect optical crosstalk due to ambient incident light from a point source at infinity that is spatially conjugated to a 27 μm mesa APD pixel aligned with the camera optic axis and located in the center of the 1024x1024 FPA. The results from Figs. 41-42, are used to calculate in Fig. 43, the fraction of the light incident at the APD aligned with the camera optic axis and located in the center of the 1024x1024 FPA, that is transmitted at points *T*1, *T*2 and *T*3 following reflections in the sapphire substrate, when the focal ratio setting *f*/# = 5.6.

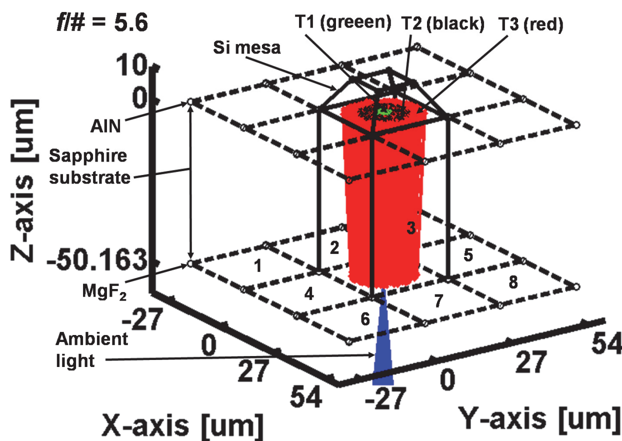


Fig. 42. 3-D ray tracing shows minimal crosstalk for *f*/# = 5.6.

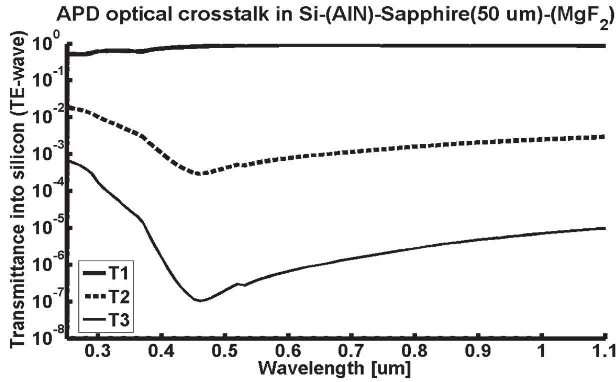


Fig. 43. Indirect crosstalk for 50 μm thick sapphire and *f*/# = 5.6.

The Figs. 44-45 show light propagation paths for randomly generated optical k-vectors emitted by a point source at infinity over a 250 < λ_0 < 1100 nm wavelength range, transmitted into the sapphire substrate and undergoing multiple reflections, for camera focal ratios *f*/# = 16 and *f*/# = 2.0 respectively.

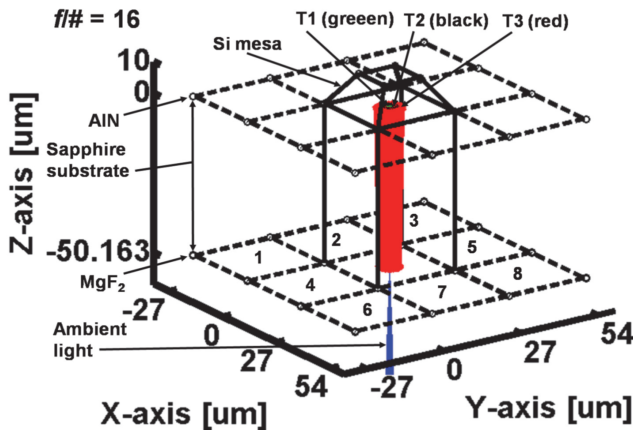


Fig. 44. 3-D ray tracing shows minimal crosstalk for *f*/# = 16.

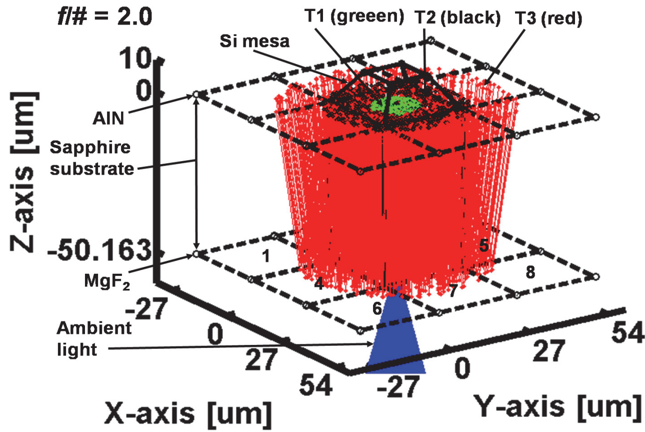


Fig. 45. 3-D ray tracing reveals indirect crosstalk for *f*/# = 2.0.

The indirect optical crosstalk due to incident illumination from a point source at infinity of a 27 µm mesa APD pixel coincident with the camera optic axis and located in the center of the 1024x1024 FPA, has been shown to be negligible in planar Si-(AlN)-sapphire substrates with $\lambda/4$-MgF_2 back-side antireflective layer, 50 µm thick sapphire and no microlenses. Although multiple reflections in the sapphire substrate occur for both the APD emitted light and the ambient incident illumination from a point source at infinity, the effect of the latter can be minimized by setting a higher camera focal ratio of *f*/# = 5.6 for example, to ensure that the multiple points of transmittance *T*1, *T*2 and *T*3 occur within the area of the illuminated 27 µm mesa APD. The same spatial confinement of multiply reflected optical k-vectors cannot be implemented as readily for the APD emitted light.

4.2 Indirect optical crosstalk from light incident on back-illuminated, sapphire waveguide; Si-(AlN/$SiN_{0.62}$)-sapphire

To study the nature of indirect optical crosstalk in APD-FPAs fabricated on planar, back-illuminated, Si-(AlN/a-$SiN_{0.62}$)-sapphire substrates with $\lambda/4$-MgF_2 back-side antireflective layer without microlenses, due to incident illumination from a point source located at infinity as shown in Fig. 39, a Monte Carlo modeling and simulation approach is used similar to Sec. 4.1. A 3-D Cartesian coordinate system can be defined where the z-axis represents the optic axis of the camera system as shown Fig. 39, and the camera lens with focal length F = 21 cm is located in the x-y plane at $z = -(F + d_{SAPPHIRE} + d_{a\text{-}SiN_0.62} + d_{AlN} + d_{MgF2})$. The 1024x1024 APD-FPA with 27 µm mesa pixels is located at z = 0 cm. Figure 46 shows a 3x3 array of 27 µm mesa APD detectors and Fig. 47 shows the points of light transmittance *T*1, *T*2 and *T*3 in the sapphire substrate due to multiple reflections, for an optical k-vector incident to the F = 21 cm camera lens with focal ratio setting *f*/# = 5.6, from a point source located at infinity.

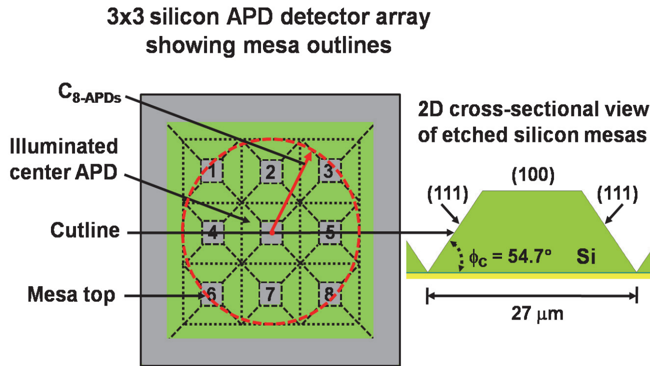


Fig. 46. 3x3 array showing eight immediately adjacent APDs.

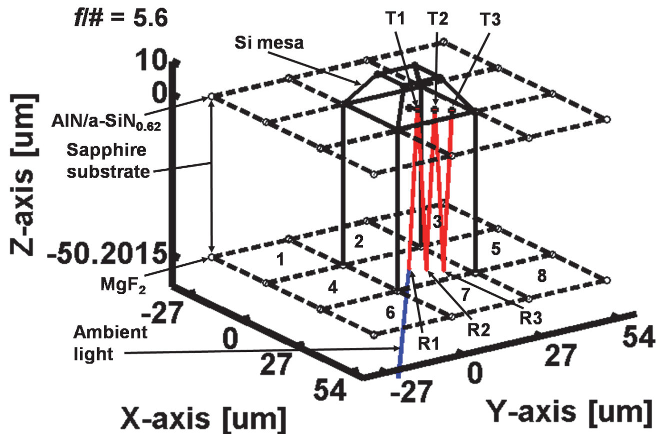


Fig. 47. 3-D ray tracing shows multiple reflections for *f*/# = 5.6.

In Fig. 48, 3-D ray tracing is used to calculate paths of light propagation for the randomly generated optical k-vectors from a point source at infinity over a 250 < λ_0 < 1100 nm wavelength range, transmitted into the sapphire substrate and undergoing multiple reflections, for a camera focal ratio *f*/# = 5.6. In Fig. 48, even after three reflection cycles, the points of transmittance at *T*3 occur inside the 27 μm mesa pixel base area for the APD, which is aligned with the camera optic axis and located in the center of the 1024x1024 FPA. Therefore, a camera focal ratio setting *f*/# = 5.6 produces negligible indirect optical crosstalk due to ambient incident light from a point source at infinity, that is spatially conjugated to a 27 μm mesa APD pixel aligned with the camera optic axis and located in the center of the 1024x1024 FPA. The results from Figs. 47-48, are used to calculate in Fig. 49, the fraction of the light incident at the APD aligned with the camera optic axis and located in the center of the 1024x1024 FPA, that is transmitted at points *T*1, *T*2 and *T*3 following reflections in the sapphire substrate, when the focal ratio setting *f*/# = 5.6.

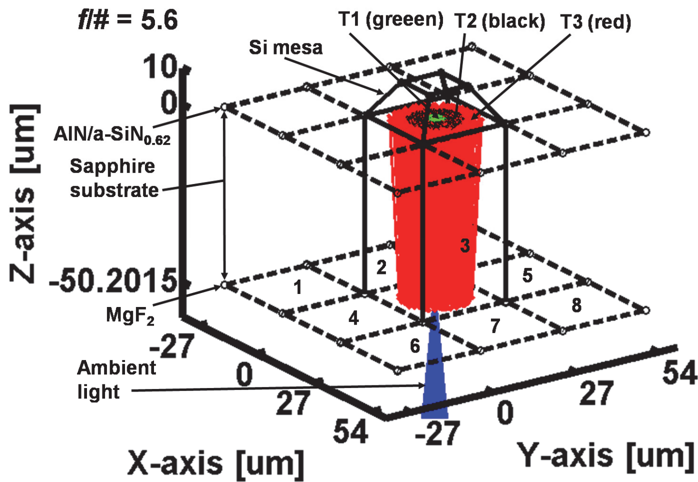


Fig. 48. 3-D ray tracing shows minimal crosstalk for $f/\#$ = 5.6.

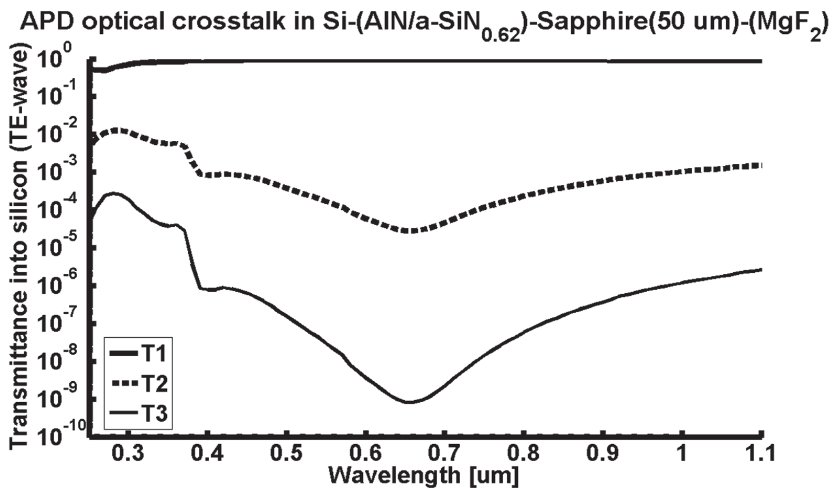


Fig. 49. Indirect crosstalk for 50 μm thick sapphire and $f/\#$ = 5.6.

The Figs. 50-51 show light propagation paths for randomly generated optical k-vectors emitted by a point source at infinity over a 250 < λ_0 < 1100 nm wavelength range, transmitted into the sapphire substrate and undergoing multiple reflections, for camera focal ratios $f/\#$ = 16 and $f/\#$ = 2.0 respectively.

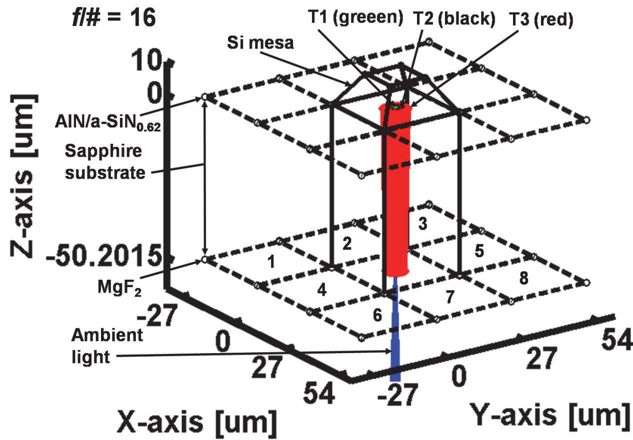


Fig. 50. 3-D ray tracing shows minimal crosstalk for $f/\#$ = 16.

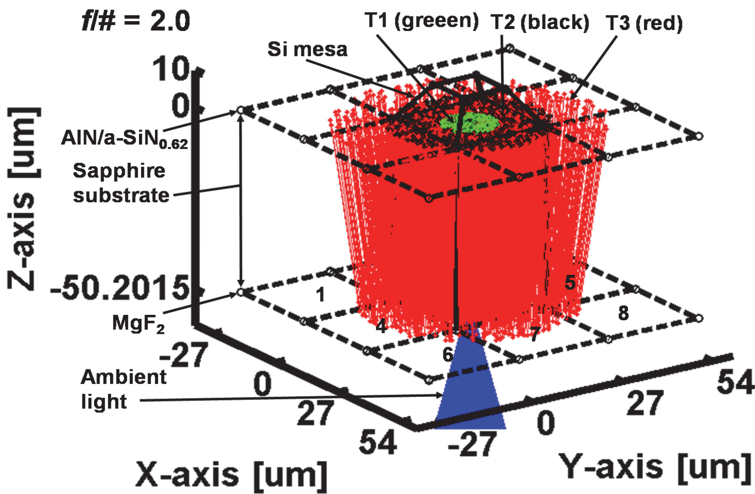


Fig. 51. 3-D ray tracing reveals indirect crosstalk for $f/\#$ = 2.0.

The focal ratio setting $f/\#$ = 2.0 in Fig. 51 produces some, although minimal indirect optical crosstalk due to multiple reflections in the sapphire substrate since the points of transmittance at $T1$ occur inside the 27 µm mesa pixel base area for the APD aligned with the camera optic axis and located in the center of the 1024x1024 FPA.

5. Sensitivity of silicon-on-sapphire mesa APDs with indirect optical crosstalk

The sensitivity or signal-to-noise ratio (SNR) of a back-illuminated, wide dynamic range 27 µm mesa APD detector comprising a 1024x1024 pixel focal plane array (FPA),

fabricated using Si-(AlN)-sapphire or Si-(AlN/a-$SiN_{0.62}$)-sapphire substrates with $\lambda/4$-MgF_2 back-side antireflective layer, can be degraded by indirect optical crosstalk. Previous calculations have shown that the back-illuminated, 27 µm mesa APD detector fabricated on Si-(AlN)-sapphire substrate with $\lambda/4$-MgF_2 back-side antireflective layer without a microlens and unaffected by optical crosstalk, will be capable of imaging with high sensitivity and wide dynamic range using dual linear and Geiger-mode, over the full range of natural illumination conditions from AM 0 in space to a cloudy moonless night. (Stern & Cole, 2010) In this section, the sensitivity of the back-illuminated, 27 µm mesa APD fabricated without a microlens, is calculated for the lowest level of natural illumination of 0.0001 lux at the camera lens, occurring on a cloudy moonless night. The sensitivity calculation considers the increased noise level in the detector due to indirect optical crosstalk from APD emitted light described in Sec. 3. The indirect optical crosstalk from ambient illumination of pixels described in Sec. 4, however, will not be considered in the detector sensitivity calculation due primarily to improved spatial confinement of optical k-vectors for $f/\# \geq 5.6$, compared to the APD emitted light in Sec. 3. The expressions in Eqs. (14-15) yield the wide dynamic range, 27 µm mesa APD detector signal-to-noise ratio with optical crosstalk for the linear and Geiger-mode optical receivers respectively, assuming the optical crosstalk current is independent from the detector photocurrent and dark current.

$$SNR_{APD-linear-receiver} = \frac{\left\langle i_p(t) \right\rangle^2}{\sigma_{i_p}^2 + \sigma_{i_D}^2 + \sigma_{i_{CT}}^2 + \sigma_{i_{en}}^2} \quad (14)$$

$$SNR_{APD-Geiger-receiver} = \frac{\left\langle i_p(t) \right\rangle^2}{\sigma_{i_p}^2 + \sigma_{i_D}^2 + \sigma_{i_{CT}}^2} \quad (15)$$

In Eqs. (14-15), the signal is given by the square of the mean detector photocurrent $i_p(t)$, and the noise is given by the sum of the variances of the photocurrent, dark current, optical crosstalk current and electronic readout circuit current, which are all assumed to be independent. The signal-to-noise ratios given by Eqs. (14-15) were previously calculated by Stern for the wide dynamic range 27 µm mesa APD without a microlens using Si-(AlN)-sapphire substrate with $\lambda/4$-MgF_2 back-side antireflective layer in the absence of optical crosstalk. (Stern & Cole, 2010) The expressions for the photoelectron current variance, dark current variance and electronic readout circuit noise were also derived. (Stern & Cole, 2010) After late dusk, the wide dynamic range 27 µm mesa APD can be operated in the non-linear Geiger-mode without saturating the maximum count rate of the detector, for a 50 ns quench time. (Kumar et al., 2004) In the Geiger-mode, the optical receiver sensitivity is calculated according to Eq. (15) where the contribution of the electronic circuit noise has been eliminated by direct photon-to-digital conversion in the detector pixel. (Ghioni et al., 1996)

In Eqs. (14-15), it is assumed that the indirect optical crosstalk current due to APD emitted light i_{CT} is independent from the photocurrent i_p and from the dark current i_D in the 27 µm mesa APD detector, thereby allowing the variances $\sigma_{i_p}^2$, $\sigma_{i_D}^2$ and $\sigma_{i_CT}^2$ to be added as

shown. The assumption of independence for i_p, i_D and i_{CT} in the 27 μm mesa APD detector is valid if the total multiplied electron flux $\Phi_{CT\text{-}TOT}$ given by Eq. (12), due to optical crosstalk from the neighboring pixels is given by the sum of even k terms (k = 0, 2, 4 …). The variance of the crosstalk current $\sigma_{i_CT}^2$ in Eqs. (14-15), due solely to even k term contribution to the total multiplied crosstalk electron flux $\Phi_{CT\text{-}TOT}$, has a similar form to the photocurrent variance $\sigma_{i_p}^2$. (Stern & Cole, 2010) The inclusion of odd higher order k terms (k = 1, 3, 5 …) in calculating Eq. (12), would require Eqs. (14-15) to be modified, to account for correlations between the photoelectron current and dark current with the crosstalk current in the 27 μm mesa APD. Figure 52 shows the SNR of 27 μm mesa APD detectors fabricated on Si-(AlN)-sapphire and Si-(AlN/a-$SiN_{0.62}$)-sapphire substrates with λ/4-MgF_2 back-side antireflective layer, calculated according to Eq. (15), using a fixed Geiger-mode gain <G> = 1 x 10^6 and including only the k = 0 term also given by Eq. (10), for calculating the total multiplied electron flux $\Phi_{CT\text{-}TOT}$, as given by Eq. (12).

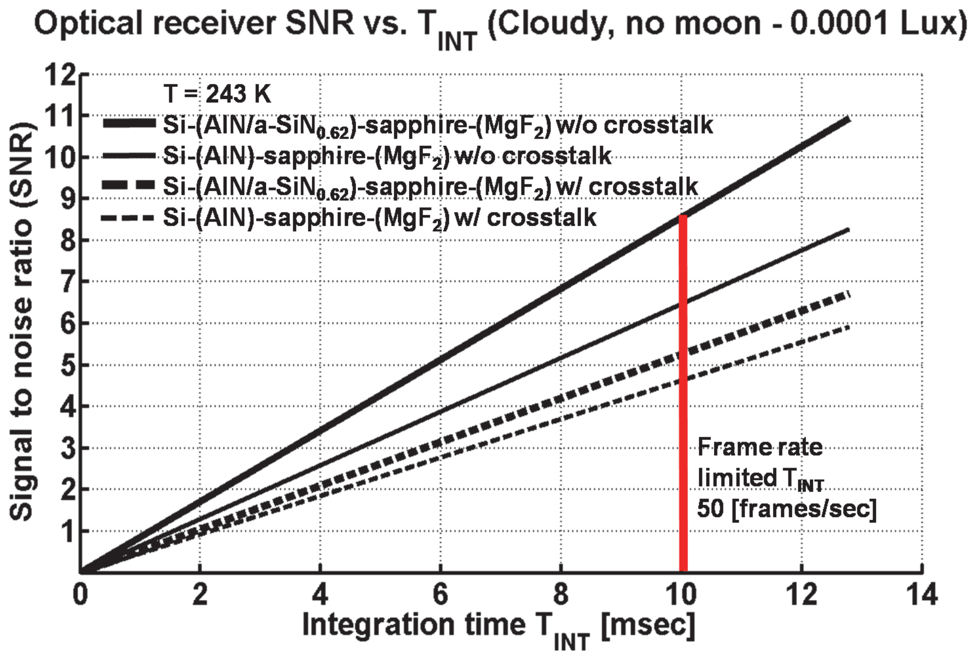


Fig. 52. SNR of the Geiger-mode APD imager on a cloudy moonless night and 0.0001 lux illumination. The Geiger-mode APD has 8-bit resolution and a quench time of 50 ns.

It is evident in Fig. 52, that indirect optical crosstalk in back-illuminated, high resolution, 27 μm mesa APD arrays fabricated on Si-(AlN)-sapphire and Si-(AlN/a-$SiN_{0.62}$)-sapphire substrates with λ/4-MgF_2 back-side antireflective layer without microlenses, reduces the detector sensitivity from SNR ≈ 6 to SNR ≈ 4 and from SNR ≈ 8 to SNR ≈ 5 respectively. Despite indirect optical crosstalk, the Geiger-mode APD-FPA will be capable of imaging at

the lowest level of natural illumination of 0.0001 lux at the camera lens at 50 frames per second with 8-bit resolution.

6. Conclusion

The detailed analysis in this chapter has confirmed an important result, namely, that the sensitivities of the 27 µm mesa APD detectors comprising back-illuminated, silicon-on-sapphire FPAs without monolithic microlenses, are not degraded by indirect optical crosstalk phenomena to the level that would prevent high sensitivity imaging, even for the highest (Geiger-mode) gain regime of operation. Thus, for all but the most demanding imaging applications, monolithic sapphire microlenses are not required. This result is important because sapphire microlens fabrication and in general any fabrication step adds to the complexity and cost of manufacturing. The calculations and analysis in this chapter have referred primarily to silicon APD detector arrays fabricated using novel, back-illuminated silicon-on-sapphire substrates, however, the substrate technology will inherently support not only silicon, but also epitaxially grown $Si_{(1-X)}Ge_X$ APD detector arrays. The family of high transmittance silicon-on-sapphire substrates represents a key enabling technology for the next generation of ultrasensitive, solid-state, high quantum efficiency and high resolution avalanche detector arrays. As the novel substrate technology is developed commercially, new as yet to be defined and designed imager concepts will emerge with it.

7. References

Akil, N., Kerns, S.E., Kerns, D.V., Hoffmann, A. & Charles, J.P. (1998). "Photon generation by silicon diodes in avalanche breakdown," *Applied Physics Letters*, 73(7), 871.

Akil, N., Kerns, S.E., Kerns, D.V., Hoffmann, A. & Charles, J.P. (1999). "A Multimechanism Model for Photon Generation by Silicon Junctions in Avalanche Breakdown," *IEEE Transactions on Electron Devices*, 46(5), 1022-1028.

Charbon, E. (2008). "Towards large scale CMOS single-photon detector arrays for lab-on-chip applications," *Journal of Physics D: Applied Physics*, 41(9), 4010.

Gardeniers, J.G.E., Tilmans, H.A.C. & Visser, C.C.G. (1996). "LPCVD silicon-rich silicon nitride films for applications in micromechanics, studied with statistical experimental design," *Journal of Vacuum Science Technology A*, 14(5), 2879-2892.

Ghioni, M., Cova, S., Zappa, F. & Samori, C. (1996). "Compact active quenching circuit for fast counting with avalanche photodiodes," *Review of Scientific Instruments*, 67(10), 3440-3448.

Guerrieri, F., Tisa, S. & Zappa, F. (2009). "Fast Single-Photon Imager acquires 1024 pixels at 100 kframes/s," *Proc. of SPIE Vol. 7249, 72490U.*

Imthurn, G.P., Garcia, G.A., Walker, H.W. & Forbes, L. (1992). "Bonded silicon-on-sapphire wafers and devices," *Journal of Applied Physics*, 72(6), 2526-2527.

Kumar, P., Kwiat, P., Migdall, A., Nam, S.W., Vuckovic, J. & Wong, N.C. (2004). "Photonic Technologies for Quantum Information Processing," *Quantum Information Processing*, 3(1-5), 215-231.

Lahbabi, M., Ahaitouf, A., Abarkan, E., Fliyou, M., Hoffmann, A., Charles, J.P., Bhuva, B.L., Kerns, S.E. & Kerns, D.V. (2000). "Analysis of electroluminescence spectra of silicon junctions in avalanche breakdown using an indirect interband recombination model," *Applied Physics Letters*, 77(20), 3182.

Lau, S.S., Matteson, S., Mayer, J.W., Revesz, P. & Gyulai, J. (1979). "Improvement of Crystalline Quality of Epitaxial Silicon Layers by Ion-Implantation Techniques," *Applied Physics Letters*, 34(1), 76.

Lowe, A.J., Powell, M.J. & Elliot, S.R. (1986). "The electronic properties of plasma deposited films of hydrogenated amorphous SiN_X ($0<x<1.2$)," *Journal of Applied Physics*, 59(4), 1251.

Makino, T. (1983). "Composition and Structure Control by Source Gas Ratio in LPCVD SiN_X," *Journal of Electrochemical Society*, 130(2), 450-455.

Manasevit, H.M. & Simpson, W.J. (1964). "Single-Crystal Silicon on a Sapphire Substrate," *J. of Appl. Phys.*, 35(4), 1349-1351.

Niclass, C., Rochas, A., Besse, P.A., Popovic, R.S. & Charbon, E. (2005). "CMOS Imager Based on Single Photon Avalanche Diodes," IEEE 13th Internl. Conference on Solid-State Sensors, Actuators and Microsystems, Seoul, S. Korea, 1, 1030-1034.

Park, S.H., Jeon, H., Sung, Y.J. & Yeom, G.Y. (2001). "Refractive sapphire microlenses fabricated by chlorine-based inductively coupled plasma etching," *Applied Optics*, 40(22), 3698-3702.

Plummer, J.D., Deal, M.D. & Griffin, P.D. (2000). [*Silicon VLSI Technology*], Prentice Hall, Upper Saddle River NJ.

Rech, I., Ingargiola, A., Spinelli, R., Labanca, I., Marangoni, S., Ghioni, M. & Cova, S. (2008). "A New Approach to Optical Crosstalk Modelling in Single-Photon Avalanche Diodes," *IEEE Photonics Technology Letters*, 20(5), 330-332.

Rosler, R.S. (1977). "Low Pressure CVD Production Processes for Poly, Nitride and Oxide," Solid State Technology, 20(4), 63.

Saleh, B.E.A. & Teich, M.C. (2007). [*Fundamentals of Photonics*], John Wiley and Sons, New York, 246-253, 755, 789.

Stern, A.G. & Cole, D.C. (2010). "High-sensitivity, wide-dynamic-range avalanche photodiode pixel design for large-scale imaging arrays," *Journal of Electronic Imaging, Special Section on Digital Photography*, 19(2).

Stern, A.G. & Cole, D.C. (2008). "Design of a back-illuminated, crystallographically etched, silicon-on-sapphire avalanche photodiode with monolithically integrated microlens, for dual-mode passive & active imaging arrays," *Proc. of SPIE Vol. 7153, 71530Z.*

Stern, A.G. et al. (2003). "Intensity-imaging focal plane array using Geiger-mode avalanche photodiodes," *Solid State Research Report MIT Lincoln Laboratory*, (4), 23-31.

Sze, S.M. (1981). [*Physics of Semiconductor Devices 2nd Edition*], John Wiley & Sons, New York.

Tauc, J. (1974). [*Amorphous and Liquid Semiconductors*], Plenum Press, New York, 173.

Verlaan, V., Bakker, R., van der Werf, C.H.M., Houweling, Z.S., Mai, Y., Rath, J.K. & Schropp, R.E.I. (2007). "High-density silicon nitride deposited at low substrate temperature with high deposition rate using hot wire chemical vapour deposition," *Surface & Coatings Technology*, 201, 9285-9288.

Part 4

Extended Topics of Photodiodes

13

Single Crystal Diamond Schottky Photodiode

Claudio Verona
Dip. di Ing. Meccanica, Università di Roma "Tor Vergata", Roma
Italy

1. Introduction

Thanks to its extreme optical and electronic properties, diamond appears to be a promising semiconducting material for photon detection. Its wide band-gap, 5.5 eV, results in a very low leakage current and its electronic properties as high carrier mobility allow fast time response (J. E. Field, 1979). Besides, it has a large breakdown electric field (~10 V/µm), a low dielectric constant (i.e. low capacitance), chemical inertness and low intrinsic carrier density, which makes cooling for noise reduction unnecessary (J.Prins, 1997). Its extreme radiation hardness is well known and another interesting feature, again related to the wide band-gap, is its selective sensitivity to radiation with wavelengths shorter than 225 nm (visible-blind detectors) (J.F. Hochedez et al., 2002). Several attempts have been made to build up UV detectors from natural or synthetic diamonds grown by Chemical Vapour Deposition (CVD). A detector often reported in literature is the photoresistor (A. Balducci e al., 2005; T. Teraji et al., 2004) having a planar structure and consisting of a photoconductive diamond film with metal electrodes placed on the top surface. It can operate only with external voltage applied and the signal is affected from secondary electron emission, which is known to strongly affect the detection properties in the UV and EUV spectral regions. A different geometry reported is a polycrystalline sandwiched photodiode structure (V.I. Polyakov et al., 1998, L. Thaiyotin et al., 2002) with a contact on the diamond growth surface and a backside contact on the silicon substrate. However, the CVD diamond performance is limited in this case by the polycrystalline structure due to defect states in the band gap introduced by the grain boundaries (R. D. McKeag&R. B. Jackman, 1998, L. Barberini, 2001), which affects the photoelectric properties and alters the detection characteristics. On the other hand, detector grade natural diamonds are extremely rare and expensive, while high pressure high temperature (HPHT) diamonds have their performance strongly worsened by defects and impurities (E. Pace et a., 2000). A great effort is therefore being devoted to produce device-grade Single Crystal Diamond films (SCD) by homoepitaxial CVD growth on low-cost diamond substrates (S. Almaviva et al., 2009, 2010a). A few years ago, at the University of Rome "Tor Vergata" laboratories, CVD single crystal diamond films were used to obtain a new class of detectors with a layered structure. Thanks to the combination of boron doped and intrinsic single crystal diamond films, together with the possibility to easily build Schottky junctions on intrinsic diamond by thermal evaporation of the metal contacts, it has been possible, by using simple multilayered a p-type/nominally intrinsic diamond/metal layered structures, to obtain high quality and highly reproducible devices which can be effectively used for detection (UV and X-rays) photons.

In this chapter, the fabrication and characterization of two Schottky photodiodes, based on p-type/intrinsic diamond/metal (PIM) Schottky junction in two different operative configurations is reported (S. Almaviva et al., 2010b) . One detector has been designed in transverse configuration with a semitransparent metallic contact evaporated on the diamond surface, while the second one operates in a planar configuration with interdigitated fingers on the diamond surface. One set of fingers is made of aluminium and the second one is made of p-type diamond.

2. CVD single crystal diamond films production (detector grade)

The reactor for Microwave Plasma Enhanced Chemical Vapour Deposition (MPECVD) diamond growth, used at the Mechanical Engineering Department of the University of Rome “Tor Vergata” laboratories is illustrated in Fig.1. This apparatus is used for the deposition of diamond films in the present work. This deposition chamber was built by developing the necessary parameters to obtain films of diamond having the best physical characteristics, together with a high degree of reproducibility results.

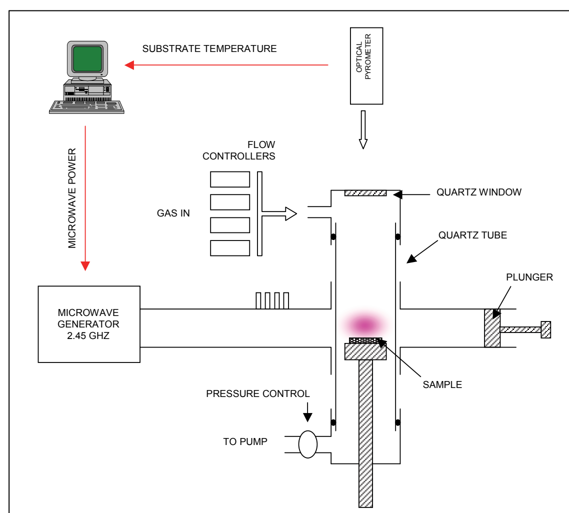


Fig. 1. General scheme of the apparatus for CVD diamond films deposition

The apparatus for CVD diamond deposition is composed by a tubular microwave CVD reactor and the growth chamber consists essentially in a quartz tube into which the precursory gases flow, controlled by four flowmeters, (a mixture 100 H_2/1 CH_4 sccm for the intrinsic growth and 90÷70 H_2/ 2 CH_4/ 10÷30 H_2-B_2H_6 sccm for the doped one) at a stationary fixed pressure (typically 120÷130 mBar).

The tube is put across a waveguide connected to a commercial 2.45 GHz MW generator (Sairem® magnetron, 2kW maximum MW power) and the typical microwave power for both SCD growths is 500÷550 W. The quartz tube acts then as a resonant cavity for the microwaves and a stable plasma burns at the centre of the tube, coupling to the sample holder, that is properly connected to a water cooling system, and thus to the diamond substrate on which the homoepitaxial deposition process takes place. The position of the

plasma can be modified through a "plunger" (a sliding short circuit) that terminates the waveguide and its coupling with the substrate can be optimized by means of a 3-stub tuner put between the MW generator and the quartz tube. The temperature of the growing film is monitored in real-time through an infrared optical pyrometer and can be maintained constant during the whole CVD process thanks to a proper software which follows the growth and varies properly the MW power supplied by the generator. In this type of reactor, microwaves give energy to the electrons of the plasma, which, at his time, give energy to the gas through molecular collisions, heating it. The chemical and physical reactions, which follow this heating, allow the formation of carbon that is deposited on diamond substrate. The chemistry of the diamond deposition is described in the next paragraph. It's worth underline that diamond growth is obtained only if growth parameters (gas composition, pressure, plasma temperature, microwave power density, etc) are chosen in the appropriate way. In particular, the grow rate of diamond with this method vary from 0.2 μm/h to 10 μm/h, depending on growth parameters and chamber geometry.

3. P-type/ intrinsic diamond/ metal (PIM) Schottky photodiode

Schottky photodiodes are based on synthetic SCD films produced in our laboratories by using MPECVD technique previously explained.
The nominally intrinsic diamond is deposited by using a completely separated apparatus in order to avoid any boron contamination. In fact, the presence of impurities in the intrinsic SCD active layer would determine a drastic worsening of the resulting performances, in terms of increasing of the dark current, temporal instability, memory effects, priming, slow response times, worsening of the spectroscopic performances, etc. All CVD films were principally characterized by X-ray diffraction, cathodoluminescence and Scanning Electron Microscopy, confirming the single crystal homoepitaxial deposition and the good crystal quality of the grown samples.
The single crystal p-type diamond was grown keeping constant the microwave power, while the intrinsic layer was grown keeping constant the temperature. The typical growth conditions are reported in Table 1.

Typical growth parameters	**Intrinsic diamond**	**p-type diamond**
Substrate	(100) HPHT type Ib	
Plasma composition	CH_4 - H_2	CH_4-H_2- B_2H_6
Gas flow rate	*1-100 sccm*	*2-100-10÷30 sccm*
Microwave power	*500÷600W*	*~500W*
Temperature	*720 °C*	*450 °C --->650°C*
Pressure	*120 mbar*	*150 mbar*
Thickness rate	*1μm/h*	*2μm/h*

Table 1. Typical growth parameters

The purity of CH4 and H2 gases were 99.9995% and 99.9999% respectively. Boron doping was performed by adding dyborane-hydrogen gas mix (100 ppm B_2H_6 in hydrogen) to the source gases. In the following section, the fabrication process of both photodiodes with different structures is reported.

3.1 Transverse configuration

The diamond photodiode in transverse configuration consist of a multilayered structure obtained by a two step deposition process. A conductive boron doped diamond homoepitaxial layer (see Fig.2 (b)), used as a backing contact, is deposited, at first, by Microwave Plasma Enhanced Chemical Vapour Deposition (MWPECVD) on a commercial low-cost synthetic High Temperature High Pressure (HPHT) <100> type Ib single crystal diamond (SCD) substrate, 4 × 4 × 0.5 mm^3 in size and approximately 400 μm thick. The boron concentration was estimated by fitting the Resistivity - Temperature curves, obtaining values in the range 10^{20} cm^{-3}. After that a nominally intrinsic diamond layer, which operates as the detecting region, is homoepitaxially grown on the doped one (see Fig.2 (c)). Its thickness can vary from a few microns up to more than 200 microns. Due to the small penetration depth of the UV radiation in the 10–200 nm spectral range (D. Palik., 1991), the detecting region of diamond has a thickness of approximately 2 μm. In the case of the soft X-ray detection, a higher thickness can be chosen (about 30 μm). The nominally intrinsic diamond is deposited by using a completely separated apparatus in order to avoid any boron contamination.

After the growth, each diamond layer has been oxidized by isothermal annealing at 500 °C, for 1h in air, in order to remove the H_2 surface conductive layer. Finally, a circular metal electrode, about 3 mm diameter is deposited on the diamond surface by thermal evaporation and annealed silver paint was utilized in order to provide an almost ohmic contact to the B–doped layer (see Fig.2 (d)).

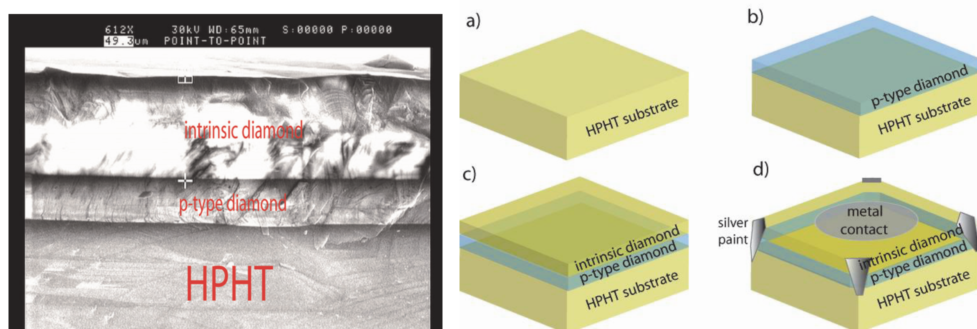


Fig. 2. Diamond deposition process of the PIM detector

Such a structure acts as a p-type/ intrinsic/ metal (PIM) Schottky Barrier Diode: the metal creates a Schottky junction with the intrinsic diamond, which acts as a drift-layer, and the SCD p-type layer, holes injector, determines the unipolarity of the device under direct polarization. The device, therefore, operates in typically reverse biased mode, with a negative voltage on the boron doped contact and the metal top contact grounded.

3.2 Planar configuration

In this section, it is report on the fabrication process of extreme ultraviolet (EUV) photovoltaic single crystal diamond Schottky diodes based on metal / intrinsic / p-type diamond junction developed at the University of Rome "Tor Vergata" operates in planar configuration with an interdigitated contact structure on the growth surface of the intrinsic

diamond layer. One set of fingers is made of aluminium and the second one is made of p⁺-type diamond. Two steps of a standard photolithographic technique are used for the fabrication process (see Fig.3) for the diamond detector with interdigitated finger electrodes (IDT-PIM in the following). First, an intrinsic diamond layer is homoepitaxially grown by MWPECVD on a commercial HPHT single crystal diamond substrate. As previously mentioned, annealing in air is employed in order to remove the surface conductive layer of the as-grown diamond film.

After the annealing process, p-type diamond interdigitated fingers are selectively grown on the top of the intrinsic diamond layer by using a patterned Cr plasma-resistant coplanar mask (Fig.3 (a)-(d)). After removal of the chromium mask by wet etching, the interdigitated Al electrode is fabricated by using a second mask which is aligned to the pattern previously obtained.

The Al fingers are patterned by standard lift-off photo-lithographic technique and by thermal evaporation on the CVD intrinsic diamond surface (Fig.3 (e)-(g)). The width and the gap between two fingers are both 20µm. A SEM image of the IDT-PIM device in Fig.3. The IDT-PIM is simply mounted in an sample holder for UV measurements. In this case, the measured photocurrent of IDT-PIM detector can contain both photoemission current and photoconductive current. The photoemission contribution contains electron emission arising from Al fingers and from p-type diamond exposed to the UV irradiation.

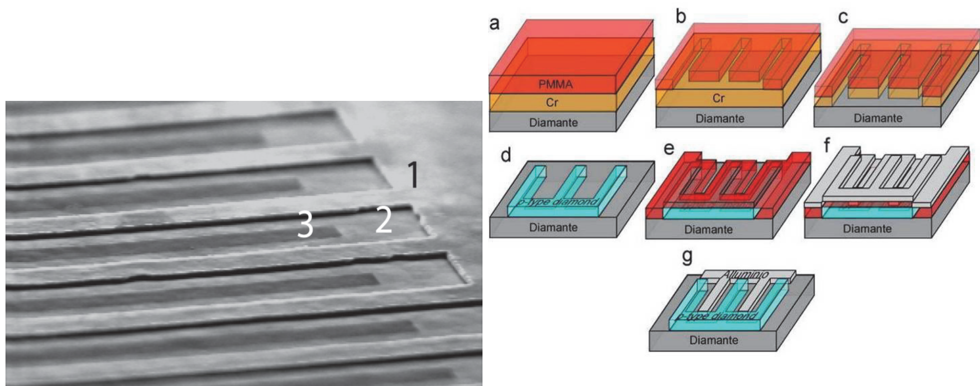


Fig. 3. Fabrication process of the IDT-PIM photodetector and SEM image of the IDT-PIM device: (1) p-type diamond, (2) intrinsic diamond and (3) Al contact.

4. Electrical characterization

The investigation of the physical principle of operation of such photodiode through the study of their electrical detection properties is reported in this paragraph. The physics of the devices is clearly based on the existence of a Schottky barrier, as demonstrated by the fact that the detectors are able to operate even with no external bias applied. The aim of the electrical characterization of the devices is, therefore, to analyze the Schottky barrier parameters, measured using the I-V curves from which the barrier height can be extracted. Moreover, the study of metal/p-type diamond contacts to obtain a good ohmic contact is also reported.

4.1 Ohmic contact on p-type diamond

High quality ohmic contacts are one of the moswt frequently encountered problems in the development of the right material for electronic devices applications. The properties of such electrical contacts directly contribute to the active devices performance. In general the requirements for ohmic contacts can be summarized as follows (M. Werner, 2003):

- low contact resistivity
- good adhesion
- high thermal stability
- high corrosion resistance
- bondable top-layer
- about zero voltage drop

As the low doped diamond films exhibited semiconducting behaviour, it was necessary to study the metal/diamond contacts to verify if they obey Ohm's law. Most contacts to common semiconductors are depletion contacts which result mainly from the action of surface states. They can show, however, an ohmic behaviour with a linear current–voltage characteristic on degenerated doped semiconductors. In the case of a depletion contact, the contact resistivity varies exponentially with the Schottky barrier height. The ohmic behaviour of the depletion contact can be achieved either when the barrier height is small, so that the charge carriers can easily overcome the barrier (thermionic emission) or when the charge carriers are able to surmount the depletion region by quantum-mechanical tunnelling.

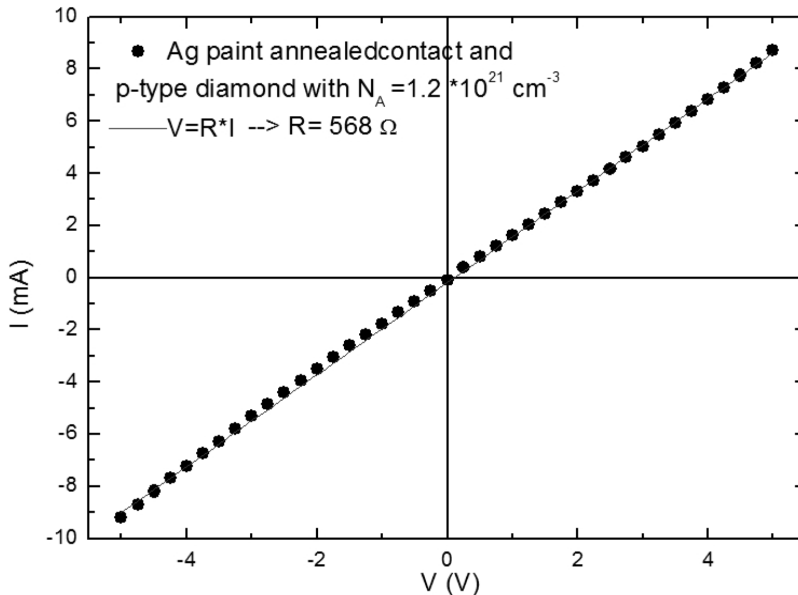


Fig. 4. I-V characteristic of annealed silver paint/p^+-type diamond junction

The depletion layer width of a metal/semiconductor contact is proportional to the square root of the reciprocal doping concentration ($W \sim N_A^{-1/2}$) (S.M. Sze, 1981). Consequently, the

depletion layer width decreases with the increase in doping concentration and the tunnelling probability increases. So a good ohmic contacts are obtained by heavily doping the p-type diamond layer (doping levels much larger than $10^{20}cm^{-3}$). The resulting layer p^{+}, which is highly doped by B, was metalized by silver paint annealed at 500°C for 10 min. The I-V characteristic is reported in Fig.4 where is also reported the specific resistance calculated by ohm's law.

4.2 Shottky contact on intrinsic diamond layer

The electrical characterization of the metal/intrinsic diamond Schottky junction of the devices was performed at room temperature in a vacuum chamber with a background pressure of 10^{-4} mbar by measuring the current–voltage (I–V) characteristics by using a Keithley 6517A pico-ampere meter.

The I-V characteristic was obtained by applying a voltage to the metal contact while the p-type diamond layer is earthing. Fig.5 shows the typical I-V characteristic of the diamond Schottky photodiodes. When the p-type rectifyng contact is reverse biased by connecting the metal to positive terminal, holes are repelled from the interface and the bands are away bent down. The potential barrier for the holes is increase, as is the width of the depletion region. The resulting net current is very low (reverse biased). If instead the metal is connected to the negative terminal, then forward biasing results as the holes are attracted toward the metal interface (forward biased).

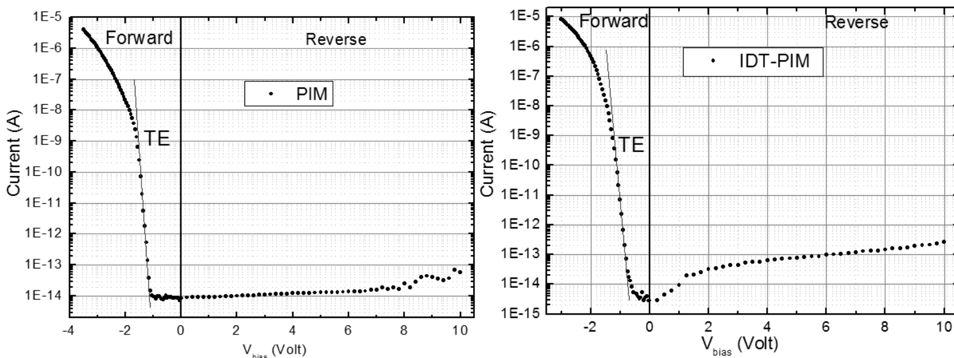


Fig. 5. Typical I-V characteristic of the PIM device

In Fig.5, it's clearly seen the different behaviour of reverse and forward current.

When a negative voltage (forward voltage) is applied on the metal electrode a hole current starts flowing from the p-type diamond, via the nominally intrinsic diamond region, towards the Schottky contact. The rectification behaviour of the both photodiodes is observed with a very high rectification ratio of about 10^{8} at ±3V. For values of $|V_B| < |V_{on}|$, where V_{on} is "turn- on voltage" that in figure is about - 1 V , the forward current is due to generation-recombination effects and leakage superficial current and it's similar at the reverse current. Increasing the forward bias, in the region between approximately -1 V and -1.6 V forward voltage (V_F), the current rises exponentially with V_F.

In this region the forward current density (J_F) is well described by the thermionic emission (TE) theory. The thermionic emission theory by Bethe is derived from the assumptions that

the barrier height is much larger than kT, thermal equilibrium is established at the plane that determines emission, and the existence of a net current flow does not affect this equilibrium. Bethe's criterion for the slope of the barrier is that the barrier must decrease by more than kT over a distance equal to the scattering length. The resulting current flow will depend only on the barrier height and not on the width, and the saturation current is not dependent on the applied bias. Then the current density of majority carriers from the semiconductor over the potential barrier into the metal is expressed as (M. Brezeanu et al., 2007):

$$J_F = J_S\left[\exp\left(\frac{qV_F}{nkT}\right)-1\right]$$
$$J_S = A^*T^2\exp\left(\frac{-q\phi_{BI}}{kT}\right) \tag{1}$$
$$A^* = \frac{m_p^*}{m_0}A^0$$

where n is the ideality factor (n≥1 and it informs the experimental I-V characteristic deviates from the behaviour SBD ideal (n = 1)), T the absolute temperature (Kelvin), k the Boltzmann's constant, J_S the saturation current density, A_0 the Richardson's constant ($120.173 Acm^{-2}K^{-2}$), A* the Richardson's effective constant, m0 and mp* electron mass and effective mass hole in diamond ($m_p^*=0.7\ m_0$) and Φ_{BI} the Schottky barrier height. From the exponential fit of the I-V characteristic, it is possible to estimate the saturation current density J_S and the ideality factor n. Substituting the values obtained from the fit in the following equation

$$\Phi_{BI} = \frac{K_BT}{q}\ln\left(\frac{A^*T^2}{J_S}\right) \tag{2}$$

it's possible estimate the Schottky barrier heigh. The values obtained for IDT-PIM and PIM photodiodes are 1.65 eV and 1.8 eV respectively.

5. Extreme UV characterization

The photodiodes have been tested over the extreme UV spectral region from 20 to 120 nm, using He and He Ne DC gas discharge as radiation sources and a toroidal grating vacuum monochromator (Jobin Yvon model LHT 30) with a 5Å wavelength resolution. The dimension of the optical aperture is $0.25 \times 6.00\ mm^2$; a manual shutter is used to switch on and off the UV radiation. The experimental apparatus of UV characterization is reported in the following picture.
The photoresponse measurements have been performed in a vacuum chamber, at a pressure of 0.03 mbar. By using a three (X-Y-Z) dimension mechanical stage powered by stepper motors, it is possible to locate the photodetector in front of the beam light and to compare its response with that of a calibrated NIST silicon photodiode (http://www.ird-inc.com) placed in the same position, which measures the absolute photon flux. A raster scansion of the beam light was performed on the detector surface so to position the photodetectors where their response has a maximum (see Fig.6(b)).

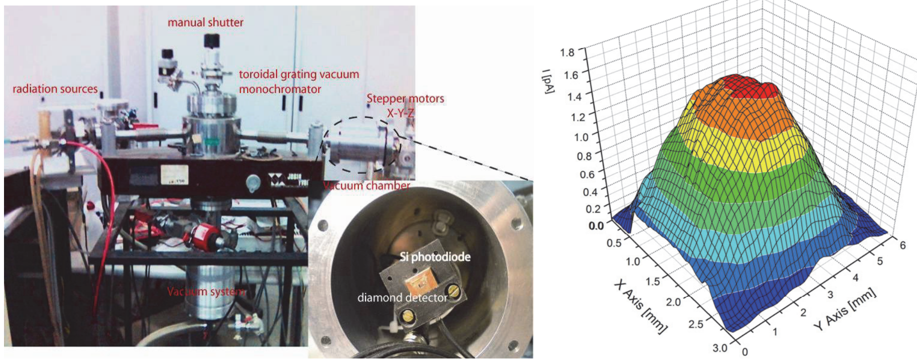


Fig. 6. a) Extreme UV characterization system, b) Raster scansion of the beam light

A hole, 2 mm in diameter, is used to collimate the radiation on the sensitive area of the detectors and to obtain the same illuminated area on the silicon photodiode. The photocurrent is measured by an electrometer (Keithley 6517A), using the internal voltage source.
Because of different geometry adopted by the two devices, they are measured differently. The PIM detector is encapsulated in a copper/vetronite shielded housing with a 2 mm pinhole. In such housing, the Al contact is grounded and the photocurrent is measured from p-type diamond so that the signal is not affected by the eventual presence of secondary electron emission current from the illuminated contact.

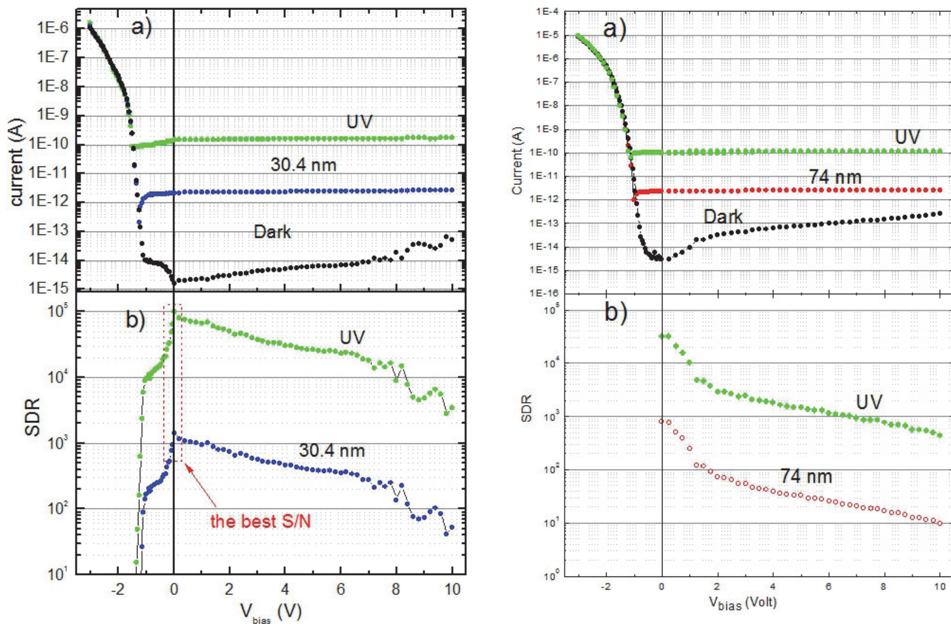


Fig. 7. a) I-V characteristic in dark and in light of PIM detectors and b) signal to dark current ratio (SDR)

The IDT-PIM is simply mounted in an sample holder for UV measurements with the same 2 mm pinhole. In this case, the measured photocurrent of IDT-PIM detector can contain both photoemission current and photoconductive current. The photoemission contribution contains electron emission arising from Al fingers and from p-type diamond exposed to the UV irradiation. The typical current – voltage (I-V) characteristics in dark current and under irradiation have been measured at room temperature of two detectors are shown in Fig. 7.
The devices operate in the reverse bias mode because when operating in the forward bias mode, the photocurrent is masked by the dark current. The dark current is very low (<0.1 pA) below about + 10 V, as expected for a metal/diamond rectifying contact.
The photocurrent vs. applied voltage is also reported in the same figure when the device is exposed to UV radiation and 30.4 nm (He lines) and 73 nm (Ne line). The device shows a photocurrent response even at zero voltage bias, exploiting the internal junction electric field. The photocurrent is almost constant with increasing positive voltage, while the dark current increases by about two orders of magnitude. Remarkably, thus, the best signal-to-dark current (SDR) ratio (see Fig.7 (b)) is obtained at zero bias voltage, so that in the following, the devices have been operated with no external bias voltage applied.

5.1 Temporal response

Temporal response measurements upon exposure to UV radiation have been performed according to the following procedure: at first the dark current value was recorded for several seconds keeping the light shutter closed, until the steady state value had been reached; then the shutter was opened and the photocurrent was measured. Finally, the shutter was closed again until the dark current reached the initial value, before starting a new measurement run. The detectors time response, upon exposure to UV radiation, have been measured by opening and closing a manual shutter during the acquisition. The temporal response of the tested devices is reported in Fig. 8 (a) under UV illumination of the He-Ne DC gas discharge radiation source.

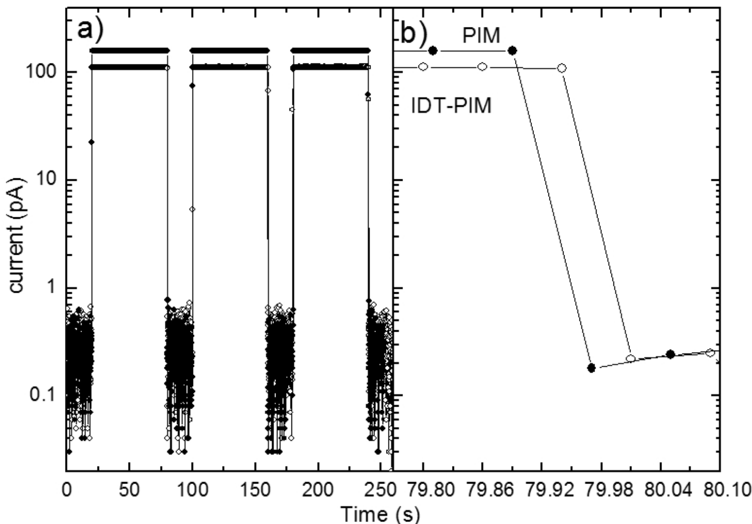


Fig. 8. a) Temporal responses under illumination of He-Ne DC gas discharge radiation source. b) The magnification of fall time of the both devices.

The response is reproducible and no undesired effects such as persistent photocurrent and priming or memory effects, which are often observed in diamond UV detectors (C. E. Nebel et al., 2000, A. De Sio et al., 2005, M. Liao et al., 2008), are observed. negligible. However, it is obtained only after the very first irradiation: the device, just mounted, reaches the described performance only after a pre-irradiation time of few minutes Fig.8 (b) shows rise and fall times of the signal of about 60 ms, which corresponds to the acquisition rate of the used electronic chain.

5.2 Linearity

A useful detector is expected to exhibit linear response with photon flux, i.e. a constant responsivity up to a saturation point where space charge effects prevail and no more electron-hole pairs can be collected under illumination. The calibration of linear detectors and related electronics is much simpler. The linearity of the photodetectors have been investigated varying the current intensity of the plasma. The photocurrent (I_{ph}) measured vs. the incident optical power (P), under irradiation of He-Ne gas discharge radiation is shown in Fig.9. We used a power law: $I_{ph} = A + B \bullet P^C$ to fit the data. Here A is the offset corresponding to the dark current, B is the photosensitivity (provided C=1) and C is a linearity coefficient. The graph shows the measured data as well as the fitting function. In this spectral region, both photodetectors shows remarkably good linearity, C being 1 in all cases within the error.

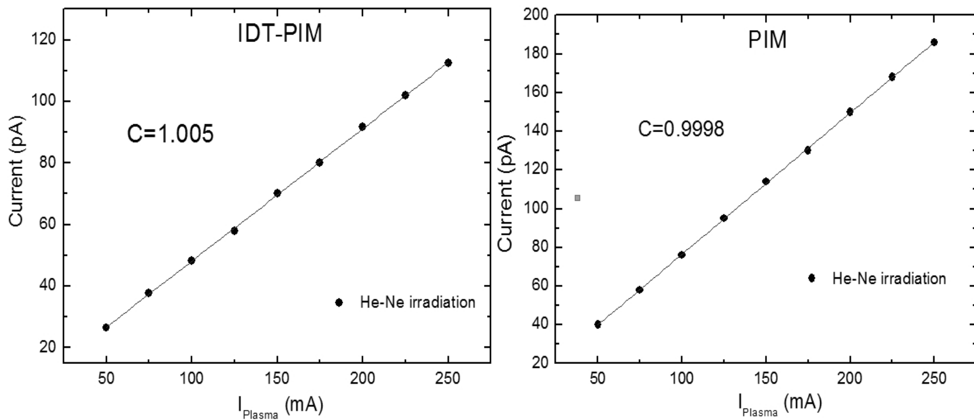


Fig. 9. Linearity of IDT-PIM and PIM photodiodes.

5.3 Extreme UV spectroscopy

The normalized emission spectra of a DC discharge He and He-Ne lamp measured in unbiased mode by the detectors are reported in Fig.10.

All spectral lines are clearly resolved and observed with a good signal to noise ratio, demonstrating the high photodetection capabilities of the CVD single crystal diamond grown in the extreme UV spectral region. All spectral lines are classified by the NIST Atomic Spectra Database Lines Form from the following website:

http://physics.nist.gov/PhysRefData/ASD/lines_form.html.

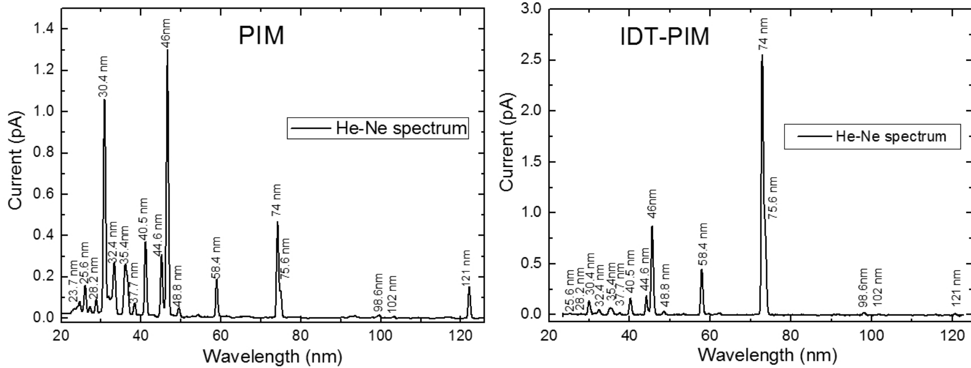


Fig. 10. He-Ne emission spectrum measured by the two devices.

In particular, the low intensity lines of the He-Ne spectrum in the wavelength range 20-30 nm are easily resolved by PIM detector

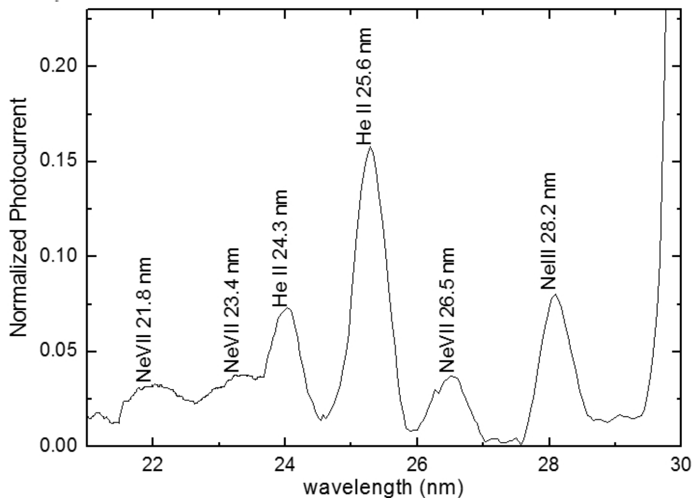


Fig. 11. He-Ne spectrum measured by PIM detector in the range 20-30 nm

5.4 Responsivity and external quantum efficiency

The absolute spectral response of the PIM detectors is measured by comparison with a calibrated photodiode exposed to the same source on the same optical area of about 1 mm2. The spectral responsivity, expressed in amperes per watt (A/W), is defined as the photocurrent per unit incident optical power and can be evaluated from the relationship $R_d = R_{Si}\ I_d\ /I_{Si}$ where R_{Si} is the responsivity of the calibrated silicon photodiode at a given wavelength, I_{Si} and I_d are the photocurrents measured by the silicon photodiode and the diamond detector, respectively.

The responsivities of both photodiodes are reported in Fig.12. The responsivity of the PIM device decreases monotonically as the wavelength increases until about 80 nm while at 120

nm an increased value is observed. At 98 nm the signal is below the noise level so that only an upper limit can be provided. However, the presence of a minimum in the responsivity around 100 nm can be clearly deduced from Fig.12.

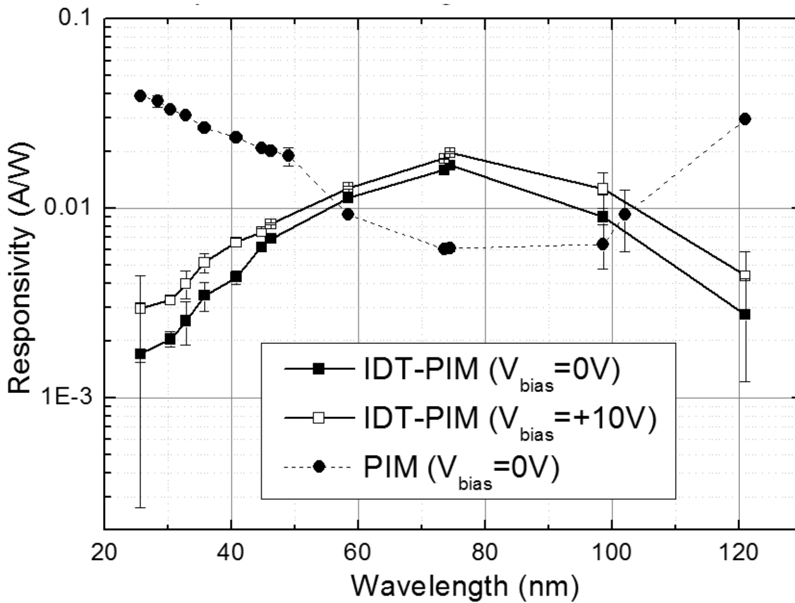


Fig. 12. Responsivity of the both devices.

The responsivity of the IDT-PIM detector is much lower than that of the PIM detector at short wavelength (below 50 nm) showing a maximum at about 73 nm. The increased sensitivity of the IDT-PIM device at intermediate wavelength could be probably ascribed to the contribution of photoemission current as already reported in the literature (T. Saito et al, 2006). For both the devices the absolute responsivity measured at around 50 nm is comparable to the best results reported in the literature for diamond based EUV detectors (A. BenMoussa et al., 2006).

The External Quantum Efficiency (EQE) spectrum, estimated by: EQE = 1240•Rd / λ[nm], is reported in Fig.7 for the PIM devices.

As mentioned above, the photocurrent measured by IDT-PIM detector includes the contains both photoconductive current and photoemission current, arising from secondary electron escape from Al fingers, which also depends on the wavelength (J. Ristein et al, 2005, W. Pong et al., 1970). On the contrary, in the encapsulated PIM device the illuminated contact is grounded and the current flowing from the boron doped layer is not affected by secondary electrons contribution. Moreover, the more homogeneous electric field configuration of the PIM device allows a simple analysis of the detection process.

In order to investigate the effect of the metallic Schottky contact upon the detection performance of the PIM devices, different semitransparent metals (thickness < 10nm) have been thermally evaporated on the oxidized surface of single crystal CVD intrinsic diamond layers.

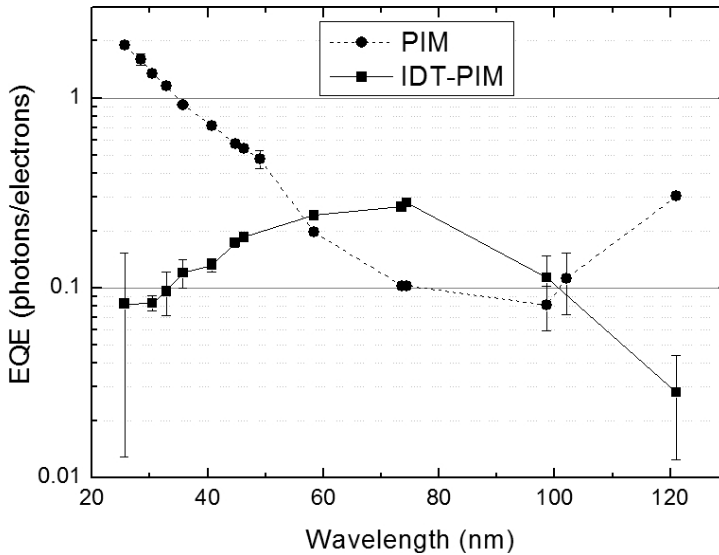


Fig. 13. External quantum efficiency EQE of the two photodiodes between 20 and 120 nm.

The absolute spectra responsivity curves versus different meal contacts of the devices are shown in Fig.14. All the devices have a maximum of the responsivity at lower wavelengths and a sharp cutting edge for longer wavelengths while at around 120 nm an increased value is observed. The lowest responsivity, between 50 ÷ 100 nm, has been measured for the device having Cr as an electrode. The device having Ag and Pt contacts shows rather similar trend of the responsivity, whereas Al contact shows the best results in the UV performances.

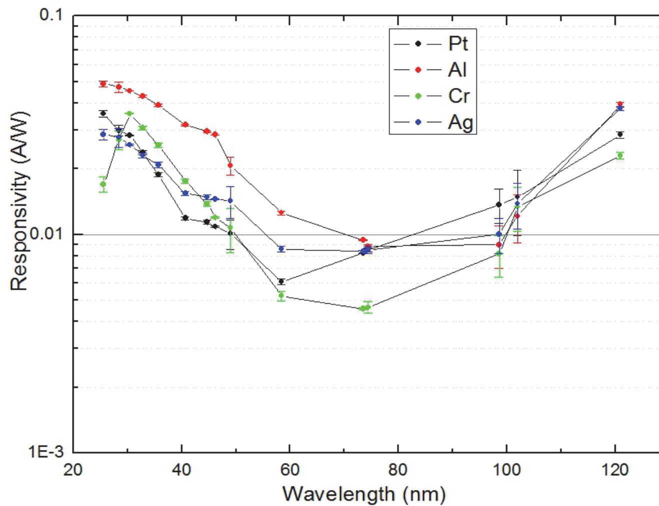


Fig. 14. External quantum efficiency EQE of the PIM devices between 20 and 120 nm as a function of the type of the metallic contact.

5.5 UV/visible rejection ratio

The photoconductive response was tested over a wide spectral range, extending from the extreme UV (EUV) up to the visible. The 210–500 nm range was investigated using an Optical Parametric Oscillator (OPO) 5 ns pulsed laser (Opolette laser by Opotek). The laser beam was scattered by an optical diffuser in order to prevent signal saturation of the electronic chain and the diamond detector was placed 10 cm away from the diffuser. A 500 MHz Le Croy WaveRunner 6050 digital oscilloscope was used to acquire the output signal.

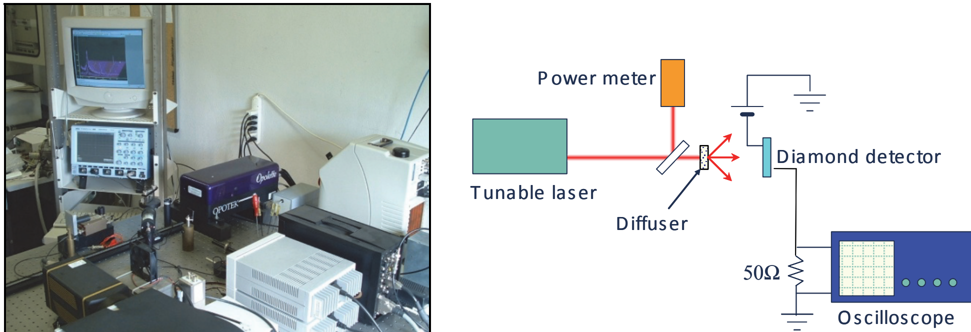


Fig. 15. Optical Parametric Oscillator and experimental set up.

Two different connection configurations were used:

i. Direct recording of the detector output by the digital oscilloscope
ii. Integrated measurement by an Ortec142A charge preamplifier.

The signal provided by a pyroelectric power meter was used to normalize the diamond detector output, in order to take into account the wavelength dependence of laser pulse amplitude and the intrinsic fluctuations of the beam intensity.

The visible-blind properties of the photodetectors were tested by measuring the photoresponse at different wavelengths in the 210–500 nm range.

In Fig.16 (a) the device responsivity of the PIM detector is reported as a function of the incident laser radiation wavelength, normalized to the pyroelectric power meter signal. A 3 orders of magnitude variation was measured when moving across the band gap wavelength of 225 nm. Such a drop increases up to 5 orders of magnitude when the UV to visible rejection ratio is considered. It should be stressed that a very stable and reproducible response was observed in the whole energy range and irradiation memory or pumping effects were not observed.

In addition, a linear increase in the photoresponse as a function of calculated radiation intensity was observed measuring the output signal at decreasing device distances from the optical diffuser.

The time response at 220 nm of the investigated PIM detector is reported in Fig. 16 (b). As clearly seen in the Fig.16 (b), the device response to a laser pulse at 220 nm, measured through a bias Tiee and recording by the digital oscilloscope (Le Croy 500MHz), shows an exponential decay time constant of about 100ns. The reason of this trend of output response is due to electrical circuit of the device. In fact, an RC circuit, the value of the time constant is equal to the product of the circuit resistance and the circuit capacitance. Therefore, taking into account the depletion capacitance measured by C-V curves of about 100pF and the resistance of p-type diamond film ~1kΩ, the time constant result to be $\tau = 100$ns.

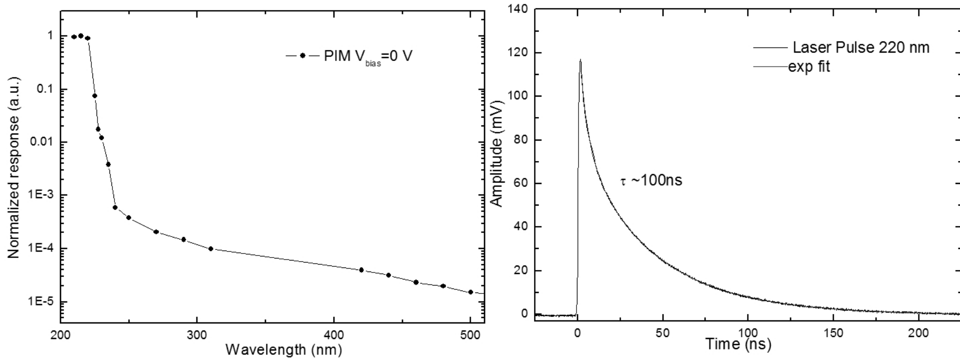


Fig. 16. a)Normalized responsivity of PIM device as a function of the incident laser radiation wavelength. b) The device response to laser pulses directly obtained by the digital oscilloscope.

The visible-blind properties of the IDT-PIM device were also tested by measuring the photoresponse at different wavelengths in the 210–500 nm range. In this region, the spectral response shows a visible/UV rejection ratio of about 4/5 orders of magnitude, as clearly seen in Fig.17(a) .Moreover, the time response at 220 nm of the investigated detector is reported in Fig.17 (b). The Fig.17(b) shows the device response to a laser pulse at 220 nm, which have a full width at half maximum (FWHM) of about 25 ns, and the time response is faster than that of PIM detector. In fact, in this case, the parallel capacitance of the photodiode is very low, about 15pF.
Interdigitated structure, therefore, can be optimized in order to build a ultrafast XUV detector, for time resolution.

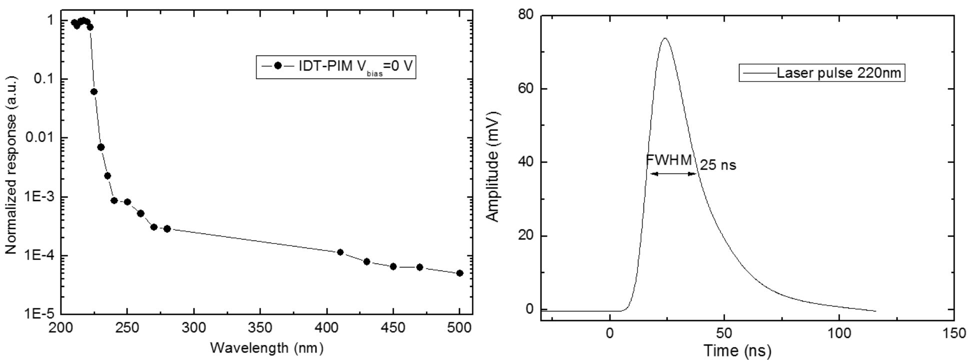


Fig. 17. a) Normalized responsivity of IDT- PIM device as a function of the incident laser radiation wavelength. b) The device response to laser pulses directly obtained by the digital oscilloscope.

6. Conclusion

Two detectors were fabricated at the University of Rome "Tor Vergata" with a structure that acts as a metal/intrinsic/p-doped diamond photovoltaic Schottky diode. The two detectors

operate in different configurations: one in transverse geometry and the other one in planar configuration.
We have measured the electrical characteristics and tested the performance under continuous vacuum UV photon irradiation of the two devices. A general result of our experiments is that diamond detectors are very sensitive devices showing show a very low dark current and very good signal-to-noise ratio. The responses are reproducible and undesired effects such as persistent photocurrent, priming or memory effects are negligible for both devices. The response time could is very fast and it is much lower than the acquisition rate of the used electronic chain (~ 60 ms). These results indicate the high quality of our CVD diamond grown for UV applications.
The responsivity and the EQE of the two devices show an opposite behaviour as a function of the radiation wavelengths due to the different operative configurations. In particular the PIM detector is more efficient at lower wavelengths and present a drop of sensitivity at approximately 100 nm. The IDT-PIM is less efficient at low wavelength and has a maximum efficiency at about 74 nm.
The visible-blind properties of the photodetector were also tested by measuring the photoresponse at different wavelengths in the 210–500 nm range. A 3/4 orders of magnitude variation was measured by diamond based detectors when moving across the band gap wavelength of 225 nm. Moreover, the spectral response shows a visible/UV rejection ratio of about 5 orders of magnitude for both photodiodes. Finally, the device response to laser pulses at 220 nm is different in two cases due to the different electrical circuit of the two devices. In particular, the time response of IDT-PIM detector is faster than that of PIM detector.

7. Acknowledgment

The devices studied in this chapter were developed by the group of Rome University "Tor Vergata" composed by Prof. Marco Marinelli, Prof. Enrico Milani, Dr Gianluca Verona-Rinati, Dr Giuseppe Prestopino and myself. They have made possible the writing of this chapter.
Moreover, I'd like to thank the staff of "O. M. Corbino" Institute of Acoustics (IDAC) of CNR who give me the possible to perform the photolithography techniques used to realize the devices developed in this chapter.

8. References

J. E. Field, *Properties of Diamond* , Academic Press, London, (1979).
J.Prins, *Applications of diamond films in electronics in "The Physics of Diamond"*, A. Paoletti and A. Tucciarone (editors), IOS Press, Amsterdam, (1997).
J.F. Hochedez, J. Alvarez, F.D. Auret, P. Bergonzo, M.-C. Castex, A. Deneuville, J.M. Defise, B. Fleck, P. Gibart, S.A. Goodman, O. Hainaut, J.-P. Kleider, P. Lemaire, J. Manca, E. Monroy, E. Munoz, P. Muret, M. Nesladek, F. Omnes, E. Pace, J.L. Pau, V. Ralchenko, J. Roggen, U. Schuhle, C. Van Hoof, *Diamond Relat. Mater*. 11 (2002), 427.
A. Balducci, M. Marinelli, E. Milani, M.E. Morgada, A. Tucciarone, G. Verona-Rinati, M. Angelone, M. Pillon, *Appl. Phys. Lett*. 86 (2005), 193509.
T. Teraji, S. Yoshizaki, H. Wada, M. Hamada, T. Ito , *Diamond Relat. Mater*. 13 (2004), 858.

V.I. Polyakov, A.I. Rukovishnikov, N.M. Rossukanyi, A.I. Krikunov, V.G. Ralchenko, A.A. Smolin, V.I. Konov, V.P. Varnin, I.G. Teremetskaya, *Diamond Relat. Mater.* 7 (1998), 821.

L. Thaiyotin, E. Ratanaudompisut, T. Phetchakul, S. Cheirsirikul, S. Supadech, *Diamond Relat Mater* 11 (2002), 442.

R. D. McKeag and R. B. Jackman, *Diamond Relat. Mater.* 7 (1998), 513.

L. Barberini, S. Cadeddu, and M. Caria, *Nucl. Instrum. Methods* 460 (2001), 127.

E. Pace, A. Vinattieri, A. Pini, F. Bogani, M. Santoro, G. Messina, S. Santangelo, Y. Sato, *Phys. Status Solidi, A Appl. Res.* 181 (2000), 91.

S. Almaviva, Marco Marinelli, E. Milani, G. Prestopino, A. Tucciarone, C. Verona, G. Verona-Rinati, M. Angelone, M. Pillon, *Diamond Relat. Mater.* 18 (2009), 101.

S. Almaviva, Marco Marinelli, E. Milani, G. Prestopino, A. Tucciarone, C. Verona, G. Verona-Rinati, M. Angelone, M. Pillon, I. Dolbnya, K. Sawhney and N. Tartoni, *J. Appl. Phys.* 107 014511 (2010).

S. Almaviva, Marco Marinelli, E. Milani, G. Prestopino, A. Tucciarone, C. Verona, G. Verona-Rinati, M. Angelone, M. Pillon *Diamond and Related Materials*, v 19, n 1, p 78-82, January 2010.

D. Palik., *Handbook of Optical Constants of Solids II*, Academic Press, New York (1991).

M. Werner, *Semicond. Sci. Technol.* 18 (2003) S41-S46.

S.M. Sze, *Physics of Semiconductor Devices*, John Wiley and Sons (WIE) (1981)

M. Brezeanu, T. Butler, N. Rupesinghe, S. J. Rashid, M. Avram, G. A. J Amaratunga, F. Udrea, M. Dixon, D. Twitchen, A. Garraway, D. Chamund, and P. Taylor, *IEEE Proc.: Circuits Devices Syst.* 1, 380 (2007).

website: http://www.ird-inc.com.

C. E. Nebel, A. Waltenspiel, M. Stutzmann, M. Paul, and L. Schäfer, *Diamond Relat. Mater.* 9, (2000), 404.

A. De Sio, E. Pace, *Nucl. Instr. Methods A* 552 (2005), 203

M. Liao, Y. Koide, J. Alvarez, M. Imura, J.P., *Physical Review B* 78 (2008), 045112.

T. Saito, K. Hayashi, H. Ishihara and I. Saito, *Metrologia* 43, (2006), S51.

A. BenMoussa, A. Theissen, F. Scholze, J.F. Hochedez, U. Schuhle, W. Schmutz, K. Haenen, Y. Stockman, A. Soltani, D. McMullin, R.E. Vest, U. Kroth, C. Laubis, M. Richter, V. Mortet, S. Gissot, V. Delouille, M. Dominique, S. Koller, J.P. Halain, Z. Remes, R. Petersen, M. D'Olieslaeger, J.M. Defise , *Nucl. Instr. Methods A* 568 (2006), 398.

J. Ristein, W. Stein, L. Ley, *Diamond and Relat. Mat.* 7, 626 (1998).

W. Pong, R. Sumida, G. Moore, *J. Appl. Phys* 41, 1869 (1970).

14

GaN Based Ultraviolet Photodetectors

D. G. Zhao and D. S. Jiang
State Key Laboratory on Integrated Optoelectronics,
Institute of Semiconductors, Chinese Academy
of Sciences, Beijing
P. R. China

1. Introduction

The band gaps of the III–nitrides are large and direct, which are 0.7 eV for InN, 3.4 eV for GaN, and 6.2 eV for AlN (Jain et al., 2000; Wu et al., 2002), consequently There are many areas where conventional semiconductors including Si and GaAs cannot be used. For example, short wavelength light emitters are required for full color display, high density information storage, and under water communication. However, III–nitrides are particularly suitable for applications in these areas. A great success in device applications has been obtained so far, including the realization of high brightness blue GaN based light emitting diodes (LED) and long lifetime GaN based laser diodes (LD) (Nakamura, 1998).
On the other hand, III-nitrides are one of the most promising materials for the fabrication of high-sensitivity visible-blind ($\lambda \leq 365$nm) and solar-blind ($\lambda \leq 280$nm) ultraviolet (UV) photodetectors, which have extensive applications in flame detection, secure space-to-space communication, and ozone layer monitoring. Various types of GaN-based photodetectors have been realized, including p-i-n and Schottky barrier photodetectors, solar-blind ultraviolet photodetector focal plane arrays, and UV avalanche photodiodes (McClintock et al., 2005; Zhao et al. 2007a; Cicek et al., 2010). The fabrication of GaN-based photodetectors were reviewed in some articles previously(Muñoz et al.,2001). Since the quality of GaN materials plays a key role in determining the performance of GaN UV photodetectors, in this chapter, firstly the growth and properties of GaN materials are introduced, then the device technology and fabrication are presented, finally a conclusion is drawn.

2. GaN material growth and ultraviolet photodetector's fabrication

The GaN-based materials used for device applications investigated in this chapter are grown on the c-plane sapphire substrate by metalorganic chemical vapor deposition (MOCVD). The ammonia (NH_3), trimethylgallium (TMGa), trimethylaluminum (TMAl) and Silane (SiH_4) have been used as N, Ga, Al, and Si precursors, respectively. H_2 has been used as the carrier gas. The quality of the thin films is mainly characterized by the double x-ray diffraction (DCXRD) and photoluminescence (PL). The full width at half maximum (FWHM) of DCXRD ω-scan rocking curves is obtained using a Rigaku SLX-1AL x-ray diffractometer. A 325 nm He–Cd laser is employed as excitation light in the measurement of the PL spectra.

2.1 Material growth for GaN based UV photodetectors

Many attempts were made to synthesize GaN crystals during the period 1930–1960, but good quality crystals could not be grown for a long period of time, mainly because there are large lattice mismatch and thermal mismatch between the III–nitrides and widely-used heteroepitaxial substrates (Jain et al., 2000). The beginning of the growth of good quality epilayers was made by the two-step method (Amano et al., 1986; Nakamura, 1991). In the following, we will firstly introduce the MOCVD growth of GaN epilayers, then the AlGaN growth will be discussed with the parasitic reaction between the TMAl and NH_3 in MOCVD, finally the defects and related yellow luminescence of GaN films are studied.

2.1.1 GaN material growth using the two-step method

The two-step growth method of GaN epilayer with low-temperature AlN buffer layers by MOCVD is studied. (Zhao et al., 2004). The growth procedure is as follows: Firstly the AlN buffer layer is grown on sapphire substrate at 600℃ and annealed in a temperature ramp, then a GaN epilayer about 2.5 μm thick is deposited on the AlN buffer layer at 1080℃. The real-time *in situ* optical reflectivity measurements are employed to monitor the whole growth stages of GaN materials.

Fig. 1 shows the traces of *in situ* optical reflectivity measured from the two GaN epilayers samples A and B grown on a 20 nm thick AlN buffer layer with different annealing processes during the temperature elevation after the growth of low-temperature AlN buffer layer. The annealing time of AlN buffer layer used in the growth of two samples A and B are 1000 second and 300 second, respectively. In Fig. 1(a) and (b), two traces are divided into three parts corresponding to three growth stages of GaN deposit on low-temperature AlN buffer layer as follows: (i) the low-temperature AlN buffer layer deposition, (ii) temperature ramp and anneal of the AlN buffer layer, (iii) the growth of GaN epilayers. The differences in the surface evolution processes during the growth of samples A and B are observed. In the initial growth stage of sample A where GaN epilayer is deposited on AlN buffer layer with a 1000 second annealing time, the surface of GaN layer becomes rough and the intensity of the *in situ* optical reflectivity decreases, then the surface of GaN layer turns to be optically smoother step by step, it means the lateral growth and coalescence of GaN islands emerge (Han et al., 1997), at last the quasi two-dimensional growth of GaN layer occurs. An oscillation of the reflectivity intensity with large and equal amplitude is well observed. However, the growth procedures of sample B deposited on AlN buffer layer with a 300 second annealing time shows a different kind of trace in Fig. 1(b). The surface roughing of GaN islands does not clearly appear. There is nearly no change in the intensity of *in situ* optical reflectivity during the starting period of the growth of GaN epilayer, as shown by the arrow in Fig. 1(b). As shown in Table 1, sample A has a narrower FWHM of x-ray rocking curve and a higher electron mobility, it seems that the longer annealing time of low-temperature AlN buffer layer tends to promote a lateral growth of GaN islands, and the quality of GaN epilayers is improved. It also suggests that the lateral growth of GaN islands is helpful to decrease the edge threading dislocations, since the FWHM of x-ray ω-scan rocking curve for (0002) and (10-12) planes represents indirectly the density of screw and edge threading dislocations (Heying et al., 1996; Heinke et al.,2000).

Samples	AlN Buffer layer		XRD FWHM (arcmin)		Electron Mobility (cm^2/Vs)
	Annealing Time (s)	Thickness (nm)	(0002)	(10-12)	
A，E	1000	20	6.9	11.2	360
B	300	20	8.1	19.1	142
C	1000	45	10.2	28.6	73
D	1000	30	7.2	13.9	217
F	1000	16	…	…	…

Table 1. Growth condition and characterization result of GaN samples.

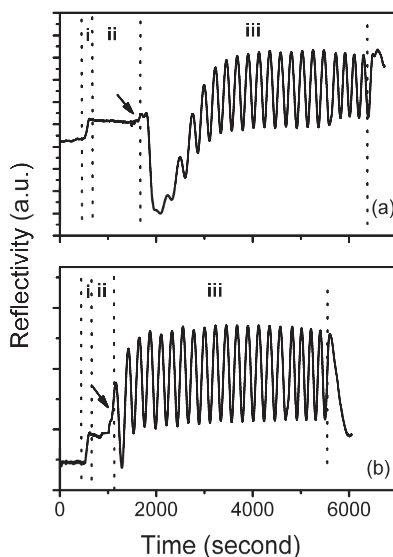


Fig. 1. The traces of *in situ* optical reflectivity measurements for the three stages in the whole growth process of GaN epilayers on low-temperature AlN buffer layer with different annealing time: (a) 1000 s (b) 300 s.

It is found that not only the annealing time, but also the thickness of low-temperature AlN buffer layer has an enormous influence on the quality of GaN epilayers. The traces of *in situ* optical reflectivity measured from GaN epilayers growth on low-temperature AlN buffer layer with different thickness are shown in Fig. 2(a)-(d), where the dashed lines denote the start of GaN epilayers growth. For the four samples, the same 1000 second annealing time of low-temperature-grown AlN is employed but the thickness of low-temperature AlN buffer layer is different. They are 45nm, 30nm, 20nm, and 16nm for sample C, D, E (where sample E and sample A is the same sample with different names) and F, respectively. It can be seen from Fig. 2 that there exist a lot of differences in the reflectivity curves measured during the initial stage of the growth process of GaN epilayers. Nearly no any growth of GaN islands (surface roughing process in the initial growth stage) and their coalescence (lateral growth) is observed in the starting period of growth process of sample C which is deposited on the 45 nm AlN buffer layer. There is a little growth of GaN islands and their coalescence in the

growth process of sample D which is deposited on the 30 nm AlN buffer layer. With thinner AlN buffer layer, however, the obvious lateral growth of GaN islands is observed in the growth process for sample E, as shown in Fig. 2 (c). The growth process of sample F is stopped as shown in Fig. 2(d), indicating that the AlN buffer layer is too thin to lead to the coalescence of GaN islands and to start a quasi two-dimensional growth. In this case the quality of GaN epilayers becomes very bad. As shown in Table 1, the optimal thickness of AlN buffer layer in this growth condition is around 20nm. A too thick or too thin AlN buffer layer will lead to the deteriorated quality of GaN epilayers.

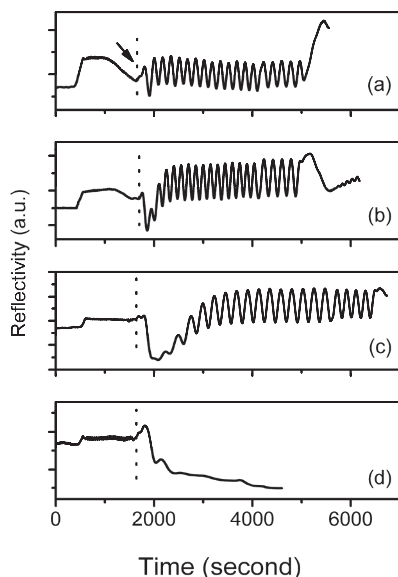


Fig. 2. The traces of in situ optical reflectivity measurements for the whole growth process of GaN epilayers on low-temperature AlN buffer layer with the same 1000 s annealing time and different thickness: (a) 45 nm, (b)30 nm, (c) 20 nm, (d) 16 nm. The dashed lines denote the start of GaN epilayer's growth.

In order to gain further insight to the effect of AlN buffer layer on the quality of GaN layers, three AlN buffer layers *a*, *b*, and *c* with growth stopped just before the growth of high temperature GaN epilayers [indicated by arrow in Fig. 1(a), (b) and Fig. 2(a)] are prepared and examined by atomic force microscopy (AFM). These AlN buffer layers are grown under the same growth conditions as those in samples A, B, and C, i.e. *a* (1000 second annealing, 20nm thickness), *b* (300 second annealing, 20nm thickness), and *c* (1000 second annealing, 45nm thickness), respectively. The surface morphology of these AlN buffer layers is shown in Fig. 3(a), (b) and (c), respectively. They are quite different, and the sample *a* has the largest grain size and the lowest nuclei density as shown in Fig. 3. It is known that the GaN epilayer A which is grown on the AlN buffer layer *a* has the best quality, implying that the quality of GaN epilayers is closely related to the surface morphology of AlN buffer layer. Because the GaN islands in the initial growth stage will coalescence quickly if the AlN buffer layer has small grain size and high nuclei density, as a result, a lot of formed dislocations will go through the GaN epilayers, leading to a deteriorated quality. On the

other hand, the quality of GaN epilayers deposit on AlN buffer layer with large grain size and low nuclei density will be much better, since the lateral growth and coalescence of GaN islands will be prolonged, which leads to an increased volume of defect-free columnar domains and improve the crystal quality (Han et al., 1997). Of course, when the AlN buffer layer has too large grain size and too low nuclei density, it will take a too long time for the lateral growth of GaN islands, the quality of GaN epilayers will also become bad. Therefore, it is concluded that the quality of GaN epilayers is dependent on the grain size and the nuclei density of AlN buffer layer. A long time annealing and a suitable thickness of AlN buffer layer are very important to the growth of high quality GaN epilayers.

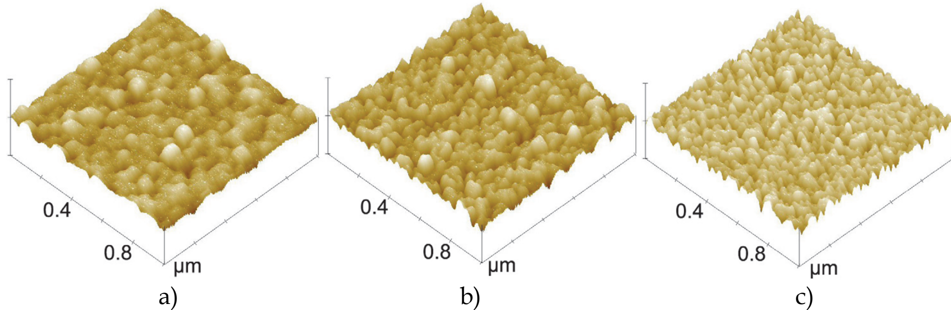


Fig. 3. AFM images of the surface morphology of the AlN buffer layers, where a, b, and c are for the growth of samples A, B, and C, respectively

Besides the growth condition of low-temperature AlN buffer layer, the V/III ratio in the initial growth stage has an important influence on the quality of a GaN epilayer grown by MOCVD, and the quality of GaN epilayer could be improved by employing a lower V/III ratio in the initial growth stage and intentionally prolonging the island coalescence process (Zhao et al., 2007b, 2009a). After optimizing the growth conditions, high-mobility MOCVD-grown n-type GaN films of about 4μm in thickness were reported (Zhao et al., 2006a). The electron mobility at room temperature was as high as 1005 cm^2/Vs at an electron concentration of 1.1×10^{16} cm^{-3}. Fig. 4 shows the Hall measurement results of high quality GaN samples. This is one of the best values reported for GaN films grown by MOCVD.

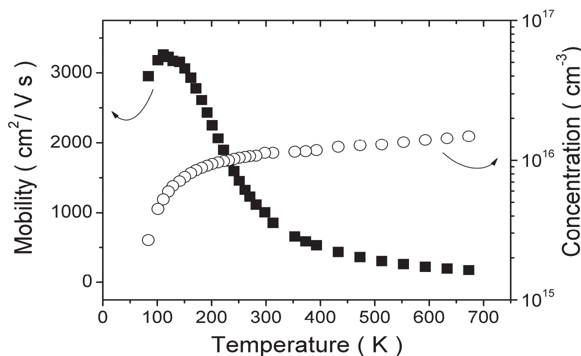


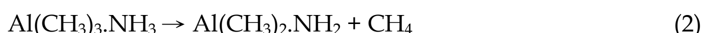
Fig. 4. The temperature-dependent electron mobility(solid square) and concentration(open circle) of high quality n-type GaN sample. The electron mobility at room temperature was as high as 1005 cm^2/Vs at an electron concentration of 1.1×10^{16} cm^{-3}.

2.1.2 Parasitic reaction between TMAl and NH_3 in MOCVD

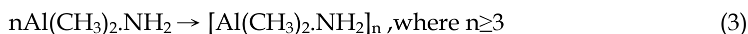
AlGaN materials are important for producing solar-blind ultraviolet photodetectors. However, the parasitic reaction of TMAl and NH_3 occurring in the vapor phase is much more serious than that of TMGa and NH_3 (Mihopoulos et al, 1998), and it has an important influence on the growth of AlGaN materials. We will discuss parasitic reaction of TMAl and NH_3 in MOCVD and study how to reduce its effect in this section (Zhao et al., 2006b).

Fig. 5 shows the relationship between the growth rate of AlN at different reactor pressure of 50, 200 and 500 Torr in MOCVD system, respectively. It is found that with the increase of growth temperature, the growth rate of AlN increase slowly when the reactor pressure is as small as 50 Torr, while it instead decreases with increasing growth temperature if the reactor pressure is raised to 200 Torr, and it decreases even more rapidly if the reactor pressure is 500 Torr. Usually it is difficult to grow AlN material at low temperature because of the difficulty of decomposing NH_3. However, the growth rate of AlN at high growth temperature is much slower than at low temperature when a relatively high reactor pressure is adopted during the growth. Such abnormal growth rate dependence on growth temperature is attributed to the parasitic effect of NH_3 and TMAl. It is known that there is a boundary layer above the substrate in the reaction chamber, and the gas precursors have a residence time in the MOCVD process. The process related to parasitic reaction of TMAl and NH_3 in the vapor phase can be described as follows (Mihopoulos et al, 1998),

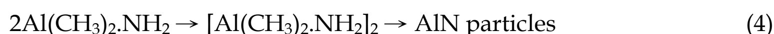
$$Al(CH_3)_3 + NH3 \rightarrow Al(CH_3)_3.NH_3 \quad (1)$$



$$Al(CH_3)_3.NH_3 \rightarrow Al(CH_3)_2.NH_2 + CH_4 \quad (2)$$



$$nAl(CH_3)_2.NH_2 \rightarrow [Al(CH_3)_2.NH_2]_n \text{ ,where } n \geq 3 \quad (3)$$



$$2Al(CH_3)_2.NH_2 \rightarrow [Al(CH_3)_2.NH_2]_2 \rightarrow AlN \text{ particles} \quad (4)$$

$$[Al(CH_3)_2.NH_2]_3 \rightarrow AlN \text{ particles} \quad (5)$$

During the growth, the AlN particles formed by parasitic reaction according to the equation (4) and (5) are carried away from the deposition zone and do not contribute to the growth (Mihopoulos et al, 1998). It is reported that the dimers (i.e. $[Al(CH_3)_2.NH_2]_2$) have sufficient energy to lose methyl groups to form AlN particles at high temperature (Mihopoulos et al, 1998). Therefore, if parasitic reaction plays an important role, much more AlN particles will be engendered with the increase of growth temperature, and the growth rate of AlN will decrease. Such effect is significant under a high reactor pressure. However when the reactor pressure is as low as 50 Torr, the surface kinetically-limited mechanism is competed with the gas phase parasitic reaction and becomes dominated, the growth rate of AlN is expected to increase with the growth temperature as is expected. From the experimental result it is deduced that the parasitic reaction depends on the reactor pressure besides the growth temperature, i.e. it dominates when the reactor pressure is 200 Torr and 500 Torr, and is less significant at 50 Torr. It is reasonable that the chance for TMAl and NH_3 to come into contact increases at higher reactor pressures, and the formation of trimers (i.e. $[Al(CH_3)_2.NH_2]_3$) and higher n-mers($[Al(CH_3)_2.NH_2]n$ for n>3) is enhanced by the increasing reactor pressure, as a result, the parasitic reaction between TMAl and NH_3 becomes weak at a low reactor pressure.

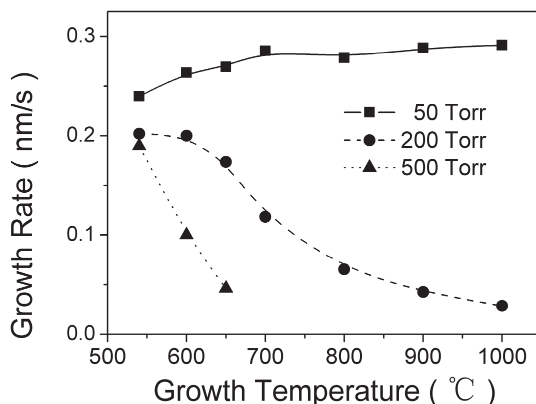


Fig. 5. The measured growth rate of AlN versus the growth temperature, the reactor pressures adopted in the growth are 50, 200 and 500 Torr, respectively.

The dependence of the AlN growth rate on the flux of NH_3 is also studied, and a very interesting phenomena are observed as shown in Fig. 6. It is found that whatever the other growth parameters are taken, the growth rate of AlN always decreases with the increase of the flux of NH_3 in the range shown in Fig. 6. The parasitic reaction is expected to be responsible for the phenomena. It has been reported that NH_3 can reduce the potential energy barrier of trimers and higher n-mers formation (Nakamura et al., 2000). Therefore, with the increase of the flux of NH_3, the quantities of trimers and higher n-mers which do not contribute to the growth will increase, and the growth rate of AlN will decrease. It is found that the growth temperature has an effect on the changing speed of growth rate with increasing flux of NH_3. The data measured at 800 °C and 540 °C are denoted by squares and circles respectively in Fig.6. It can be seen that the growth rate of AlN decreases rapidly at higher growth temperature, which supports the conclusion that higher growth temperature enhances parasitic reaction as suggested by the result of Fig. 5. In addition, with the increase of the flux of NH_3, parasitic reaction is enhanced at higher growth temperature, therefore the growth rate of AlN decreases rapidly. On the other hand, the variation of the growth rate with respect to the flux of NH_3 is also influenced strongly by the reactor pressure, as shown by the two curves in the lower part of Fig. 6 which are measured at the flux of 500 Torr and 50 Torr, respectively. It indicates that at a higher reactor pressure, more trimers and higher n-mers are formed with the increase of NH_3 flux due to parasitic reaction. As a result, the growth rate of AlN decreases more rapidly with the increase of NH_3 flux. Taking the above result into account, it is concluded that the parasitic reaction can be reduced by decreasing the growth temperature, reactor pressure and the flux of NH_3.

The further research results confirm that low reactor pressure can weaken parasitic reactions and lead to a higher Al content in AlGaN film (Deng et al., 2011). On the other hand, the enhancement of the surface mobility of Al is especially important for high quality AlGaN layers (Zhao et al., 2006b, 2008a). The migration-enhanced MOCVD is intended to carry on for improving the surface dynamic behavior of Al atoms in the growth (Zhang et al., 2002).

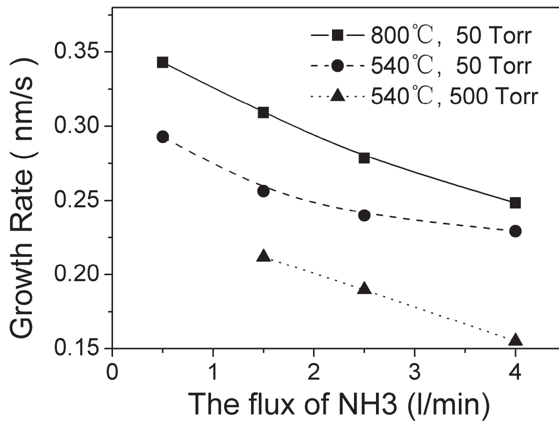


Fig. 6. The measured growth rate of AlN versus the flux of NH_3. Either the growth temperature or reactor pressure are different for the three curves.

2.1.3 Yellow luminescence and related defects in GaN materials

There are many problems of GaN materials are still hung in controversy till now. It is well-known that there is a yellow luminescence band centered at 2.2−2.3eV in the PL spectra of GaN. The mechanism behind it is still under disputation (For example, see Reshchikov & Morkoc, 2005).We find that the edge dislocations play an important role in enhancing the yellow luminescence of n-type GaN samples (Zhao et al.,2006d).

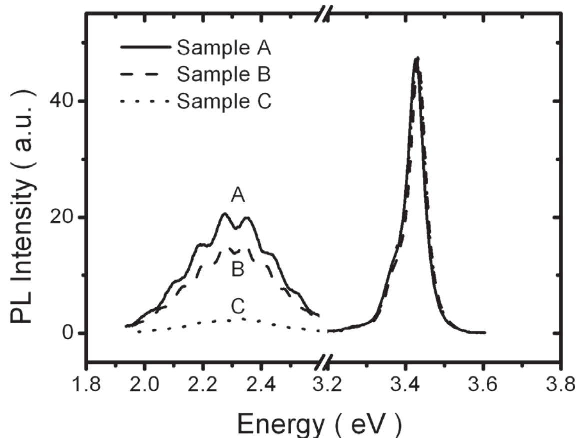


Fig. 7. Room temperature PL spectra of GaN samples A, B, and C of series I, which have the same Si doping, where the intensity of the near-band-edge luminescence peak at 3.42 eV is normalized.

Two series of GaN samples were prepared in our experiment. There were ten GaN samples in series I, where different growth conditions were used but the same SiH_4 flux of 0.72 nmol/min was employed during the materials growth. Fig. 7 shows the room

temperature(RT) PL spectra of three typical samples, A, B, and C in this series. The DCXRD characterization results of these three samples are listed in Table 2. As shown in Fig. 7, the broad band centered at 2.3 eV is the yellow luminescence. The undulations of the yellow luminescence band are formed by light interference. The intensity of the near-band-edge luminescence peak at 3.42 eV is normalized for the three PL spectra in Fig. 7. The integrated intensity ratios of the yellow luminescence band to the near-band-edge emission (I_{YL}/I_{BE}) are 2.14, 1.20, and 0.27 for samples A, B, and C, respectively. As shown in Table 2, sample A has the widest DCXRD FWHM at the (102) plane, and sample C has the narrowest DCXRD FWHM at the (102) plane. Correspondingly, sample A has the most intense yellow luminescence band, while sample C has the least. The relative intensity of yellow luminescence band seems to be strongly dependent on the FWHM at the (102) plane.
The solid square symbols in Fig. 8 show the dependence of the I_{YL}/I_{BE} value on the DCXRD FWHM at the (102) plane for the all ten samples of series I. The figure clearly indicates that the relative intensity of the yellow luminescence band increases as the DCXRD FWHM at the (102) plane increases. However, we have also found that the relative intensity of yellow luminescence exhibits a random fluctuation when the DCXRD FWHM at the (002) plane increases (not shown here). It has been reported that the GaN samples with narrower DCXRD FWHM at the (002) plane have lower screw dislocation densities, while the GaN samples with narrower FWHM at the (102) plane have lower edge dislocation densities (Heying et al., 1996; Heinke et al., 2000). Therefore, the above-mentioned results suggest that the intensity of the yellow luminescence band in n-type GaN is not influenced by the density of screw dislocations, but is strongly related to the edge dislocations.

Sample	DCXRD FWHM (arcsec)		SiH_4 Flux (nmol/min)	n (cm^{-3})	I_{YL}/I_{BE}(a.u.)
	(002)	(102)			
A	428	835	0.72	8.8×10^{16}	2.14
B	382	605	0.72	2.1×10^{17}	1.20
C	425	490	0.72	2.6×10^{17}	0.27

Table 2. Characterization results and Si doping condition of n-type GaN samples in series I

Hall measurements were carried out to gain further insight into the behavior of edge dislocations in GaN. Fig. 8 shows the dependence of net carrier concentration of the n-type GaN samples of series I (hollow triangle symbols) on the DCXRD FWHM at the (102) plane. We observed that the carrier concentration decreased with the increase of the FWHM at the (102) plane. This is surprising because the same Si doping is used during the growth of this series of GaN samples. This implies that edge dislocations may influence the net carrier concentration of GaN samples. It is well known that there are many dangling bonds along the edge dislocation lines, and they can introduce deep acceptor centers which may capture electrons from the conduction band in n-type semiconductors (Read, 1954; Podor, 1966). It has been proved that the negatively charged acceptors introduced by the dislocation line act as scattering centers in n-type GaN (Ng et al., 1998; Look & Sizelove, 1999). Therefore, the decrease of free electron concentration shown in Fig. 8 can be attributed to the compensation effect from the increasing acceptor levels introduced by the edge dislocations. This result confirms that the edge dislocations introduce acceptors in n-type GaN samples.
It is widely accepted that the yellow luminescence is caused by the transition of donor-acceptor (DA) pairs (For example, Reshchikov & Morkoc, 2005). The average distance

between donors and acceptors plays a key role in determining the luminescence efficiency of the radiative recombination in the yellow luminescence band. It is found that there are many positively charged donors surrounding the negatively charged edge dislocation core in Si-doped GaN (Krtschil et al., 2003). The transition probability of electrons and holes between donors and acceptors will be higher around edge dislocations as the average distance of DA pairs becomes shorter and the overlap integral between electron and hole wavefunctions becomes larger, which will lead to an enhancement of the luminescence efficiency of the yellow band by the edge dislocations.

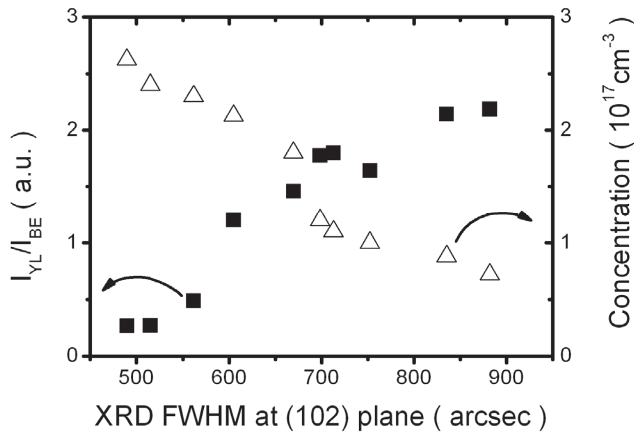


Fig. 8. The ratio between integrated intensity of the yellow luminescence band to near-band-edge emission (I_{YL}/I_{BE}) (solid squares) and the net carrier concentration (hollow triangles) vs. DCXRD FWHM at the (102) plane for the ten GaN samples of series I .

In order to study the source of donor species involved in the yellow luminescence, a different Si doping experiment was performed in which the series II GaN samples were studied. The structural quality of the series II GaN samples was high, the DCXRD FWHM at the (002) and (102) planes of these samples as narrow as 180 arcsec was obtained. Fig. 9 shows the PL spectra of samples D, E, F, and G, for which the SiH_4 fluxes used during MOCVD growth were 0, 0.05, 0.22, and 0.45 nmol/min, respectively. The corresponding electron concentrations are $5.0\times10^{15}cm^{-3}$, $1.5\times10^{16}cm^{-3}$, $8.8\times10^{16}cm^{-3}$, and $1.8\times10^{17}cm^{-3}$, respectively. In Fig. 9, the PL intensity is normalized according to the intensity of the near-band-edge emission. The ratio I_{YL}/I_{BE} increases as the SiH_4 flux increases, which is 0.15, 0.23, 0.32, and 0.35 for samples D, E, F, and G, respectively. It is well known that Si impurity atoms act as shallow donors in GaN and can increase the electron concentration in GaN samples. The PL results indicate that Si impurity has an influence on the relative intensity of the yellow luminescence band. This is reasonable if the Si donors are involved in the yellow luminescence in n-type GaN as a component of the related donor-acceptor pairs, although the multiple origins of the yellow luminescence are proved to co-exist (For example, Reshchikov & Morkoc, 2005). Due to the high quality of the GaN samples of series II which have a very low background carrier concentration as small as 5.0×10^{15} cm^{-3} for the unintentionally-doped GaN and a low dislocation density, a better observation of the correlation between Si doping and relative intensity of the yellow luminescence in these Si-

doped samples becomes possible. For example, as shown in the insets of Fig. 9, the PL spectrum of sample G measured at 10K demonstrates a sharp peak of donor-bound exciton luminescence at the energy of 3.490eV with an FWHM as narrow as 1.6 meV.

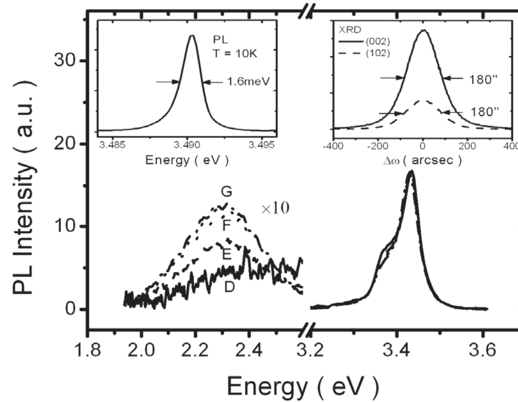


Fig. 9. RT PL spectra of GaN samples D, E, F, and G in series II with different Si doping, which have the same DCXRD FWHM at the (102) plane as narrow as 180 arcsec. In the figure the intensity of the near-band-edge luminescence peak at 3.42 eV is normalized. The insets are the ω-scan rocking curve of DCXRD(right) and 10K PL spectrum(left) of sample G.

The above results suggest that the yellow luminescence is effectively enhanced by the DA pair transitions from donor impurities such as Si to acceptors around the edge dislocations in n-type GaN. It is also found that the edge dislocation and Si impurity play important roles in linking the blue and yellow luminescence bands in n-type GaN films (Zhao et al., 2009b). In addition, there is some relationship between the yellow luminescence band and electron mobility of n-type GaN, but it is not a simple one (Zhao et al., 2007c). In fact, even the intensity of yellow luminescence is often used as an indicator of material quality for GaN, it does not have any monotonous correlation with the electron mobility of GaN.

2.2 Fabrication of GaN based UV photodetectors

On the basis of material growth, the fabrication of GaN based UV photodetectors are studied. Firstly the relationship between defects and performance of Schottky barrier photodetectors is investigated, then the fabrication of p-i-n high performance UV avalanche photodiodes are introduced, finally some new GaN photodetector devices are proposed.

2.2.3 Influence of defects on the responsivity of GaN Schottky barrier UV photodetectors

In this section, the effect of defects on the responsivity of GaN Schottky barrier UV photodetectors is investigated (Zhao et al.,2007a). A schematic diagram of the Schottky barrier photodetector structure is shown in Fig. 11. The top transparent Schottky barrier was formed on the n^--GaN layer using Ni/Au (3nm/3nm) metal film, and the bottom Ohmic contact to the n^+-GaN layer was made with Ti/Al/Ti/Au metal. Table 3 lists the carrier concentration in the n^--GaN layer and the DCXRD FWHM of three samples A, B, and C.

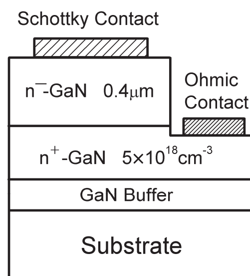


Fig. 10. Schematic diagram of GaN Schottky barrier ultraviolet photodetectors

Samples No.	FWHM of DCXRD (arcsec)		n^--GaN Layer
	(002)	(102)	
A	190	195	Si-doped, n≈1.6×10^{16}cm^{-3}
B	290	320	Si-doped, n≈1.0×10^{17}cm^{-3}
C	230	245	undoped, n≈2.0×10^{16}cm^{-3}

Table 3. The main difference in the structure parameters and epilayer XRD characterization result of three photodetector samples

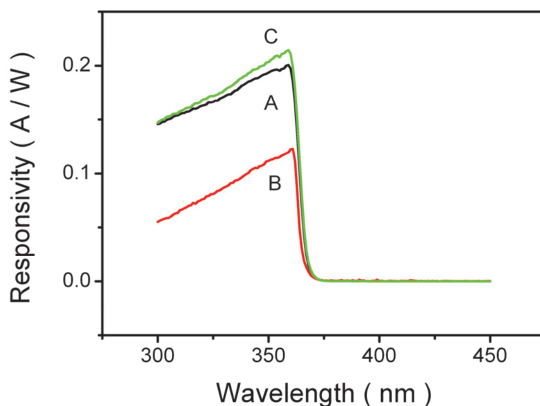


Fig. 11. The zero bias spectral response of Schottky barrier photodetector samples A, B and C (See text), where sample C uses undoped GaN layer as n^--GaN layer, samples A and B use lightly Si-doped ones .

We have investigated photocurrent spectra of many Schottky barrier photodetectors with different n^--GaN/n^+-GaN structure. Most of them show a high responsivity near the band edge of GaN, which decreases with increasing photon energy due to the reduction of penetration depth of light and the increase in surface recombination of photogenerated carriers. Fig. 11 depicts the spectral response at zero bias for three typical ultraviolet photodetector samples, A, B and C. The n^--GaN layers in samples A and B are doped with Si, but have different carrier concentrations n and DCXRD FWHM values, as shown in Table 3. The peak responsivity is

about 0.20, 0.12 and 0.21 A/W for samples A, B, and C, respectively. Considering the n^--GaN layers of sample A and B are both Si-doped, we firstly investigated the mechanism behind the different responsivity for these two photodetectors. After the device simulation, it is confirmed that the difference in the electron concentration can not be mainly responsible for the difference in responsivity between these two Schottky photodetectors (Zhao et al.,2007a). Actually, sample B has a higher dislocation density as shown by the DCXRD FWHM and a lower responsivity, indicating that dislocations can play a role in decreasing the responsivity of GaN photodetector. A lot of research work have discussed the behavior of dislocations in GaN, and they have confirmed that the dislocation cores are normally negatively charged in n-type GaN (Krtschil et al., 2003). It is reasonable to conceive that dislocation-induced acceptor levels can increase the recombination probability of photogenerated electron-hole pairs and thus reduce the responsivity of GaN ultraviolet photodetectors (Zhao et al., 2006d). Therefore, normally the responsivity of GaN based photodetectors can be improved through decreasing the dislocation density of the epilayer structure.

It is noted that the properties of different kinds of n^--GaN layer, i.e., undoped or Si-doped, may cause a difference in the responsivity of GaN Schottky barrier photodetectors. Fig. 13 shows a comparison of the spectral responses between samples A and C, where the n^--GaN layer of sample C is not Si-doped. Compared with sample A, sample C shows an unexpectedly higher peak responsivity of 0.21 A/W even though it has a higher dislocation density. It suggests that the major difference between the two samples is caused by the Si doping of n^--GaN layer. The positron annihilation experiment has confirmed that the lightly Si doping can increase the concentration of Ga vacancies in comparison with undoped n-type GaN, and lead to a decrease in the minority carrier diffusion length (Zhao et al., 2006e). It is thus proposed that the photo-generated holes can be trapped by the Ga vacancies. The n^--GaN layer in sample C has a lower concentration of Ga vacancies than in samples A and B. Since the photocurrent in the investigated Schottky barrier photodetector mainly comes from the drift current, the concentration of movable photogenerated holes in the depletion region has a strong influence on the photocurrent. The Ga vacancies will trap photogenerated carriers and increase their recombination probability in the depletion region, leading to a serious reduction of responsivity. The Ga vacancies may play an important role in the reduction of responsivity of GaN Schottky barrier photodetectors. It is also found that the Ga vacancies may lead to an increase of leakage current of GaN Schottky barrier photodetetors (Zhao et al., 2010a). Of course, it is also possible that other defects such as Ga vacancies complexes or unknown defects due to the Si doping will increase the recombination of photogenerated electron-hole pairs and reduce the photocurrent.

From the above-mentioned experiment results, using an undoped n^--GaN layer with a lower dislocation density is obviously necessary to achieve a higher responsivity in the GaN Schottky barrier photodetectors. In addition, the unexpected decrease in measured responsivity are observed in a specific GaN Schottky barrier photodetector at high reverse bias voltage (Zhang et al., 2008). A model is proposed which explains the phenomena and suggests that the choice of load resistance is also important for the application of GaN UV photodetectors.

2.2.4 Fabrication of GaN avalanche photodiodes

In this section, the fabrication and characterization of GaN-based p-i-n avalanche photodiodes (APD) with large active area are discussed (Liu et al., 2009), which is suitable

for the very weak ultraviolet radiation detection. The cross section and plan view schematic diagrams of a typical GaN p-i-n APD are shown in Fig.13. A 350nm thick SiO_2 layer was deposited on the mesa surface by plasma enhanced chemical vapor deposition (PECVD) to protect the mesa and to passivate the damage induced by dry etching. A Ni (5nm)/Au (5nm) p-contact layer was formed by electron-beam evaporation and lift-off, then Ti/Al/Ti/Au (15nm/250nm/50nm/150nm) multilayer contact layers were evaporated on both n-type layer serving as an Ohmic electrode and thin Ni/Au contact as a top interconnect pad.

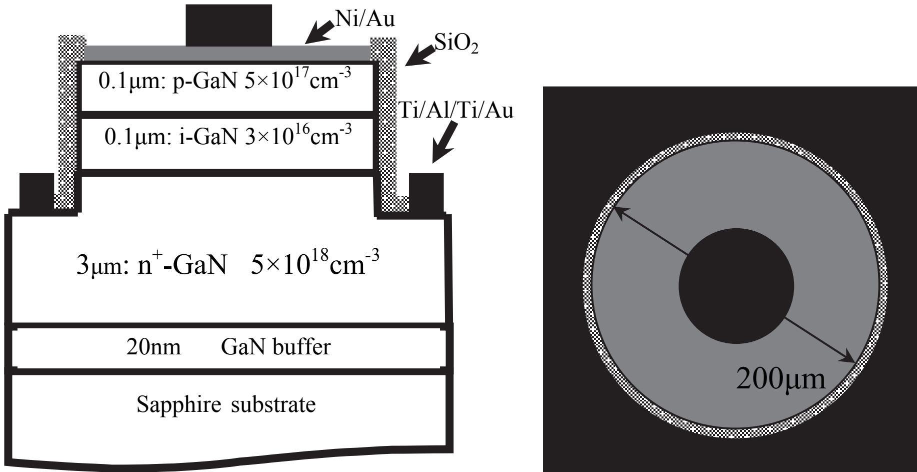


Fig. 12. The cross section and plan view schematic diagrams of the GaN APD

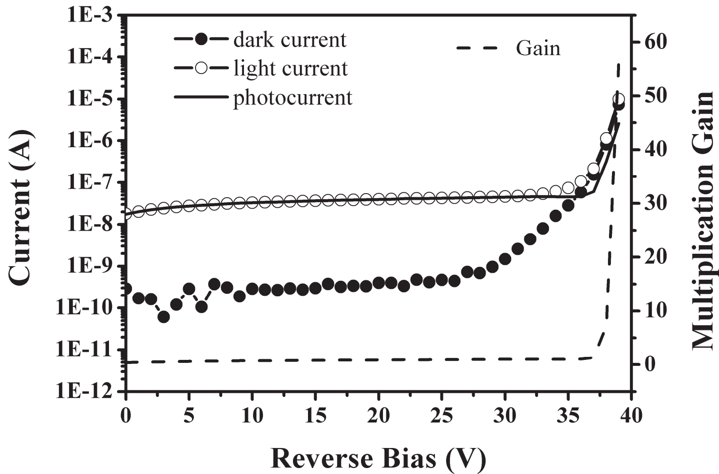


Fig. 13. The dark current and light (λ=360nm) current measured in the APD with large device area. The solid line represents the calculated photocurrent. The right-hand axis indicates the normalized gain with the unity gain at -35V as shown by dashed line.

The solid circles in Fig. 13 shows the current voltage (I-V) characteristics in dark at room temperature. The dark current for 200 μm-diameter devices with the reverse bias smaller than 20V was on the order of ~ 0.05 nA, which is down to the measurement limit of Keithly 2400 sourcemeter. This very low leakage current of GaN APD with relatively large area was ascribed to the low defect density of the material [The FWHM values of both (0002) and (10-12) XRD were found to be as narrow as ~200 arcsec, the density of screw and edge dislocations is estimated to be about 8×10^{7} cm^{-2} and 7.3×10^{7} cm^{-2}, respectively.], as well as the low plasma-induced etch damage of the mesa and the high quality of dielectric passivation. The dark current exhibits a sudden increase at about -38V, as shown in Fig. 14, indicating that the device reaches avalanche breakdown. It means that the impact ionization across the band gap occurs under the high electric field in the depletion region of p-n junction. This process is nondestructive for device. We can prove the reproducibility by performing many times of I-V scan. The breakdown behaviors of the device can be repeated with nearly the same leakage current and the consistent breakdown voltage characteristics. Therefore, the sidewall or local breakdown which may degrade the device is effectively eliminated. Under reverse bias stress near breakdown no light emission from dispersed spots induced by microplasmas was observed. It suggested that the breakdown was uniform in the device. The maximal electric field at the onset avalanche breakdown voltage (V_b=-38V) was estimated to be about 2.9 MV/cm. We attribute this high breakdown field to the optimized growth conditions and the sophisticated device fabrication process which are able to give rise to the uniform breakdown over the entire junction area.

The light current of APD was measured under λ=360nm illumination. As shown in Fig. 13, it is found that the photocurrent (the difference between light current and dark current represented by the curves with hollow circles and solid circles, respectively) takes only very little variation until the device reaches breakdown at high reverse voltage. If the reverse bias increases beyond 38V, the photocurrent increases dramatically. In the obtained light current curves no any significant change is observed even several times of I-V scan are performed, demonstrating a stable multiplication gain of the avalanche process. Then the avalanche gain can be taken by normalizing the photocurrent with setting the low bias flat region as the unity photocurrent gain. Finally the p-i-n APD with a large device area of 200μm-diameter has shown a stable avalanche multiplication gain larger than 57 at a reverse bias of 39V. We can conclude that with optimized fabrication procedures (controlled defects density of wafer, suitable structural design, and sophisticated device process), uniform bulk breakdown and stable avalanche multiplication gain can be achieved in APDs with large area grown on c-plane sapphire substrates by MOCVD.

2.2.5 New structure GaN UV photodetectors

An interesting device structure of GaN photodetectors which can be used for judging the wavelength of incident ultraviolet light is proposed (Zhao et al., 2008b). The device structure is composed of two back-to-back conventional GaN UV photodiodes with very different spectral response between 240 and 360 nm. The wavelength of incident light can be identified by measuring the photocurrent ratio of these two photodiodes, which is not dependent on the incident light power, but just strongly dependent on its wavelength.

Fig. 14(a) shows the schematic diagram of the proposed GaN photodetector structure. It is composed of two integrated back-to-back photodiodes, one with *p-n* and another with *p-i-n* structure. The first photodiode (PD_1), a *p-n* GaN structure, consists of electrode 1, n^{-}-GaN

layer 1, p-GaN layer and electrode 2. The second photodiode (PD_2) is a p-i-n GaN structure, which consists of electrode 2, p-GaN layer, n^--GaN layer 2 (i-GaN layer), n^+-GaN layer, and electrode 3. The photosensitive area of PD_1 and PD_2 is nearly the same when the incident light illuminates the semitransparent electrode of the front side of the device, as shown in Fig. 14(a). The characteristics of the two photodiodes and the electrode transmittance function are calibrated in a preliminary step as they will play a part in the measurements to be performed. The equivalent circuit of the device is shown in Fig. 14(b). Obviously PD_1 and PD_2 form an integrated back-to-back structure. The spectral response of these two photodiodes is simulated, and the designed structural parameters for PD_1 are as follows: the top n^-- GaN layer 1 and the p-GaN layer are 0.2μm and 0.3μm, respectively, while the carrier concentrations of n^-- GaN layer 1 and p-GaN layer are $1\times10^{16}cm^{-3}$ and $5\times10^{17}cm^{-3}$, respectively. The spectral response of PD_1 is shown by the solid line in the inset of Fig. 15. It is seen that the quantum efficiency of PD_1 (QE_1) only slightly decreases with the decrease of wavelength. It is worthwhile noticing that the carrier concentration of the p-GaN layer is relatively high and the doping concentration of the n^-- GaN layer 1 is low in the designed structure. The latter layer is completely depleted, even when there is a low Schottky barrier between the electrode 1 and n^-- GaN layer 1. As a result, the electric field in the n^-- GaN layer is so strong that most of photo-generated electron-hole pairs can be swiftly swept out and contribute to the photocurrent even when the absorption is stronger in the shorter wavelength region. Namely the PD_1 is not much influenced by the surface states. Therefore, the spectral response of PD_1 is quite flat in the wavelength region between 360 and 240 nm.

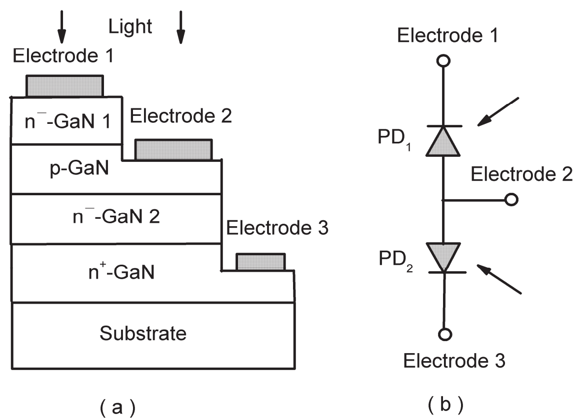


Fig. 14. (a) The schematic diagram of the structure for proposed wavelength-characterizing GaN photodetector; (b) The equivalent circuit of the proposed GaN photodetector

The spectral response of PD_2 is different from that of PD_1. The design of structure parameters of PD_2 is as follows: the n^-- GaN layer 2 and the p-GaN layer are 0.3μm and 0.2 μm, respectively. The carrier concentrations of n^-- GaN layer 2, n^+-GaN layer and p-GaN layer are 1×10^{16} cm^{-3}, 5×10^{18} cm^{-3} and 5×10^{17} cm^{-3}, respectively. As shown by the dashed line in Fig. 15, it is found that the quantum efficiency of PD_2 (QE_2) decreases rapidly with the decrease of wavelength. We can explain the phenomenon as follows: the absorption

coefficient above the band gap of GaN is as high as in the order of 10^5 cm^{-1}, the penetration depth is smaller than 0.2μm in the short-wavelength region (Zhao et al., 2006e). The depletion region is distributed mainly on the side of n^-- GaN layer 2 instead on the *p*-GaN side. Therefore, most of photo-generated carriers in *p*-GaN layer must diffuse to the electric field region (or depletion region) before they can contribute to the photocurrent. When the wavelength of incident light becomes shorter, the absorption of GaN becomes larger, and most of photons are absorbed by the *p*-GaN layer. A large part of photo-generated carriers in the *p*-GaN layer recombine before they reach the depletion region. So the quantum efficiency of the *p-i-n* GaN ultraviolet photodiode decreases rapidly with the decrease of wavelength, as shown in the inset of Fig.16 by the dashed line.

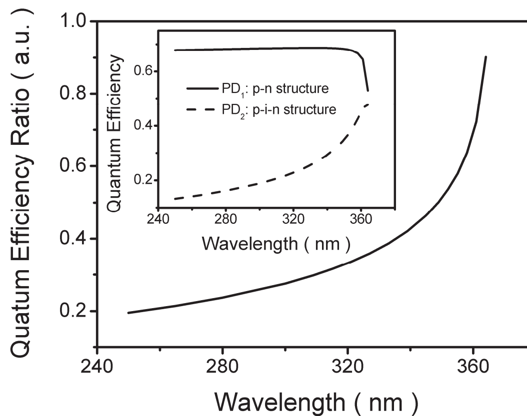


Fig. 15. The ratio of quantum efficiency of PD_2 to PD_1 (QE_2/QE_1) vs. wavelength of incident ultraviolet light. The inset shows the simulated spectral responses of PD_1 and PD_2 which are shown by the solid line and dashed line, respectively.

It is noted that for the proposed GaN device, the designed spectral response of PD_1 and PD_2 must be very different. The spectral response of PD_1 is flat while the slope of spectral response of PD_2 is much steeper, the latter should be dependent on the wavelength as much as possible. It is why PD_1 is a p-n structure while PD_2 is a p-i-n structure. For PD_2, the thickness and doping concentration of the p-GaN layer are carefully chosen. This layer is relatively thick and has a higher carrier concentration to ensure that a large part of the p-GaN layer is not depleted. Instead, the n—- GaN layer 2 has a lower doping concentration. Therefore the main part of n—- GaN layer 2 is a depletion region with strong electric field. Actually, most of incident photons are absorbed by the un-depleted p-GaN layer region and then arrive to the n—- GaN layer 2 by diffusion to form photocurrent at the junction. However, the p-GaN layer should not be too thick, otherwise a lot of photo-generated carriers will not be able to reach the n—- GaN layer 2 by diffusion, leading to a low quantum efficiency of PD_2. There is an electric circuit employed to measure the quantum efficiency ratio of PD_1 to PD_2 (Zhao et al., 2008). After two logarithm circuits and a subtraction circuit, the quantum efficiency ratio of PD_2 to PD_1 could be obtained (Kato et al., 1983), then the wavelength of incident light can be determined, since the quantum efficiency ratio of these two photodiodes is strongly dependent on the wavelength but not on the power of incident light.

In addition, A new structure of GaN based Schottky barrier ultraviolet photodetector is proposed, in which a thin n-type AlGaN window layer is added on the conventional n^--GaN/n^+-GaN device structure (Zhou & Zhao, 2007). The simulation result shows that the new structure can reduce the negative effect of surface states on the performance of Schottky barrier GaN photodetectors and then improve the performance. It is also found that an N-face growth structure can improve the quantum efficiency while a Ga-face strcuture may cause the decrease of quantum efficiency near the band-edge energy of p-i-n heterostructure UV photodetector (Li et al., 2004). In addition, a new method to estimate the hole concentration of p-type GaN by analyzing the spectral response of p-n^+ GaN UV photodetectors is suggested (Zhao et al., 2010b). It is demonstrated that the device simulation is very helpful for the fabrication and designing of new structures of GaN UV photodetectors.

3. Conclusion

In summary, both GaN materials and III-nitrides based UV photodetectors have been introduced. In the aspect of GaN and AlGaN materials, the growth mechanism and the yellow luminescence band related defects are investigated. In the aspect of GaN UV photodetectors, the fabrication and performances of Schottky barrier photodetectors, and the fabrication and characterization of p-i-n structure APDs are studied. We hope all these contents may be helpful for the readers and accelerate the development of GaN ultraviolet photodetectors.

4. Acknowledgment

The authors acknowledge the support from the National Science Fund for Distinguished Young Scholars (Grant No. 60925017), the National Natural Science Foundation of China (Grant Nos. 10990100, 60836003 and 60776047), and the National High Technology Research and Development Program of China (Grant Nos. 2007AA03Z401, 2004AA31G240 and 2001AA313050).

5. References

Amano H., Sawaki N., Akasaki I., and Toyoda Y. (1986). *Applied Physics Letters*, Vol. 48, No.5, pp.353-355

Cicek E., Vashaei Z., McClintock R., Bayram C., Razeghi M. (2010). *Applied Physics Letters*, Vol. 96, No.26, p.261107

Deng Y., Zhao D. G., Le L. C., Jiang D. S., Wu L. L., Zhu J. J., Wang H., Liu Z. S., Zhang S. M., Yang H., and Liang J. W. (2011). *Journal of Alloys and Compounds,* Vol.509, No.3, pp.748-750.

Han J., Ng T. B., Biefeld R. M., Crawford M. H., and Follstaedt D. M. (1997). *Applied Physics Letters*, Vol. 71, No.21, pp.3114-3116.

Heinke H., Kirchner V., Einfeldt S., and Hommel D. (2000). *Applied Physics Letters*, Vol.77, No.14, pp.2145-2147.

Heying B., Wu X. H., Keller S., Li Y., Kapolnek D., Keller B. P., DenBaars S. P., and Speck J. S. (1996). *Applied Physics Letters*, Vol.68, No.5, pp.643-645.

Kato H., Kojima M., and Yoshida A. (1983). *Review of Scientific instruments,* Vol.54, No.6, pp.728-732.

Krtschil A., Dadgar A., and Krost A. (2003). *Applied Physics Letters,* Vol.82, No.14, pp.2263-2265.

Li N., Zhao D. G. and Yang H. (2004). *Solid State Communication,* Vol.132, No.10, pp.701-705.

Liu W. B., Zhao D. G., Sun X., Zhang S., Jiang D. S., Wang H., Zhang S. M., Liu Z. S., Zhu J. J., Wang Y. T., Duan L. H. and Yang H. (2009). *Journal of Physics D: Applied Physics,* Vol.2, No.1, p.015108.

Look D. C. and Sizelove J. R. (1999). *Physical Review Letters,* Vol.82, No.6, pp.1237-1240.

McClintock R., Mayes K., Yasan A., Shiell D., Kung P., Razeghi M. (2005). *Applied Physics Letters*, Vol.86, No.1, p.011117

Mihopoulos T. G., Gupta V., and Jensen K. F. (1998). *Journal of Crystal Growth,* Vol.195, No.1-4, pp.733-739.

Muñoz E., Monroy E., Pau J. L., Calle F., Omnès F., and Gibart P. (2001). *Journal of Physics: Condensed Matter,* Vol.13, No.32, pp.7115-7137

Nakamura K, Makino O, Tachibana A, Matsumoto K. (2000). *Journal of Organometallic Chemistry,* Vol611, No.1-2, pp.514-524.

Nakamura S. (1991). *Japanese Journal of Applied Physics*, Vol. 30, No. 10A, pp.L1705-L1707

Nakamura S. (1998). *Science*, Vol.281, No.5379, pp. 956-961

Ng H. M., Doppalapudi D., Moustakas T. D., Weimann N. G., Eastman L. F. (1998). *Applied Physics Letters,* Vol.73, No.6, pp.821-823.

Jain S. C., Willander M., Narayan J., and Van Overstraeten R. (2000). *Journal of Applied Physics* , Vol.87, p.965-1006

Podor B. (1966). *Physica Status Solidi,* Vol.16, No.2, p.K167.

Reshchikov M. A. and Morkoc H. (2005). *Journal of Applied Physics,* Vol.97, No.6, p.061301.

Read W. T. (1954). *Philosophical Magazine,* Vol.45, No.367, pp.775-796

Wu J., Walukiewicz W., Yu K. M., Ager J. W., Haller E. E., Lu H., Schaff W. J., Saito Y., Nanishi Y. (2002). *Applied Physics Letters*, Vol.80, No.21, pp.3967-3969

Zhang J. P., Khan M. A., Sun W. H., Wang H. M., Chen C. Q., Fareed Q., Kuokstis E., and Yang J. W. (2002). *Applied Physics Letters*, Vol.81, No.23, pp.4392-4394.

Zhang S., Zhao D. G., Jiang D. S., Liu W. B., Duan L. H., Wang Y. T., Zhu J. J., Liu Z. S., Zhang S. M., and Yang H. (2008). *Semiconductor Science and Technology,* Vol.23, No.10, p.105015.

Zhao D. G., Zhu J. J., Liu Z. S., Zhang S. M., Yang H. and Jiang D. S. (2004). *Applied Physics Letters*, Vol. 85, No. 9, pp. 1499-1501

Zhao D. G., Yang H.,Zhu J. J., Jiang D. S., Liu Z. S., Zhang S. M., Wang Y. T., and Liang J. W., (2006a). *Applied Physics Letters,* Vol.89, No.11, p.112106.

Zhao D. G., Zhu J. J., Jiang D. S., Yang H., Liang J. W., Li X. Y., and Gong H. M. (2006b). *Journal of Crystal Growth,* Vol.289, No.1, pp.72-75.

Zhao D. G., Liu Z. S., Zhu J. J., Zhang S. M., Jiang D. S., Yang H., Liang J. W., Li X. Y., and Gong H. M. (2006c). *Applied Surface Science,* Vol.253, No. pp. 2452.

Zhao D. G., Jiang D. S., Yang H., Zhu J. J., Liu Z. S., Zhang S. M., Liang J. W., Li X., Li X. Y., and Gong H. M. (2006d). *Applied Physics Letters,* Vol.88, No.24, p.241917.

Zhao D. G., Jiang D. S., Yang H., Zhu J. J., Liu Z. S., Zhang S. M., Liang J. W., Hao X. P., Wei L., Li X., Li X. Y., and Gong H. M. (2006e). *Applied Physics Letters,* Vol.88, No.25, p.252101.

Zhao D. G., Jiang D. S., Zhu J. J., Liu Z. S., Zhang S. M., Liang J. W., Yang H., Li X., Li X. Y., Gong H. M. (2007a). *Applied Physics Letters,* Vol.90, No.6, p.062106.

Zhao D. G., Jiang D. S., Zhu J. J., Liu Z. S., Zhang S. M., Yang H. and Liang J. W. (2007b). *Journal of Crystal Growth,* Vol. 303, No. 2, pp.414-418.

Zhao D. G., Jiang D. S., Zhu J. J., Liu Z. S., Zhang S. M., Liang J. W. and Yang H. (2007c) . *Journal of Applied Phyics,* Vol.102, No. 11, p.113521.

Zhao D. G., Jiang D. S., Zhu J. J., Liu Z. S., Zhang S. M., Yang H., Jahn U., and Ploog K. H. (2008a). *Journal of Crystal Growth,* Vol.310, No.24, pp.5266-5269.

Zhao D. G., Jiang D. S., Zhu J. J., Liu Z. S., Zhang S. M. and Yang H. (2008b). *Semiconductor Science and Technology,* Vol.23, No.9, p.095021.

Zhao D. G., Jiang D. S., Zhu J. J., Guo X., Liu Z. S., Zhang S. M., Wang Y. T., and Yang H. (2009a). *Journal of Alloys and Compounds,* Vol.487, No.1-2, p.400-403.

Zhao D. G., Jiang D. S., Zhu J. J., Liu Z. S., Wang H., Zhang S. M., Wang Y. T., and Yang H. (2009b). *Applied Physics Letters,* Vol.95, No.4, p. 041901.

Zhao D. G., Zhang S., Liu W. B., Hao X. P., Jiang D. S., Zhu J. J., Liu Z. S., Wang H., Zhang S. M., Yang H., and Wei L. (2010a). *Chinese Physics B,* Vol.19, No.5, p.057802.

Zhao D. G., Jiang D. S., Zhu J. J., Wang H., Liu Z. S., Zhang S. M., and Yang H. (2010b). *Journal of Alloys and Compounds,* Vol.492, No.1-2, p.300-302.

Zhou M. and Zhao D. G. (2007). *Chinese Physics Letters,* Vol.24, No.6, pp.1745-1748.

15

Quantum Dot Composite Radiation Detectors

Mario Urdaneta[1], Pavel Stepanov[1], Irving Weinberg[1],
Irina Pala[2] and Stephanie Brock[2]
[1]Weinberg Medical Physics LLC
[2]Wayne State University
USA

1. Introduction

Inspired by experimental high-energy physics experiments, the first radiation detectors in positron emission tomography (PET), computer tomography (CT), and gamma-cameras were built of scintillators combined with vacuum phototubes (e.g., photomultipliers, photodiodes). Fifty years later, the scintillators/photomultiplier approach has matured, but still has intrinsic limitations. Vacuum phototubes are relatively bulky. Photomultipliers require high voltage (e.g., 0.5-2.5 kV). Vacuum phototubes are delicate because of fragile glass or quartz windows (a requisite for light to enter the vacuum tube) and fine-gap electrode structures (e.g., dynodes, grid and anode) suspended within the vacuum. Photocathodes can be irreversibly damaged if the photomultiplier is powered under normal lighting conditions. Most types of vacuum photodetectors are sensitive to external magnetic fields and therefore require magnetic shielding for certain environments (e.g., inside Magnetic Resonance Imaging systems). Vacuum phototube aging is often another challenge, because over time, vacuum inside photomultipliers tube degrades, resulting in performance degradation (e.g., increasing noise).

Today, the scintillator/vacuum phototube combination is being replaced by the scintillator/solid state photomultiplier combination (e.g., Multi-Pixel Photon Counter by Hamamatsu or Silicon Photomultipliers by SensL). Solid state photomultipliers address some of the limitations of vacuum phototubes. For example, solid state photodetectors have small size, do not require high voltage, are more robust mechanically, and are compatible with strong magnetic fields. However, they bring new challenges: higher noise and very high sensitivity to temperature and supply voltage variations. Even more importantly, solid state photodetectors do not address main challenges in radiation detection (e.g., the need for better sensitivity and better energy and spatial resolution), because the approach still depends on the same scintillators and thus involves a multi-step process for converting radiation to signal: using scintillators to convert radiation to visible light, then transporting the light into the photodetector, which finally converts light into electrical signals. A better alternative is the use of direct-conversion radiation detectors.

Direct conversion detectors are detectors in which radiation is converted directly into electrical signals. Most commonly, such detectors are made of semiconductors. There is usually an electrical bias applied to the semiconductor. Photons that interact with the semiconductor generate electron-hole pairs. Moving electrons and holes generate electrical signal in the form of increased semiconductor current. The simplest example of such a radiation detector is a silicon diode. While direct radiation detectors offer potential benefits of improved energy and spatial resolution, they have reduced detection efficiency (or "stopping power") as compared to many scintillators, especially for high energy radiation (e.g., radiation with energy above 100 keV).

In order to minimize patient radiation dose in medical imaging, it is helpful to increase stopping power, since stopping power is inversely related to the dose required for obtaining high quality patient images. High stopping power also enables the detector length to be reduced, improving spatial resolution (due to lower depth-of-interaction error) (Nassalski et al., 2007). Employing composite solid-state detectors with high stopping power as components would significantly reduce size, weight, and power requirements for imaging systems, and decrease the dose required to perform high-quality radiological examinations. As an example, we previously published a design for a PET-enabled glove, which would be difficult to implement using the current generation of vacuum phototubes (Wong et al., 2006). Detection efficiency is proportional to the fifth power of the effective atomic number (for photoelectric absorption). Therefore semiconductor type materials with high effective atomic number have the potential for developing efficient direct-converting radiation detectors.

Using the effective atomic numbers for various materials, Table 1 lists an estimate of the relative dose that would be required to collect satisfactory images using diagnostic radiologic equipment (Jackson and Hawkes, 1981).

In order to utilize the electrical signal produced by radiation (e.g., electron-hole pairs), one needs to be able to transport charge carriers through the semiconductor into electrodes at the edges of the detector. Therefore, a semiconductor with high mobility and lifetime for charge carriers is needed for radiation detection applications. For common semiconducting materials, high atomic number and good charge transport properties are not generally found together. For example, silicon has excellent charge transport properties but a low atomic number, $Z = 14$, which makes the silicon efficient only for low energy radiation detection (e.g., radiation with energy ≤ 10keV). The current generation of solid-state direct-conversion radiation detectors utilizes cadmium zinc telluride ("CZT", effective atomic number $Z = 48$), or cadmium telluride (effective atomic number $Z = 50$) (Liu et al., 2000). Unfortunately, elements with higher stopping powers (e.g., Pb, $Z = 82$) do not form compounds with the transport properties (e.g., mobility-lifetime product) that would be favorable for direct-conversion detectors (Perkins et al., 2003).

One material that would be attractive as a direct conversion material is lead sulfide (PbS, with an effective atomic number of $Z = 77$) because it is a semiconductor and because its relatively populated electron cloud increases the likelihood of photon-electron interaction. The bulk form of PbS has a small band gap (0.2 and 0.41 eV for 4 K and 293 K respectively) (Hoffmann and Pentel, 2000), and this results in a large dark current (due to thermally generated charge carriers). It is possible to engineer the material to have a larger band gap by making quantum dots (QD) of the material (Steigerwald and Brus, 1990). Engineering the

band gap enables PbS to be used as a practical material in radiation detection when the noise levels is of concern (e.g., spectroscopy, low photon count conditions). The effective band gap of a quantum dot (Fig. 1) depends on the quantum dot size, and the solution to the Schrödinger equation for the quantum dot excited state can be approximated to relate the quantum dot size and its band gap (Nedelikovic et al., 1993). We have examined the use of QDs in host matrices that combine the transport properties of the host material with the band-gap and stopping power properties of PbS QDs.

Material	Effective Atomic Number	Relative Dose Requirement
Lanthanum bromide	47	100%
Cadmium zinc telluride	48	95%
Sodium iodide	50	86%
Cadmium telluride	50	86%
Cesium iodide	54	70%
Lutetium yttrium orthosilicate	66	43%
Lead sulfide	77	29%

Table 1. Expected Dose Reduction. Estimated patient dose required for collection of equal signal-to-noise ratios, in molecular imaging systems utilizing equally-long samples of radiodetector materials, in comparison with lanthanum bromide (one of the best scintillator options available). The effective atomic number for composite materials was calculated using published methods (Taylor et al., 2009). The relative dose requirement is computed by taking the fifth power of the ratios of effective atomic numbers (to get the relative efficiency at detecting gamma radiation via photoelectric absorption) and then taking the square root (to get the relative noise, which goes as the root of the number of photons collected), and then inverting to get the relative dose requirement. For PET, the relative dose requirement would be even more pronounced than in Table 1, since the efficiency is squared in a coincident measurement. The numbers shown illustrate the motivation for increasing the atomic number.

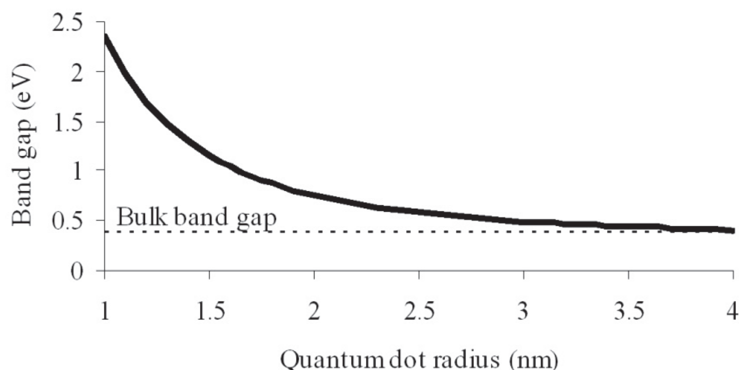


Fig. 1. Relationship between the size of lead sulfide (PbS) quantum dots and their electronic band gap.

2. Approach and methodology

Previous approaches to quantum dot radiation detectors have used indirect conversion of the radiation. In these QD-enabled detectors, the quantum dots act as scintillators to generate light pulses upon impingent radiation. The pulses are then recorded by photomultipliers (Campbell and Crone, 2005). Instead of the indirect approach, we chose to pursue a direct-conversion route, which holds the promise of better energy resolution. Our initial effort was inspired by photovoltaic research suggesting the use of organic semiconductors as host materials to quantum dots that sensitized the material to wavelengths of interest (Schwenn et al., 2005). Although these prior efforts were able to produce quantum dot/organic semiconductor films 1 µm thick, such thicknesses would only be appropriate for capturing low-penetration radiation (e.g., visible light or alpha particles). Thicker films would be necessary in order to stop incoming radiation of high enough energies to be of interest to the medical or defense fields. Additionally, organic semiconductors have reduced charge transport performance (as compared to inorganic semiconductors), and deteriorate as a result of both oxygen and the impinging radiation they are supposed to detect. We therefore pursued a novel approach involving porous and micromachined silicon as a matrix for a quantum dot composite material. The approach of using silicon as a matrix is easier to implement than using organic semiconductors, because of a reduced requirement for an oxygen-free environment during fabrication of the detector.

The combination of quantum dots and porous silicon resulted in the prototype detector schematically shown in Figure 2. In some respects, the detector is very similar in geometry and operation to other direct conversion radiation detectors: it is comprised of a planar semiconductor material with electrodes on the top and bottom faces; a reverse bias on the electrodes depletes the semiconductor material and electrons and holes are collected at the electrodes. The main difference is that the detecting portion of the device is nano-engineered to maximize the production and collection of electrons and holes in the presence of high energy radiation.

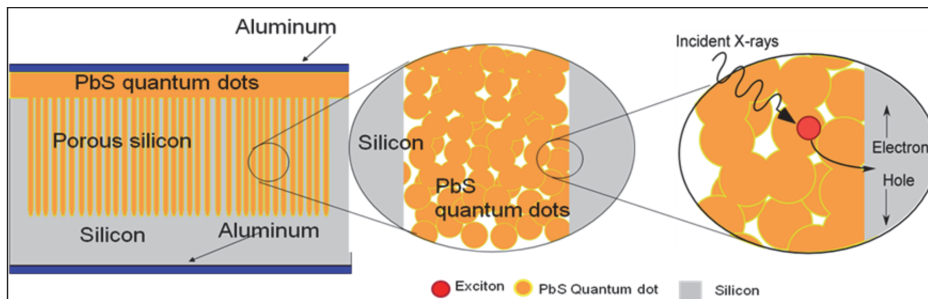


Fig. 2. Quantum dot/silicon device. Device design on left, illustrates that when incident photons interact with the lead sulfide quantum dots, excitons are produced, which in turn are separated into an electron and hole at heterojunctions (i.e., junctions between dissimilar materials). An externally applied electric field draws the electrons and holes towards aluminum electrodes, which causes an increase in the current through the device.

The generation and disassociation of excitons is a critical step (Sambur et al., 2010) in the detection of radiation in quantum-dot based devices. Excitons are disassociated into electrons and holes when they meet a heterojunction (a junction of two materials with different electronic structures) because of the sudden electrostatic field at the interface of the two materials. The resulting electrons and holes can be manipulated using electric fields as in traditional semiconductor detectors. It is of critical interest to maximize the probability of exciton to electron-hole-pair conversions in order to have good conversion efficiencies. Once electrons and holes are generated, these charge carriers have to travel to the readout electrodes. Lead sulfide, in both bulk and quantum dot forms, has sub-optimal charge mobilities and lifetimes. Making matters worse, our interest in radiation detection means that we desire detectors with large charge carrier travel paths (we need thick detectors so that they can stop radiation of very high energy, e.g., 4.4 MeV). Thus, charge transport is a critical factor in the realization of detectors for high-energy radiation. We address both challenges, charge transport and exciton disassociation, by using porous silicon as a host material for the PbS quantum dots.

Porous silicon is a nanostructured material that consists of silicon (in crystalline form, typically) which has been electrochemically treated in a hydrofluoric acid-rich solution to have pores of size ranging between one and hundreds of nanometers (Foll et al., 2002, Lehmann et al., 2000). We processed a silicon wafer in order to obtain porous silicon with straight holes, of diameter 100 nm, normal to the surface of the silicon wafer and along the crystal direction <100>. Figure 3 shows an overhead view of one of our silicon wafers after such processing. The depth of the pores is a function of processing conditions and the properties of the silicon. Pores up to 1 mm deep have been reported (Holke and Henderson, 1999). In the experimental results presented we used a porous silicon layer 20 µm thick, and used additional micromachining techniques to achieve effective layers 700 µm thick.

The quantum dots that were employed in the experiments presented were manufactured in-house using solution-phase methods under inert conditions (Hines and Scholes, 2003). The quantum dots were capped with oleic acid and dispersed in hexane. Each batch of quantum dots was characterized using photoluminesence and photoabsorbance data. Typically, absorbance and emission peaks occur around 800 nm, corresponding to a band gap of about

1.3 eV (Figures 4A and B). This band gap is consistent with a quantum dot size of 2 – 3 nm in diameter, which is consistent with both transmission electron microscopy (TEM) images of the quantum dots (Figure 4C) and literature models. Powder X-ray diffraction (PXRD) data revealed that the quantum dots adopted the cubic PbS crystal structure (Figure 4D).

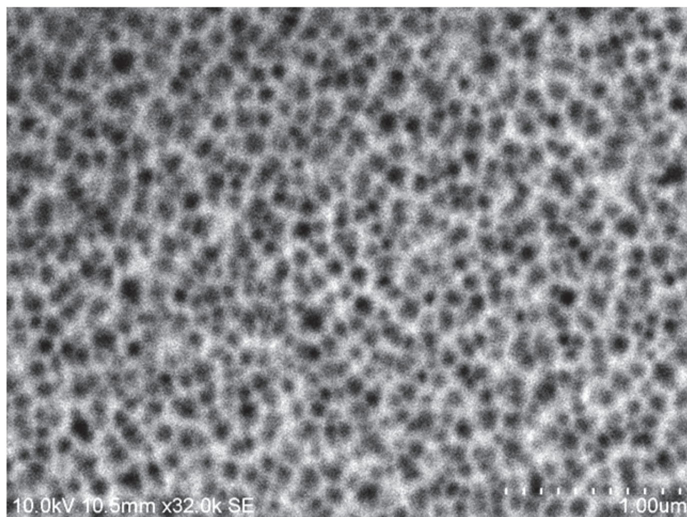


Fig. 3. Overhead view of the porous silicon used in the radiation detectors described. The holes are normal to the surface of the wafer, along the <100> crystal direction, which is preferentially weak to the attack of the anodization process.

The quantum dots were loaded into the porous silicon by capillary action. The surface of freshly treated porous silicon is hydrophobic, and is wetted by the organic solvent solution in which the quantum dots are dispersed. We verified that the quantum dots entered the entire depth of the pores by cleaving a PbS loaded section of porous silicon and performing electron energy dispersive spectroscopy on the cross-section. An image of such cross-section is shown in Figure 5. Distinct regions of elemental silicon, elemental lead, and a region that includes both (the quantum dot-laden portion of the porous silicon) are clearly visible, with lead reaching the full depth of the region of porous silicon. Once loaded, aluminum was evaporated on top of the quantum dot layer, which makes up one of the two electrical contacts in the detector (the other contact consists of aluminum evaporated on the opposite side of the silicon wafer).

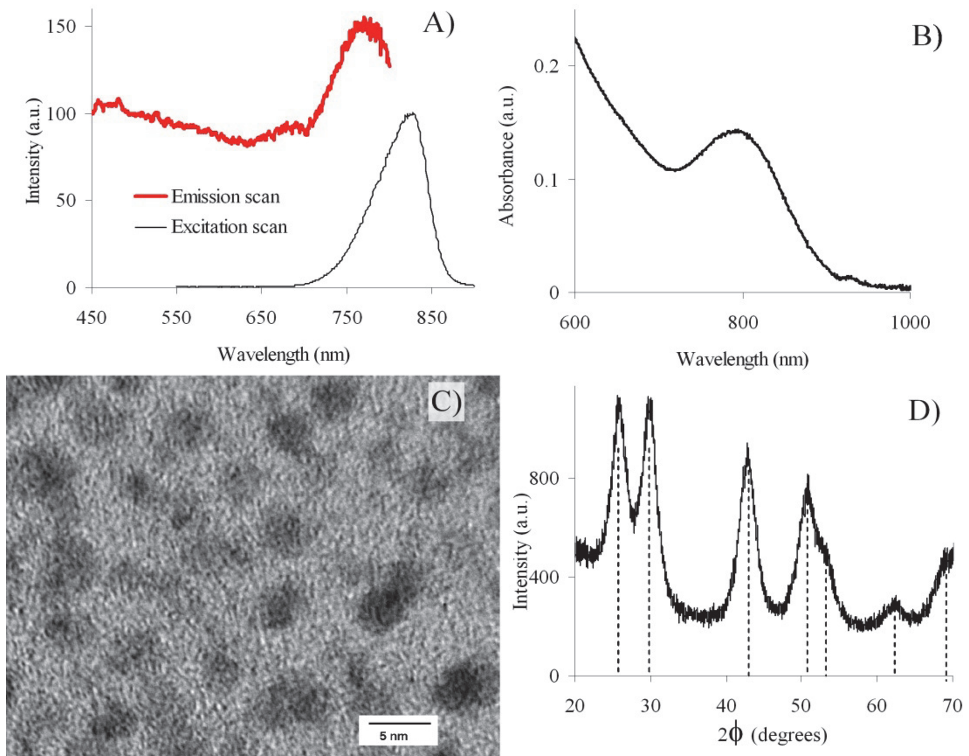


Fig. 4. The characterization of the quantum dots used in the radiation detectors include (A) photoluminesence, which in this case shows a peak at 827 nm, (B) absorbance, which in this case shows an absorption peak at 794 nm, (C)TEM images, showing quantum dots of various sizes, and (D) PXRD data of the PbS nanoparticles, showing the cubic lead sulfide structure.

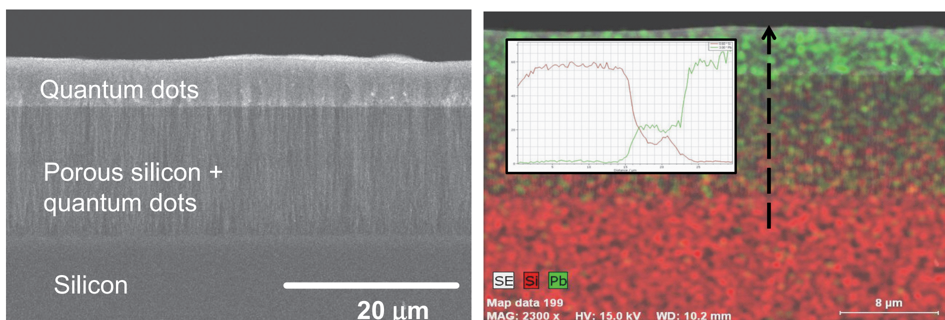


Fig. 5. (Left) Cross-sectional backscattered electron micrograph of the active region of the detector. (Right) Energy dispersive spectroscopy map of the same region showing the distribution of lead (green) and silicon (red). Regions of pure PbS, non-porous silicon, and the composite PbS/porous silicon (mixed green-red) are evident. The insert shows the proportions of silcon and lead along the length of the dashed arrow.

The detector was reverse biased at 1.6V and connected to an operational amplifier in the classical configuration illustrated in Figure 6.

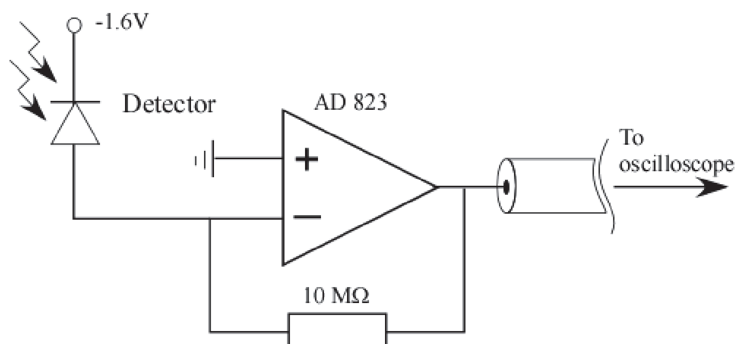


Fig. 6. Connection diagram of the detector during testing and operation.

3. Experimental results

In order to test the detector under relevant clinical conditions, we exposed it to x-rays produced by a CT scanner at various energy levels. A typical detector response is shown in Figure 7. One can see the effect of the attenuation of the x-rays reaching the detector after they have passed through the various materials that make up the detector housing (Figure 7 insert shows the CT-acquired image of the detector housing, with the detector itself pointed by the arrow). The detector also showed good radiation hardness, by having a response that remained the same even after exposure to over 11 Gy of x-ray radiation, equivalent to 3000 mammograms.

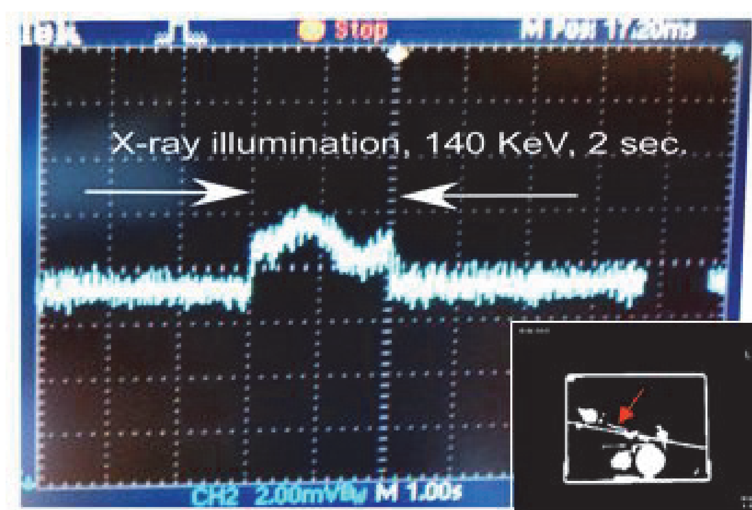


Fig. 7. Radiation Response Characterization. X-ray response of composite material, unchanged after exposure to equivalent of 3000 x-ray mammograms.

The detector response to the intensity of the radiation beam is linear, as evidenced in Fig. 8. Such linearity is required for clinical radiography.

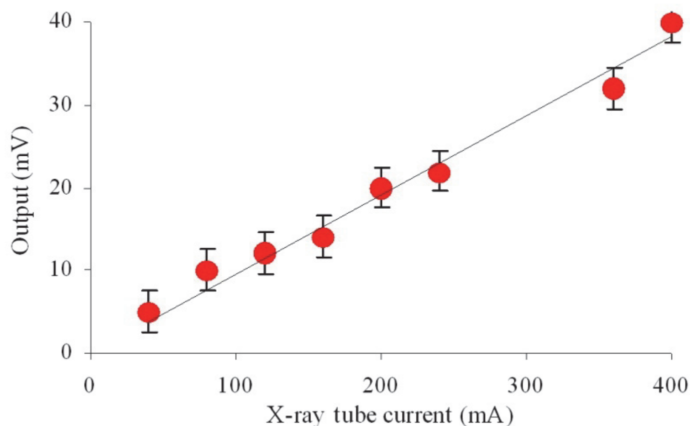


Fig. 8. Detector output response to x-rays of 120 keV at various intensities (tube currents).

The detection of radiation is intimately related to the ability of the detector material to stop the radiation. We expected our detector to attenuate x-ray radiation better than common radiation direct-conversion materials (e.g., Si, CdTe, CZT), and we verified this by blocking the radiation window of a bismuth germinate orthosilicate (BGO) scintillator coupled to a photomultiplier with a PbS/pSi detector (no housing). We exposed the detector to x-rays at various energies while monitoring the photomultiplier output, in order to obtain the relative attenuation (with and without the sample).

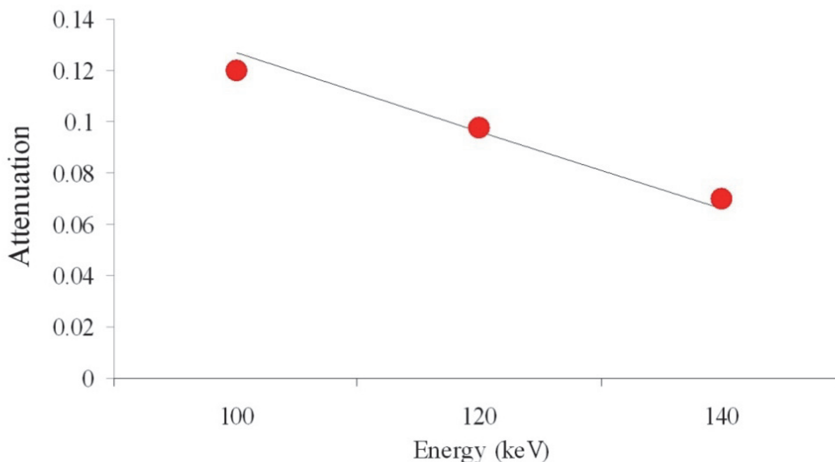


Fig. 9. X-ray Attenuation. We compare experimental attenuation values (circles) of layers of composite material to expected NIST-derived attenuation coefficients (black line). The linear attenuation coefficient was higher than CdTe, CZT.

Figure 10 shows measurements of mobility-lifetime product and resistivity from decay lifetime measurements. Mobility-lifetime (ML) was calculated to be 10^{-4} cm^2/V, using the relation

$$ML = (d^2/V)*(\tau_{lifetime}/\tau_{transition});$$

where $\tau_{lifetime}$ is lifetime for charge carries and $\tau_{transition}$ is time needed for charge carries to go across the device thickness d under applied bias voltage V. Mobility-lifetime measurements were obtained using published methods (Schwenn et al., 2005), in response to a fast (250 ns) laser pulse of infrared radiation (900 nm), and to ambient light. Resistivity measurements, measured by recording the leakage current while the detector is in reverse bias in the dark, show that the average value is 3 x 10^{10} Ω-cm. These values are within acceptable ranges for radiation detection (Owens and Peacock, 2004).
The detector also shows response to visible radiation (in addition to infrared), as would be expected based on the band gap of the PbS quantum dots employed. We measured the detector response to white light (white-LED flash light) as a function of bias, and found a strong dependance on illumination, especially in the forward bias condition. An example of detector response to light, as a function of bias, is shown in Figure *11*.

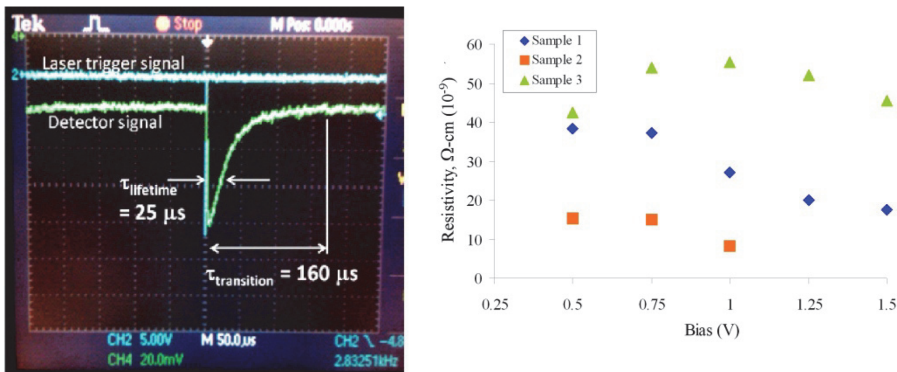


Fig. 10. (Left) Mobility-lifetime measurement using published methods (Schwenn et al., 2005), in response to a fast (250 ns) laser pulse of infrared radiation. (Right) Resistivity of the PbS/porous Si material as a function of reverse bias.

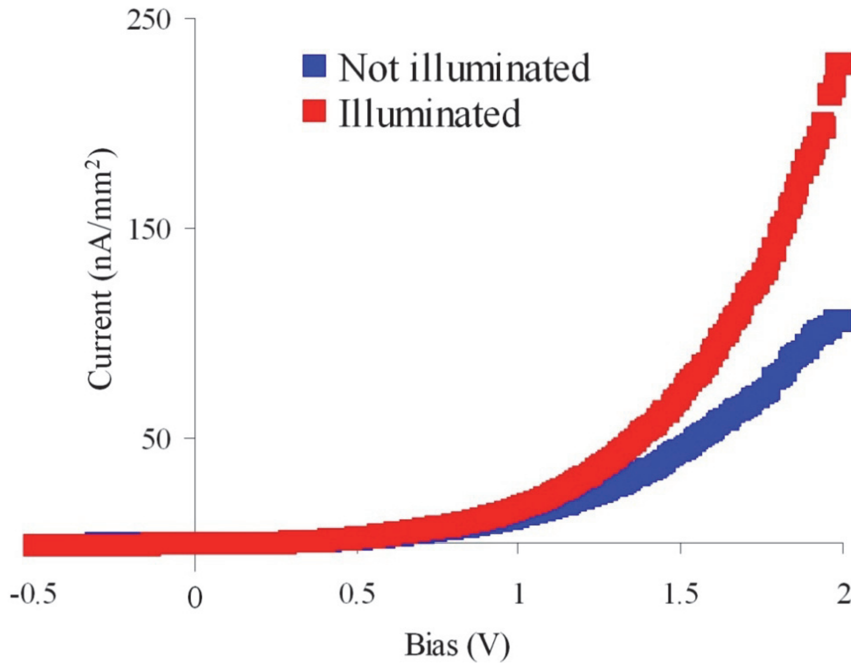


Fig. 11. Visible light sensitivity: The I-V curve shows response to visible light (white LED).

It is important to note that each experiment included control detectors consisting of silicon pieces treated the same way as the detectors but without being loaded with quantum dots. None of the control devices showed any response to any type of radiation.

4. Discussion

The detector presented is the result of the combination of two nano-engineered materials. This combination is notable as an important technical milestone: unlike classic rules concerning composite radiodetection materials (Owens and Peacock, 2004), it may now be possible to separate the problem of charge transport from that of charge conversion.

In this work, we presented the use of porous silicon as a host for quantum dots in a radiation detector. An attractive aspect of working with porous silicon is that the crystalline structure of the silicon remains intact, and thus the charge transport characteristics of the silicon remain relatively unaffected. In addition to improved charge transport, we expected that porous silicon would aid in increasing the conversion fraction of excitons to electron-hole pairs. This is because excitons travel a limited range (~20 nm) before they recombine and dissipate their energy, with lifetime of about one microsecond (Moreels et al., 2009). Thus, only the quantum dots whose distance from a heterojunction within an exciton's range will result in the appearance of electronic signals from the composite detector. Quantum dots deposited as a planar film, however thick the film is, will only yield signals coming from the quantum dots in the volume within exciton range of the silicon-QD interface (Winder and Sariciftci, 2004). In contrast, QDs deposited within a porous silicon matrix will be able to yield much larger signals (for a given device footprint) because the volume of QDs within exciton range of the QD-silicon interface is much greater. For the case of pores of about 100 nm in diameter, most of the QDs in the pore will be within exciton range of the heterojunction. Thus, the effective layer of quantum dots that are active participants in detection grows dramatically (i.e., from 20 nm to >20 µm).

This new nanotechnology-enabled paradigm promises to confer flexibility to future material designers, who will be able to increase effective atomic number and band-gap with less concern about impairing the material's ability to transport charge (e.g., mobility-lifetime product). The detector involves innovations at multiple scales of fabrication, from the nano-level (in the production of quantum dots), to the micro-level (in the fabrication of silicon micro-structures to accommodate the quantum dots and to amplify signals), to the macro-level (as imaging devices are designed to take advantage of favorable absorption properties). Under this new paradigm, it is easy to imagine the incorporation of quantum dots with different band-gap properties into a single silicon system, capable of detecting broad swaths of the electromagnetic spectrum. The concept of a single detector to sense multiple spectral domains (infrared, visible, x-ray) might be attractive to designers of night-vision goggles and viewfinders, who currently use separate cameras to image visible and infrared features for surveillance uses by soldiers and first-responders. The ability to place amplifiers and buffers on the same wafer as the sensor is very important to designers of focal-plane arrays, who are currently forced to hybridize detectors with bump-bonding methods that limit spatial resolution and are vulnerable to temperature changes (due to differing thermal-expansion characteristics of the bonded strata). The use of a silicon platform also satisfies designers who wish to put as much digital logic as possible in close proximity to the detectors, as has been done for silicon photomultipliers. An illustration of the flexibility of the platform is shown in Figure 12.

Currently, many PET, SPECT, CT scanners and digital mammography and radiography systems employ small scintillating crystals to detect radiation. Such crystals take weeks to grow, devour considerable energy, and may deplete supplies of rare elements. As an example, lutetium orthosilicate (widely used in the PET industry) melts at 2,150 °C, requiring iridium crucibles which must be recast frequently. The low-cost quantum dot methods we are investigating (which might utilize dip vats or spray processes to create and deposit the the QDs) could potentially replace the need to grow crystals.

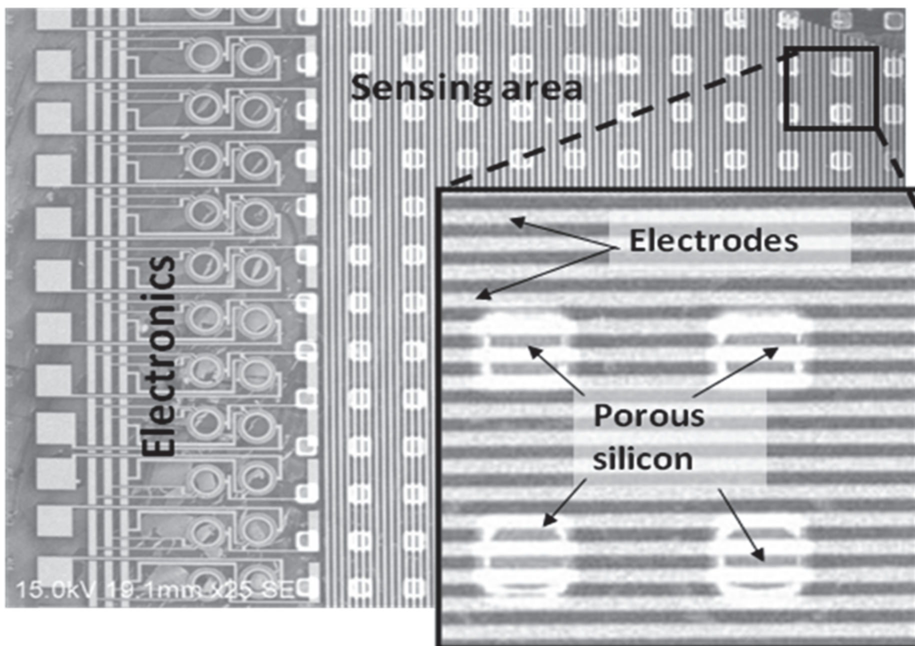


Fig. 12. Electronic test structures. By masking the material before creating the pores, we could include electronic features on the same wafer as the sensor.

5. Conclusions

This project represents an early application of nanotechnology to radiation detection. The innovation promises to reduce radiation dose and health care costs and improve radiological device performance. In addition to the diagnostic radiology market, the platform technology may be useful for homeland security and broad-spectrum surveillance for the consumer and defense markets, with potential cross-fertilization to the solar power industry.

6. Acknowledgment

We gratefully acknowledge funding from the National Cancer Institute (SBIR R43CA138013), and the assistance of Drs. Pamela Abshire and Elisabeth Smela of the University of Maryland, and Dr. Andrew Watt of the Material Science Department at Oxford University.

7. References

Campbell, I. H. & Crone, B. K. 2005. Quantum-Dot/Organic Semiconductor Composites For Radiation Detection. *Adv. Mat.*, 18, 77-79.

Foll, H., Christophersen, M., Carstensen, J. & Hasse, G. 2002. Formation And Application Of Porous Silicon. *Mat. Sci. Engr. R,* 39, 93 - 141.

Hines, M. A. & Scholes, G. D. 2003. Colloidal PbS Nanocrystals With Size-Tunable Near-Infrared Emission: Observation Of Post-Synthesis Self-Narrowing Of The Particle Size Distribution. *Adv. Mat.*, 15, 1844-1849.

Hoffmann, E. & Pentel, P. 2000. PbS: From Solids To Clusters. *International Symposium On Structures And Dynamics Of Heterogeneous Systems*, 148-155.

Holke, A. & Henderson, H.T. 1999. Ultra-deep anisotropic etching of (110) silicon. *J. Micromech. Microeng.*, 9, 51-57.

Jackson, D. F. & Hawkes, D. J. 1981. X-Ray Attenuation Coefficient Of Elements And Mixtures. *Physics Reports,* 70, 169-233.

Lehmann, V., Stengl, R. & Luigart, A. 2000. On The Morphology And The Electrochemical Formation Mechanism Of Mesoporous Silicon. *Mat. Sci. Engr. B,* 69 - 70, 11 - 22.

Liu, J., Nelson, W. & Seefred, R. 2000. Response Calculations Of The CdZnTe Detector Using EGS4. SLAC.

Moreels, I., Lambert, K., Smeets, D., De Muynck, D., Nollet, T., Martins, J., Vanhaecke, F., Vantomme, A., Delerue, C., Allan, G. & Hens, Z. 2009. Size-Dependent Optical Properties Of Colloidal PbS Quantum Dots. *ACS Nano,* 3, 3023 - 3030.

Nassalski, A., Kapusta, M., Batsch, T., Wolski, D., Mockel, D., Enghardt, W. & Moszynski, M. 2007. Comparative Study Of Scintillators For PET/CT Detectors. *IEEE Transactions In Nuclear Science* 54, 3-10.

Nedelikovic, J. M., Patel, R. C., Kaufman, P., Joyce-Purden, C. & O'Leary, N. 1993. Synthesis And Optical Properties Of Quantum-Size Metal Sulfide Particles In Aqueous Solution. *J. Chem. Ed.,* 70, 342 - 344.

Owens, A. & Peacock, A. 2004. Compound Semiconductor Radiation Detectors. *Nuclear Instrumentation And Methods In Physics Research A,* 531, 18-37.

Perkins, J., Krawczynski, H. & Dowkontt, P. Characterizing Imarad CZT Detectors With Time Resolved Anode And Cathode. Proceedings Of The 13th International IEEE Workshop On Room-Temperature Semiconductor X- And Gamma-Ray Detectors 2003.

Sambur, J., Novet, T. & Parkinson, B. A. 2010. Multiple Exciton Collection In A Sensitized Photovoltaic System. *Science,* 330.

Schwenn, P. E., Watt, A. A. R., Rubinsztein-Dunlop, H. & Meredith, P. Lead Sulfide Nanocrystal/Conducting Polymer Solar Cell. SPIE, 2005.

Steigerwald, M. L. & Brus, L. E. 1990. Semiconductor Crystallites: A Class Of Large Molecules. *Acc. Chem. Res.,* 23, 183 - 188.

Taylor, M. L., Franich, R. D., Trapp, J. V. & Johnston, P. N. 2009. Electron Interaction With Gel Dosimeter: Effective Atomic Numbers For Collisional, Radiative, And Total Interaction Processes. *Radiation Research,* 171, 123-126.

Winder, C. & Sariciftci, N.S. 2004. Low bandgap polymers for photon harvesting in bulk heterojunction solar cells. *J. Mater. Chem.,* 14, 1077-1086.

Wong, K. H., Gruionu, L. G., Cheng, P., Abshire, P., Saveliev, V., Mun, S. K., Cleary, K. & Weinberg, I. N. PETglove: A New Technology For Portable Molecular Imaging. SPIE Medical Imaging Conference 2006.

16

HgCdTe Heterostructures Grown by MBE on Si(310) for Infrared Photodetectors

Maxim Yakushev et al.*
A.V. Rzhanov Institute of Semiconductor Physics
Siberian branch of the RAS
Russian Federation

1. Introduction

A high spatial resolution of infrared (IR) thermal-imaging systems is determined by the amount of pixels of the IR focal plane array (FPA), an increase in which leads to an increase in FPA sizes. To realize such IRFPAs based on mercury cadmium telluride (MCT) alloys, a photosensitive material with a high lateral uniformity of the composition is required. Recently, considerable effort has been directed to the development of the growth of heteroepitaxial MCT structures by molecular beam epitaxy (MBE) on Si large in diameter substrates [Reddy et al., 2008].

Matrix IR FPAs are produced by element-by-element hybrid assemblage of a matrix of photosensitive elements based on MCT and a Si multiplexer with the help of In columns. For cooled IR FPAs, the problem of destruction of a hybrid assemblage can arise because of the difference in the thermal expansion coefficients of the photosensitive element and a Si multiplexer.

Correspondingly, the larger the IR FPA format, the larger the device size and the more pronounced the effects associated with a difference in thermal expansion coefficients. The use of MBE-grown MCT heterostructures on the Si substrate allows us to solve the problem of the service life of the IR FPA upon its cooling from room temperature to cryogenic temperatures.

The selection of the substrate orientation during the MBE growth of MCT is governed by a low incorporation coefficient of the Hg atoms into the crystal lattice and, as a consequence, by a high pressure of Hg vapors during the growth [Sivananthan et al., 1986]. It was established that the MCT can epitaxially grow on the (111)B surface at the lowest pressure of Hg vapors. However, the (111) plane in MCT is a twin plane, which leads to a low structural quality of the HgCdTe (111) layers because of the formation of large number of twins and stacking faults. In 1988, Koestner and Schaake [Koestner and Schaake, 1988] showed that MCT growth on the (112)B surface is possible at low Hg vapor pressures without intense twin formation, which determined this orientation as basic for the development of the growth processes of various MCT structures for IR FPAs on various substrates. However,

*Vasily. Varavin, Vladimir Vasilyev, Sergey Dvoretsky, Irina Sabinina,
Yuri. Sidorov, Aleksandr Sorochkin and Aleksandr Aseev
A.V. Rzhanov Institute of Semiconductor Physics Siberian branch of the RAS Russian Federation

the (112) orientation is sensitive to insignificant variations in the growth conditions, which determines a narrow range of optimal values for the growth of MCT with a minimal defect density [Ryu et al., 2004]. Nevertheless, intense studies and development of equipment are being carried out for MBE growth of the MCT heterostructures on large in diameter Si (112) substrates. It is shown that such structures can be used for fabrication of high quality IR FPAs in a spectral range right up to 10 μm [Carmody et al., 2008].
We studied the MBE growth of MCT heterostructures on GaAs substrates [Yakushev et al., 2009]. It was shown that the empirically determined substrate orientation (310) provides growth of high quality MCT films without stacking faults and twin lamellae also at low Hg pressures, but considerably broadens the range of optimal growth conditions.
In this study, we examined the growth of MCT heterostructures on Si (310) substrates with a diameter as large as 100 mm for IR FPAs of the spectral range of 3–5 μm and the formation and characteristics of p–n junctions and photoelectric parameters of IR photodetectors with different formats.

2. Growth process of HgCdTe/Si(310) heterostructures

HgCdTe/Si(310) heterostructures were grown using the multichamber ultrahigh vacuum MBE installation "Ob'" [Sidorov et al., 2000]. For in situ monitoring of preepitaxial preparation and growth, we used reflection high-energy electron diffraction (RHEED) and single-wave ellipsometry at wavelength λ = 632.8 nm. P-Si:B (with resistivity of 10 Ω cm) wafers 76 and 100 mm in diameter oriented along the (310) plane were used as substrates.
Before charging into a vacuum system, Si substrates were treated by the standard RCA procedure [Kern and Puotinen, 1970]. As a result, the surface was passivated with a thin SiO_2 layer. At the last stage of chemical treatment, the substrates were immersed into a 1% aqueous HF solution to remove SiO_2 and hydrogenate the surface [Fenner et al., 1989]. The substrates were charged into the installation from a sealed box in atmosphere of dry nitrogen.
Preepitaxial vacuum annealing was performed in two stages. Preliminary annealing was carried out to remove physically adsorbed contaminants. Then the sample was heated to 550–600°C in the arsenic flow. After exposure to the As_4 vapors for 15 min and cooling, the Si substrate was transported to the growth chamber of the buffer layers.
The lattice mismatch between the Si substrate and MCT is ~19%. It is eliminated by the introduction of ZnTe and CdTe buffer layers. A ZnTe layer 0.01 μm thick was grown on Si at 200–240°C. The beam equivalent pressure (BEP) for Zn was higher than the BEP for Te_2 by a factor of 20–40. A CdTe layer 6–8 μm thick was grown on ZnTe/Si at 280–320°C. The BEP for Cd was higher than the BEP for Te_2 by a factor of 3–5. MCT layers were grown on the obtained "alternative" CdTe/ZnTe/Si(310) substrate by the process described in detail in [Sidorov et al., 2000].
A specially constructed chamber was used to obtain MCT MBE heterostuctures. Usually, a substrate is rotated to obtain a uniform composition. In this case, flux densities are averaged over the area. This effect allows to obtain required uniformity. Unfortunately, it is very difficult to use precision ellipsometric methods for composition controlling while the substrate is being rotated.
To uniform the density of molecular flows over the area of the substrate the coaxial arrangement of the molecular sources is realized [Blinov et al., 1997a, 1997b]. Molecular sources with circular lenses were developed. These sources provide highly uniform

molecular flows over a large area with a minimum expense of evaporated material. The calculation shows that the homogeneity of the composition should not exceed 0.0002 cm^{-1} in the exact alignment of sources which is an order of magnitude above the most stringent requirements for uniformity. High homogeneity of flows allows to exclude rotation of the substrate and carry out continuous monitoring of the growth process.
As a result, obtained MCT MBE have high uniformity of distribution of composition over the area. For the layers with x = 0.3–0.35, the maximal variation in composition x over the surface of a sample 76.2 mm in diameter does not exceed 0.002. Such variation corresponds to the deviation of a wavelength cut-off of photosensitivity at 77 K by less than 0.1 µm (Fig. 2), which provides a high uniformity of parameters of large-size FPAs.

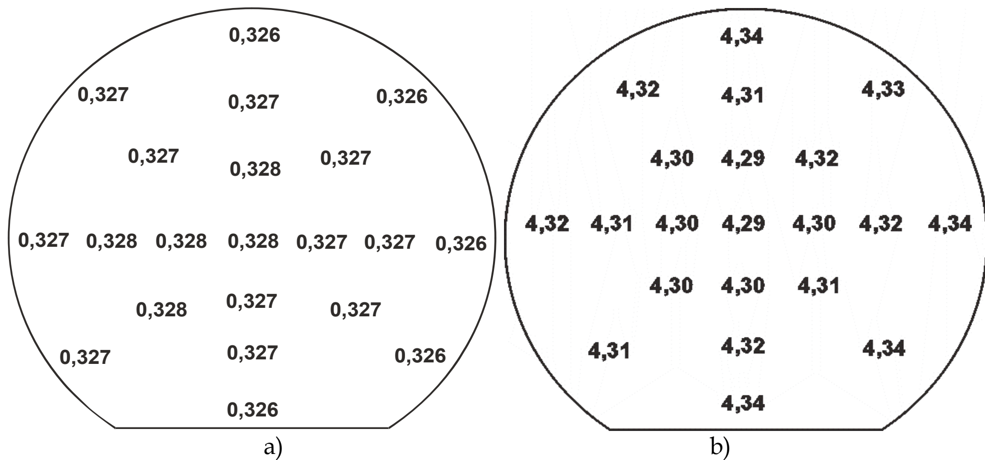


Fig. 1. Distribution of composition (a) and distribution of wavelength cut-off (microns) at 77K (b) over the area of HgCdTe/Si(310).

Due to the continuous monitoring of the growth process by *in situ* ellipsometry, MCT layers can be given composition profile throughout the thickness, such as the working layer of constant composition and graded-gap layers. Figure 2 illustrates the change in composition throughout the thickness of a typical MBE MCT heterostructure with graded-gap layers measured by the *in situ* ellipsometry. When the composition of the working layer X_{CdTe} = 0.22 the boundaries of this layer are graded-gap layers in which the CdTe content rises to the surface and to the boundary with the buffer layer. Graded-gap layers with high content of CdTe can be used for surface passivation [Bhan et al., 1996]. Increasing the width of the bandgap at the heterointerfaces and at the film surface creates built-in fields. These fields brush aside the non-equilibrium carriers from the surface which has a high rate of recombination. There is reason to believe that in this way the effective lifetime of nonequilibrium carriers can be increased [Remesnik et al., 1994;Buldygin et al., 1996;Voitsekhovsky et al., 1996].

3. Structural defects in HgCdTe/Si(310) heterostrucutures

There are three types of defects in MCT layers grown on silicon substrates: threading dislocations, stacking faults and antiphase domains. In addition, macroscopic V-defects may

exist. V-defects are complex entities that contain the area of broken structures, twin lamellae and stacking faults (SF).

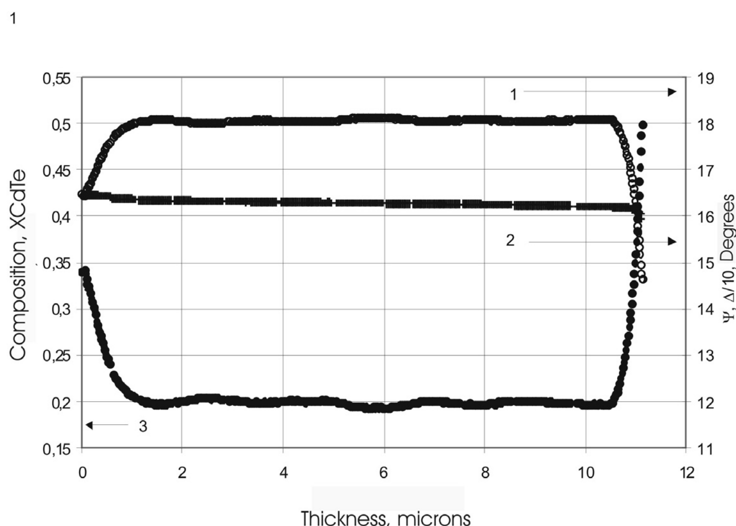


Fig. 2. Results of in situ ellipsometric measurements: 1 and 2- ellipsometric angles ψ and Δ respectively; 3- composition profile.

Studies of defects in the bulk of the films were carried out by transmission electron microscopy (TEM) and selective etching. TEM analysis was performed using the electron microscope JEM-4000EX (JEOL). Samples for TEM were prepared in the form of thin foils parallel to the growth surface, and in the form of cross sections.

To obtain transparent to the electron beam thin foils in a plane parallel to the surface of the film an original technique was developed. A sample (HgCdTe/CdTe/ZnTe/Si heterostructure) was placed in a hot KOH solution and held until complete dissolution of the silicon substrate. Then the film (HgCdTe/CdTe/ZnTe) was washed and made thinner by etching from the growth surface or from the opposite side in a solution of bromine (1.5%) in methanol. Using this method of separation of film from the substrate it is possible to investigate the structure of the film close to the surface and near the heterojunction.

Following solutions were used for selective etching: 10ml HNO_3 + 20 ml H_2O + 4g $K_2Cr_2O_7$ + 1,5 mg $AgNO_3$ (etchant E-Ag1) [Sidorov et al., 2000] and 5g CrO_3 + 3 ml HCl + 15 ml H_2O (etchant Schaake) [Kern and Puotinen, 1970]. Etchant E-Ag1 was used to detect defects in CdTe layers, and Schaake etchant for detecting defects in CdHgTe layers. As a result of selective etching, etch pits appeared on the surface. Etch pits had a characteristic shape, different for different types of defects. Dislocations were revealed in the form of elongated triangles for CdTe and in the form of points for HgCdTe. Antiphase boundaries were identified as lines of arbitrary shape, and stacking faults in the form of parallel straight lines. The density of surface macroscopic defects was measured using an optical microscope with a built-in CCD camera combined with a personal computer. This hardware and software system can automatically scan the entire surface of the heterostructure and determine the lateral distribution of defect density and defect size over the sample area.

3.1 Antiphase domains

Antiphase domains (APD) are characteristic defects for any heterostructure where the substrate is a semiconductor with a diamond lattice (Si, Ge) and the grown layer is a binary compound with the crystal lattice of sphalerite (GaAs, CdTe). The most detailed mechanism of the formation of antiphase domains and methods of obtaining of single-domain layers are considered for the case of heteroepitaxy of GaAs/Si and GaAs/Ge [Fenner et al., 1989]. The appearance of antiphase domains in GaAs layers grown on Si substrates is due to the monatomic of silicon surface. If the silicon surface contains steps of monatomic height and deposition on the terraces begins, for example, with arsenic then the second monolayer of gallium will be a continuation of the arsenic monolayer on the overlying terraces (Fig. 3a). This situation corresponds to the formation of antiphase boundaries which will lead to antiphase domains. Thus, it is necessary that the surface of the sample had steps of a diatomic height to lack of antiphase domains. Since arsenic in a wide range of conditions is adsorbed on the surface of silicon in an amount equal to one monolayer [[Blinov et al., 1997a]], the presence of diatomic steps is necessary and sufficient requirement for GaAs/Si heterostructures without APD.

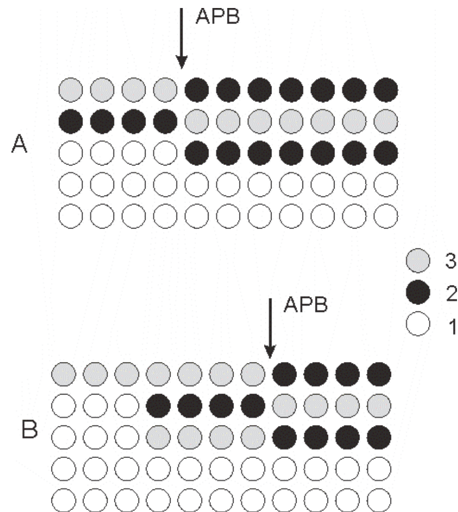


Fig. 3. Formation mechanism of antiphase boundaries. A – in case of monoatomic steps on a surface. B – in case of discontinuous absorption layer. 1 – substrate atoms, 2 – V or VI group atoms, 3 – II or III group atoms.

In case of ZnTe/Si heterostructures, the presence of diatomic steps on the silicon surface is not enough to grow films without APD. Modern technologies for high-quality structures HgCdTe/CdTe/Si require a preepitaxial annealing of the substrate in As_4 vapours [Blinov et al., 1997b; Bhan et al., 1996; Remesnik et al., 1994]. Clean Si surface is actively cooperating with the residual atmosphere of the vacuum system, in particular with tellurium [Buldygin et al., 1996], forming the centers which act as nuclei for defects of crystal structure. Adsorption of As passivates the surface of Si since excessive (relative to the silicon) valence electrons of As are completely saturating surface bonds of Si [Voitsekhovsky et al., 1996]. As a result, the amount of residual contamination on the surface of the substrate decreases

which leads to an increase in crystalline perfection of the heteroepitaxial layers. It was shown for ZnSe on Si (100) [Kuleshov et al., 1985] that the arsenic monolayer prevents the formation of amorphous layer $SiSe_x$ and significantly reduces the density of twin lamellae and stacking faults in the layers of ZnSe. It was established experimentally that the absence of As in CdTe/Si (211) heterojunction leads to polycrystalline growth [Buldygin et al., 1996].
Neither Zn nor Te is absorbed in the form of a continuous layer on the surface of As/Si (310) or clean Si (310) [Wang et al., 1976]. Tellurium is adsorbed in the form of separate islands. Zinc in the absence of Te vapors has an infinitesimal rate of insertion which begins to approach 1 only if the vapor phase contains tellurium atoms. In such circumstances, there is the possibility of the simultaneous presence on the terrace of nuclei, the lower layer of which is formed from the atoms of zinc and tellurium. This effect will inevitably lead to the formation of APB even if the terraces on the surface of the substrate are separated by diatomic steps (Fig. 1b).
It is evident from the above arguments that the formation of antiphase domains in HgCdTe/CdTe/ZnTe/Si (310) is determined not only by the structure of the substrate surface but also by the conditions of formation of the ZnTe/Si (310) heterojunction.
Using scanning tunneling microscopy and high and low energiy electron diffraction methods, the influence of vacuum annealing on the morphology of surfaces of hydrogenated Si (310) was investigated.
In a wide temperature range (500 - 1250^{0}C) surface of Si (310) has strong relief, roughness of which according to the STM was 0.15 - 0.3 nm (Fig. 4).
Nevertheless, according to the LEED there are equidistant steps of diatomic height on such a surface. The same diffraction patterns were observed by LEED for each of our samples after annealing at temperature range from 500^{0}C to 1250^{0}C. Qualitative changes of the diffraction patterns did not occur with increasing annealing temperature. Annealing at 1250^{0}C led to an increase in the brightness of reflections. An example of a diffraction pattern from the Si(310) is shown in Figure 5. Rows of reflections along the [-130] can be seen. The distance between the rows of reflections indicated by letter **a**. It can be seen that certain reflexes are split in two ones. The distance between the paired reflections is denoted as **b**. Such diffraction patterns are characteristic for a system of equidistant steps. There are two things confirming presence of steps. Firstly, the splitting reflections, and secondly, the characteristic "transfusion" of paired reflections when the primary beam energy is being varied [Kuleshov et al., 1985]. The ratio **a / b** within the measurement error is the ratio of sides of the unit cell of a smooth surface of the Si (310) (**a / b** = 1.63). The distance between the paired reflections corresponds to the larger side of the unit cell of Si (310). Based on this we can conclude that such diffraction patterns obtained from the stepped surface of Si (310) with the distance between steps equals to the size of the larger side of the unit cell of this surface. Using simple geometric calculations, we can see that in this case the step height is equal to two interplanar distances for the Si (100).
It was also found that it is necessary to achieve low concentration of residual contaminants on the surface for increasing the percentage of surface formed by diatomic steps and, accordingly, reducing the probability of formation of antiphase domains. Thus, when the concentration of oxygen and carbon on the surface is more than 5% of a monolayer (according to Auger spectroscopy), there are basically monatomic steps on Si (310). This circumstance imposes very high demands on the procedure of preepitaxial preparation and loading the substrate into the vacuum system.

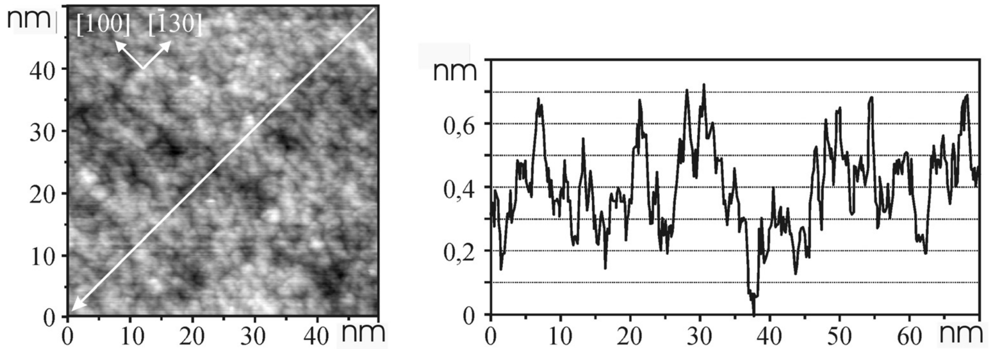


Fig. 4. STM-image of Si(310) surface after annealing at 850°C.

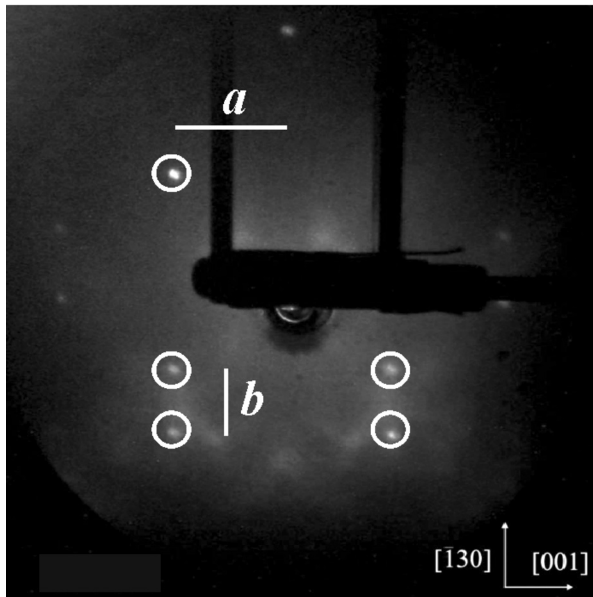


Fig. 5. Diffraction pattern of Si(310) surface after a desorption of a passivation layer at 800°C. *a* - the distance between the rows of reflections; *b* – the distance between the paired reflections.

However, despite the presence of diatomic steps on Si(310), antiphase domains could occur in HgCdTe/CdTe/ZnTe/Si(310). Figure 6 shows TEM - images of CdTe surface containing the domains obtained in the pole (100). Pictures of microdiffraction received from adjacent domains are identical indicating that there is no rotation of crystal lattices in the relevant fields. As can be seen in Fig.4c, mutually perpendicular stripes with a period of approximately 18 nm elongated along [110] are observed at high magnification in the adjacent domains. Spot contrast observed along the stripes might arise due to decorating of the most active sites of CdTe growth surface. The nature of the selected linear irregularities as well as the nature of the decorating particles will not be discussed in this paper. Analysis

of images at the poles (301) and (100) allows to say that the direction of leased lines coincides with the traces of intersection of (110)-planes with the surface of the film. Since the surface of CdTe in the MBE growth conditions are usually terminated by one sort of atoms, the fact that leased lines in the domains are mutually perpendicular (at observation at the pole (100)) means that the domains are antiphase domains.

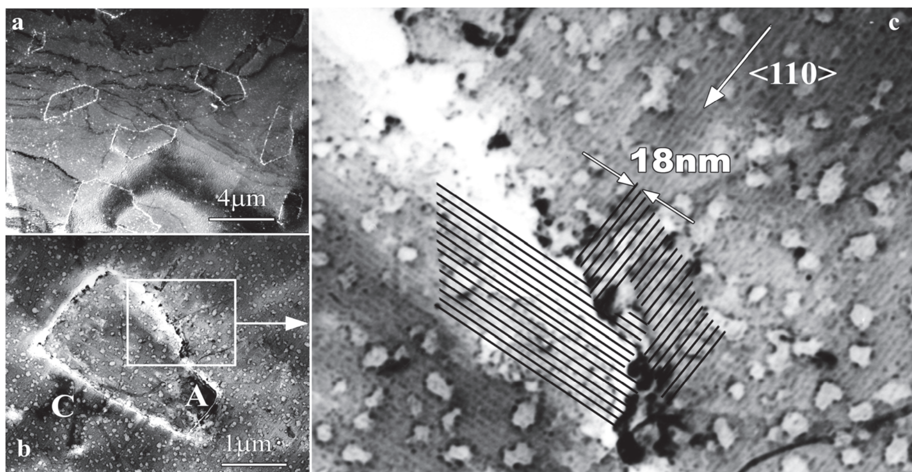


Fig. 6. TEM-images of the subsurface area of CdTe/ZnTe/Si(310) heterostucture with antiphase domains obtained in (100) pole (a), (b); (c) – boundary between two domains.

It can also be seen on Fig.6c that the antiphase boundary (APB) is a layer with structural damage. This allows us to identify the antiphase domains by chemical etching in a selective etchant followed by observation under an optical microscope.

CdTe/ZnTe/Si(310) heterostructures with CdTe thickness of 6 - 8 μm and ZnTe thickness of 0.01 - 0.02 μm were grown for establishing the connection between growth conditions of ZnTe/Si(310) and the density of APB. Only the growth conditions of ZnTe were varied, CdTe growth conditions and preepitaxial preparation processes were the same. The grown structures were etched using E-Ag1 etchant.

Comparing the growth conditions of ZnTe with the results of selective etching it was found that there are optimum conditions for the formation of ZnTe/Si(310) heterointerface without antiphase boundaries (see Figure 7a). Such conditions were the following. The substrate temperature was 200 - 220^{0}C with the flux of Te_2 molecules equivalent to pressure (5 - 20) • 10^{-8} Torr and the flux of Zn atoms equivalent to pressure (1 - 10) • 10^{-6} Torr. Deviation from these conditions namely a decrease of Zn flux or an increase of substrate temperature leads to the appearance of antiphase boundaries. Figure 7 shows pictures of the surface of CdTe/ZnTe/Si(310) after selective etching. Samples presented in Figure 7 were obtained under identical conditions except for growth temperature of ZnTe, which amounted to 200, 240, and 280^{0}C (Figures 7a, 7b and 7c, respectively). It is evident from the presented figures that the heightened over optimal temperature of growth of ZnTe leads to the heightened density of APB. A similar sequence pattern was observed in the case when the growth temperature and vapor pressure of tellurium molecules remain unchanged but vapor pressure of zinc atoms reduced. Increasing the flux of Zn atoms above the optimum does not cause the formation of APB.

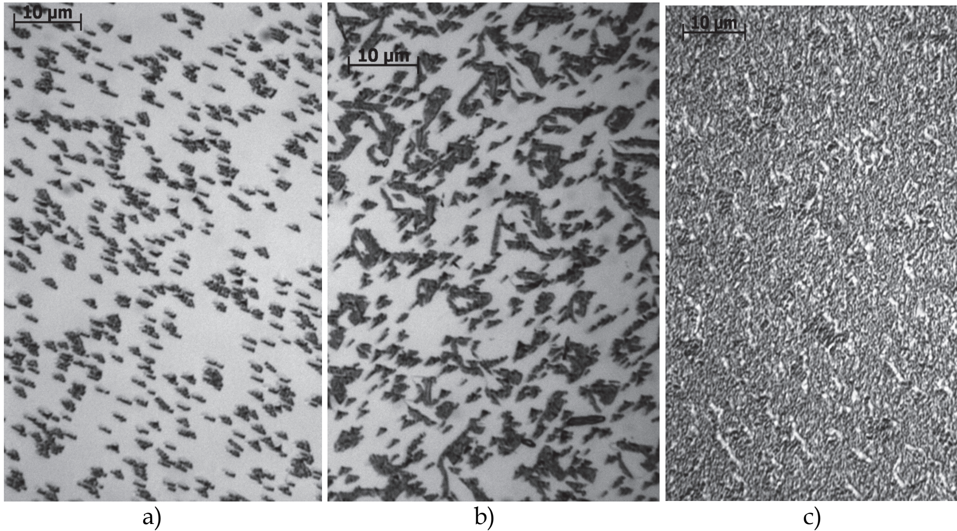


Fig. 7. Etch pits on the surface of CdTe/ZnTe/Si(310) heterostructures grown at different substrate temperatures: 200^{0}C(a), 240^{0}C(b), 280^{0}C(c).

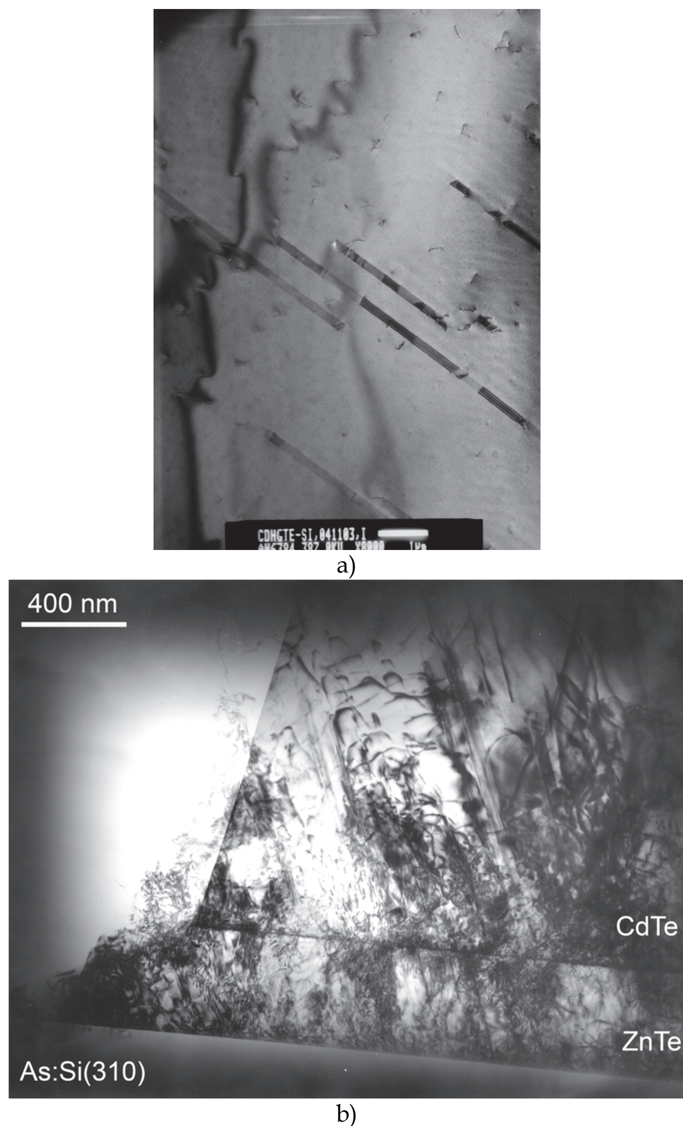
Growth temperature of ZnTe is 300^{0}C and fluxes of both components have the equivalent pressure of 10^{-6} Torr. As can be seen the optimal conditions of growth of ZnTe/Si (310) heterojunction differ substantially from optimal conditions of growth of ZnTe thick layers. The main difference is that in order to suppress the formation of APB it is necessary to decrease the temperature of the substrate and increase the vapor pressure of zinc atoms in the initial moment of growth - that is to create conditions to facilitate the adsorption of zinc. There are two types of antiphase domains for ZnTe/Si heterojunction which differ by alternating atomic layers in the heterojunction - Si-As-Zn-Te and Si-As-Te-Zn. The result can be interpreted in two ways. Firstly, we can assume that the sequence of atomic layers Si-As-Zn-Te is more favorable than Si-As-Te-Zn. Secondly, we can assume that since neither Zn, nor Te is absorbed in the form of a continuous layer we can increase the probability of one of two possible atomic configurations by making a vapor pressure of one component in 1 - 2 orders of magnitude higher than a vapor pressure of the other component.

A number of heterostructures was grown under conditions where the vapor pressure of Te_2 over an order of magnitude higher than the vapor pressure of Zn to verify the second assumption. However, ZnTe layers grown under these conditions had a polycrystalline structure according to RHEED.

The obtained results confirm the conclusion of a preference for one atomic configuration over another. This is apparently due to an excess of valence electrons at the heterointerface in case of realization of the atomic configuration Si-As-Te-Zn.

3.2 Stacking faults

Our studies revealed that there are multilayer stacking faults in HgCdTe/CdTe/ZnTe/Si(310) These defects predominantly have a subtracting type with the density of 10^5 - 10^7 cm^{-2}. (Fig. 8a). Stacking faults lie in closely spaced parallel planes (111) intersecting the plane (310) at an angle of 68.58 degrees. Stacking faults nucleate at ZnTe/Si(310) interface and grow through the entire thickness of the film to its surface (Fig. 8b).



a)

b)

a) – planar image of the subsurface area. b) – cross-section of CdTe/ZnTe/Si(310).

Fig. 8. Stacking faults in HgCdTe/CdTe/ZnTe/Si(310) heterostructure revealed by TEM.

Comparison of patterns of selective etching with TEM images of MCT film allowed to establish the relationship between the shape of etch pits and the microstructure of the film. It was established that the stacking faults revealed in the form of parallel straight lines after the selective etching (Fig. 9a). Ability to identify stacking faults by selective etching can get express information about the density, crystallography, and the distribution of stacking faults over the surface area.

In case of crystals with a sphalerite lattice stacking faults may lie in four {111} planes forming the pyramid with the (310)-plane. Intersection line of planes {111} with the (310)-plane form a quadrangle. It's diagonals lie along the [-130] and [001] (Fig.9b). Sides of this quadrangle are not parallel that allows to distinguish stacking faults occurring in each of four {111} planes. It is observed that there is an anisotropy in the distribution of stacking faults relative to the crystallographic directions [-130] and [001]. According to TEM and selective etching the density of stacking faults in the planes (111)A and (111)B differ by several orders of magnitude, and the lateral size - several times.

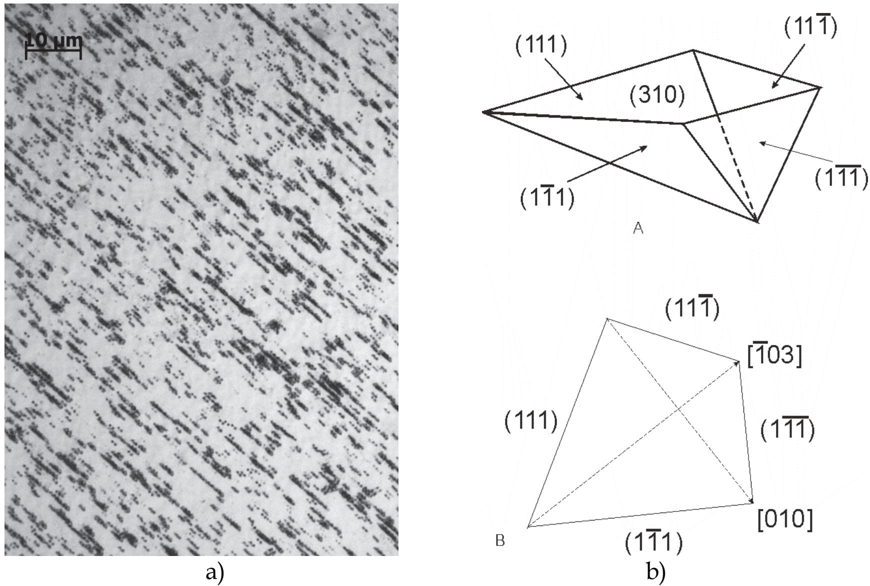


Fig. 9. (a) - stacking faults in HgCdTe/CdTe/ZnTe/Si(310) in optical microscope after selective etching; (b) - pyramid formed by {111} planes and the quadrangle formed by the intersection of the pyramid with the (310).

There are several reasons for the formation of stacking faults in heteroepitaxy of semiconductors with a sphalerite lattice. It must be taken into account that stacking faults occur only in one of four possible {111} planes in HgCdTe/CdTe/ZnTe/Si(310). The proposed model must explain not only the formation of stacking faults but also their anisotropy. Two {111}-type planes form an angle of 43.09^0 with the (310)-plane and the other two - 68.58^0. Stacking faults lying to the boundary at a smaller angle will have a larger area than stacking faults lying at a larger angle with the same thickness of the grown layer. Accordingly, they will have a large excess energy and their formation will not be profitable. The two remaining stacking faults lie in (111) planes with different polarity. It remains to understand why one of the two different polar planes is preferable to another.

The most detailed investigation of formation of stacking faults was performed for ZnSe/GaAs(100) heterostructure. It was established that the maximum density of stacking faults observed in the case when the growth of ZnSe on substrates of GaAs (100) occurs in three-dimensional mechanism [Sidorov et al., 1996]. Stacking faults are formed on the facets

(111) that occur on the slopes of three-dimensional islands [Aoki et al., 2003]. Polarity of (111) plane in which stacking faults lie is given by the vapor pressures of Zn and Se atoms. Growth in excess of zinc leads to the appearance of stacking faults in the (111)A plane whereas growth under conditions of selenium excess causes the formation of stacking faults in the (111) plane [Sabinina et al., 2005]. Stacking faults in $A^{III}B^{V}$ layers grown on Si substrates also have the growth nature [Sidorov et al., 2001].
The lattice mismatch between film and substrate for ZnTe/Si (310) heterostructure is f = 12.3%. Therefore, the formation of stacking faults can be caused by misfit strain in a heterojunction [Rheenen et al., 2006] (formation of partial dislocations at the initial stage of stress relaxation). An indirect confirmation of this fact is, firstly, the type of stacking faults (nearly all stacking faults have subtraction type and formed by sliding of Shockley partial dislocations) and, secondly, the formation of stacking faults occurs in closely spaced parallel planes (111).
The disappearance of stacking faults after annealing may also be associated with a sliding of Shockley partial dislocations in the planes of stacking faults but the sign of these dislocations should be opposite to that of already existing Shockley dislocations. Detailed analysis of the reasons and mechanisms of the disappearance of stacking faults will be carried out in our further studies.
It is the fact for heteroepitaxy of semiconductors with a sphalerite lattice that the stacking faults lie mainly in (111) planes of the same polarity. A model was proposed in [Rogalski, 2003] to explain the anisotropy of stacking faults in ZnSe/ZnS/GaAs(001) heterostructure. According to this model the anisotropy of stacking faults is due to different slip rates of α and β dislocations. After reaching the critical thickness of the growing layer dislocation loops are formed. More mobile α-dislocation quickly reach the interface and form misfit dislocations. Less mobile β-dislocations are split into partial dislocations during the motion to heterojunction. This effect leads to the appearance of stacking faults lying in (111) planes of the same polarity. Without discussing the legality of such an approach to explain the anisotropy of stacking faults we note that in the case of ZnTe/Si(310) heterostructure the critical thickness of pseudomorphic layer is so small that the formation of misfit dislocations occurs in the first monolayer of ZnTe. In this case, the formation of partial misfit dislocations and associated stacking faults may take place. Although the stress is not fully relaxed and misfit dislocations continue to be imposed for some time of growth.
Our previous studies demonstrated that the growth of ZnTe on Si (310) occurs in three-dimensional mechanism and rate of growth is limited by the formation and growth of nuclei [Haakenaasen et al., 2002]. Therefore, in our opinion, in addition to the deformation mechanism there is also the growth mechanism of the formation of stacking faults. Such a mechanism is realized when a coalescence of three-dimensional islands occurs at the initial stage of growth just as it is described in [Aoki et al., 2003; Sidorov et al., 2001]. As it was established in this paper the formation of antiphase domains may be suppressed when the vapor pressure of Zn is more than ten times greater than the vapor pressure of Te. Creating such conditions we seem to do the preferred appearance of (111) facets of the same polarity which leads to the observed anisotropy of stacking faults.
Confirmation of the existence of the growth stacking faults in the HgCdTe/CdTe/ZnTe/Si(310) is the observed correlation between the density of stacking faults in the surface layer of HgCdTe and growth conditions of ZnTe (Fig. 10). A multilayer heterostructure with a total thickness of layers more than 10 microns was chosen as an

object of study due to the following reasons. The method of selective etching was chosen to control stacking faults because of its quickness. Also, this method allows to work with samples of a larger area as compared with the TEM method. We were unable to find a selective etchant revealing stacking faults in ZnTe and CdTe layers. The method of selective etching can be used only in HgCdTe layers. Mismatch of crystal lattices of the substrate and the heterostructure layers is very large so decreasing of a layer thickness is expressed in increasing of the density of threading dislocations until the impossibility to count the density of defects (both dislocations and stacking faults). In addition, the practical interest for production of devices based on MCT heterostructures is the defect density in the absorber layer.

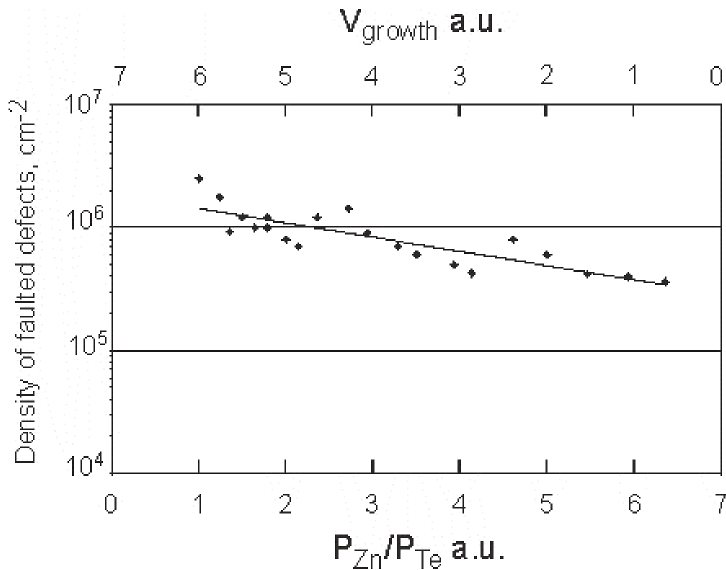


Fig. 10. The dependence of the stacking faults density on growth conditions of ZnTe for HgCdTe/CdTe/ZnTe/Si(310) heterostructures.

The source temperature of Te was varied while growing these heterostructures. Source temperatures of Zn and substrate were kept constant. Regimes of preepitaxial annealing, CdTe buffer layer growth and the growth of MCT were the same. Thickness of the ZnTe layer was in situ controlled by a single-wave ellipsometer. The growth of zinc telluride was stopped at 20 nm. The vapor pressure of Zn was constant and more than ten times greater than the vapor pressure of Te. The growth time of a layer was directly proportional to the relative of vapor pressures of Zn and Te (P_{Zn}/P_{Te}) and inversely proportional to the growth rate of ZnTe (Vgrowth). Figure 10 shows that the density of stacking faults decreases with increasing of relative vapor pressure of Zn and decreasing of growth rate. It was assumed that the growth rate affected the density of stacking faults. As it was mentioned, our previous study showed that the growth rate of ZnTe on Si (310) is limited by the formation and growth of nuclei. A lower rate of growth expresses in a lower density of nuclei (as evidenced by in situ ellipsometric measurements [Vilela et al., 2005]) and correspondingly lower density of nucleation centers of stacking faults.

Our studies have shown that annealing of HgCdTe/CdTe/ZnTe/Si(310) heterostructures in an inert atmosphere at a temperature of 200^{0}C - 250^{0}C for 5 - 10 hours leads to the disappearance of stacking faults in the whole volume of HgCdTe layer. We performed a series of annealings for a more complete understanding of the mechanism of the disappearance of stacking faults. CdTe/ZnTe/Si (310) heterostructures with thicknesses of 6 - 8 μm were annealed in different environments and at different temperatures. Annealings were performed in atmospheres of tellurium, cadmium, and in an inert atmosphere. It was found that annealing in tellurium atmosphere at 350^{0}C leads to the disappearance of stacking faults in the layers of CdTe (according to TEM stacking faults were absent in the entire volume of the heterostructure). Annealing in atmosphere of cadmium and in an inert atmosphere did not lead to reduction of density of stacking faults in CdTe/ZnTe/Si(310) heterostructures.

3.3 Threading dislocations

Crystal structure defects and, especially, misfit dislocations are necessary equilibrium elements of heteroepitaxial structures which provide the stress relaxation. Therefore, avoiding the generation of dislocations during heteroepitaxy is impossible in most cases and sometimes impractical because it can lead to a strong macrodeformation of an epitaxial layer until its destruction. In this regard, one of the important problems of heteroepitaxy is not the elimination of structural defects in general but the creation of optimum distribution by volume of the heterostructure. First of all, we mean the maximum reduction in the number of structural defects in the working area. These defects are the worst thing for electrical and optical properties of the material [Wang and Farlane, 1976].
Density of threading dislocations (N_{DS}) in the epitaxial film decreases from the interface to the surface. This phenomenon can be described by the equation:

$$\frac{dN_{DS}(x)}{dx} = -aN_{DS}(x) - bN_{DS}^2(x) \qquad (1)$$

where x - coordinate along an axis oriented normal to the surface. The interface corresponds to x = 0, and the surface x = h, a and b-constants [Tashikawa and Yamaguchi, 1990].
Dislocations can interact with point defects, locally strained areas, other dislocations during the growth. As a result, they can change directions of propagation and reach a border of the structure. This mechanism corresponds to the first term on the right side of equation (1). The second term corresponds to the reaction between pairs of dislocations with identical Burgers vectors when approaching them to distance of interaction L^0. As a result of the reaction dislocation loops are formed and their subsequent intergrowth in the epitaxial layer becomes impossible (the constant b is proportional to L^0).
The solution of equation (1) has the form:

$$N_{DS}(x) = [\left(N_{DS}(0)^{-1} + b/a\right)exp(ax) - b/a]^{-1} \qquad (2)$$

The interaction of dislocations is the dominant process at small thicknesses (h <50 μm) typical for the MBE and the solution (1) takes the form:

$$N_{DS}(x) = \left[N_{DS}(0)^{-1} + bx\right]^{-1} \qquad (3)$$

Effect of $N_{DS}(0)$ on the value of $N_{DS}(x)$ is relatively small especially when h> 1 μm. The parameter b is determined by the growth conditions. If growth conditions are standard then the curves NDS (x) for various systems are sufficiently close to one another. This is confirmed experimentally in [Sheldon et al., 1988] for InAs / GaAs, GaAs / Ge / Si, GaAs / InP heterostructures grown by MOCVD. In this case, NDS (x) ~ $10^9/x$ at h> 0.5 microns (here the dimension of the N_{DS} (x) – cm^{-2}, x – thickness in microns).
Molecular-beam epitaxy, as a rule, provides the N_{DS} values in surface regions of films at 10^8-10^6 cm^{-2} at h> 5 microns.
At the present time the method of selective etching is used to identify the threading dislocations. The density of etch pits is used as a parameter of structural quality in most studies of photodiodes on the basis of MCT. Figure 11 shows the density distribution of etch pits in layers of CdTe grown on Si (310).
Dependence of the density of etch pits on the layer thickness is satisfactorily described by the expression (3), b = (7.0-9.0)*10^{-5}. That is, the final density of threading dislocations in CdTe / Si (310) heterostructures is determined by reactions between pairs of dislocations with identical Burgers vectors.

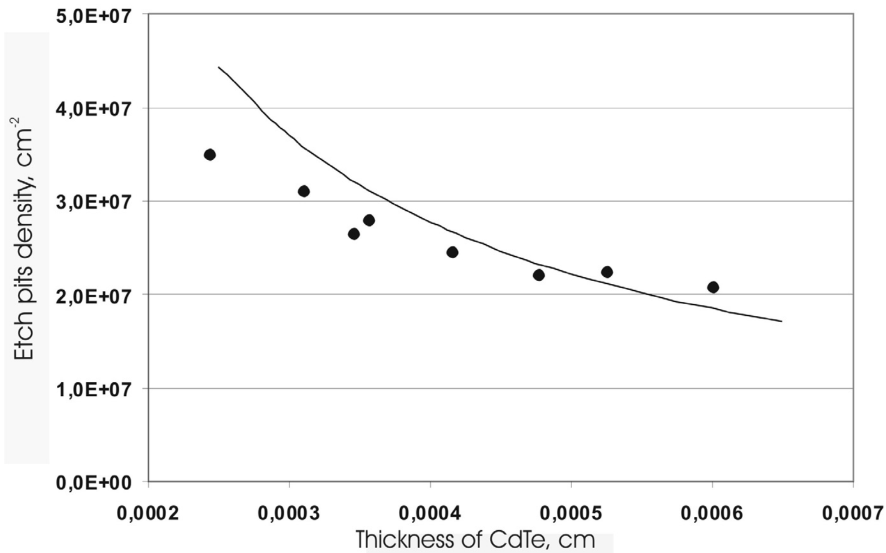


Fig. 11. Etch pits density distribution throughout the thickness of CdTe grown on Si(310). Dots – experimental results, full line – calculated results.

It can be seen from the data presented in Figure 11 that the thickness of the film CdTe h = 5 - 7 microns provides the dislocation density in the surface region $N_{DS} \approx 10^7$ cm^{-2} when grown on Si (310). Further reduction of N_{DS} without increasing the thickness of the film requires serious efforts.

3.4 V-defects

The mercury vapor pressure is less than 10^{-3} Torr at typical growth temperatures of MCT MBE (180-200^{0}C) which does not match the definition of the conditions of molecular beam

epitaxy - molecules or atoms of the deposited material must reach the substrate without collisions with other atoms or elements of the chamber construction.

Thermodynamic analysis shows that the MBE growth of HgCdTe films is carried out in conditions where two phases - $HgTe_{cr}$ and Te_{cr} - are stable [Sidorov et al., 1996]. Figure 12 shows the calculated dependence of supersaturation on the deposition temperature for HgTe (curve 1) and Te (curve 2) at a mercury pressure of 10^{-3} Torr and deposition rates 1 μm / hour. As the temperature decreases below a critical level (T_1) the crystallization of tellurium becomes possible while the crystallization of mercury telluride even impossible. With further temperature decreasing (below T_2) the formation of crystalline mercury telluride is thermodynamically possible but the possibility of deposition of elemental tellurium is also saved.

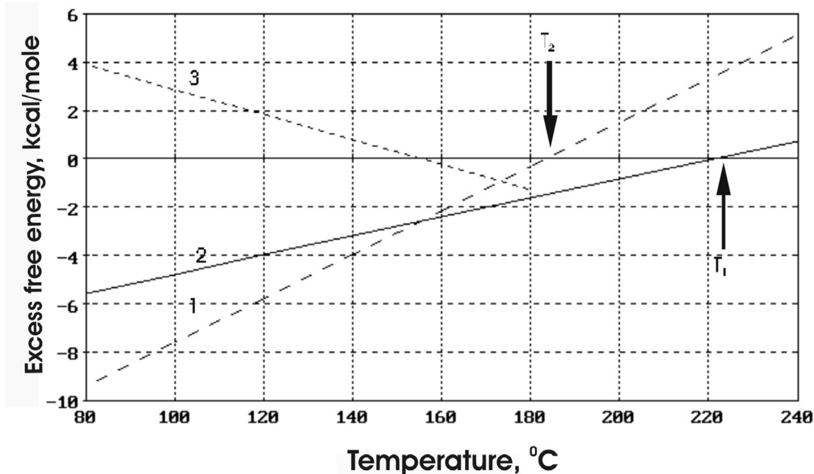


Fig. 12. Temperature dependence of superstauration. Curve 1 – for HgTe (deposition is possible at $T<T_2$). Curve 2 – for Te (deposition is possible at $T<T_1$).

The only stable phase in the temperature range $T_1> T> T_2$ is tellurium. The formation of a polycrystalline film of tellurium is observed at these temperatures by RHEED in situ. HgTe and Te phases are stable simultaneously when $T <T_2$. A predominant formation of one of the phases is determined solely by the kinetics of formation of the corresponding phases when there is a thermodynamic probability of formation of several phases. There is reason to believe that HgTe phase has a higher rate of forming. This fact is possible due to the relatively large vapor pressure of mercury and is indicated by experimental results. During the crystallization of tellurides supersaturation of tellurium is decreased. In the extreme case, when the formation of HgTe is close to equilibrium the vapor pressure of tellurium drops to a value

$$P_{Te2(eff)} = K_{dissHgTe} / P_{Hg}$$

Then the effective supersaturation of Te (Fig. 12, curve 3) decreases.

Thus, if the formation of the HgTe phase does not meet kinetic barriers, the probability of the formation of elemental tellurium phase is reduced. If, however, the crystallization of

HgTe is hampered, the probability of the formation of elementary tellurium increases. The main problem is the fact that tellurium evaporates and reaches the surface of the growing MCT film as a diatomic molecule. The process temperature is so low that tellurium which did not react with mercury and cadmium can not reevaporate.
Figure 13 schematically shows the main possible processes occurring at the surface with the participation of tellurium. Molecules of tellurium involved in two processes: the dissociation of molecules and the crystallization of a perfect MCT film and crystallization of tellurium as a separate phase when the dissociation process does not have time to occur. In the last case, the formation of tellurium phase on the surface breaks the crystal growth of MCT and leads to the avalanche multiplication of defects in the accordance with aforesaid the difficulties in crystallization of MCT in the defect sites increase the possibility of formation of elemental tellurium. As a result, formation of specific threading defects, so-called V-defects (or voids) [Aoki et al., 2003], takes place. These defects avalanchely grow to the surface. Such defects are hallmark patterns of MCT grown by MBE.

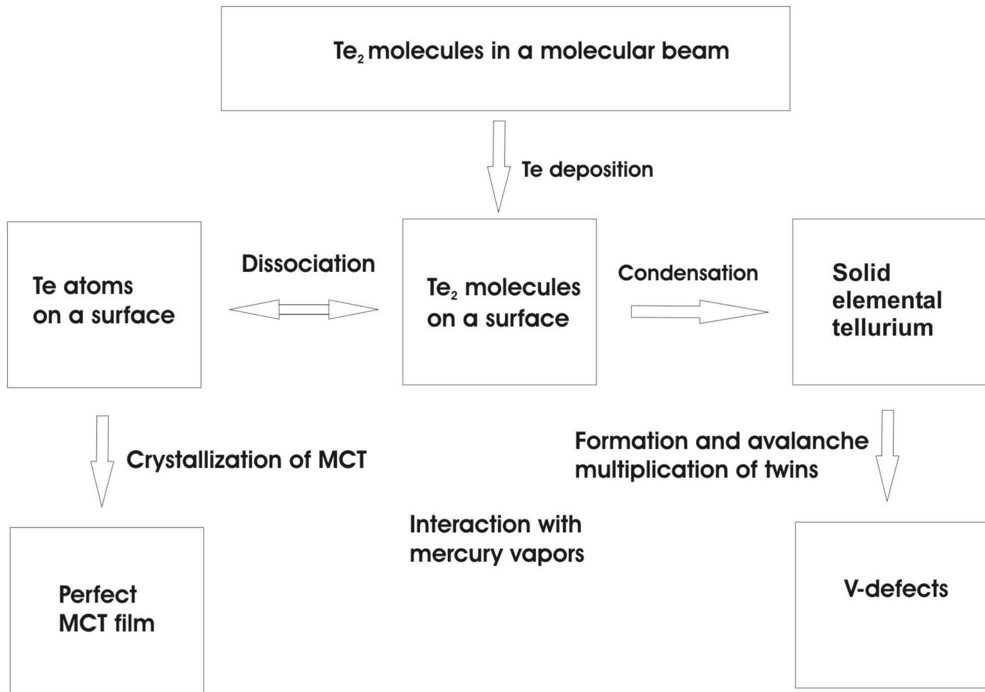


Fig. 13. Processes involving tellurium occurring at the surface during the growth of MCT.

For modern practical device applications MCT MBE with a density of V-defects ~ 10^3 cm^{-2} is used.
The process of growth of MCT film with a low density of V-defects requires precise maintenance of the growth conditions and high surface quality of the buffer layer. At non-

optimal growth conditions like lack or excess of mercury, original inhomogeneity of the substrate surface (relief or a high density of defects which can be linked together) take place there is the possibility of irreversible deterioration of surface and structure of MCT during MBE. Also one of the causes of V-defects is the perturbations of the relief [Sabinina et al., 2005].

Comparison of results of selective etching with a density of macroscopic V-defects allowed to establish a correlation between the density of V-defects and the density of antiphase domains in HgCdTe / Si (310) heterostructures.

Optimized conditions of preepitaxial preparation processes of the substrate and growth of ZnTe and CdTe buffer layers allow to obtain HgCdTe/Si(310) heterostructures without antiphase boundaries. Optimization of the growth process and the absence of antiphase boundaries have reduced the density of morphological V-defects to a value of ~ 1000 cm^{-2}. Also, these defects have the uniform distribution over the surface (Fig. 15a). Fig. 15b shows the appearance of MCT MBE 100 mm in diameter. The structure surface is the mirror-smooth and allows to create photosensitive elements by planar technology.

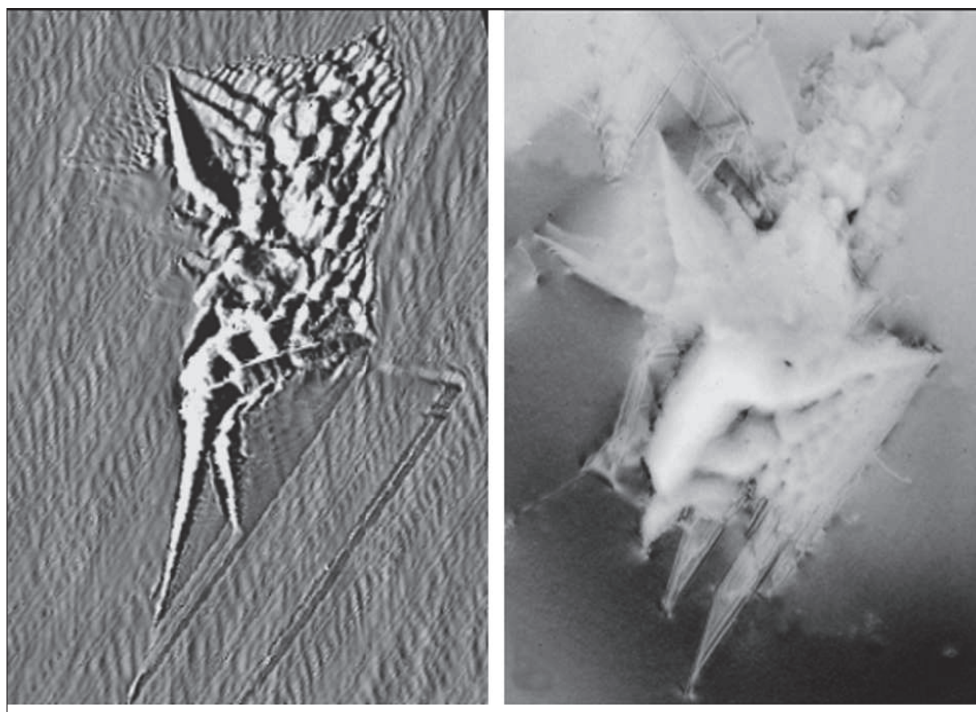
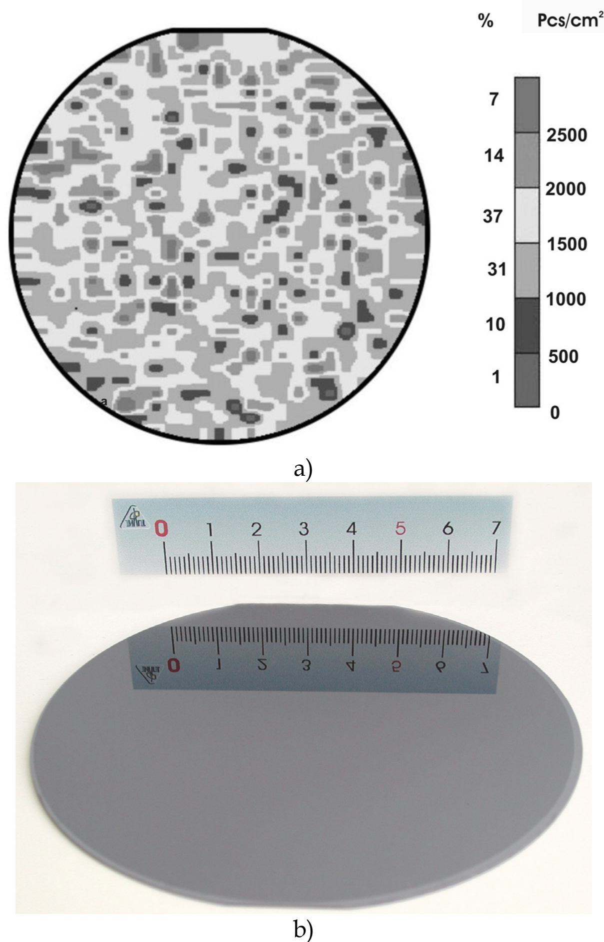


Fig. 14. Typical AFM (a) и TEM (b) 12×12 μm^2 images of a V-defect consisting of stacking faults, twin lamellae и defect structure area on the surface of HgCdTe(310).



a)

b)

Fig. 15. Distribution of V-defects over the area (a) and appearance (b) of HgCdTe/Si(310)

4. Electrophysical characteristics of HgCdTe grown on Si(310) substrates

As-grown undoped MCT films have n-type conductivity regardless of the substrate (GaAs or Si). Structures of p-type conductivity were obtained by isothermal annealing in helium atmosphere at an annealing temperature 230^{0}C, mercury temperature 30^{0}C and the duration of annealing of 20 hours. Ampoules filled with gas (hydrogen or helium) were used for the heat treatments. The dependence on the results of annealing of the type of gas were not observed. Ampoule was placed in a two-zone furnace. One zone is intended to heat a reservoir of mercury, and the second - to heat the sample. Conversion to a p-type conductivity is reversible. Annealing at 230^{0}C and more than 180^{0}C mercury temperature gives again n-type conductivity.
Carrier concentration in n-type films are in the range of 1×10^{14} cm^{-3} to 1×10^{15} cm^{-3} regardless of the composition of grown layer. Calculations of equilibrium concentrations of the donor

centers introduced as intrinsic point defects and impurities show that, in MBE, the equilibrium concentration of donor centers does not exceed the level of 10^7-10^{10} cm^{-3}. Model of nonequilibrium dissolution of defects in the MCT, taking into account a deviation from equilibrium, predicts the increase in concentration of antisite tellurium to values 10^{14}-10^{15} cm^{-3}. Experimental facts on the influence of annealing conditions on the properties of MCT films allow to suggest the presence of mobile acceptor centers with variable concentration. Donor centers can also be presented. Their concentration depends on the growth conditions. It was found that the major donor centers in the films of MCT grown by MBE, apparently, are the tellurium atoms in antisite positions [Sidorov et al., 2001].

The values of electron mobility and lifetime of photoexcited carriers in MCT layers with different compositions can vary by almost two orders of magnitude. The values of electrical parameters of HgCdTe/Si heterostructures with different composition at 77 K are shown in Table 1.

Composition X_{CdTe}	Carrier concentration cm^{-3}	Mobility $cm^2/(Vs)$	Minority lifetime
X=0.22 n-type	$(1\text{-}10)\times10^{14}$	30000-70000	0.2-1.0 μs
X=0.22 p-type	$(5\text{-}15)\times10^{15}$	200-400	10-20 ns
X=0.3 n-type	$(1\text{-}10)\times10^{14}$	15000-30000	5-15 μs
X=0.3 p-type	$(5\text{-}15)\times10^{15}$	200-300	35-50 ns

Table 1. Electrical characteristics of HgCdTe/Si at 77K

The majority mobility and minority lifetime of charge carriers in heterostructures CdxHg1-xTe/Si are somewhat lower than in the MCT layers grown on lattice-matched substrates. Especially noticeable difference is observed for the n-type conductivity.

It was established on the example of heterostructures with composition x = 0.3 that the density of stacking faults and misfit dislocations influence on the mobility of electrons in the structures. Figure 16 shows the corresponding dependences.

It is seen that the mobility depends weakly on the density of dislocations. Dependence of the carrier mobility on the density of stacking faults can be divided into three areas. When the density of stacking faults is less than 2.5×10^6 cm^{-2} (area 1 on the chart) it is possible to obtain values of carrier mobility close to the theoretical maximum for MCT with composition x = 0.3 (40000 $cm^2V^{-1}s^{-1}$). When the density of stacking faults is in the range from 2.5×10^6 cm^{-2} to 5.5×10^6 cm^{-2} (area 2 on the chart), the carrier mobility varies from sample to sample in a wide range and can take both high enough and low values. Apparently, the density of stacking faults still not large enough to degrade the electrical properties of structures and other factors that limit mobility have high influence. When the density of stacking faults is more than 5.5×10^6 cm^{-2} (area 3 on the chart), the high mobility of the carriers are not observed. We can say that such a high density of stacking faults leads to the degradation of electrical properties. At the same time, it is clear that it is possible to obtain structures with high carrier mobility close to the theoretical limit despite the presence of stacking faults.

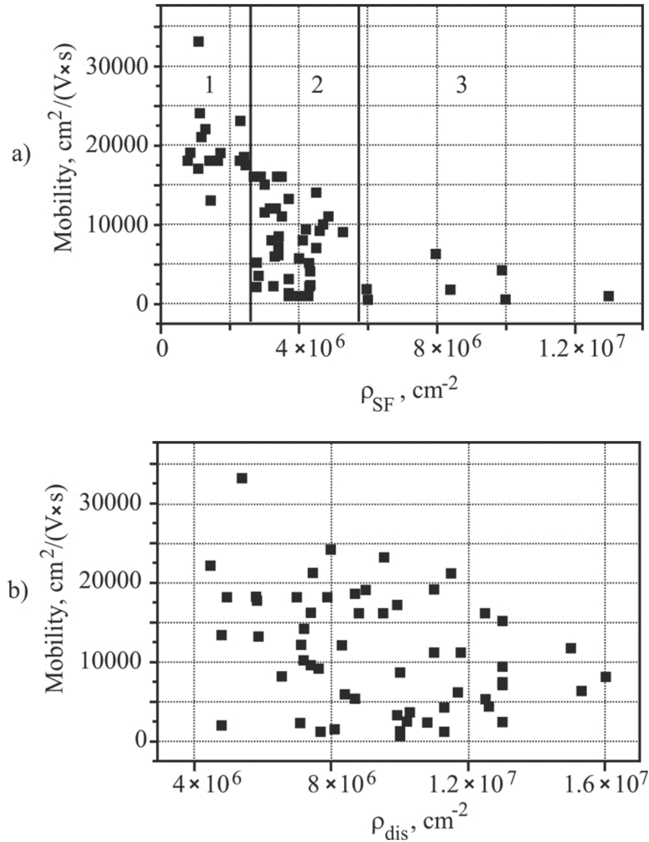


Fig. 16. Dependence of charge carrier mobility on stacking faults density (a) and misfit dislocations density (b).

5. Properties of photodiodes based on HgCdTe/Si(310) heterostuctures

5.1 Mid-wavelength spectral range

Photosensitive arrays 320 × 256 with a step of 30 μm and 640 × 512 with pixel size of 25 μm for the spectral range of 3–5 μm were fabricated from the p-type MCT structures with x = 0.29–0.33 using ion implantation of boron, and their characteristics were measured.

The current–voltage (I–V) characteristics, differential resistance, and ampere–watt sensitivity of photodiodes were measured in a nitrogen cryostat. The measurements were performed for a sample of a matrix photosensitive element with In bumps. One electric contact was constantly connected to a base layer of the photosensitive array, while the second contact was formed via lowering a mobile probe onto a selected photodiode. A photocurrent was measured under illumination from a background at 293 K from the side of In bumps through a ZnSe-based cryostat window (aperture angle θ was 36°).

Figure 14a shows the dependence of the dark current (I_d) under a bias voltage of –100 mV on the inverse temperature for a diode fabricated of the structure with a composition x = 0.328. It is seen that, in a temperature range of 160–300 K, the variation in the dark current is

proportional to n_i^2 and is determined by the diffusion mechanism [Rheenen et al., 2006]. In a temperature range of 140–160 K, I_d is proportional to n_i and is caused by generation–recombination processes in the depletion region.

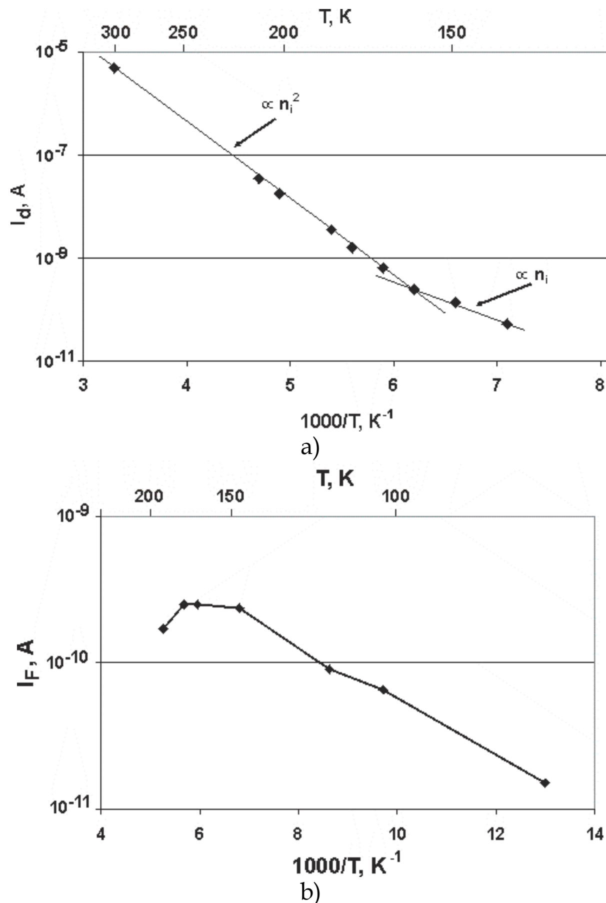


Fig. 14. Dependence of the dark current (a) and photocurrent (b) on the inverse temperature at -100 mV for photodiodes based on $Hg_{1-x}Cd_xTe$ with x=0.328. Dots – experimental results. Full lines – calculated dependencies.

Photocurrent I_F (Fig. 14b) is peaked in a temperature range of 160–180 K. At temperatures lower than 160 K, the photocurrent increases as the temperature is increased, which agrees with other published data. According to [Kuleshov et al., 2005], the diffusion length of minority charge carriers in the MCT-based photodiodes with x = 0.31 continuously increased in a temperature range from 50 to 210 K due to an increase in the lifetime, while the photocurrent is proportional to the diffusion length. The peak of the photocurrent at 150–180 K can be attributed to the effect of variation in the band gap as temperature is changed which leads to a shift of the absorption edge to shorter wavelengths. As the temperature increases from 77 to 200 K, the long-wavelength photosensitivity region

decreases from 4.3 to 4 μm and the photocurrent from the background illumination decreases by a factor of 2. An increase in the diffusion length is insufficient to compensate for a decrease in the photocurrent, which leads to the emergence of a peak at 160–180 K in the temperature dependence of the photocurrent.

Comparing the plots of the dark current and photocurrent and assuming that the photocurrent increases by a factor of 2 upon illumination without shadowing with the In bump and is 0.5 nA, we can find the temperature of equality of the dark current and photocurrent, which is $T \approx 170$ K. Thus, above 170 K, the background-limited mode is not realized.

To characterize the photodiodes, the product of differential resistance at a zero bias (R_0) by its optical area (A) R_0A is often used. The value of R_0 is determined directly from the measured I–V characteristics. To evaluate A, let us use the dependence of the photocurrent on the density of the photon flow and collection area of the photogenerated charge carriers [Rogalski, 2003]:

$$I_F = \eta q Q(\theta) A \tag{4}$$

where η is the quantum efficiency (the number of electron–hole pairs generated by an incident photon), q is the elementary charge, A is the collection area of the photogenerated carriers, and $Q(\theta) = Q(2\pi)\sin 2(\theta/2)$ is the density of the photon flux in the aperture angle θ from an absolutely black body with temperature T =293 K in a wavelength range from 0 to $\lambda_{1/2}$ μm, where, in turn,

$$Q(2\pi) = (2\pi c\lambda^{-4}) / \exp((hc / \lambda kT) - 1) \tag{5}$$

It was taken into account in calculations that the illumination with the background light is equivalent to the use of a black body with emissivity of 0.95, η = 0.7, and the ZnSe window used in the experiment can reflect up to 30% of the incident flux.

Taking into account the measured values of the photocurrent and using formulae (1) and (2), we can determine the collection area of photogenerated carriers A, which was in a range of 100–200 μm^2 for all samples. Physically, this area is a ring around the In bump (we assume that the bump itself is opaque); then, via the addition of the area of the In bump to the obtained value (we assume that the In bump is circular with a radius of 10 μm), we obtain the optical area of the p–n junction A. Multiplying A by the value of differential resistance at the zero bias R_0, we obtain the value of R_0A.

Another important parameter of both the diode itself and its material is the diffusion length of charge carriers. Assuming that the p–n junction is of circular geometry and subtracting its metallurgical radius from the optical radius of the p–n junction, we can determine the diffusion length of charge carriers. In this case, we can take into account that the geometry size of the p–n junction exceeds the size of the window for doping (a circle 10 μm in diameter) due to Hg diffusion by 2–3 μm [Haakenaasen et al., 2002].

The measured values of the photocurrent and differential resistance under the zero bias for the photodiodes fabricated from MCT of various compositions as well as the values of the diffusion length and R_0A calculated based on these parameters are listed in the Table 2.

The obtained estimated values of the diffusion lengths are smaller compared with the results of [Kuleshov et al., 1985] approximately by a factor of 2. The cause of this may be that the authors of [Kuleshov et al., 1985] grew MCT on the lattice-matched CdZnTe substrates and the density of structural defects was considerably lower than for the layers discussed in our study. A decrease in the defect density can lead to an increase in the lifetime and

mobility of photogenerated carriers and, as a consequence, to an increase in the diffusion length. Our values of R_0A do not exceed the results given in publications concerned with MCT photodiodes grown on Si substrate [Vilela et al., 2005].

Sample	X_{CdTe}	$\lambda_{1/2}$, μm	Photocurrent, A	R_o, Ohm	R_oA, Ohm·cm^2	$L_{diff.}$, μm
MCT090316	0.327	4.3	$2.0 \cdot 10^{-11}$	$4 \cdot 10^{11}$	$1.7 \cdot 10^{6}$	4.5
MCT081023	0.328	4.3	$3.5 \cdot 10^{-11}$	$3 \cdot 10^{11}$	$1.5 \cdot 10^{6}$	5.5
MCT090305	0.289	5.2	$2 \cdot 10^{-10}$	$2 \cdot 10^{11}$	$1.0 \cdot 10^{6}$	5.7
MCT090302	0.293	5.1	$1.5 \cdot 10^{-10}$	$2 \cdot 10^{11}$	$9.7 \cdot 10^{5}$	5.4

Table 2. Photoelectric properties of photodiodes based on MCT heterostructures with different composition

5.2 Long-wavelength spectral range

Photosensitive arrays 288 × 4 of standard topology [Vasiliev et al., 2004] with a detector pitch in scan of 43 μm for the spectral range of 8–12 μm were fabricated from the p-type MCT structures using ion implantation of boron, and their characteristics were measured. Pixel size is 28×25 μm.

The dark current-voltage characteristic of the real photodiode is formed as a result of the superposition of several components caused by different mechanisms. Currently, there are the following mechanisms:

- Diffusion current
- Generation-recombination current
- Tunneling current (band-to-band tunneling and trap-assisted tunneling)
- Surface currents

Diffusion current is a fundamental mechanism of charge transport in photodiodes based on p-n junctions. The total density of the diffusion current is determined by the contribution of the electron and hole currents from both sides of the junction.

Generation-recombination current can exceed the diffusion current, especially at low temperatures, although the width of the space charge region is much smaller than the diffusion length of carriers. Generation rate in the depletion region strongly depends on the applied bias voltage and may greatly exceed the rate of generation in the bulk material.

Tunneling current is caused by electrons tunneling directly through the junction from the valence band into the conduction band (direct tunneling) or through the trap levels in the junction region. The latter is a two-stage process in which the first phase is the thermal transition between one zone and a trap, and the second is the tunneling between a trap and the other zone. Tunneling process in this case occurs at lower fields compared to the direct band-to-band tunneling as electrons tunnel at a shorter distance. The tunneling current depends strongly on the band structure, the applied bias voltage, the effective dopant concentration and weekly on the temperature and the shape of the barrier in the junction.

The most controversial contribution to the formation of the dark current of a real diode is made of surface effects. Component of the dark current associated with the surface may depend not only on the type of passivation layer and method of its coating but also on the quality and composition of the MCT material.

Each component of the current depends on voltage and temperature in its own way. Many researchers suggest that only one mechanism is dominant in a particular bias range . This

method of analysis of current-voltage characteristics is not always correct. The best solution is the numerical simulation of the superposition of current components at different temperatures and bias voltages using the experimental data.
Diffusion current in n+-p junction can be described by the following equation:

$$I_{dif} = \frac{qAn_i^2}{N_a}\left[\frac{kT}{q}\frac{\mu_e}{\tau_e}\right]^{1/2} C\left(\exp\left(\frac{qV}{kT}\right)-1\right) \tag{6}$$

where A – area of the junction, N_a – acceptor concentration in lightly doped p-region, n_i – intrinsic concentration, μ_e and τ_e – electron mobility and lifetime respectively, V – bias voltage, C - factor associated with the surface recombination rate S.
Formula (6) describes the electron current from the p-type region in n-type region. If need to take into account the hole current from the n-type region in p-type region the parameters of the electrons should be replaced by the relevant parameters of the holes.
The following equation (reverse bias) was used to calculate the current component due to generation in the depleted region [Schoolar et al., 1992]:

$$I_{g-r}[V<0] = \frac{qAn_iW_{dep}V}{V_t\tau_{g-r}} \tag{7}$$

where τ_{g-r} – generation-recombiantion lifetime depending on a trap concentration N_t, $V_t=(V_{bi}-V)$ – the full potential of the junction, V_{bi} – built-in voltage of the junction, W_{dep} – width of the depletion region. In case of forward bias:

$$I_{g-r}[V>0] = \frac{2An_iW_{dep}kT}{V_t\tau_{g-r}}\sinh\left(\frac{qV}{2kT}\right) \tag{8}$$

Another component of the dark current is tunneling through the traps. According to [Gopal et al., 2001]:

$$I_{tat} = qAN_tW_{dep}W_cN_c \tag{9}$$

where N_t – trap concentration, W_cN_c – tunneling probability.
Current of band-to-band tunneling can be described by the following equation [Gopal et al., 2001]:

$$I_{btb} = \frac{\sqrt{2m_e}\,q^3E(V-V_{bi})A}{4\pi^3\left(\frac{h}{2\pi}\right)^2\sqrt{E_g}}\exp\left(-\frac{\pi\sqrt{m_eE_g^3}}{2\sqrt{2}qE\frac{h}{2\pi}}\right) \tag{10}$$

where m_e – electron effective mass, h - Planck's constant, E – electric field in depletion region.
Boron is not electrically active impurity in our p-n junction. Ion implantation with boron causes the area of mercury enrichment at the surface of p-type MCT due to radiation defects. Mercury diffuses from this region into the bulk and restores the original n-type conductivity at a certain depth filling the vacancies in the metal. Based on this mechanism of the formation of p-n junction, measured characteristics of MCT layers were used for calculating

the parameters of the photodiode. Characteristics are listed in Table 1. The thickness of the n-layer was taken equal to 2 microns. Due to the lack of detailed information on the energy of traps in the material we consider a single Shockley-Read-Hall level in the mid of bandgap which is quite widely used method [Rogalski, 2003]. Capture cross sections of electrons and holes are assumed to be $\sigma_n = 10^{-16}$ cm^2 and $\sigma_p = 10^{-17}$ cm^2 respectively.
Experimental and calculated current-voltage characteristics for diodes fabricated from MCT with x = 0.231 (λc = 9.06 μm) are presented in Fig. 15.

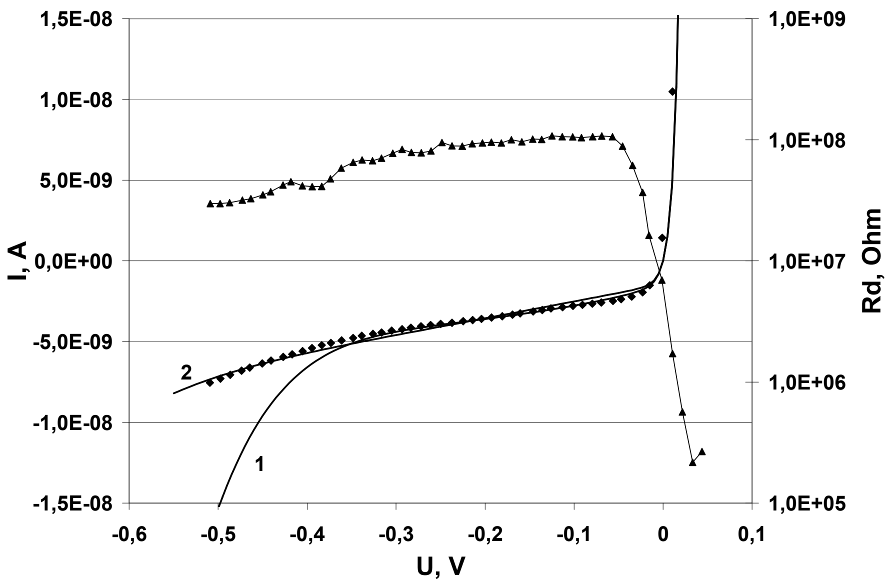


Fig. 15. Current-voltage and differential resistance-voltage characterisitics for a typical diode with λ_c=9.06 μm. Dots - experimental results; full lines – calculations.

Diffusion current is a fundamental value that does not depend on the applied reverse bias. Contribution of the diffusion current in the dark current is negligible and is about 10^{-11} A for given parameters of the MCT material. The band-to-band tunneling is largely dependent on the concentration of charge carriers on both sides of the junction and the applied bias voltage. Varying the concentration of electrons and holes within reasonable limits shows that the contribution of this mechanism in the total dark current can also be neglected at low bias voltages up to -0.6 V.
The main contribution to the total amount of dark current in reverse bias less than 0.6 V is determined by two mechanisms: the generation in the depletion region and tunneling through traps. The generation-recombination current is also dependent on the concentration of traps. The contribution of other mechanisms is negligible.
The dependence of the dark current on the reverse bias shown in Fig. 15 has a pretty strong inclination that can mean the predominance of the generation-recombination current. That in turn means a high concentration of traps. Unfortunately, we will inevitably obtain high tunneling currents in reverse bias over 0.4V if substituting the concentration of traps provides the observed slope in the calculation. This effect is not observed in the

experimental dependence (curve 1). Varying the other parameters used in the calculations (density, composition, thickness) within the measurement error does not allow to obtain a satisfactory agreement between the experimental and calculated data. Hence, there is at least one other parameter besides the concentration of traps that has a significant influence on the behavior of current-voltage characteristics.

The experimental curves can be reliably described if assuming that a shunt resistance R_{sh} which obeys Ohm's law is parallel connected to the diode. In this case, the total current will be determined by the following expression:

$$I_{sum} = I_{dif} + I_{g-r} + I_{tat} + I_{btb} + \frac{V}{R_{sh}} \qquad (11)$$

Calculated current-voltage characteristic (CVC) obtained from the formula (11) (R_{sh} = 1.1 × 10^8 ohms, and N_t = 2.6 × 10^{15} cm^{-3}) corresponds to curve 2 in Figure 15. It is seen that the calculated CVC repeats the experimental with this approach.

The nature of leakage through R_{sh} remains open and requires further study. We assume that the shunting of the p-n junction is due to threading dislocations. The R_0 value was determined directly from the I–V curves, while the A value was evaluated from the measured photocurrent as described in [Gopal et al., 2001]. Figure 16 presents a plot of the R_0A versus cut-off wavelength λ_c in the photodiodes based on HgCdTe/Si(310) heterostructures operating at 77 K. As can be seen, the R_0A product of these photodiodes in a 8–12 μm wavelength range is below the upper values calculated assuming the limitation by thermal generation (curve 1), while significantly exceeding the values determined for a regime limited by the background noise (curve 2) [Rogalski, 2003]. Thus, the obtained data are indicative of a high quality of the material, which is a necessary prerequisite for the development of multielement IRFPAs.

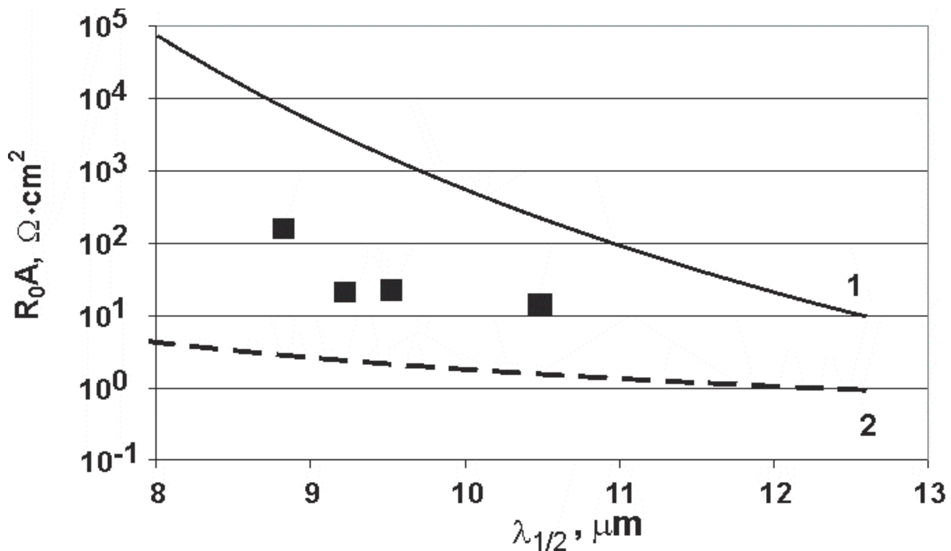


Fig. 16. Dependence of R_0A product on cutoff wavelength for HgCdTe/Si(310)-based photodiodes at 77 K.

6. Properties of HgCdTe/Si(310-based) photodetector arrays

6.1 Mid-wavelength spectral range

Infrared focal-plane arrays of formats of 320 × 256 and 640 × 512 elements for a spectral range of 3–5 μm were fabricated based on photodiode photosensitive elements and a Si multiplexer by the method of hybrid assembly through In bumps, and their characteristics were measured. The temperatures of the sample, background and absolute blackbody were 78, 293, and 501 K, respectively. The measurements were performed in an aperture angle of 56°, the pixel output rate was 2.0 MHz, and the integration time was 640 μs.

The measurements showed that the amount of defect elements in FPAs does not exceed 2.5%, while for the best IR FPAs it is 1%. We assumed that the defect elements are photodiodes with response differing from the average value for more than by 35%, while the threshold irradiation exceeded the average value by a factor of larger than 3. The defect elements are uniformly distributed over the FPA area and do not form clusters in the central part (Fig. 17). The average values and deviation of the volt sensitivity and the noise-equivalent difference in temperature (NEDT) (Fig. 18) are close to the limiting values for these measurement conditions.

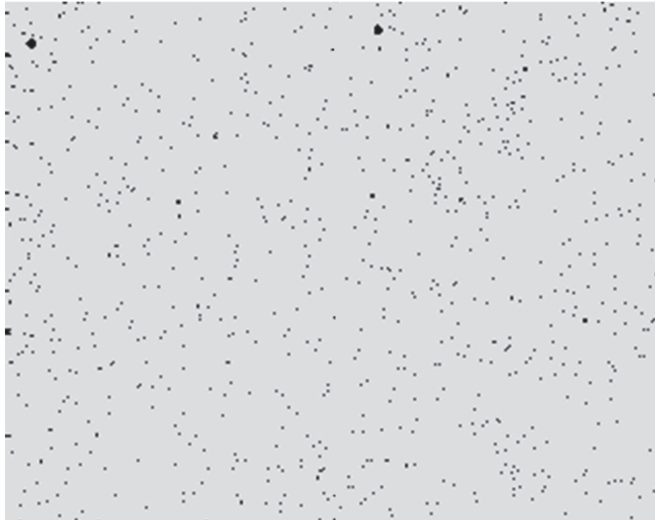


Fig. 17. Topograph of defect elements of 320 × 256 array with c $\lambda_{1/2}$ (77K) = 5.2 μm. The number of defect elements is 0.4% (19 pieces) in the central part of the format 80 × 64 elements

The thermal image obtained using a model of a thermal vision channel based on a photodetector devices of sizes of 320 × 256 and 640 × 512 elements is visually observed in real time, and a characteristic temperature distribution over a human face is observed in the image (Fig. 19).

We studied the effect of temperature cycling from 77 to 300 K on the parameter of the IR FPA ($\lambda_{1/2}$(77 K) = 4.3 μm) based on MBE-grown MCT heteroepitaxial structures on a Si substrate of a size of 320 × 256 elements. The dependences of NEDT and the number of defect elements on the number of the temperature-variation cycles are presented in Fig. 20.

It is seen that, taking into account the measurement error, the average value of NEDT was almost invariable after more than 2500 cycles. The number of defect elements insignificantly increased from 2.25 to 2.9% after the first 400 cycles and was further invariable. The presented results show high stability of the IR FPAs to thermocycling.

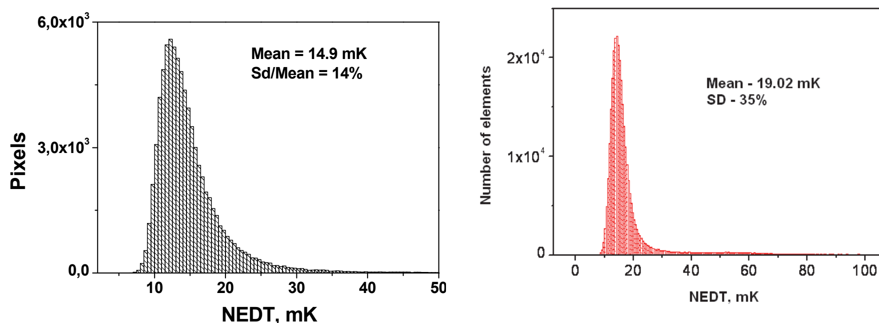


Fig. 18. NEDT histogram for MCT-based 320×256 array with $\lambda_{1/2}$ (77K) = 5.2 μm (a) and 640×512 array with c $\lambda_{1/2}$ (77K) = 4.1 μm (b).

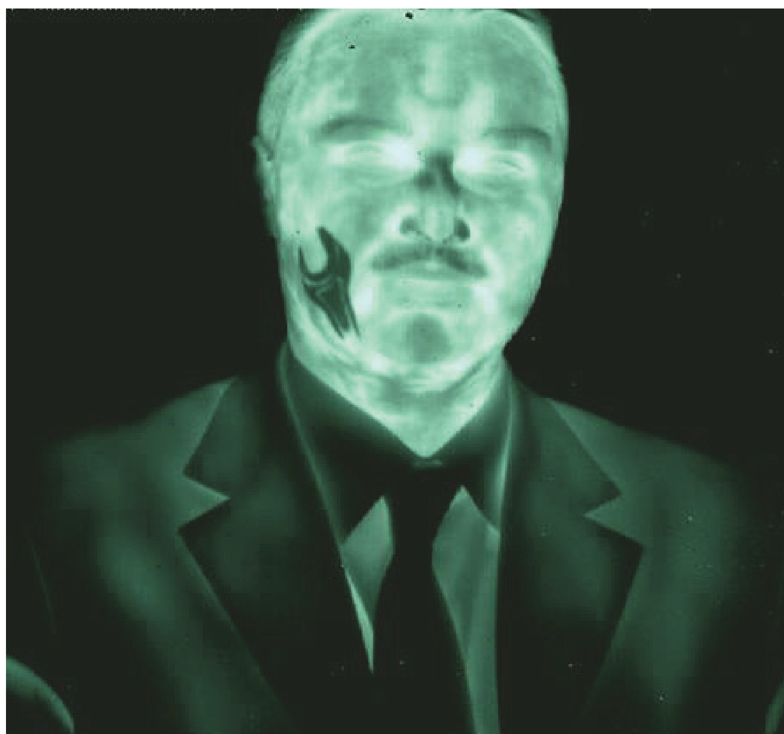


Fig. 19. Thermal image obtained from 640×512 array based on HgCdTe/CdTe/ZnTe/Si(310)

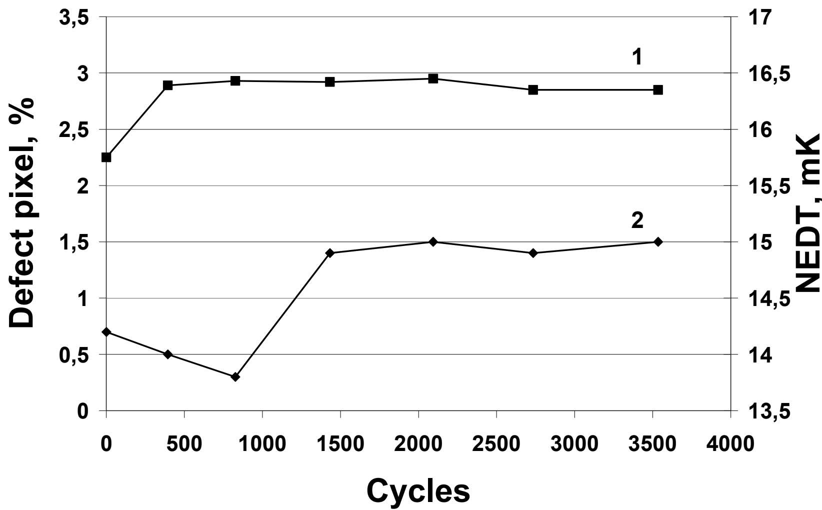


Fig. 20. Dependencies of number of defect elements (1) and NEDT (2) on number of thermocycles from 77 to 300K.

6.2 Long-wavelength spectral range

Using these photosensitive arrays in a hybrid assembly with silicon multiplexers, IRFPAs of the 288 × 4 format were manufactured for a wavelength range of 8–12 μm [λc(77 K) = 9.5 μm].

The IRFPAs of the 288 × 4 format were provided with a silicon multiplexer possessing an original scheme and special design, the distinctive features of which are the fully digital control via parallel and serial interfaces, possibility of deselecting any defect cell, bidirectional scanning of pixels, and possibility of testing the analog parameters.

The multiplexer was manufactured using a commercial 1-μm CMOS technology with two metal and two poly-Si levels [Sizov et al., 2006]. The gate and spacer oxide layers were 40 and 90 nm thick, respectively, with the corresponding specific capacitances of 8.65×10^{-4} and 3.8×10^{-4} pF/μm^2. The direct injection was provided by a subdoped n-channel transistor with a threshold voltage of ~0.7 V and a channel length increased to 2.4 μm, which ensured a spread of the bias voltage on the diodes not exceeding 10 mV.

The multiplexer was subdivided into four identical blocks, each of 72 × 4 channels multiplexed to four outputs. The deselect trigger and output device make possible the exclusion of defect diodes and 8-fold variation of the gain. A commutator ensures alteration of the scan direction and direct access to diodes bypassing the TDI tract. A charge-sensitive amplifier provides the charge/voltage conversion and a read-out integrated circuit (ROIC) ensured the gain and storage of analog signals during multiplexing. A charge capacity of the proposed multiplexer is greater than 2.5 pF at a nonlinearity not exceeding 2%.

The performance of this multiplexer with respect to the main functional and electrical characteristics is close to those of a BD TL015-XX-V3 (Sofradir) multiplexer that is employed in Pluton LW 288 × 4 FPAs.

The parameters of IRFPAs were measured at a device temperature of 77 K, a background temperature of 293 K, and a blackbody temperature of 501 K. The measurements were

performed at an aperture angle of 30°, the pixel output rate of 4.0 MHz, and a signal integration time of 18 µs. The tests were carried out with a reject filter having a long-wavelength cut-off 8.0 µm.

Figure 21 shows distributions of the main parameters of IRFPA of the 288 × 4 format based on CMT heteroepitaxial structures grown by MBE on a Si(310) substrate with

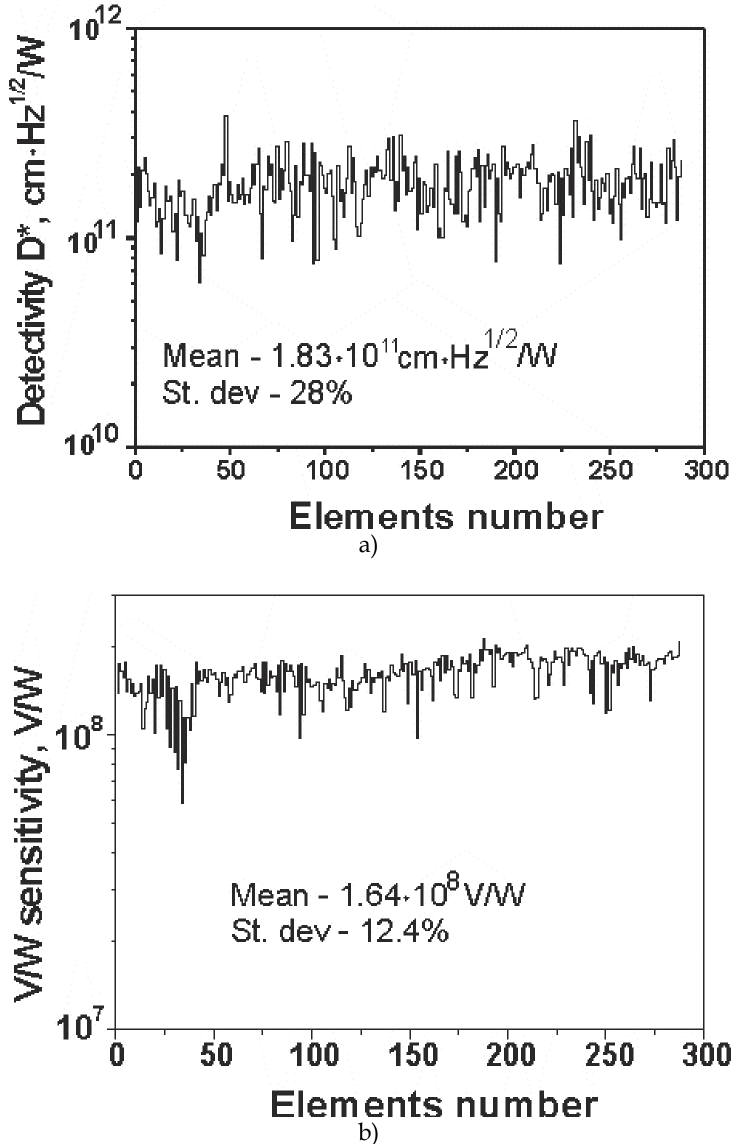


Fig. 21. Topogram of the specific detectivity (a) and voltage sensitivity (b) of 288×4 FPA's channels (λ_c (77K) = 9.5 мкм).

λc(77 K) = 9.5 μm. As can be seen, all 288 channels are photosensitive and their characteristics are not inferior to those of the analogous IRFPAs based on MCT grown on lattice-matched CdZnTe substrates. Indeed, the mean specific detectivity (D*) of the proposed 288 × 4 IRFPA based on HgCdTe/Si(310) heterostructure amounts to 1.83×10^{11} cm*Hz$^{1/2}$/W at a standard deviation of 28%, while the voltage sensitivity is 1.64×10^{8} V/W at a standard deviation of 12.4%. Analogous parameters reported for the IRFPAs based on lattice-matched MCT/CdZnTe heterostructures [Reddy et al., 2008] exceed these values by no more that 5%.

When the linear IRFPAs are employed in technical imaging systems, additional requirements are imposed on the homogeneity of parameters of the elements. In long-wavelength FPAs, channels with a specific detectivity below 5×10^{10} cm*Hz$^{1/2}$ /W are conventionally classified as defect elements. According to this criterion, the presented 288 × 4 IRFPA based on HgCdTe/Si(310) heterostructure has no defect channels, since the minimum specific detectivity is about 6×10^{10} cm*Hz$^{1/2}$/W. Another criterion is the voltage sensitivity, which must fall within ±30% of the mean value. In this respect, the proposed IRFPA has 12 defect channels. However, if the boundaries of admissible variation of the sensitivity are expanded to range from +30% to –60%, then the proposed device is also free of defect channels with respect to sensitivity.

It is believed that the proposed linear IRFPA is characterized by a high stability of parameters with respect to temperature cycling from 77 to 300 K. Indeed, tests that have been previously carried out for 288 × 4 IRFPAs based on HgCdTe/Si(310) heterostructures for a 3–5 μm wavelength range showed that they were highly stable with respect to temperature cycling and their photoelectric parameters remained unchanged upon 1250 cooling/heating cycles.

In conclusion, the present investigation showed that undoped hole-type HgCdTe layers with x =0.23 grown by MBE on Si(310) substrates ensure high photoelectric parameters (limited by background radiation) of IRFPAs of the 288 × 4 format for long-wavelength (8–12 μm) IR spectral range.

7. Conclusion

Investigations of growth processes of MCT MBE on Si(310) substrates for 3rd generation IRFPAs were carried out.

It is shown that the optimization of processes of surface preparation and growth conditions allows to obtain MCT MBE on Si (310) without antiphase domains. Optimization of the growth process and the absence of antiphase boundaries allowed to reduce the density of morphological V-defects to ~ 1000 cm^{-2}.

A technology for the device quality undoped p-type MCT was developed.

It was demonstrated that HgCdTe/CdTe/ZnTe/Si(310) hetrostructures could be used to create reliable, resistant to thermal cycling IRFPAs for the spectral ranges 3-5 and 8-12 μm.

8. References

Aoki T., Chang Y., Badano G., Zhao J., Grein C., Sivananthan S., and David J. Smith., (2003) *J. Electron. Mater.*, Vol. 32, p. 703

Bhan R.K., Dhar V., Chaudhury P.K. et al. (1996) *Appl. Phys. Lett.*, , Vol. 68, No. 17, pp.2453-2454

Blinov V.V., Dvoretsky S.A., Sidorov Yu.G. (1997) Patent of Russian Federation №2071985. Priority from 11.01.1993. Registration 20.01 1997. Bulletin №2 from 20.01.97. (in Russian)

Blinov V.V., Goryaev E.P., Dvoretsky S.A., et al. (1997) Claim for invention № 95102853/25, priority from 01.03.95. Positive solution from 20.08. 1997. (in Russian)

Buldygin A.F., Vdovin A.V., Studenikin S.A., et al. (1996) *Avtometriya*, No. 4, pp.73-76 (in Russian)

Carmody M., Pasko J.G., Edwall D., Piquette E., Kangas M., Freeman S., Arias J., Jacobs R., Mason W., Stoltz A., Chen Y., and Dhar N.K. (2008). Status of (LWIR) HgCdTe-on-Silicon FpA Technology. *J. Electron. Mater.*, Vol. 37(9), p. 1184

Fenner D.B., Biegelsen D.K., Bringans R.D. (1989) Silicon surface passivation by hydrogen termination: A comparative study of preparation methods. *J.Appl. Phys.*, Vol. 66, p.419

Gopal V., Singh S.K. and Mehra R.M. (2001). Excess dark currents in HgCdTe p+-n junction diodes, *Semicond. Sci. Technol.*, Vol. 16, pp. 372-376

Haakenaasen R., Moen T., Colin T., Steen H., and Trosdahl-Iversen L. (2002) Depth and lateral extension of ion milled pn junctions in CdHgTe from electron beam induced current measurements. *J. Appl. Phys.*, Vol. 91, p. 427.

Kern W., Puotinen D.A. (1970) Cleaning solutions based on hydrogen peroxide for use in silicon semiconductor technology. *RCA rev.*, Vol. 31, p. 187

Koestner R. J., Schaake H.F. (1988). Kinetics of molecular-beam epitaxial HgCdTe growth. *J. Vac.Sci.Technol.*, Vol. A 6, No. 4, p. 2834

Kuleshov V.F., Kuharenko Yu.A., Fridrihov S.A., et al. (1985) Spectroscopy and electron diffraction in the study of solid surfaces, Nauka, Moscow (in Russian)

Reddy M., Peterson J.M., Lofgreen D.D., Franklin J.A., Vang T., Smith E.P.G., Wehner J.G.A., Kasai I., Bangs J.W., and Jonson S.M., (2008). MBE Growth of HgCdTe on Large-Area Si and CdZnTe Wafers for SWIR, MWIR and LWIR Detection, *J. Electron. Mater.*, Vol. 37, No. 9, p. 1274.

Remesnik V.G., Mischenko A.M., Mihaylov N.N. (1994) *Invention*, №20. Patent of Russian Federation №2022402, priority from 30.10.94, Bulletin № 20 from 30.10.94. (in Russian)

Rheenen A.D. van, Syversen H., Haakenaasen R., Steen H., Trosdahl-Iversen L. and Lorentzen T. (2006) Temperature dependence of the spectral response of lateral, MBE-grown, ion-milled, planar, HgCdTe photodiodes. *Phys. Scr.* ,Vol. T126, p.101.

Rogalski A. (2003) *Infrared Detectors (Electrocomponent Science Monographs, Volume 10),* CRC Press

Ryu Y.S., Song B. S., Kang T.W., Kim T.W. (2004) Dependence of the structural and the electrical properties on the Hg/Te flux-rate ratios for $Hg_{0.7}Cd_{0.3}Te$ epilayers grown on CdTe buffer layers, *J. Mater. Sci.*, Vol. 39, p. 1147

Sabinina I.V., Gutakovsky A.K., Sidorov Yu.G., Latyshev A.V. (2005) Nature of V-shaped defects in HgCdTe epilayers grown by molecular beam epitaxy. *J. Crystal Growth*, Vol. 274, p. 339

Schoolar R., Price S., Rosbeck J. (1992). *J. Vac. Sci. Technol.*, Vol. *B* 10, pp. 1507–1514

Sheldon P., Jones K.M., Al-Jassim M.M., Yacobi B.G. (1988) Dislocation density reduction through annigilation in lattice-mismatched semiconductors grown by MBE. *J.Appl. Phys.*, Vol.63, No.11, pp.5609-5611.

Sidorov Yu.G., Varavin V.S., Dvoretsky S.A. et al. (1996) *In Growth of Crystals*, Vol.20, pp.35-45.

Sidorov Yu.G., S.A. Dvoretsky, Mihaylov N.N., Yakushev M.V., Varavin V.S., Antsiferov A.P. (2000). Molecular beam epitaxy of narrow-band materials CdHgTe. Equipment and technology. *Opticheskiy zhournal*, Vol. 67, No.1, p. 39 (in Russian)

Sidorov Yu.G., S.A. Dvoretsky, Mihaylov N.N., Varavin V.S., (2001) Physico-chemical and technical basis of molecular-beam epitaxy of $Cd_xHg_{1-x}Te$. In: *Infrared focal plane arrays*, Nauka, Novosibirsk (in Russian)

Sivananthan S., Chu X., Reno J., Faurie J.P. (1986). *J. Appl. Phys.*, Vol. 60 (4), p. 1359

Sizov F. F., Vasil'ev V.V., Suslyakov A.O., Reva V.P., Golenkov A.G. (2006) 4×288 Readouts and FPAs Propeties, *Optoelectron review*, Vol. 14., pp. 67–74.

Tashikawa M., Yamaguchi M. (1990) Film thickness dependence of dislocation density reduction in GaAs on Si substrates. *Appl. Phys. Lett.*, Vol. 56, No.5, pp.484-486.

Vasiliev V. V., Klimenko A.G., Marchishin I.V., Ovsyuk V. N., Talipov N. Kh., Zakhar'yash T.I, Golenkov A. G., Derkach Yu. P., Reva V. P., Sizov F. F., Zabudsky V. V. (2004) MCT heteroepitaxial 4×288 FPA. *Infrared Physics&Technology*, Vol. 44, pp. 13-23,

Vilela M.F., Buell A.A., Newton M.D., Venzor G.M., Childs A.C., Peterson J.M., Franklin J.J., Bornfreund R.E., Radford W.A., and Johnson S.M. (2005) Control and Growth of Middle Wave Infrared (MWIR) HgCdTe on Si by Molecular Beam Epitaxy. *J. Electron. Mater.*, Vol. 34, No. 6, p.898.

Voitsehovsky A.V., Denisov Yu.A., Kohanenko A.P., et al. (1996) *Avtometriya*, Vol. 4, pp..51-58 (in Russian)

Wang C.C., Me Farlane S.H. *(*1976) Crystal growth and defect chracterization of heteroepitaxial III-V semiconductor films, *Thin Solid Films*, Vol. 31, No. 3, p. 323-332

Yakushev M.V., Babenko A.A., Sidorov Yu.G. (2009) Effect of substrate orientation on the growth conditions of HgTe films grown by molecular beam epitaxy, *Neorganicheskie materialy*, Vol. 45, No. 1, p. 15 (in Russian)