

Instrumentation on silicon detectors: from properties characterization to applications

N. Dinu

► To cite this version:

N. Dinu. Instrumentation on silicon detectors: from properties characterization to applications. Instrumentation and Detectors [physics.ins-det]. Université Paris Sud - Paris XI, 2013. <tel-00872318>

HAL Id: tel-00872318

<https://tel.archives-ouvertes.fr/tel-00872318>

Submitted on 29 Oct 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Paris-Sud 11

LAL-13-192
9 Octobre 2013

MEMOIRE D'HABILITATION A DIRIGER DES RECHERCHES

Instrumentation on Silicon Detectors:
from properties characterization to applications

Nicoleta DINU

Soutenance le 9 Octobre 2013, devant le jury composé de :

Yves Charon

Christian Joram (Rapporteur)

Christophe de La Taille

Abdenour Lounis

Adam Para (Rapporteur)

Thomas Patzak (Rapporteur)

Achille Stocchi (Président)

Remerciements

Je remercie tout d'abord le LAL et tous ses responsables, à tous les niveaux, de m'avoir accueilli au sein de ce grand Laboratoire de Recherche du CNRS / IN2P3, initialement en tant que CDD, puis, après concours, comme titulaire sur un poste d'Ingénieur de Recherche.

Mais je n'aurais peut-être pas eu le courage d'entreprendre ce travail, si le nouveau Directeur, Achille Stocchi, n'avait lui-même encouragé les Ingénieurs à se lancer dans cette voie de "l'Habilitation à Diriger des Recherches", et ne m'avait donné les moyens de mener à bien ce mémoire. En effet la synthèse et la rédaction des résultats de plus de dix années de recherche et d'expérimentation, dans des domaines en constante évolution, ne constitue pas une mince tâche.

Ce document, n'est bien évidemment pas le fruit d'un travail solitaire, mais celui d'une équipe, plus précisément de plusieurs équipes successives, tant en Italie (FBK/Irst et INFN -Pisa) qu'en France (LAL, IMNC, GeMAC, IEF) et aux USA (Fermilab). Je voudrais donc, remercier tous ceux qui, par leur minutieux travail, ont permis à cet ouvrage de voir le jour.

Au sein de ces diverses équipes, chaque collaborateur se retrouvera, en tête de chaque chapitre, en fonction de ses contributions respectives à diverses phases ou activités développées dans ce document. ...Ainsi cela "obligera" le lecteur à feuilleter le document et pas seulement à lire ces remerciements... ! *(J'espère n'avoir oublié personne ...mais on oublie toujours quelqu'un ... qu'il(s) me pardonnent par avance !)*

Je remercie Christian Joram (CERN), Adam Para (Fermilab), et Thomas Patzak (APC) d'avoir accepté la lourde tâche d'être Rapporteurs de ce travail, ainsi que Christophe de La Taille, Abdenour Lounis et Yves Charon d'avoir accepté de faire partie du jury et bien sûr Achille Stocchi d'avoir présidé ce jury. Cela constitue pour moi un véritable honneur.

Mais je tiens également à étendre mes remerciements à Françoise Marechal notre Bibliothécaire, qui a beaucoup œuvré pour me trouver des documents plus ou moins exotiques, nécessaires aux recherches bibliographiques, à Dominique Longieras qui a assuré la reproduction et la fabrication du document, et plus généralement toutes les personnes des services administratifs qui m'ont aidé.

Merci également à mes parents et à ma sœur, qui, même de très loin, m'ont encouragée et soutenue de leur grand amour, pour l'accomplissement de ce mémoire. Mes parents auraient été très fiers d'assister à la soutenance, comme ils l'ont été il y a 12 ans, lors de la soutenance de ma thèse de Doctorat en Roumanie.

Enfin, Jean-Jacques, tu as tellement compté pour moi.... Je voudrais te remercier pour ta présence de tous les jours à mes côtés, pour m'avoir remonté le moral dans tous les moments difficiles

survenus depuis mon arrivée en France et surtout, d'avoir accepté que la "petite chambre d'amis" soit devenue mon deuxième lieu de travail pendant plusieurs mois pour la rédaction de ce mémoire.... !

Je terminerai par un extrait des *Pensées* du plus célèbre poète Roumain Mihai Eminescu (1850-1889):
« *Viata este un bine pierdut pentru cel care nu a trait-o cum ar fi dorit* »

Que je traduirai, sachant que tous n'ont pas "*fait Roumain*" en 1ère langue... par :
« *La vie est un bien perdu pour celui qui ne l'a pas vécu comme il l'aurait voulu* ».

Preface

Based on the remarkable properties of p - n junction, silicon detectors allow us to measure the characteristics of interactions with the matter produced by visible light photons or charged particles. However, the progress of such detectors addresses the continuous refinement of two domains of experimental techniques, closely interconnected one to the other:

- on one side, the development of experimental tools for the characterization and understanding of the detector characteristics itself and the related physical phenomena;
- on the other side, the development of more sophisticated detection systems to be used in different applications.

In this dissertation, the interconnection between these two branches of the experimental techniques will be shown to be of great importance. In particular, my personal contribution in this field for the case of two types of silicon detectors will be described:

- **PART A** is oriented to the study of the physical parameters of different photon detectors and their application to medical imaging. The most known devices will be firstly reviewed. Then, the experimental techniques as well as the characteristics of the more recent devices called Geiger-Mode Avalanche Photodiode and Silicon Photomultiplier (SiPM) will be described in detail. In the last two sections, the characteristics of more complex detection systems based on SiPM's and used in two medical imaging applications will be shown.
- **PART B** is related to the experimental methods used for the doping profiles measurements, profiles which are extremely important for the understanding of the p - n junction properties used in silicon detectors. After a short description of the experimental techniques, the results of the doping profiles measurements performed in the case of ATLAS planar pixel detectors will be given. The importance of these measurements for the fabrication process control and the calibration of the Technology-Computed Aided Design simulations (TCAD) will be shown. Such simulations have been subsequently used to predict the behavior of new planar pixel sensors with improved geometrical and radiation hardness characteristics for the upgrade of the ATLAS inner detector at high-luminosity LHC.

PART A

Silicon detectors for visible light detection and medical imaging applications

Table of contents: PART A

1. Introduction	9
2. Review: from Photodiode to Avalanche Photodiode	11
2.1. PN Photodiode	11
2.2. PIN Photodiode	15
2.3. Avalanche Photodiode (APD)	16
2.4. Summary : advantages and drawbacks of PN, PIN and APD's	21
References on PN, PIN and APD's	22
3. Geiger-Mode Avalanche Photodiode (GM-APD)	23
3.1. Physics principle	23
3.2. Characteristics and performances	27
3.2.1. Static characteristics	28
3.2.2. Dynamic characteristics	31
3.2.3. Photon detection efficiency	39
3.2.4. Timing resolution	45
3.3. Summary : advantages and drawbacks of GM-APD's	47
References on GM-APD's	47
4. Silicon Photomultiplier (SiPM)	51
4.1. Physics principle	51
4.2. Characteristics and performances	52
4.2.1. Static characteristics	53
4.2.2. Dynamic characteristics	55
4.2.3. Photon detection efficiency	62
4.2.4. Timing resolution	67
4.2.5. Thermal effects – first part	71
4.2.6. Thermal effects – second part	76
4.3. SiPM arrays	82
4.3.1. Discrete SiPM arrays	83
4.3.2. Monolithic SiPM arrays	84
4.4. Summary: advantages and drawbacks of SiPM's	85
References on SiPM's	86

5. Arrays of SiPM for Medical Applications	89
5.1. Arrays of SiPM's for high resolution small animal PET scanner	90
5.1.1. The principle of PET	90
5.1.2. Detection features for PET systems	93
5.1.3. PET detector modules	95
5.1.4. Could SiPM's array be implemented as photon detector for PET systems?	98
5.1.5. Evaluation of SiPM's arrays for PET scanner	105
5.1.6. Summary on SiPM's arrays for small animal PET scanner	114
5.2. Arrays of SiPM's for radio-guided surgery of tumors	115
5.2.1. The principle of radio-guided surgery	115
5.2.2. Performances parameters of intra-operative systems	116
5.2.3. Review of intra-operative gamma detection systems	117
5.2.3.1. Intra-operative gamma probes	117
5.2.3.2. Intra-operative gamma imaging camera	119
5.2.4. SIPMED project - Silicon Photomultiplier for medical imaging	122
5.2.4.1. Characteristics of the SiPM's arrays for SIPMED camera	122
5.2.4.2. SiPM's arrays implementation for SIPMED camera	126
5.2.4.3. Preliminary evaluation for gamma imaging	128
5.2.5. Summary on SiPM's arrays for radio-guided surgery	129
References on SiPM's arrays for medical imaging	130
6. Summary and perspectives of PART A	133
Annex 1: Reverse I-V characteristic for ideal and practical diode; relationship with the GM-APD characteristics	169
Annex 2: Temperature dependence of reverse I-V characteristic; relationship with the GM-APD characteristics	177

1. Introduction

The purpose of any visible light detector is to convert electromagnetic radiation into an electronic signal – ideally one that is proportional to incident light intensity. A particular requirement of many applications to photon detectors is that to be able to detect very low light intensities, down to the single photon.

Because their compactness and robustness, semiconductor materials like silicon (Si) and germanium (Ge) or compound semiconductors like Gallium Arsenide (GaAs), Indium Gallium Arsenide (InGaAs), Gallium Arsenide Phosphate (GaAsP) and Gallium Phosphate (GaP) are very often used today as sensitive medium for visible light detection. Among these, the Si is the most widely used since it can benefit from fast technological developments of silicon electronics industry.

Following the applications requirements, silicon photons detectors showed an important evolution starting from classical Photodiodes, well known today, up to a more recently device called Silicon Photomultiplier (i.e. arrays of Geiger-Mode Avalanche Photodiodes) well adapted for medical imaging applications but also calorimeter readouts in high-energy physics experiments. This evolution step was of course not obvious and in between, two other detectors have been developed: the Avalanche Photodiodes and the Geiger-Mode Avalanche Photodiodes.

A review of the physics principle and the main characteristics of PN&PIN Photodiodes and Avalanche Photodiodes will be presented in the second Section.

My pioneering work on photon detectors started around 10 years ago, working as post-doctoral fellow in the framework of the DASIPM collaboration between two Italian institutes: Istituto Nazionale di Fisica Nucleare (INFN) and Fondazione Bruno Kessler (FBK-irst). DASIPM was focusing on Applications of SiPM's in Medical, Space and High Energy Physics. During that period I participated to the design and technological developments of the first Geiger-Mode Avalanche Detectors (GM-APD) developed at FBK-irst, working in particular to the characterization and the analysis of the physical phenomena related to these detectors. Therefore, the third Section will relate the physical principle of GM-APD detectors as well as their characteristics based on the measurements performed on GM-APD devices developed at FBK-irst.

At the end of 2006 I obtained a research engineer position at the Laboratory of Linear Accelerator (LAL) in Orsay where I had the chance to continue my work on photon detectors, in particular on SiPM's build in different technologies as well as on arrays of SiPM's and their applications in medical imaging. We have build at LAL calibrated experimental set-ups able to test single SiPM characteristics, starting from static and dynamic characteristics in the dark up to the measurement of the photon detection efficiency and timing resolution in light conditions.

Since the temperature is influencing some of the important characteristics of SiPM detectors, I started a detailed program of device analysis over a very wide range of temperatures and this project is under development in collaboration with Fermi National Accelerator Laboratory.

More complex detection systems build from arrays of SiPM's and coupled to multi-channels read-out electronics represents two medical imaging projects in which I was and I continue to be directly involved performing technical work as well as project management.

The fourth and fifth Sections will describe in detail my personal work performed at LAL: primary, the analysis of physical phenomena related to SiPM detectors, in particular their room temperature characteristics as well as their temperature dependence, and secondly the SiPM's arrays for medical applications: (1) small animal PET system in collaboration with Pisa University and FBK-irst; (2) SIPMED project in collaboration with l'Institut de l'Imagerie et Modelisation en Neurobiologie et Cancérologie d'Orsay (IMNC) and Hospital Lariboisiere.

2. Review: from photodiode to avalanche photodiode

2.1. PN photodiode

A photodiode is a semiconductor device consisting in a p-n junction working in reverse biased mode.

The design of a generic PN silicon photodiode is presented in Fig. 2.1. A n-type silicon is used as substrate (silicon doped with impurities as P, As) and a p-type impurity (i.e. B) is shallow diffused at the active surface in order to obtain the p-n junction (usually the thickness of the p-type layer is approximately 1 μm or less). The contact pads are then deposited on the front active area on defined areas, and on the back side, completely covering the device. The active area is then passivated with an antireflective coating to reduce the reflection of the light for a specific range of wavelengths (as we know, the silicon is a material which can act as a mirror, reflecting more than 30% of the incident light). This antireflective coating is usually made from layers of SiO_2 and Si_3N_4). The non-active area on the top is covered with a thick layer of silicon oxide. By controlling the thickness of the p and n-type layers as well as the doping concentration, the speed and the spectral response of the photodiode can be controlled.

When a photon with sufficient energy (i.e. greater than the band gap energy of 1.12 eV for Si which corresponds to a wavelength shorter than 1100nm) strikes the photodiode, it is absorbed and it excites an electron from the valence band to the conduction band, producing an electron-hole pair. Such electron-hole pair creation occurs throughout the p-layer, depletion layer or n-layer material as a function of the photons wavelength: short wavelength photons are absorbed in the p-type layer and the long wavelength photons are absorbed in the n-type layer.

Since the p-n junction electric field is mostly in the depletion region, the photo-carrier generation is mostly confined to the depletion region. Fig. 2.2 shows the cross-section of a p-n junction and possible locations of electron-hole pair generation.

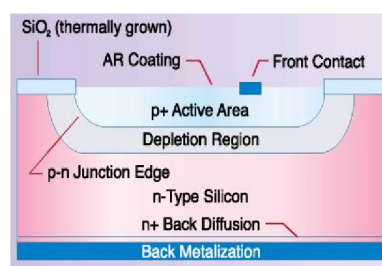


Fig. 2.1 Design of a generic PN photodiode

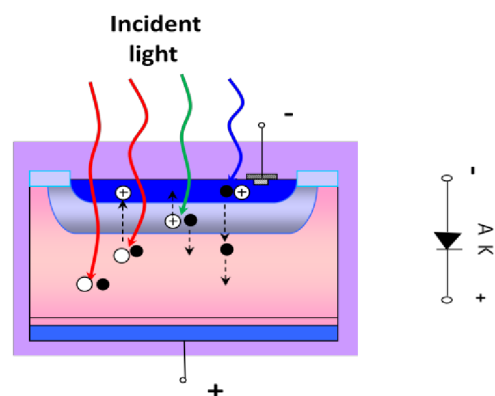


Fig. 2.2 Cross-section of a reversed biased p-n junction showing different regions where the photon generates electron-hole pairs, and direction of each carrier (black dot = electron; white dot = hole)

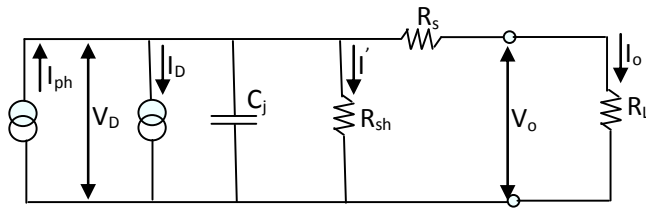
There are three possible locations where an electron-hole pair can be generated:

- In the depletion region, where the electron and hole quickly drift in opposite directions because of presence of the electric field in this region. Since recombination does not take place in the depletion region, each photo-carrier adds a charge q (considering generation rate of 1) to the current collected to the device electrodes;
- Far away from the depletion region. Here the pair recombines after some time without contributing to the signal;
- Close to the depletion region, where the electron and hole have a chance to diffuse to the depletion region by random diffusion. An electron coming from the p-side or a hole coming from the n-side can quickly be transported across the junction and therefore contribute the charge of q to the external circuit.

The collection of the created charges gives rise to the “photocurrent” which is proportional to the intensity of the incident light.

Equivalent circuit

A Si photodiode can be represented by a current source in parallel with an ideal diode (Fig. 2.3). The current source represents the current generated by the incident radiation (proportional to the amount of light), and the diode represents the p-n junction. In addition, a *junction capacitance* (C_j) and the *shunt resistance* (R_{sh}) are in parallel with the other components. *Series resistance* (R_s) is connected in series with all components.



I_{ph} : current generated by the incident light
(proportional to the amount of light)
 I_D : diode current
 V_D : voltage across the diode
 C_j : junction capacitance
 R_{sh} : shunt resistance
 R_s : series resistance
 R_L : load resistance
 I_o : output current
 I' : shunt resistance current
 V_o : output voltage

Fig. 2.3 Equivalent circuit for the silicon photodiode

Current-voltage characteristic

The photodiode current can be described in a first approximation by the Shockley ideal diode law:

$$I_{diode} = I_{sat} \left(e^{\frac{q V_{BIAS}}{n k_B T}} - 1 \right) = I_{sat} \left(e^{\frac{V_{BIAS}}{n V_T}} - 1 \right) \quad 2.1$$

where I_{sat} is the reverse bias saturation current, V_{BIAS} is the bias voltage across the diode, V_T is the thermal voltage, n is the ideality factor, also known as the quality factor or sometimes emission coefficient. The factor n typically varies from 1 to 2, depending on the fabrication process and semiconductor material and in many cases is assumed to be approximately equal to 1. The thermal voltage V_T is approximately 25.85 mV at 300K. At any temperature it is known constant defined by: $V_T = k_B T / q$, where $k_B = 1.38 \times 10^{-23}$ J/K is the Boltzmann constant, T is the absolute temperature (273K=0°C) and $q = 1.6 \times 10^{-19}$ C is the electron charge.

The ideal Shockley equation can give only a qualitative agreement with reverse current-voltage characteristics of real devices. The departures from the ideal situation are described in details in the Annex 1.

The dark IV characteristic is represented by the P_0 curve in Fig. 2.4.

Illuminating the photodiode with light, the IV curve will be shifted by the amount of photocurrent $-I_{ph}$ proportional to the photon flux (curves P_1 , P_2 and P_3 in the Fig. 2.4). Thus:

$$I_{total} = I_{sat} \left(e^{V_d / (nV_T)} - 1 \right) - I_{ph} \quad 2.2$$

As a function of the bias voltage condition, a photodiode can operate either in photovoltaic mode (unbiased) or in photoconductive mode (few volts reverse bias voltage). In the photoconductive mode, a very linear response is achieved.

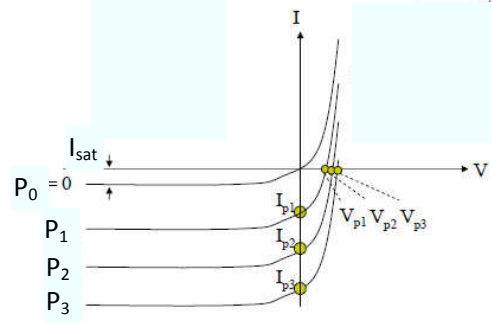


Fig.2.4 Current-voltage characteristic for a photodiode in dark (P_0) and light conditions ($P_3 > P_2 > P_1$)

Optical characteristics

The parameter measuring the effectiveness of the conversion of light power into electrical current in a photodiode is called *responsivity* R_λ , and it is defined as the ratio of the photocurrent I_{ph} to the incident light

power P_{opt} at a given wavelength: $R_\lambda = \frac{I_{ph}}{P_{opt}}$. It varies with the wavelength of the incident light as well as applied bias (it increases slightly with applied reversed bias) and the temperature. A typical spectral responsivity of several types of diffused photodiodes produced by UDT is presented in Fig. 2.5.

The quantum efficiency of a photodiode is defined as the fraction of the incident photons that contribute to photocurrent. It is related to the responsivity by:

$$QE = R_\lambda \cdot hc / \lambda q = 1240 \cdot R_\lambda / \lambda \quad 2.3$$

where h is the Planck constant, c is the light speed, q is the electron charge, R_λ is the responsivity in A/W and λ is the wavelength in nm.

A Si photodiode is considered linear if the generated photocurrent increases linearly with the incident light power. In other words, linearity exhibits the consistency of responsivity over a range of light power. Non-linearity of less than $\pm 1\%$ is specified over 6-9 decades for planar diffused photodiodes. The lower limit of the photocurrent linearity is determined by the noise or dark current and the upper limit by the diode series resistance and the load resistance. As the photocurrent

increases, first the non-linearity sets in, gradually increasing with increasing photocurrent, and finally at saturation level, the photocurrent remains constant with increasing incident power. The linearity range can slightly be extended by applying a reverse bias to the photodiode.

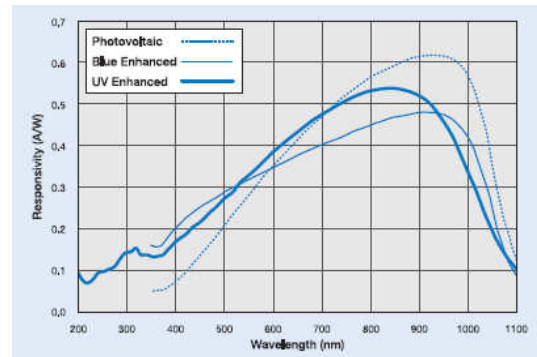


Fig. 2.5 Typical spectral responsivity for different types of planar diffused photodiodes (UDT sensors)

Noise

The lower limit of light detection for photodiodes is determined by the noise characteristics of the device. The *photodiode noise* i_{tn} is the sum of the *thermal noise* (or *Johnson noise*) i_{jn} and the *shot noise* i_{sn} :

$$i_{tn} = \sqrt{i_{jn}^2 + i_{sn}^2}$$

The Johnson noise is determined by the thermal generation of the charge carriers in the shunt resistance of the photodiode and it is expressed as: $i_{jn} = \sqrt{4k_B T B / R_{sh}}$, where $k_B = 1.38 \times 10^{-23}$ J/K is the Boltzmann constant, T is the absolute temperature in degrees Kelvin, B is the noise measurement bandwidth and the R_{sh} is the shunt resistance of the photodiode. This type of noise is the dominant current noise in photovoltaic (unbiased) operation mode.

The shot noise is related to the statistical fluctuations in both the diode dark current and the photocurrent and it is expressed as: $i_{sn} = \sqrt{2 \cdot q \cdot (I_{diode} + I_{ph}) \cdot B}$. Shot noise is the dominating source when operating in photoconductive (biased) mode.

The amplitude of these noise sources are each proportional to the square root of the measured bandwidth B so they are expressed in units of A/ $\sqrt{\text{Hz}}$.

The lower limit of the light detection for a photodiode is usually expressed as the intensity of incident light required to generate a current equal to the noise current. Essentially, this is the *noise equivalent power NEP*:

$NEP = i_{tn} / R_\lambda$ where R_λ is the photodiode responsivity in A/W and i_{tn} is the total noise. *NEP* values can vary from 10^{-11} W/ $\sqrt{\text{Hz}}$ for large active area photodiodes down to 10^{-15} W/ $\sqrt{\text{Hz}}$ for small active area photodiodes.

Temperature dependence

The photodiodes characteristics are generally affected by change in temperature. The most affected ones are shunt resistance, dark current, responsivity and to a lesser extend other parameters as junction capacitance.

A constant increase in dark current takes place with increasing temperature as a result of increasing temperatures causing electrons in the valence band to become excited, pulling them in the conduction band. In photoconductive mode, the dark current may approximately double for every 8-10°C increase in temperature. This is equivalent to a reduction of the shunt resistance R_{sh} and a subsequent increase in thermal and shot noise. In photovoltaic mode, shunt resistance may double for every 6°C decreases in temperature. The exact change is dependent on additional parameters such as the applied reverse bias, resistivity of the substrate as well as the thickness of the substrate. More details about the temperature dependence of the reverse dark current of a p-n junction are given in the Annex 2.

The responsivity of a photodiode presents variations due to change in temperature. At higher temperatures, the absorption coefficient increases, so that all wavelengths of light are absorbed in a shorter distance. This could alter the sensor's response if it cases wavelengths to be absorbed differently than they were at the reference temperature. This can be seen in Fig. 2.6 which shows the depth in Si at which ~95% of the energy at a given wavelength is absorbed. Note that as the *temperature is increased beyond 27°C, the absorption rate increases causing the absorption depth to decrease. This makes the sensor more sensitive to light at longer wavelengths.* As the temperature is lowered, the absorption depth increases, making the sensor less sensitive to light at longer wavelengths. This variation of the absorption depth with temperature is due to the inverse

relationship variation between the band gap voltage and the temperature for silicon: this means that as the temperature is increased, the band gap voltage decreases.

The band gap voltage of silicon is approximately 1.12eV at 25°C and approximately 1.11 eV at 70°C (Zeghbrouck, 2007). Therefore, a change in temperature from 25°C to 70°C, would result in a change in Si spectral response limit from 1100 nm to 1110 nm.

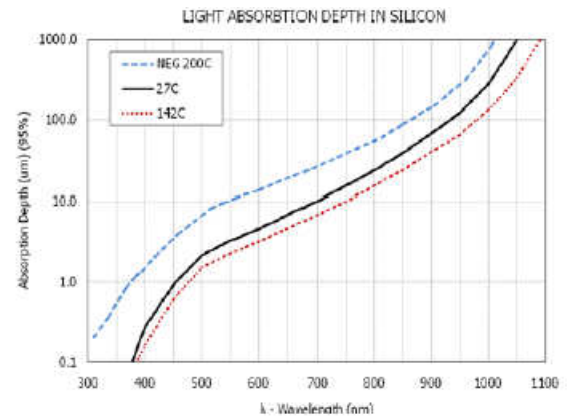


Fig. 2.6 Absorption depth of light in Si vs wavelength for three temperatures

2.2. PIN Photodiode

A PIN photodiode is a p-n junction with a wide, lightly doped “near” intrinsic semiconductor layer sandwiched between p and n layers. The p-type and n-type regions are typically heavily doped because they are used for ohmic contacts. Most of the photons are absorbed in the intrinsic region, and carriers generated therein can efficiently contribute to the photocurrent. Fig. 2.7 presents the schematic drawing of a PIN photodiode. The gray layer on the top of the sensitive area is an anti-reflective coating.

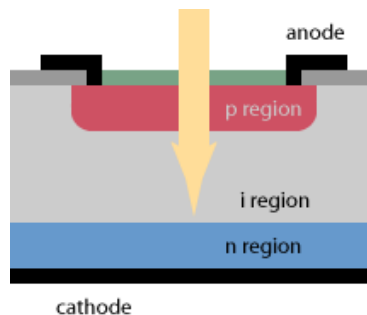


Fig. 2.7 Cross-section of a PIN photodiode

Compared to an ordinary PN photodiode, a PIN photodiode has a thicker depletion region. This allows:

- a more efficient collection of the carriers and thus a larger quantum efficiency in red and IR wavelengths; at longer wavelengths, the absorption efficiency and thus the responsivity drops sharply, but the parameters of this cut-off depend on the thickness of the “i” region
- a lower capacitance and thus a higher detection bandwidth (reduced RC delay constant); the transit time of carriers drifting across the depletion region, and the RC time response are the main contributions of the response time of photodiodes detectors.

By using an external reverse bias voltage applied over an ordinary PN diode, the depletion region can be increased improving the frequency response and the linearity. But such operating mode has the drawback of

increasing the dark current and noise levels also with danger of damaging the device by excessive applied reverse voltage. For use in applications requiring high response speed, the PIN photodiode provides not only a good response speed but excellent dark current and voltage resistance characteristics with reverse voltage applied.

2.3. Avalanche photodiode (APD)

An APD is also a p-n junction and it detects light by using essentially the same principle as the PN photodiode. The main difference between an APD and a PN photodiode is that an APD is designed to work at higher reverse bias voltage than a PN photodiode, in such way that the photo-carriers created by the incident photons will be accelerated by the high field created in the depleted region. The photo-carriers will collide with atoms in the field and secondary carriers will be produced, this process occurring repeatedly. This is known as the “avalanche” effect and thus resulting in a signal being amplified by a certain multiplication factor (i.e. gain). So this type of device is more adapted to detect lower level incident light than the photodiodes.

Of course, an ideal APD would be a good PIN photodiode with gain! In reality however, this is difficult to achieve because of the need to trade-off conflicting design requirements.

Different types of APD's are today commercially available (Webb et al., 1974, Pansart, 1997). The schematic cross-section of the most widely used APD structure called “reach-through” is presented in Fig. 2.8. The basic structural elements of an APD design include an absorption region “A”, and a multiplication region “M”. Present across region “A” is an electric field “E” that serves to separate the photo-generated holes and electrons, and sweeps one carrier towards the multiplication region. The multiplication region “M” is designed to exhibit a high electric field so as to provide internal photo-current gain by impact ionization.

The gain region must be broad enough to provide a useful gain, M , of at least few hundreds. In addition, the multiplying electric field profile must enable effective gain to be achieved at field strength below the breakdown field of the diode.

A “reverse” APD structure presenting radiation hardness and being particularly used for particle detection in high energy experiments at CERN (i.e. CMS experiment at CERN, to read-out the PbWO_4 scintillating crystals in the electromagnetic calorimeter) is presented in Fig. 2.9. The p-n junction of the CMS APD structure is formed in an epitaxial layer n^- with 40-50 μm thickness. In such structure, the photons create electron-hole pairs in the thin p-layer on the top of the device and the electrons are first multiplied in the high field region at the p-n junction and then drifted toward the n^+ electrode ($M_e=50$). Holes created behind the junction contribute little because of their much smaller ionization coefficient ($M_h=1.5$).

Such kind of APD structure is particularly advantageous in high-energy physics experiments environment because in these experiments some of the electrons and photons in an electromagnetic shower can leak out from the rear end of the calorimeter and can hit the APD photo-detector directly, depositing a significant amount of energy in the photo-detector. This “nuclear counter effect” would cause a deterioration of energy resolution and possibly a systematic shift in the average detected energy. Therefore, in the APD structure presented in the Fig. 2.9, this nuclear counter is minimized, since the electrons produced by ionizing particles and traversing the bulk of the APD device are drifted to the n^+ electrical contact without traversing the high field region, therefore without amplification. In consequence, such APD has a lower nuclear counter effect (~ 50 times) with respect to a classical PN photodiode.

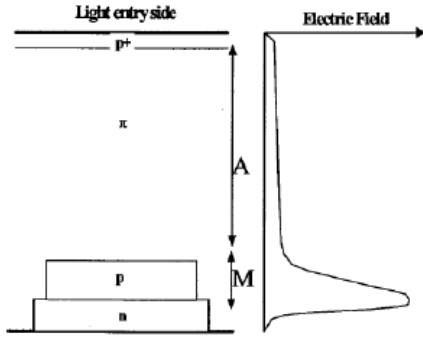


Fig. 2.8 Schematic representation of an “reach-through” APD structure; the multiplication region “M” is designed to exhibit a high electric field so as to provide internal photo-current gain by impact ionization

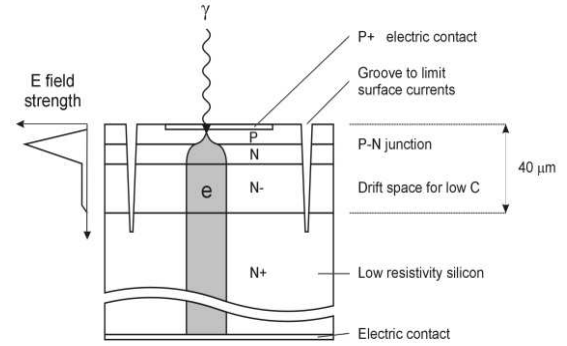


Fig. 2.9 Schematic representation of Hamamatsu “reversed” APD structure used for CMS ECAL detector and the electrical field generated by the application of the bias voltage

Gain

An APD differs from a PIN photodiode by providing internal photo-electronic signal gain. Therefore, output signal current, $I_{APD} = M \cdot R_0(\lambda) \cdot P_{opt}$, where $R_0(\lambda)$ is the intrinsic responsivity of the APD at a gain $M=1$ and wavelength λ , M is the gain of the APD and P_{opt} is the incident optical power.

The gain is a function of the APD’s reverse voltage V_{BIAS} and will vary with applied bias. A typical gain-voltage curve for two types of APD’s manufactured by PerkinElmer is shown in Fig. 2.10 (PerkinElmer, Application Note). The “standard” APD structure refers to the “reach-through” structure presented in Fig. 2.8. The gain-voltage curve of the “reversed” APD structure used for the CMS ECAL experiment is presented in Fig. 2.11.

The optimum gain depends in part on the excess noise factor F (see later for F definition) and ranges usually from $M=50$ to 1000 for silicon APD’s. For the APD used in CMS ECAL experiment, the applied bias is around 350 – 400 V allowing to obtain a gain of ~ 50 .

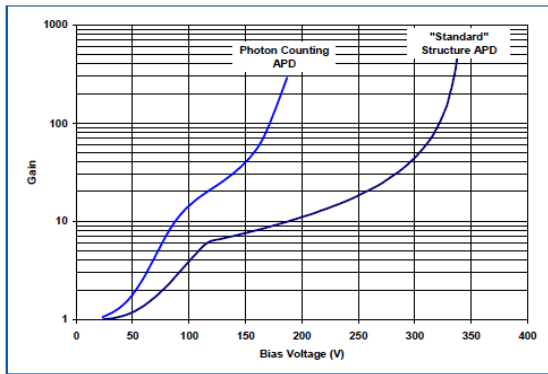


Fig. 2.10 Gain vs V_{BIAS} for two PerkinElmer APD structures; the standard one represents the “reach-through” structure; the gain as a function of the V_{BIAS} varies with the structure of the APD

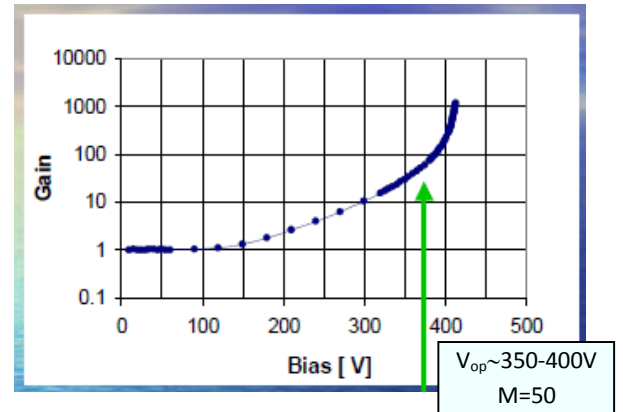


Fig. 2.11 Gain vs V_{BIAS} for the “reverse” APD structure used in CMS ECAL experiment (Musienko, 2003)

Current-voltage characteristic

The dark current for an APD can be described in a first approximation by the Shockley ideal diode law. But the Shockley equation can give only a qualitative agreement with reverse current-voltage characteristics of real device, as described in details in the Annex 1. The dark current of a real APD device is given by the sum of: surface leakage current $I_{surface}$ flowing through the interface between the PN junction and Si oxide layer existing on the diode surface and the internal current I_{bulk} generated inside of the Si substrate. The surface leakage current is not multiplied because it does not flow in the avalanche region, but the internally generated current flows in the avalanche region so it is multiplied. Therefore, the total dark current $I_{diode-total}$ produced from an APD becomes as follows:

$$I_{diode-total} = I_{surface} + I_{bulk} \cdot M \quad 2.4$$

where M is the multiplication factor. Fig. 2.12 shows how the dark current changes with the reverse voltage. It is clear that I_{bulk} increases as the reverse voltage rises. The I-V characteristic of the APD devices used for the CMS ECAL experiment is presented in Fig. 2.13. We can observe an increase of the dark current as the bias voltage increases.

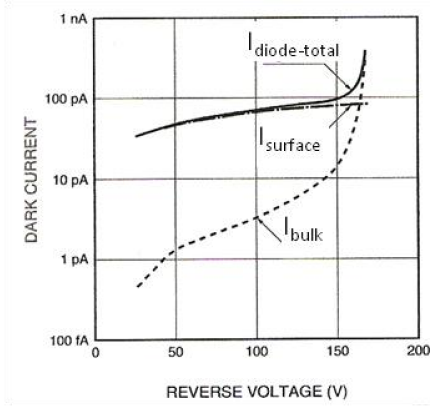


Fig. 2.12 Dark current vs reverse voltage for a classical APD device; the contributions of surface leakage current ($I_{surface}$) and the bulk generated leakage current (I_{bulk}) are showed (Hamamatsu Solid State Division)

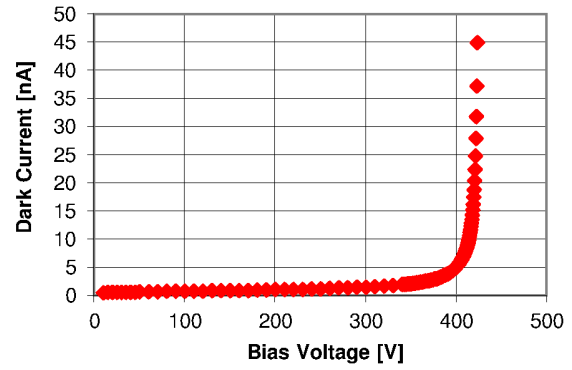


Fig. 2.13 Dark current vs bias voltage for the APD structure used in CMS ECAL experiment (Musienko, 2003)

Noise

As an APD is designed to be operated under reverse bias, sensitivity at low light levels will be limited by the shot noise and the APD's leakage current. Shot noise derives from the random statistical Poissonian fluctuations of the dark current (or signal current). Dark current shot noise is normally given by

$i_{sn} = \sqrt{2 \cdot q \cdot I_{diode} \cdot B}$ for a PIN detector, where B is the system bandwidth. This differs from an APD however, as bulk leakage current I_{bulk} is multiplied by the gain, M , of the APD, as described previously.

In addition, the avalanche process statistics generate gain fluctuations, and APD performance is degraded by the “excess noise factor” F compared to a PIN, expressed by: $F = \frac{M^2 + \sigma_M^2}{M^2}$, where M is the gain and σ_M^2 is the gain variance. Therefore the total spectral noise current for an APD in dark conditions is given by:

$$i_n = \sqrt{2 \cdot q \cdot (I_{surface} + I_{bulk} \cdot (M^2 + \sigma_M^2)) \cdot B} = \sqrt{2 \cdot q \cdot (I_{surface} + I_{bulk} \cdot M^2 \cdot F) \cdot B} \quad 2.5$$

where q is the electron charge.

Total noise from an APD in illuminated conditions will therefore equal the quadratic sum of the detector noise plus the signal shot noise. For a given optical signal power P_{opt} , this is given by:

$$i_m = \sqrt{2 \cdot q \cdot (I_{surface} + (I_{bulk} \cdot M^2 + R_0(\lambda) \cdot M^2 \cdot P_{opt}) \cdot F) \cdot B} \quad 2.6$$

In the absence of other noise sources, an APD therefore provides a signal-to-noise ratio (SNR) which is $\sqrt{F}$ worse than a PIN detector with the same quantum efficiency.

Excess Noise Factor

The avalanche photodiodes generate excess noise due to the statistical nature of the avalanche multiplication process. The “Excess Noise Factor” is generally denoted as “ F ”. As shown in the noise equation 2.6, $\sqrt{F}$ is the factor by which the statistical noise on the APD current (equal to the sum of the multiplied photocurrent plus the multiplied APD bulk dark current) exceeds that which would be expected from the noiseless multiplier on the basis of Poissonian statistics (shot noise) alone.

The excess noise factor is a function of the carrier ionization ratio, k , where k usually defined as the ratio of hole to electron ionization probabilities ($k = \alpha_h/\alpha_e \leq 1$). The excess noise factor can be calculated using the model developed by McIntyre (McIntyre, 1966) which considers the statistical nature of avalanche multiplication. The excess noise factor is given by:

$$F = k_{eff} \cdot M + (1 - k_{eff}) \left(2 - \frac{1}{M} \right) \quad 2.7$$

Therefore, the lower values of k and M , the lower the excess noise factor. The effective k -factor, k_{eff} , for an APD can be measured experimentally by fitting the McIntyre formula to the measured dependence of the excess noise factor on gain. This is best done under illuminated conditions. It may be also theoretically calculated from the carrier ionization coefficients and the electric field profile of the APD structure.

The ionization ration k is a strong function of electric field across the APD structure, and takes its lowest value at low electric fields in silicon. Since the electric field profile depends upon the doping profile, the k factor is also a function of the doping profile. Depending on the APD structure, the electric field profile traversed by a photo-generated carrier and subsequent avalanche-ionized carriers may therefore vary according to photon absorption depth. For indirect band-gap semiconductors such as silicon, the absorption coefficient varies slowly at the longer wavelengths, and the “mean” absorption depth is therefore a function of wavelength. The value of k_{eff} , and gain, M , for a silicon APD is thus a function of wavelength for a given doping profiles.

The McIntyre formula can be approximated for a $k < 1$ and $M > 20$ without significant loss of accuracy as:

$$F = 2 + k_{eff} \cdot M \quad 2.8$$

where $k_{eff} \approx k = \alpha_h / \alpha_e$.

The typical values for a Si “reach-through” APD structure are: the ratio of hole to electron ionization probabilities $k=0.02$, the multiplication factor $M=150$, the excess noise factor $F=4.9$.

The excess noise factor as a function of gain for the APD devices used in CMS ECAL experiment is presented in Fig. 2.14. At the operating gain of $M=50$, the excess noise factor is $F \approx 2.1$.

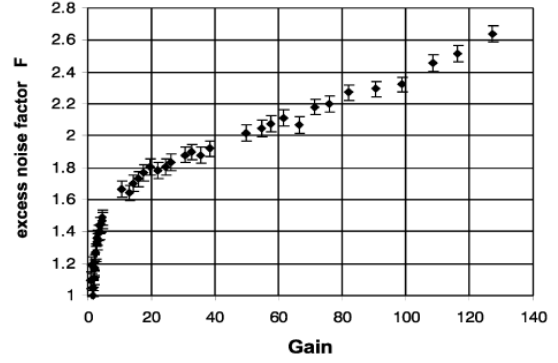


Fig. 2.14 Excess noise factor vs gain for the “reverse” APD structure used in CMS ECAL experiment (Deiters et al., 2000); for the operating gain of $M=50$ the excess noise factor is $F \approx 2.1$

Optical characteristics

The spectral response characteristics of the APD are almost the same as those of normal photodiodes if no bias voltage is applied. When a bias voltage is applied, the spectral response curve will change, because the gain changes. The spectral response of the Si APD used in CMS ECAL experiment is presented in Fig. 2.15. We can observe that these devices present a QE of $\sim 73\%$ at the wavelength of 420 nm (the peak emission of the scintillating crystals of PbWO_4 used in CMS ECAL calorimeter)

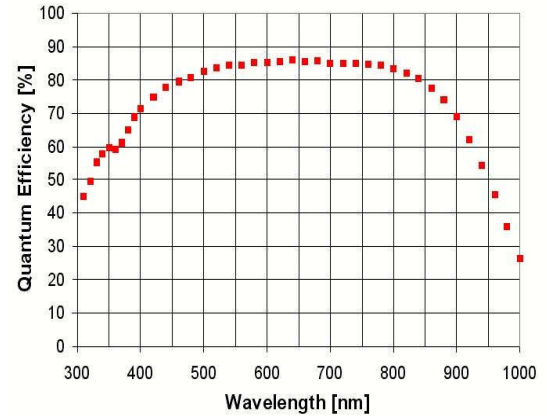


Fig. 2.15 Quantum efficiency for the “reverse” APD structure used in CMS ECAL experiment (Musienko, 2003)

Temperature dependence

In order to get a noticeable amplification, the APD device works at bias voltage near the breakdown voltage. Since the avalanche multiplication depends on the mean free path of electrons between ionizing collisions, which is temperature dependent, the APD gain varies with temperature (Fig. 2.16). This variation can be described by the temperature coefficient of the gain (Fig. 2.17):

$$\frac{1}{M} \frac{dM}{dT} = -2.2\%/^{\circ}\text{C}$$

2.9

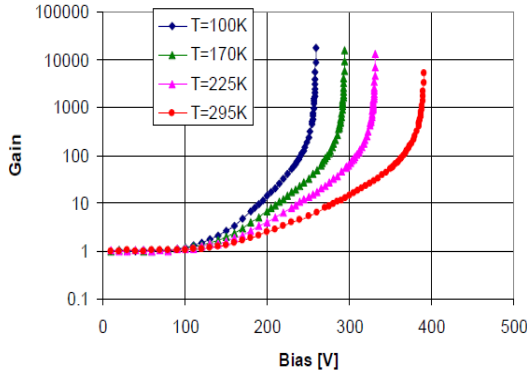


Fig. 2.16 APD gain as a function of bias voltage for four temperatures

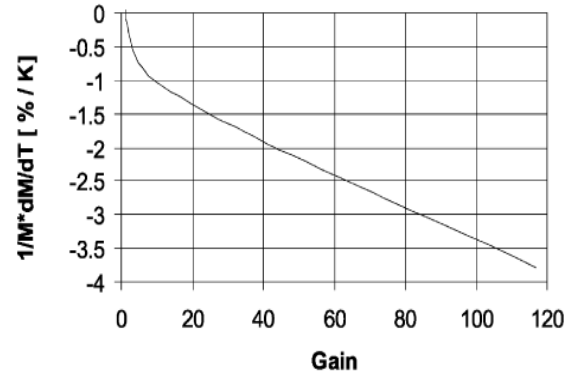


Fig. 2.17 Change of the APD gain with the operating temperature

2.4. Summary: advantages and drawbacks of PN, PIN and APD's

The PN and PIN photodiodes are the simplest, most reliable and cheapest visible light detector. They are produced today in a wide range of shapes and sizes as presented in Fig. 2.18.

The photodiodes present important advantages as: (1) operation at light levels many times higher than photomultiplier tubes; (2) they are more robust and considerably less expensive; (3) they present a high quantum efficiency for a wide wavelength range – a silicon photodiode can be used to detect radiation with wavelength from 190 nm to 1100 nm; usually anti-reflective coatings are applied to enhance the response (by up to 25%) at a required wavelength; they present very high quantum efficiencies (close to 90% for wavelengths from 350 nm to 850 nm); (4) they present a linear response; (5) they are insensitive to magnetic fields.

One important limitation of the PN and PIN photodiodes is the lack of the internal gain – an incoming photon produces only one electron-hole pair. For this reason its signal is usually so small that it needs a very low noise external amplifier.

Because the amplifier introduces some noise, photodiodes are used for measurements at relatively high light levels. They reach their highest signal-to-noise ratio when operating in high light levels because this ratio is limited only by the shot noise of the detector.



Fig. 2.18 Picture of photodiodes with different shapes and size (Hamamatsu HPK)

An APD presents an important advantage with respect to PN or PIN photodiodes by providing internal photo-electronic signal gain. Consequently it can be used at lower light levels than a photodiode. In particular for high energy experiments, it presents the advantage of a much lower (~ 50 times) nuclear counter effect with respect to PIN photodiodes.

Commercial linear-mode silicon APD's are nowadays a mature technology, with outstanding performances in terms of quantum efficiency, Excess Noise Factor and bandwidth. As an example, Fig. 2.19 a) presents a picture of Hamamatsu APD (5x5 mm² area) used to read-out the PbWO₄ crystals (Fig. 2.19 b) in the ECAL detector of the CMS experiment at CERN. Totally, 122.400 APD's have been used in the experiment.

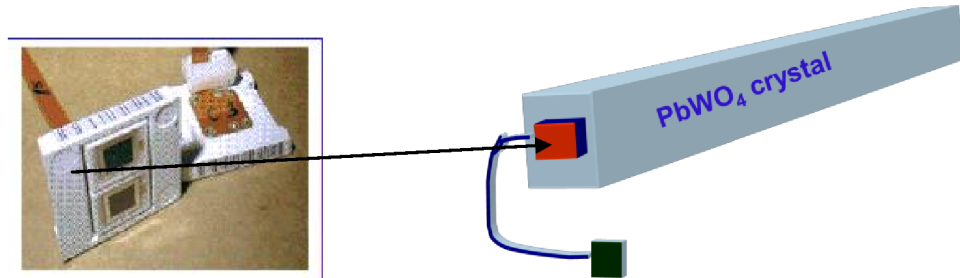


Fig. 2.19 (a) Picture of the Hamamatsu APD S8664-55 used in the CMS ECAL detector; **(b)** Picture of APD's connected to read-out PbWO₄ scintillating crystals in CMS ECAL detector

Commercial APD's can have peak quantum efficiency (QE) exceeding 80%, an excess noise factor $F \approx M^{0.3}$ for reach-through devices and as low as $F \approx M^{0.17}$ in the case of SlikTM devices fabricated by Perkin Elmer Optoelectronics. The gain is however limited to maximum 10^4 by statistical fluctuations. No single photo-electron resolution is achievable with such detectors, except at low T with slow electronics. Because of temperature variations of the APD gain, these devices have to operate in controlled temperature environment.

Only a few small APD arrays are currently present on the market, and the maximum number of devices in a single array is currently limited to 64 (8x8 module by Radiation Monitoring Devices, Inc.).

References on PN, PIN and APD's:

- Deiters, K.; Ingram, Q.; Musienko, Y.; Nicol, S.; Patel, P.; Renker, D.; Reucroft, S.; Rusack, R.; Sakhelashvili, T.; Swain, J.; Vikas, P.; (2000) Properties of the avalanche photodiodes for the CMS electromagnetic calorimeter; *Nuclear Instruments and Methods A*, 453, 223-226
- Hamamatsu Solid State Division; Characteristics and use of Si Avalanche Photodiode, *Technical information SD-28*
- McIntyre, R.J.; (1966); Multiplication noise in uniform avalanche diodes, *IEEE Transactions on Electron Devices*, ED13, 164-168
- Musienko, Y.; (2003); Advances in Avalanche Photodiodes, *Workshop of Innovative detectors for Supercolliders*, Erice, Italy
- Pansart J.P. (1997); Avalanche photodiodes for particle detection, *Nuclear Instruments and Methods in Physics Research A*, 387, 1997, 186-193
- PerkinElmer Application note, Avalanche Photodiode – A user guide, www.optoelectronics.perkinelmer.com
- Zeghbroeck, B. Van; (2007); Principals of Semiconductor Devices, University of Colorado, <http://ecee.colorado.edu/~bart/book/contents.htm>
- Webb P.P.; McIntyre R.J.; Conradi J. (1974); Properties of Avalanche Photodiodes, *RCA Review*, Vol. 35, June 1974, 234-278

3. Geiger-Mode Avalanche Photodiodes (GM-APD)

Up to here in this dissertation we have seen how two different configurations of the p-n junctions can be used to detect photons:

- a p-n junction, working at no or very low reverse bias voltage, allows to measure high light intensities (several hundreds of photons); in such device the number of charge carriers equals the number of impinging photons (the internal gain is 1) and the photon detector is called PN photodiode;
- a p-n junction, working at higher reverse bias voltage than a photodiode but below breakdown voltage and allowing to measure lower light intensities than the photodiode; in such device the number of charge carriers varies linearly with the impinging photon count (by a multiplication factor $M \sim 50$ to 500); therefore the junction is said to be in linear-mode and the photon detector is called Avalanche Photodiode (APD).

In this Section, we will show that a new configuration of the p-n junction, working at reverse bias voltage exceeding the breakdown voltage, allows the measurement of lower light intensities than an APD, going down to single photon detection. This device is called Geiger-Mode avalanche Photodiode (GM-APD) or Single Photon Avalanche Photodiode (SPAD) and it will be described in details in the following.

3.1. Physics principle

A GM-APD, like a PN photodiode or an APD, uses the pn junction functionalities to detect visible light photons. The first GM-APD structures have been proposed by Haitz (Haitz, 1964, 1965) and McIntire (McIntyre, 1961, 1966) for the studies of the micro-plasma instabilities in silicon and they are presented in the Figs. 3.1 (a) and (b). The main difference of a GM-APD with respect to the previous photon detectors is that it works at a reverse bias voltage *exceeding the breakdown voltage*.

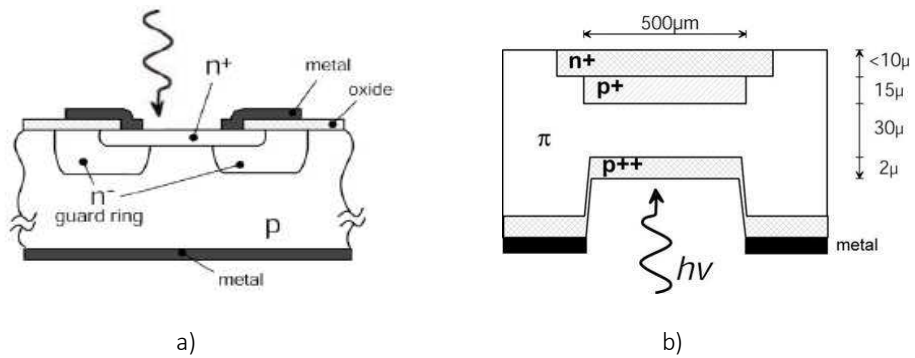


Fig. 3.1 GM-APD structures proposed for studies of micro-plasma instabilities by (a) Haitz, 1965; (b) McIntyre, 1966

The two presented structures are quite different. The one showed in Fig. 3.1 (a) represents a so-called thin-junction GM-APD; it presents the optical input window on top side (very shallow n^+ layer, $\sim 0.3 \mu\text{m}$) and the avalanche region, where the carriers find sufficient electric field to trigger an avalanche, is the n^+p depleted junction formed just below the n^+ layer (the substrate resistivity is usually $0.5\text{-}1 \Omega\text{cm}$). The breakdown region is surrounded by a deep-diffused ($5\text{-}8 \mu\text{m}$) n^- guard-ring, used to isolate and to create a small, well defined, avalanche region (i.e. avoids edge breakdown); therefore, most of the electrons generated in the neutral p region (by photon absorption and thermal generation) are drained by the wide, low electric field, n^-p junction (it has a

breakdown voltage at least a few volts higher than the active region). Almost all the electrons generated within a few microns of the n+p junction contribute to the avalanche pulses. The main features of such kind of GM-APD detectors is a breakdown voltage of 20-50 V, and small active area, with diameter from 20 to 100 μm .

The structure showed in Fig. 3.1 (b) represents a so-called thick-junction GM-APD. The region of the optical absorption, formed by the p++ layer and the π epi-layer, is separated from the avalanche region n+p+ situated on the other side of the structure; it has to be noted that the avalanche region is extending over a thickness higher than 10 μm . The main features of such structure are breakdown voltages of 200-500 V, fairly wide active area, with diameter from 100 μm to 500 μm . Such thick devices operating at high bias voltage with high avalanche current (few tens of mA), are characterized by high power dissipation (pulse peak power up to 10W). Furthermore, they are delicate and costly devices, fabricated with a dedicated technology, and therefore they are not widely used as GM-APD devices.

As mentioned previously, a GM-APD works at a reverse bias voltage exceeding the breakdown voltage. In this operation condition, the ionization rates are so high that the electrons and holes are multiplying much faster than they can be extracted in the external circuit. Therefore, the population of electrons and holes in the high field region and the associated output current increase exponentially to very high values. This growth of current is limited to a steady-state value, reached when the voltage drop across the series resistance of the diode (e.g. resistance of the neutral regions inside the silicon, space-charge resistance and contact resistance) increases enough to reduce the voltage drop across the high-field region to the breakdown voltage. If the steady-state current exceeds a minimum value (usually few tens of microamps), it continues to flow “indefinitely”. Such minimum value, below which the population of electrons and holes is sufficiently small that there is a significant probability that they will all transit out of the high-field region at some point in time without ionizing, is called *latching current* I_{latch} .

According to the model proposed by McIntyre (McIntyre, 1961), the mechanisms involved in the operation of GM-APD can be qualitatively described as in Fig. 3.2.

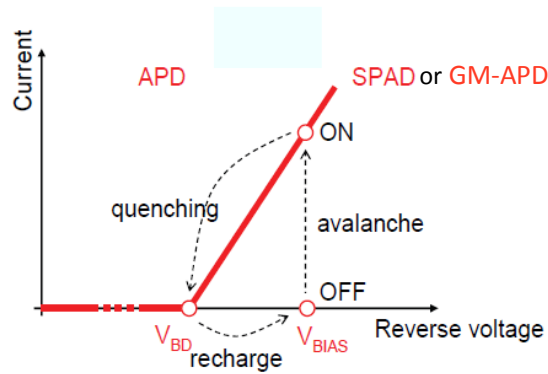


Fig. 3.2 Qualitative description of mechanisms involved in GM-APD operation

When a GM-APD is biased above the breakdown voltage (the difference between the V_{BIAS} and V_{BD} is called *excess bias voltage or overvoltage*, ΔV), it will stay in an OFF state for a given time (no current is flowing), until a carrier (electron or hole) will trigger an avalanche bringing the device into its ON state. Since not every injected or generated carrier in the depletion region will initiate an avalanche, the avalanche initiation is characterized by a *turn-on probability* P_{OI} . This parameter is defined as the probability that a carrier will initiate a current pulse that continue to grow exponentially “indefinitely” if the conditions that determine the magnitudes of the ionization coefficients in the high-field region do not change. In any real device, of course, some mechanism like contact resistance, circuit impedance, diode heating or space-charge effects, sets the current to a large self-sustaining value. Nevertheless, in order for a GM-APD to be useful as a photo-detector, the avalanche

current must be turned off by using proper *quenching mechanism*, able to reduce the bias voltage down to or below the breakdown point, and to finally restore it to its initial value, so that a new event can be detected. The device recharges to the bias voltage, and it becomes ready to detect the arrival of a new photon. The output of a Geiger-mode APD is not a linear current, as in photodiode and APD mode, but a current pulse, this explaining the dynamic behaviour of a GM-APD presented in Fig. 3.2.

The simplest mechanism used to quench the avalanche is by using a high ohmic resistor (hundreds of k Ω or more) in series with the GM-APD (Cova et al, 1996). Such quenching mechanism, presented schematically in Fig. 3.3 (a), is called *passive quenching*. Active quenching mechanisms can also be used (Cova et al. 1996): they consist on the replacement of R_Q by an active resistor made of MOS transistor and controlled by fast trigger electronics (Fig. 3.3 (b)).

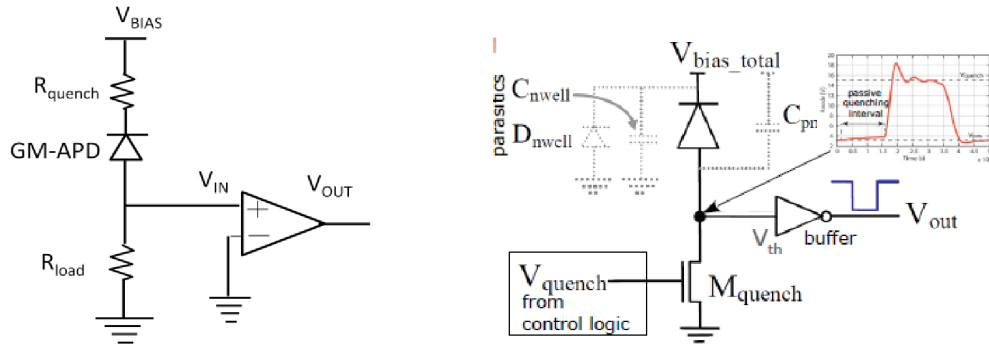


Fig. 3.3 Basic circuits of GM-APD with (a) passive quenching; (b) active quenching (Cova et al, 1996)

For a better understanding of the dynamic behavior of GM-APD, the electrical circuit model, developed by Haitz in the 1960s, will be used (Fig. 3.4 (a)). V_{BD} is the breakdown voltage, R_D and C_D the diode resistance and capacitance (R_D includes both the resistance of the neutral regions inside the silicon as well as the space charge resistance, and the contact resistance), R_Q is the quenching resistor, S is a switch with OFF/ON positions which represents the diode respectively before and when it avalanches, V_{BIAS} is the applied voltage, and V_D is the voltage across the diode (equal to V_{BIAS} when the switch S is in the OFF position: no avalanche) (Haitz, 1964). Fig. 3.4 (b) presents the time evolution of the current signal I_D and the voltage V_D across a Geiger-mode diode during the avalanche evolution, as well as the current in the external circuit I_{RQ} .

According to this model, the pre-avalanche state can be represented as a capacitance (junction capacitance, C_D) in series with the quenching resistor. Referring to Fig. 3.4 (a) this state corresponds to the switch S in the OFF condition. In this steady state (time $t < t_0$), the diode capacitance C_D is charged at $V_D = V_{BIAS} > V_{BD}$ and no current is flowing through the circuit.

When a carrier traverses the high-field region initiating an avalanche ($t = t_0$, Fig. 3.4 (b)), the new state of the system can be modeled adding to the circuit a voltage source V_{BD} and a diode series resistor R_D in parallel to the diode capacitance (the switch S is in ON position in Fig. 3.4 (a)). At t_0 a very short current spike through the diode occurs (10^{-11} s, Haitz, 1964, Senitzky, 1958) with a maximum value of $(V_{BIAS} - V_{BD})/R_D$. The diode capacitance C_D , originally charged at $V_D = V_{BIAS} > V_{BD}$, discharges exponentially through the diode series resistance R_D down to V_{BD} ($t_0 < t < t_1$). This discharge phenomenon is governed by a time constant $\tau_{\text{discharge}} = R_D C_D$.

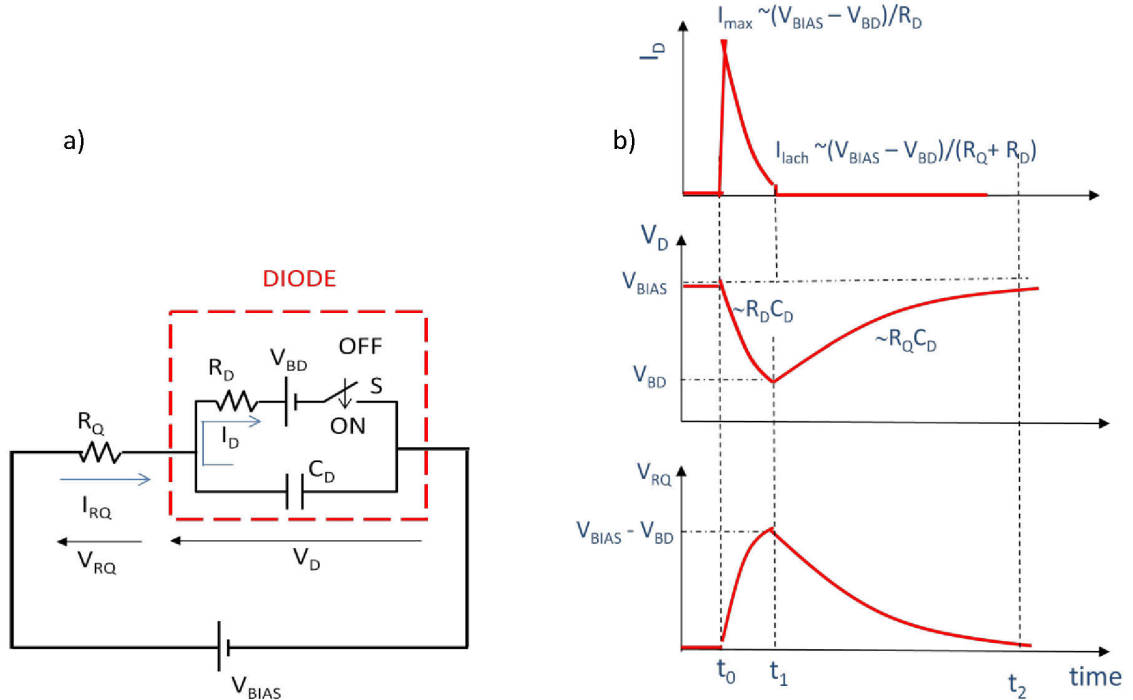


Fig. 3.4 (a) Equivalent circuit of a GM-APD (Haitz, 1964); **(b)** Time evolution of the avalanche current signal and the voltage across a Geiger-mode diode ($R_Q \gg R_D$)

As V_D is going down to V_{BD} , the voltage V_{BIAS} applied across the quenching resistor R_Q and the diode R_D causes an exponentially growing current flowing through the external circuit (i.e. quenching resistance and the diode series resistance) tending to the asymptotic value of $(V_{BIAS} - V_{BD})/(R_Q + R_D)$.

Proper value of quenching resistance R_Q is crucial to let the internal current I_D decrease to a level that statistical fluctuations may quench the avalanche. If the quenching resistor is high enough to reduce the voltage across the diode, the 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. The probability of such a fluctuation, called *turn-off probability* P_{10} , becomes significant when the diode current is below a critical value of $\sim 10\text{--}20 \mu\text{A}$ (defined as latching current, I_{latch}). However, excessively large quenching resistors lead to very long dead times (if the detector is in off state and not available to detect photons, then the sensitivity of the detector is reduced). To obtain the fastest speed of response, it is desired that the diode is quenched and the voltage restored across the detector as soon as possible. The latching current determines a strict limit on the lower value of R_Q to some hundreds of $k\Omega$. The average time needed to stop the avalanche, when this condition is satisfied, is in the sub-nanoseconds range. Therefore, the proper quenching regime can be defined for an overvoltage value ΔV in the range of few volts typically ($0 < \Delta V < (R_Q + R_D) \cdot I_{latch}$).

The value of the diode resistance R_D depends on the semiconductor device structure: it is lower than 500Ω for types with a wide area and thick depletion layer and from a few hundred ohms to some $k\Omega$ for devices with a small area and a thin junction (Cova et al., 1996).

As the avalanche process is terminated, the switch is again open and the circuit returns to its initial configuration. The capacitance previously discharged at V_{BD} , starts recharging to the bias voltage with a time constant $\tau_{recharge} = C_D R_Q$ ($t_1 < t < t_2$), and the device becomes ready to detect the arrival of a new photon. Since the R_Q is usually much higher (hundreds of $k\Omega$) than the diode series resistance (few hundreds of $k\Omega$), the $\tau_{recharge}$ is

much longer than the $\tau_{\text{discharge}}$ (voltage output can also be pick-up adding a small R_{load} between cathode of the diode and ground assuming $R_{\text{load}} \ll R_q$).

3.2. Characteristics and performances

Collaboration : M. Boscardin, A. Collini, C. Piemonte, A. Pozza, N. Zorzi, (FBK-irst, Trento, Italy)

N. D'Ascenzo, D. Herbert, GM. Collazuol, A. Del Guerra (Pisa University and INFN, Pisa, Italy)

R. Battiston (INFN, Perugia, Italy)

During my post-doctoral fellowship at FBK-irst (Trento, Italy), GM-APDs devices with integrated passive quenching resistance have been produced in the same fabrication runs like the Silicon Photomultiplier detectors (as we will see in the next Section, a GM-APD in series with its integrated passive quenching resistance on the same substrate represents the elementary cell of a Silicon Photomultiplier detector). The characteristics presented in this chapter are based on the experimental measurement of GM-APD FBK-irst devices (i.e. GM-APD in series with integrated quenching resistance). This work has been performed together with the FBK-irst group during the period 2004-2006 and published in the reference Piemonte et al., 2007. In particular, the experimental set-ups presented in this Section have been developed by my-self for the GM-APD device characterization.

The sketch of the section of a shallow-junction GM-APD produced at FBK-irst during a production run of year 2005, is shown in Fig. 3.5. The structure was build on highly doped p-type silicon substrate (i.e. $\sim 4 \times 10^{19} \text{ cm}^{-3}$) with 3-5 μm lowly doped epitaxial layer ($\sim 10^{15} \text{ cm}^{-3}$) grown on top. This latter forms the active thickness of the device since it is fully depleted at the operating voltage. The junction is created by an arsenic implantation and it is located very superficially, at about 100 nm from the top silicon surface (simulated peak concentration $5 \times 10^{18} - 1 \times 10^{19} \text{ cm}^{-3}$, arsenic energy $\sim 40 \text{ keV}$). The breakdown voltage value is fixed by a further boron implantation created underneath (peak concentration $\sim 3 \times 10^{16} - 1 \times 10^{17} \text{ cm}^{-3}$; boron energy $\sim 100 \text{ keV}$). Thus, the final structure, from top to bottom, is n+/p/ π /p+substrate.

To optimize the photon-detection efficiency in the short wavelength region of 420 - 450 nm, the doping profiles and the dielectric layers presented the following configurations:

- an anti-reflective coating, composed of two layers of 20 nm SiO_2 and 30 nm Si_3N_4 , have been deposited on top of the structure, optimizing the transmittance of the input light;
- a very abrupt junction of about 100 nm depth has been implemented by a very shallow n+ implantation of a relatively "low" doping concentration (e.g. peak value $5 \times 10^{18} - 1 \times 10^{19} \text{ cm}^{-3}$) layer and by a relatively high doping concentration of the p-type implant ($3 \times 10^{16} - 1 \times 10^{17} \text{ cm}^{-3}$) over the lowly doped epi-layer, this determining a very thin high-field region.

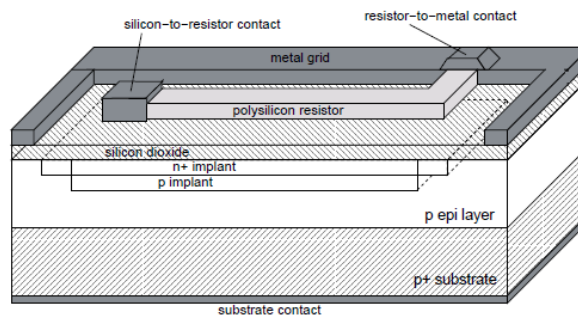


Fig. 3.5 Sketch of a GM-APD produced at FBK-IRST (Piemonte et al., 2007)

The total area of the diode is $40 \times 40 \mu\text{m}^2$. A passive quenching of the GM-APD is assured by a quenching resistance, obtained by poly-silicon implantation on the top of the junction area.

Simulated doping profiles and the electric field distribution related to presented structure are shown in Fig. 3.6 (a). The simulated Arsenic profile has been checked by using an experimental method called Secondary Ion Mass Spectrometry (more details about this technique will be given in the Part B of this dissertation). The result of SIMS measurement is presented in Fig. 3.6 (b). In close agreement with the simulation, we can observe a very shallow As profile (max depth of $\sim 300 \text{ nm}$) with a concentration peak of $\sim 3 \times 10^{19} \text{ cm}^{-3}$ located at a depth of $\sim 100 \text{ nm}$. In addition, the simulation of the absorption curves for 3 wavelengths showing the optimization of the antireflective coating for blue light detection (420 nm) are presented in Fig. 3.6 (c) (Piemonte, 2006).

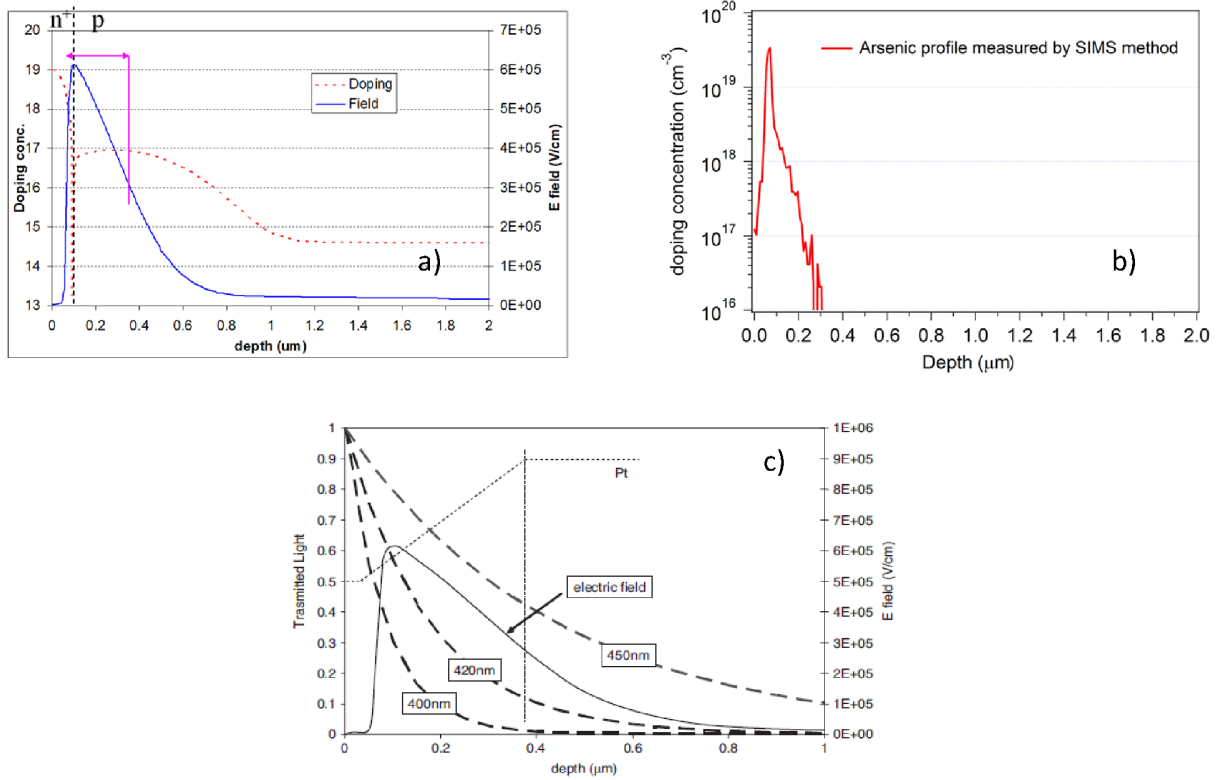


Fig. 3.6 (a) Simulated doping profiles and the electric field distribution for the GM-APD produced at FBK-irst; **(b)** measured As profile by using SIMS method; **(c)** the absorption curves for 3 wavelengths (Piemonte, 2006)

3.2.1. Static characteristics

The experimental set-up developed at FBK-irst for the static characterization of the GM-APD devices is presented in Fig. 3.7. A Karl-Suss probe-station has been used to perform electrical contacts and a HP 4156C Semiconductor Parameter Analyser has been used to bias the detector and record the current.

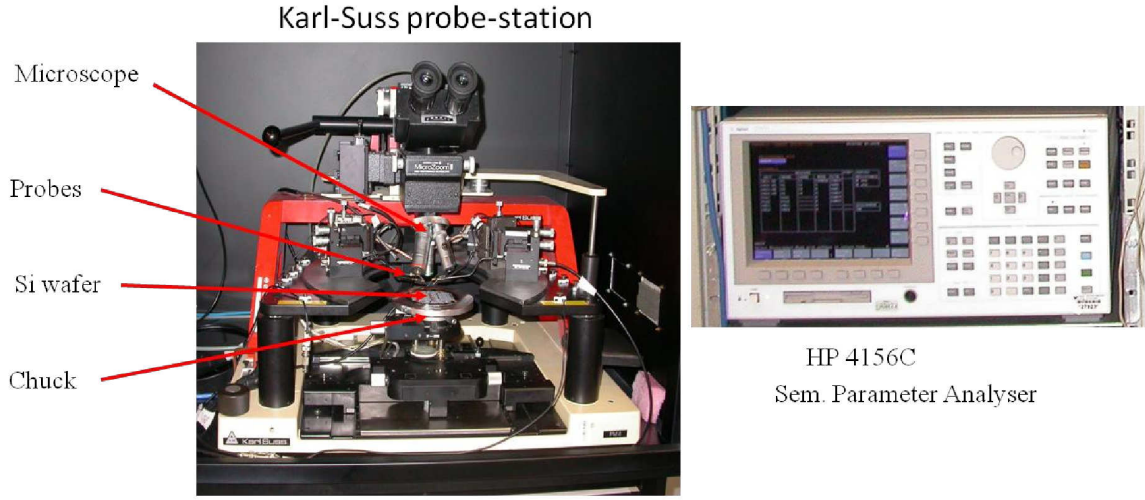


Fig. 3.7 Experimental set-up used at FBK-IRST for the static characterization of GM-APD devices

Figs. 3.8 (a) and (b) show examples of the IV plots in these two bias conditions, respectively. This characterization has been carried out in the dark, at the wafer level, using a probe station and a HP 4156C Semiconductor Parameter Analyser and then repeated on the packaged devices. So far, 20 GM-APDs have been measured.

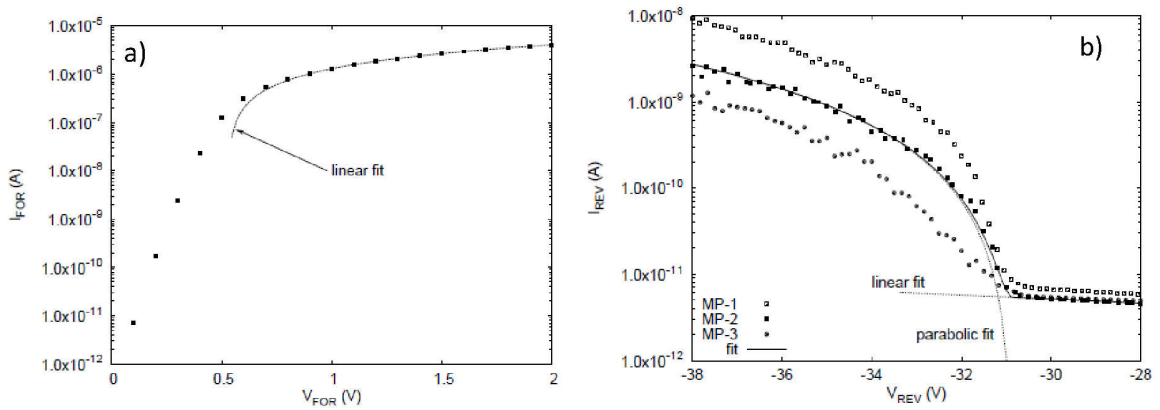


Fig. 3.8 (a) Forward and **(b)** reverse IV characteristics for GM-APDs produced in FBK-irst technology

The forward characteristic, performed in a forward bias voltage range from 0 to 2V, presents two distinct parts: one (bias voltage lower than 0.5V) in which the current is governed by the diode equivalent resistance and, thus, characterized by an exponential behavior; a second in which the quenching resistor R_Q becomes dominant and limits the current to a linear growth. From the slope of the linear part it is possible to extract the value of R_Q which is about 350k Ω .

The reverse characteristic, performed in a reverse bias voltage range from 0 to -38 V but represented in Fig. 3.8 (b) just in the breakdown range of interest, contains precious information on the functionality of the GM-APD. First of all, the breakdown voltage (V_{BD}) can be extracted. From Fig. 3.8 (b) it is clearly visible that V_{BD} is about 31V, in a good agreement with the expected theoretical values as presented in the following.

From theoretical calculations of breakdown conditions and the field dependence of the ionization rates (Sze, 1981), two formulas can be used to estimate the maximum values of the electrical field in the depletion region as well as the value of the breakdown voltage at room temperature. For Si junctions, the maximum value of the electrical field can be expressed as:

$$E_m = \frac{4 \times 10^5}{1 - \frac{1}{3} \log_{10}(N_B / 10^{16})} \quad (V/cm) \quad 3.1$$

where N_B is the ionized background impurity concentration of the lightly doped side of the junction, expressed in cm^{-3} . For an abrupt p^+-n junction, the background doping concentration N_B equals the donor doping concentration N_D of the n -type side of the junction, and for a n^+-p junction, the background doping concentration N_B equals the acceptor doping concentration N_A of the p -type side of the junction. For the case of FBK-irst devices, the GM-APD diode is a n^+-p junction, therefore the N_B equals the boron doping concentration in the p -type region of $\sim 3 \times 10^{16} \text{ cm}^{-3}$. Following this calculation we found a maximum value of the electrical field of $\sim 4.7 \times 10^5 \text{ V/cm}$, in quite a good agreement with the simulated value presented in Fig. 3.6 (a).

For Si one side abrupt junctions, the breakdown voltage V_{BD} can be expressed by the formula:

$$V_{BD} = \frac{\epsilon_0 \epsilon_{Si} E_m^2}{2q} (N_B)^{-1} \quad 3.2$$

where $\epsilon_0 = 8.854 \times 10^{-14} \text{ F/cm}$ is the permittivity of the free space, $\epsilon_{Si} = 11.9$ is the silicon dielectric constant (also referred as relative permittivity), q is the electron charge, E_m is the maximum electrical field, and N_B is the ionized background impurity concentration of the lightly doped side of the junction. Taking into account the value of the maximum electrical field previously calculated, a predicted value of $\sim 25 \text{ V}$ is expected for the breakdown voltage in the case of FBK-irst devices. The agreement with the measured value is satisfactory.

As previously mentioned, 20 devices were tested showing a very uniform behavior from this point of view. Concerning the current level below the breakdown (I_{pre-BD}), this is determined by the carriers generated both in the surface as well as in the bulk depleted region around the junction (more details about different contributions are given in the Annex 1):

$$I_{pre-BD} = I_{surface} + I_{bulk} \propto V_{bias} \quad (linear) \quad 3.3$$

In FBK-irst devices, the first component is dominant as confirmed by the fact that the dark count rate of the device operated in Geiger mode is orders of magnitude lower than the value predictable from that current level (e.g. surface charges do not contribute to breakdown avalanches).

Concerning the current above breakdown, the 20 samples showed quite high variations. As an example, at $V_{BIAS} = 38 \text{ V}$ (excess voltage $\Delta V = V_{BIAS} - V_{BD} = 7 \text{ V}$), the maximum current difference between the different GM-APD's is about a factor 10 (see lower and higher curves of Fig. 3.8 (b)). It is extremely interesting to note that the breakdown current as a function of the bias voltage is characterized by a parabolic growth. The procedure to be used to determine the exact law has to take into account that the device current above breakdown is given by the sum of the surface leakage and the breakdown currents, respectively. In order to consider only the breakdown currents, the surface current, extrapolated from a linear fit before the breakdown, has been subtracted from the device current. This pure breakdown component has then been fitted with a parabola. The linear and parabolic trends for the device MP-2 are plotted in Fig. 3.8 (b) using dashed lines. The solid line is the sum of the two curves and matches rather well the experimental data. The vertex of the parabola is located at the breakdown voltage. This law is maintained up to about 38-40V, then the current starts increasing at a much

higher rate finally reaching (at about 47V) a resistance-limited value (this part of the measurement is not represented in the Fig. 3.8 (b)).

The overall behavior of the IV characteristic above breakdown can be explained considering the dynamics of the device (more details about the dynamic parameters will be given in the next Section). The DC current is the average charge flowing through the GM-APD in 1s. In an ideal device, it is given by the charge released in one pulse (gain, G) times the number of pulses per second (dark count rate, DCR):

$$I_{post-BD} \approx q \cdot G \cdot DCR \propto q \cdot \Delta V \cdot \Delta V \quad (\text{quadratic}) \quad 3.4$$

Thus, depending on both the gain and dark count rate versus voltage behaviors, the DC current features a well defined shape. In FBK-first devices the IV curve above breakdown is parabolic because, as will be shown in the following sections, both the gain and the dark count rate grow linearly with the bias voltage. At $V_{BIAS} = 38-40V$, the asymptotic value of the current flowing through the junction ($I_{latch} = \Delta V / (R_D + R_Q)$) reaches a level for which the time needed to quench the avalanche (Haitz, 1964) becomes comparable to the recharging time constant ($R_Q \times C_D$), and, as a consequence, the amount of charge per pulse increases more than linearly. From (Haitz, 1964) it can be seen that, for $I_{latch} = 20\mu A$, the average time needed for the avalanche quenching is about 5ns and increases very quickly for larger current values. In FBK-first devices the current reaches an asymptotic value of $20\mu A$ applying an over-voltage of $\Delta V = 20\mu A \times 360k\Omega = 7V$. 38-40V is exactly the voltage at which the DC current starts growing more than parabolic. This is only a simplified explanation of the behavior of the IV plot above breakdown because the growth of the after-pulsing process with the bias voltage might anticipate the overgrowth of the DC current.

From the previous considerations it is evident that the IV characterization is of huge importance in view of a large production of GM-APD's because it allows a fast discrimination between good and bad devices. Indeed, given a certain device with a known gain, the IV curve provides precise information on: (i) the dark noise; (ii) the breakdown voltage value; (iii) the quenching resistor value.

3.2.2. Dynamic characteristics

Given the dynamic nature of the GM-APD already mentioned in the Section 3.1, a dynamic characterization of the device is necessary to measure the important parameters determining its performances when used as photon detector. Since GM-APD signals generated by an absorbed photon and by a dark generated carrier are identical, a characterization of the device in the dark has been easily performed without any light source. Therefore, the results of such characterization, performed on GM-APDs presented in Fig. 3.5 will be presented in the following.

A dedicated experimental set-up has been developed for the dynamic characterization of GM-APD devices (Fig. 3.9). It employs a GM-APD device connected to a commercial wide bandwidth amplifier and read-out by an oscilloscope. The amplifier has a bandwidth of ~ 2 GHz and an input resistance R_{load} of 50Ω (acting as a current-to-voltage converter) with a gain of 100. All presented measurements have been performed at $23^\circ C$.

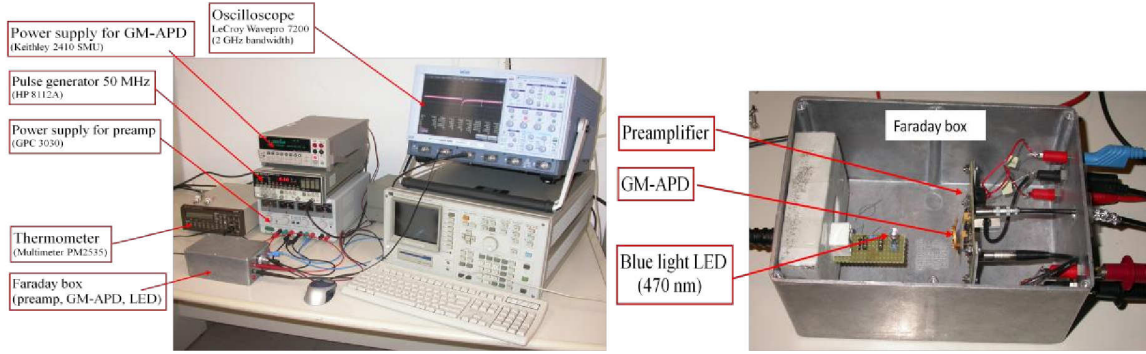


Fig. 3.9 Experimental set-up developed at FBK-IRST for the dynamic characterization of GM-APD devices

a. Signal shape

Following the equivalent circuit of a GM-APD presented in Fig. 3.4, the signal coming from such device is expected to show a very fast (hundreds of ps) leading edge, with a time constant $\tau_{\text{rise}} = R_D \times C_D$ determined both by the avalanche spreading and by the discharge of the diode capacitance C_D through the diode series resistance R_D (that includes the neutral regions resistance, the space-charge resistance of the junction, and the contact resistances), and a trailing edge, determined by the time constant $\tau_{\text{fall}} = R_Q \times C_D$ (slow exponential decay, of the order of tens to hundreds of ns). The diode capacitance C_D can be estimated from the geometry of the micro-cell applying the parallel-plate capacitance equation: $C_D = \epsilon_{\text{Si}} \epsilon_0 A / d$, where $\epsilon_0 = 8.854 \times 10^{-14}$ F/cm is the permittivity of the free space, $\epsilon_{\text{Si}} = 11.9$ is the silicon dielectric constant, A is the diode area and d is the depletion thickness. Considering a $4 \mu\text{m}$ thick epitaxial layer and a device area of $40 \times 40 \mu\text{m}$, an indicative value is 40fF.

The measured dark signal, for three bias voltages, is shown in Fig. 3.10 (a). The shape is different from what expected since it presents a very fast leading edge, but the trailing edge is composed by a fast exponential decay in addition to the expected slow exponential decay (i.e. tail). This behavior can be explained adding to the equivalent circuit of the GM-APD a parasitic capacitor C_Q in parallel to the quenching resistance (Fig. 3.11). This parasitic element is physically related to the fact that the poly-silicon quenching resistor lays on top of the junction area (see Fig. 3.5).

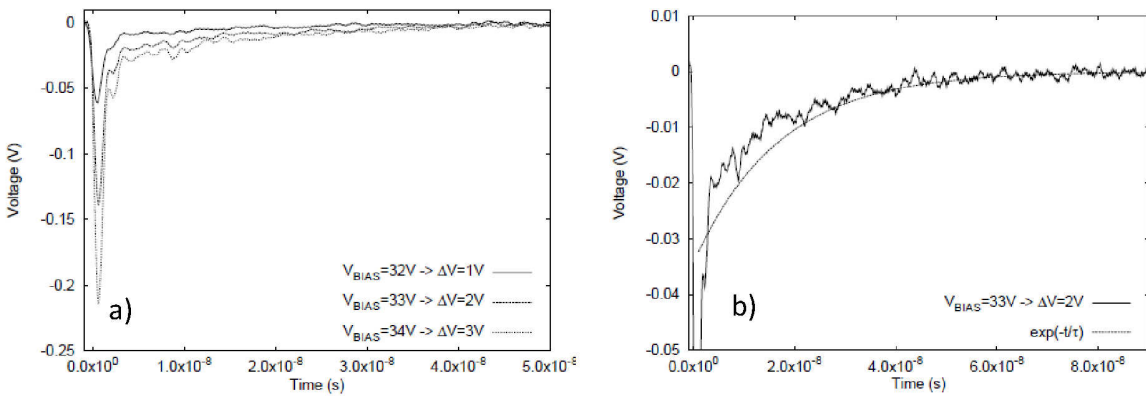


Fig. 3.10 (a) GM-APD signal at three different bias voltages **(b)** zoom of the trailing edge, showing the fast part and the slow tail; a theoretical exponential decay with the time constant $\tau_{\text{fall slow}}$ is also shown

According to this new electrical model, the common point to C_D , R_Q and C_Q is originally at V_{BIAS} . At t_0 (switch ON) a very short current spike occurs through the diode, in a time range of $\sim 10^{-11}$ s (Haitz, 1964, Senitzky, 1958) with a maximum value of $(V_{BIAS}-V_{BD})/R_D$. Subsequently, the diode capacitance C_D , originally charged at $V_D = V_{BIAS}$ discharges through the diode series resistance R_D down to V_{BD} . In the same time, the parasitic capacitor C_Q discharges from V_{BIAS} to V_{BD} through the diode series resistance R_D . The falling down voltage at the common point gives rise to an exponential growing current through the external circuit (i.e. quenching resistance) tending to the asymptotic value of $(V_{BIAS}-V_{BD})/(R_Q+R_D) \sim (V_{BIAS}-V_{BD})/R_Q$ for $R_D \ll R_Q$. Therefore the overall discharge phenomenon, represented by the fast leading edge of the signal in Fig. 3.10 (a), is governed by a time constant $\tau_{rise} \sim R_D \times (C_D+C_Q)$. For this process R_{load} has been neglected (assuming $R_{load} \ll R_Q$).

At very high frequencies or in transient conditions (i.e. during avalanche process) C_Q behaves as a low impedance path in parallel with R_Q . So, part of the avalanche current flows through this capacitance to the output. This gives the fast part of the trailing edge that appears as a spike. This additional fast component, characterized by a time constant $\tau_{fall\ fast} \sim R_{load} \times (C_Q + C_D)$, is determined by the fast recharging of the capacitances C_Q and C_D through R_{load} .

When the avalanche is quenched (switch OFF), the capacitances recharge up from V_{BD} to V_{BIAS} through $R_Q + R_{load}$. This determines the slow tail of the signal with a time constant $\tau_{fall\ slow} \sim (R_Q + R_{load}) \times (C_D + C_Q) \approx R_Q \times (C_D + C_Q)$, since R_{load} is usually much smaller than R_Q . So, two signals with two different time constants (one fast, the other slow) are superimposed building the overall signal shape.

It should be noted that GM-APDs devices characterized by a resistor not overlapping the diode area have also been designed and tested at FBK-IRST. As a matter of fact, the signal does not show the additional fast component of the trailing edge.

From the experimental point of view, we have to keep in mind that additional factors of the read-out system can influence the time constants of the measured signal, such as the finite bandwidth of the amplifier and the oscilloscope as well as the inductances and capacitances in the signal chain.

As mentioned earlier, the output signal shape presented in Fig. 3.10 (a) has been obtained by converting the current signal of the GM-APD into a voltage by means of the resistor R_{load} and a voltage amplifier (see Fig. 3.11). In this read-out configuration, R_{load} has a strong influence on the GM-APD signal shape (the signal shape presented in Fig. 3.10 (a) can be obtained only for small R_{load} of the order of 25-50 Ω). Detailed measurements and simulations of the influence of the R_{load} on the GM-APD signal shape have been performed by our Italian group (Corsi, 2006).

Fig. 3.10 (b) shows the signal of a GM-APD device biased at 33V along with an ideal GM-APD response featuring an exponential decay with a peak calculated from the equation: $V_{max} = I_{max} \times A = \Delta V / (R_D + R_Q) \times (50 \times 100)$, where A is the amplifier gain given by the product of the input load resistance (50 Ω) and the voltage gain (100). The time constant of the decay, which corresponds to $\tau_{fall\ slow} \sim R_Q \times (C_D + C_Q)$, and turns out to be ~ 15 ns. The recovery time, defined as the time needed to recharge the diode to 99% of the bias voltage, is around $5\tau_{fall\ slow} \sim 80$ ns. This is an important parameter because it gives an idea of the maximum allowed photon rate in case of an illumination with continuous light. Spice simulations presented in Fig. 3.12 shows clearly the two contributions for the output current: one fast through C_Q , one slow through R_Q .

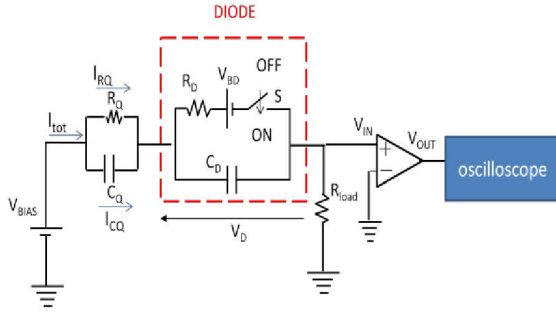


Fig. 3.11 Equivalent circuit of the GM-APD produced at FBK-irst (it includes a parasitic capacitance C_Q , in parallel to R_Q); the device is connected to a voltage amplifier via a load resistance R_{load}

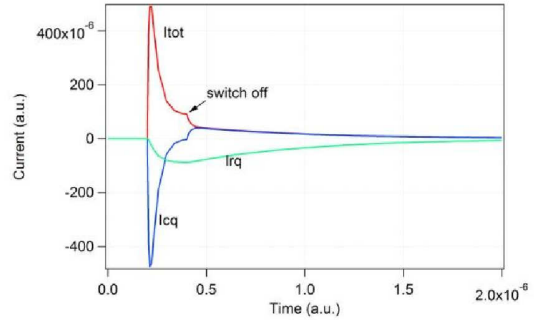


Fig. 3.12 Spice simulation of $I_{tot} = I_{RQ} + I_{CQ}$ of the circuit presented in Fig. 3.11; the spike of I_{tot} is determined by the presence of C_Q

b. Signal gain

The gain of a GM-APD is defined as the number of carriers created during an avalanche discharge. It is expressed by the formula:

$$G = \frac{Q}{q_e} = \frac{(C_D + C_Q)\Delta V}{q_e} = \frac{(C_D + C_Q)(V_{BIAS} - V_{BD})}{q_e} \quad 3.5$$

where q_e is the electron charge. As observed from the previous formula, the gain has a linear dependence of excess bias voltage ΔV in the quenching regime, but it can presents few sources of non-linearity response (non proportionality).

The GM-APD intrinsic gain fluctuations differ in nature compared to APD where the statistical process of internal amplification shows characteristic fluctuations. In the case of GM-APD, the gain fluctuations are related to breakdown voltage uniformity within the active region and to quenching time fluctuations:

$$\frac{\delta G}{G} \approx \frac{\delta V_{BD}}{V_{BD}} \oplus \frac{\delta \tau_{quench}}{\tau_{quench}} \quad 3.6$$

In addition, the after-pulses can contribute to gain fluctuations (more details in the following Section).

From the experimental point of view, the GM-APD gain can be calculated from the time integration of the device signals presented in Fig. 3.10:

$$G = \frac{Q}{q_e} = \frac{\int I(t)dt}{q_e} = \frac{\int V_{out}(t)dt}{R_{load} \cdot G_{amplifier} \cdot q_e} \quad 3.7$$

where the $\int V_{out}(t)dt$ is the area of the signal seen at the oscilloscope, the $G_{amplifier}$ is the gain of the amplifier, the R_{load} is the input amplifier load resistance and q_e is the electron charge..

The spectrum of the signal area measured on GM-APD devices produced at FBK-IRST, for an integration time of 100ns (that includes the whole signal), is shown in Fig. 3.13. Four spectra are visible, corresponding to bias voltages respectively of 1, 2, 3 and 4V above breakdown. Each distribution has a dominant peak that can be very well fitted by a Gaussian function. The centroid of the fit is the most probable charge ($Q = (C_D + C_Q) \times \Delta V$).

This is clearly visible in Fig. 3.14 showing the charge expressed in number of carriers as a function of the bias voltage along with a linear fit. The intercept with the x axis corresponds to the breakdown voltage value of $\sim 31V$ that was extracted from the IV measurement. The slope is proportional to the capacitance of the micro-cell

($C_D + C_Q$), which, in this case, is about 50fF. Since we already calculated from the geometrical parameters $C_D = 40$ fF, the slope ($C_D + C_Q$) gives ~ 50 fF, so we can conclude that C_Q is around 10 fF. The order of magnitude of the gain for such devices is typically 10^6 (Fig. 3.14).

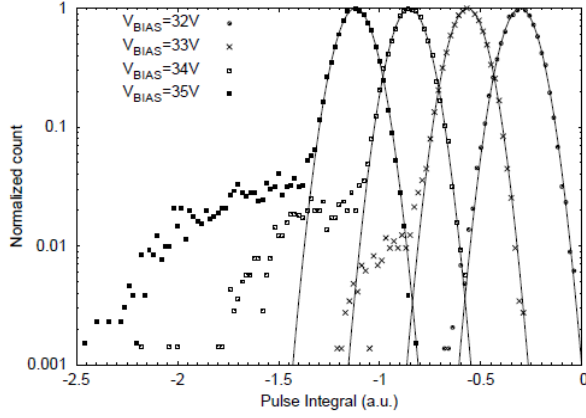


Fig. 3.13 Spectra of GM-APD signal charge (integration time of 100 ns) for 4 bias voltages

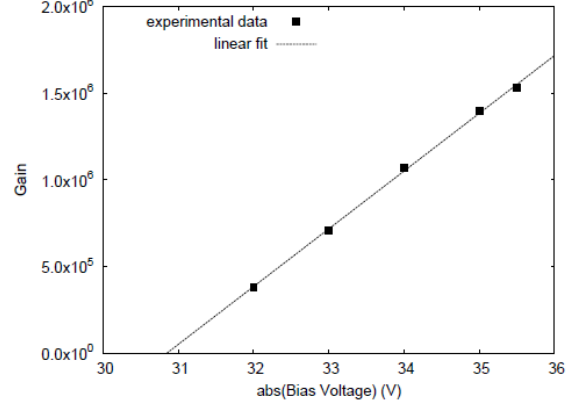


Fig. 3.14 Gain of the GM-APD as a function of the bias voltage

The variance (σ^2) of the gaussian is determined by two contributions: the noise introduced by the read-out system and the fluctuations of the GM-APD gain. The first term can be extracted from the variance of the pedestal which is not visible in the plots. The measured resolution ($\sigma/\text{peak charge}$) both including and excluding the noise contribution is shown in Fig. 3.15 as a function of the bias voltage. Since the variance is constant with V_{BIAS} , the curves decrease with a trend following the reciprocal of the charge signal growth. The resolution reaches extremely low values indicating that the device is working properly, with a good uniformity of the breakdown voltage within the active region and an effective avalanche quenching.

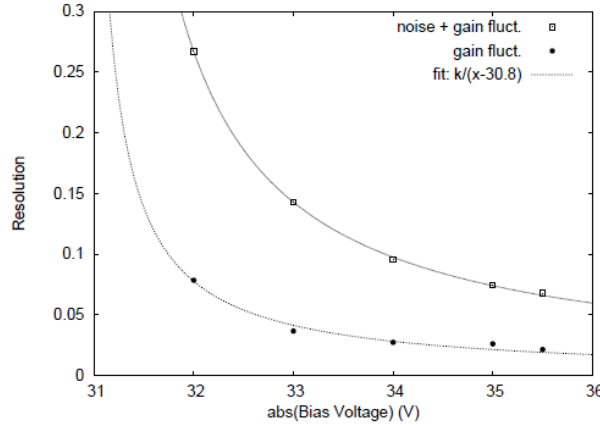


Fig. 3.15 Charge resolution (σ/peak) of the GM-APD as a function of the bias voltage

Going back to Fig. 3.13, it can be noted that the experimental data present a tail in the distribution corresponding to events with an anomalous charge increasing with the bias voltage. From an inspection of the signal shape associated to these events it was found that, within the integration time, the main pulse was always followed by a second one. These secondary correlated pulses are called after-pulses and they will be discussed in details in the Section 3.2.2 c.

c. Dark count rate

In a GM-APD, bulk generated carriers traversing the depleted region can trigger avalanches and produce output current pulses even when the device is kept in dark conditions. The per-second rate at which such dark pulses occur is called the dark count rate (DCR) in Hertz and it represents the intrinsic noise limit of the device. This parameter is fundamental for the GM-APD detector since it imposes a limit on the minimum detectable signal. The main phenomena contributing to these carrier generations are: (1) thermal generated carriers, enhanced by trap-assisted tunneling (TAT); and (2) band-to-band tunneling. These phenomena are described in details in Annex 1.

The dark count rate includes primary (uncorrelated) and secondary pulses (correlated). Primary dark pulses are mainly due to carriers thermally generated in the GM-APD depletion region. Two other contributions can participate to the increases of the dark rate as a function of V_{BIAS} : field-assisted enhancement of the emission rate from generation centers (TAT) and an increase of the avalanche triggering probability, but they are negligible with respect to thermal generated carriers.

The secondary dark pulses are related to after-pulsing effects that may strongly enhance the total dark count rate. During the avalanche of thermally generated carriers, some carriers are captured by deep trap levels in the junction depletion layer and subsequently released with a statistically fluctuating delay, whose mean value depends on the life time of the deep levels actually involved (Cova, 1991). Released carriers can retrigger the avalanche, generating after-pulses correlated with a primary avalanche pulse. The number of carriers captured during an avalanche pulse increases with the total number of carriers crossing the junction, that is, with the total charge of the avalanche pulse (gain). Therefore, after-pulses effect increases with the excess bias voltage (overvoltage).

From the experimental point of view, the GM-APD DCR can be determined by measuring the time elapsed between subsequent dark pulses. The statistical distribution of these measured delays has the characteristic exponential behavior of the independent Poisson events as presented in Fig. 3.16, for the case of a GM-APD fabricated at FBK-IRST. The exponential decay factor is the rate of the dark count events and can be equivalently evaluated by counting the number of primary pulses per second.

The primary dark count rate as a function of the bias voltage is shown in Fig. 3.17. The experimental data can be fitted by a linear function having the $y=0$ intercept at the breakdown voltage. Since the number of thermally generated carriers in the active region above breakdown is expected to be constant with the bias voltage (the epitaxial layer is already fully depleted at the breakdown point), the primary dark count rate as a function of V_{BIAS} should reflect the behavior of the triggering probability. In order to validate the assumption of constant carrier generation in the active region beyond breakdown, the sample was illuminated with a continuous constant light causing a growth of the pulse rate of about a factor of 5 with respect to dark. Also in this condition, the growth of the count rate with the bias voltage could be fitted by a linear function confirming that triggering probability increases linearly with V_{BIAS} .

Looking carefully to the Fig. 3.16, we can observe that the first bin of the distribution is much higher than the expected value. A close up of the first part of the curve, represented as a normalized distribution of the time elapsed between two consecutive pulses, is shown in Fig. 3.18. It was obtained counting only those events characterized by two peaks with a time distance not exceeding 300ns. We can observe that the probability of the second pulse decreases very quickly with time and more than 90% of the secondary pulses occur within 30ns. This spectrum is found to be independent from the bias voltage (at least up to $\Delta V = 5V$). The distribution can be fitted with three exponential decays, each corresponding to trap levels located in three different positions in the energy band-gap. It must be noted that the triggering probability of secondary carriers is not constant in the first

50ns since the diode voltage is increasing from V_{BD} to the nominal bias voltage. As a consequence, the first exponential term cannot be directly used to extract the energy level position.

From an inspection of the signal shape associated to the events plotted in the Fig. 3.18, it was found that, within the integration time, the main pulse was always followed by a second one (Fig. 3.19).

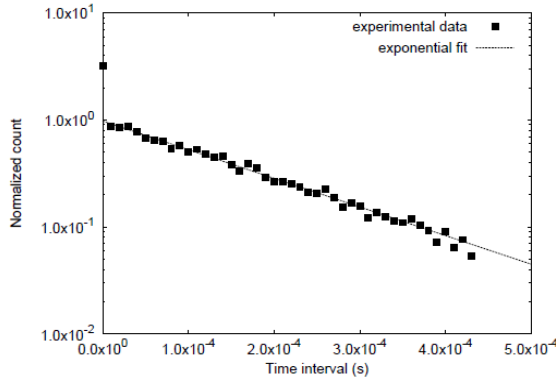


Fig. 3.16 Time distribution of dark events in the GM-APD produced at FBK-IRST at a bias voltage of 35V

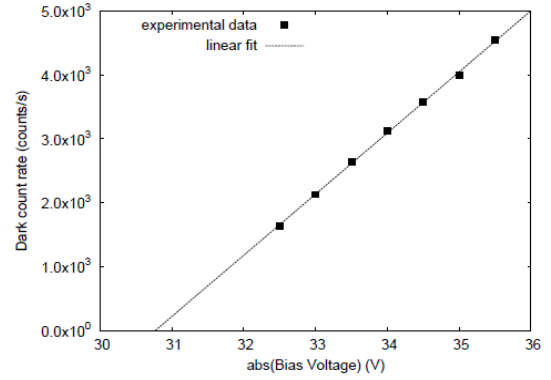


Fig. 3.17 Dark count rate of a GM-APD produced at FBK-IRST as a function of the bias voltage

These secondary correlated pulses are called after-pulses. The main pulses are exactly overlaid at time 0, whereas the secondary pulses are randomly distributed after the primary. The height of the second pulse depends on the relative position with respect to the first one, because the diode voltage is still slowly increasing towards V_{BIAS} and consequently the gain changes according to the plot in Fig. 3.14. It is interesting to note that, from the convolution of the secondary pulse peak (dashed line), a recovery time constant consistent with the one previously reported can be extracted. The relatively high number of counts characterized by secondary pulses taking place in such a short time cannot be simply attributed to normal dark count because, as shown before, its rate is in the order of few kHz. The only phenomenon that can explain this behavior is after-pulsing, for which carriers trapped during the main avalanche are released after a short time and trigger, with a certain probability, a second discharge (see Fig. 3.20).

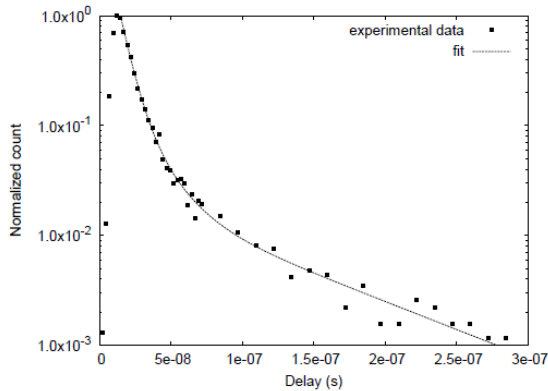


Fig. 3.18 Spectrum of the delay time from the primary pulse to the after-pulse

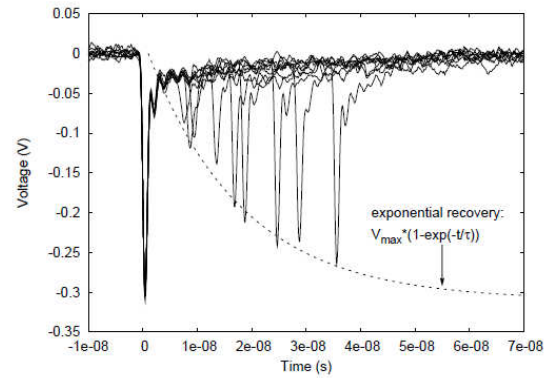


Fig. 3.19 Collection of primary pulses followed by an after-pulse measured on a GM-APD device produced at FBK-IRST

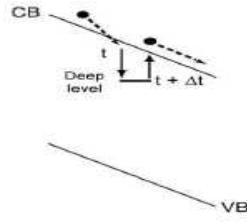


Fig. 3.20 After-pulses mechanism

The percentage of counts presenting after-pulses can be estimated from the integral spectra shown in Fig. 3.13. The Fig. 3.21 presents the normalized integral of both the spectra and the Gaussian fits reported in Fig. 3.13. The difference between the maximum levels reached by the experimental integral spectra (1 in Fig.3.21) and the corresponding integral gaussian fits is the percentage of events with after-pulse. Fig. 3.22 reports this value as a function of the bias voltage. As already evidenced, it increases with V_{BIAS} , and more in detail the growth can be fitted with a parabolic function (solid line in Fig. 3.22). This trend can be explained examining the quantities that influence the occurrence of the phenomenon. First of all, it is proportional to the number of carriers crossing the high- field region that, as already shown, grows linearly with the bias voltage. This is valid as long as not all the traps are filled during a discharge. Secondly, the probability to have an after-pulse depends on the triggering probability. Previously in the text it has been shown that this parameter grows linearly with V_{BIAS} as well. Hence, the parabolic trend observed in Fig. 3.22 is determined by the product of two quantities that grow linearly with the bias voltage.

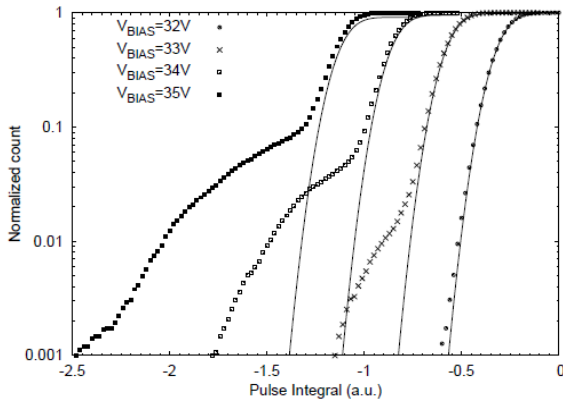


Fig. 3.21 Normalized integral spectra of the signal charge presented in Fig. 3.13

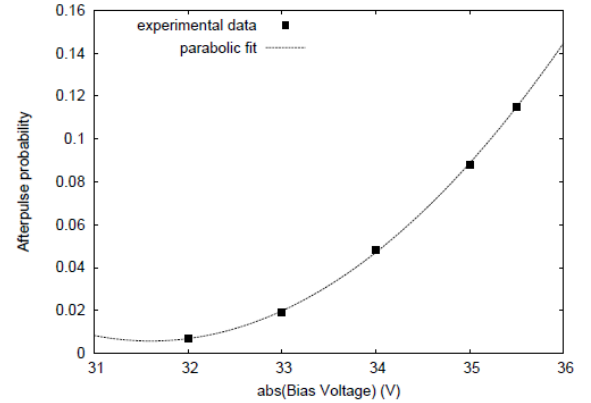


Fig. 3.22 After-pulse probability obtained from the integral spectra

Taking into account the different parameters influencing the appearance of the after-pulses, the after-pulse probability can be expressed by the following expression

$$P_a(t) = P_c \cdot \frac{\exp\left(-\frac{t}{\tau}\right)}{\tau} \cdot P_{01} \propto \Delta V^2 \quad 3.8$$

where: P_c is the trap capture probability, depending on the traps number N_t and the number of carriers flowing during the avalanche ($\propto \Delta V$); τ is the trap lifetime, depending on the energy of the trap level; P_{01} is the avalanche

triggering probability ($\propto \Delta V$). Therefore, for a low after-pulse probability, low number of traps is required. Moreover, a long recovery time as well as working at low gain can reduce artificially the presence of the after-pulses in real measurements.

d. Dead time

The dead-time is typically defined as the time a photon-counting detector needs to recover after it registers a photon and is ready to register another one. Reducing the effects of dead-time is the most direct way to achieve higher detection rates. This recovery time may be due to the physical properties of the detector and/or the pulse processing electronics.

In the GM-APD for example, the detector dead-time is related to the avalanche quenching and the return to the initial bias conditions (Haitz, 1964). However, in the case of a passively quenched GM-APD this is not strictly the case. As the device is recharged via the quenching resistor (a phase that can last from several *tens to a few hundreds of nanoseconds*), it becomes increasingly biased beyond its breakdown voltage, so that it is able to detect a new photon arrival prior to being fully restored.

The acceptable peak rate of detection for a photon counting solution is application specific. Obviously, for higher count rate we need shorter dead time of the detection system (ability to detect a new event). Therefore, the higher the count rate, the larger the non-linearity of detection.

Clearly the dead time should be kept as small, and as consistent as possible in order to achieve the highest possible dynamic range of incident photon flux and least variation in photon count output to a certain photon arrival rate. However, short dead times are often accompanied by enhanced after-pulsing probability due to inadequate trap flushing time. A solution for reducing the GM-APD non-linearity has been proposed by Golovin and Sadygov in the years '80 (Sadygov, 1997; Golovin, 1999), by connecting in parallel, on the same substrate, hundreds or thousands of cells, each cell representing a GM-APD connected in series with its passive quenching resistance. These detectors have been very much developed during last year's and they can be found today with different names in the literature: Silicon Photomultiplier (SiPM), Multi-Pixel Photon Counter (MPPC), Multi-pixels Geiger-mode Avalanche Photo-Diode (MPGM APD), Avalanche Micro-pixel Photo-Diode (AMPD) etc. (see Renker, 2005). Major part of my last year's work has been dedicated to the study of these detectors. Therefore, they will be treated in details in the Section 4.

3.2.3. Photon detection efficiency

The photon detection efficiency (PDE) of a GM-APD is defined as the ratio between the number of output counts of the device exceeding its dark count rate and the number of photons impinging on the detector surface. In order to be detected, a photon must be focalized in the active area of the detector, it should generate a primary carrier (more precisely an electron-hole pair) and the primary carrier should succeed in triggering an avalanche. Therefore, the PDE of a GM-APD depends on three parameters: the quantum efficiency QE (wavelength dependent), the triggering probability P_{01} (linked to the impact ionization rates of electrons and holes) and the geometrical efficiency FF (the ratio of the sensitive area on the total detector surface).

$$PDE = QE \cdot P_{01} \cdot FF \quad 3.9$$

Quantum efficiency

The QE represents the probability for a photon to generate an e-h pair that reaches the high field region. It is given by the product of two factors: the transmittance of the dielectric layer on top of the silicon surface and the internal QE.

The schematic representation of optical light transmittance on the surface of the GM-APD diode is presented in the Fig. 3.23 (a). Usually, it can be maximized, by implementing an anti-reflective coating (ARC) (Tyagi, 1991). The transmittance of the ARC layer is wavelength dependent. The simulation of ARC layer transmittance for stacks of different SiO₂ and Si₃N₄ thicknesses is presented in Fig. 3.23 (b). We can observe that optimum antireflective coating for blue light detection (420 nm) can be obtained for ~20 nm of SiO₂ and ~30 nm of Si₃N₄.

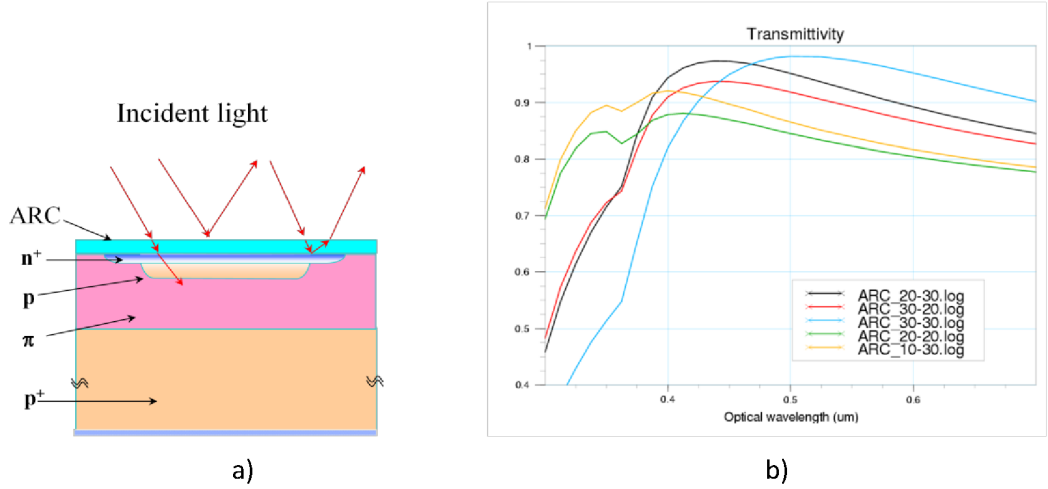


Fig. 3.23 (a) Schematic representation of optical light transmittance through the antireflective coating on the surface of the GM-APD diode; **(b)** Simulation of dielectric layer transmittance for stacks of different SiO₂ and Si₃N₄ thicknesses (C. Piemonte, not published)

The internal QE represents the probability for a photon that has passed the ARC layer to generate an e-h pair in the active thickness. An incident photon flux I_{photon} will attenuate in silicon according to the Beer-Lambert law, creating an equivalent number of electron/hole pairs:

$$I(\lambda, x) = I_{\text{photon}}(\lambda) \cdot \exp[-\alpha(\lambda) \cdot z] \quad 3.10$$

where I_{photon} is the photon flux at a wavelength λ , and α is the absorption coefficient at a given distance z into the detector. The absorption coefficient for silicon is shown in Fig. 3.24 on the main y-axis in cm⁻¹. On the alternative y-axis, the inverse of the absorption coefficient, or penetration depth, is shown in microns. To be detected in Si, the photons should have an energy exceeding the Si band-gap energy. At room temperature, E_g of Si is 1.12 eV, so the maximum detectable wavelength will be 1100 nm ($\lambda = hc/E_g = 1240/E_g$, where h is the Plank constant 6.626×10^{-34} (Js), c is the light speed 3×10^8 m/s). A thin absorption layer is suited for short wavelength detection and greater thicknesses are required to detect longer wavelengths.

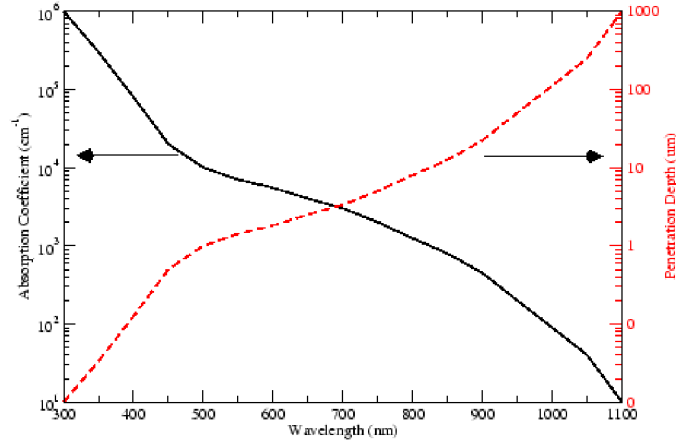


Fig. 3.24 Absorption coefficient (expressed in cm^{-1}) and penetration depth (expressed in microns) of photons in silicon

In a GM-APD diode as developed at FBK-IRST ($n^+/p/\pi/p^+$), the active layer is roughly limited on top by the undepleted n^+ layer, whereas on the bottom by the p^+ layer used for the ohmic contact or by the highly doped substrate in case of epitaxial substrates (Fig. 3.6 (a)). Indeed, when a pair is generated in those regions, there is a high probability for the electron and hole to recombine due to Auger or Shockley–Read–Hall (SRH) processes (Tyagi, 1991). For short wavelengths detection, the limiting factor is the thickness of the top n^+ doping layer. As an example, a 420nm light is almost totally absorbed in the first 500 nm of silicon (see simulation in Fig. 3.6 (b)); therefore, many of the generated e-h pairs will recombine at the surface unless the doping profile creates an electrical field that collects them immediately. To optimize the internal QE for low wavelengths, the GM-APD structure fabricated at FBK-IRST presents an extremely shallow n^+ doping layer (~ 100 nm depth), with a thin depletion region $\sim 1\mu\text{m}$ (see Fig. 3.6 (a)).

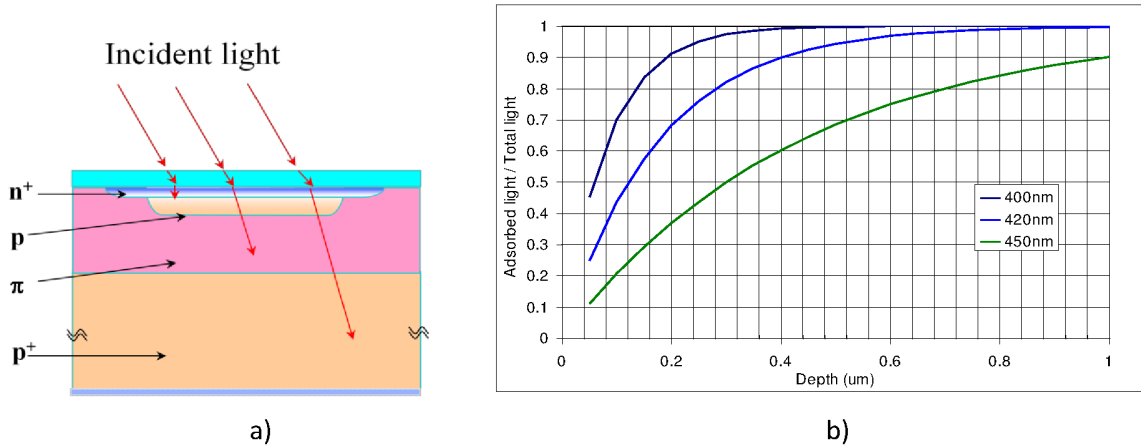


Fig. 3.25 (a) Schematic representation of light absorption in a GM-APD diode; **(b)** Simulation of absorbed light normalized to the total incident light as a function of depth in silicon (C. Piemonte, not published)

Triggering probability

Once a photon has generated an electron-hole pair, the two carriers are travelling within the diode in opposite directions: the electrons travel toward the n -side terminal and the holes toward the p -side terminal. Therefore, there is a certain probability called triggering probability P_{01} that one or both of them will generate an

avalanche when passing through the high field region. The total triggering probability can be evaluated from the following equation:

$$P_t = P_e + P_h - P_e P_h \quad 3.11$$

where P_e and P_h are the electron and hole breakdown initiation probabilities.

The breakdown initiation probabilities represent the probabilities that a chain of ionizations (produces by electrons or holes) continues to increase until the hole diode is discharged (sometimes, chains of ionizations just starts that terminates, because of carriers fluctuations to zero, resulting in a relatively low gain: $1 \sim 10^4$).

In order to understand the physical impact of the P_t (equation 3.11), an n+/p junction having a constant high-field region, extending beyond the n+ layer, is considered (see dashed line in Fig. 3.26). When a pair is generated in the left side of the high-field region [(a) case], the electron is directly collected at the n+ terminal; thus, it does not contribute to the triggering probability. The hole is forced to pass the whole high-field region and so its triggering probability is maximized and $P_t = P_{hM}$. In case of photo-generation on the right side [(c) case], the situation is symmetrical and only electrons contribute to the triggering probability, thus, $P_t = P_{eM}$. In the central region [(b) case], both carriers contribute to a different extent as a function of the interaction position and the P_t value is between P_{eM} and P_{hM} .

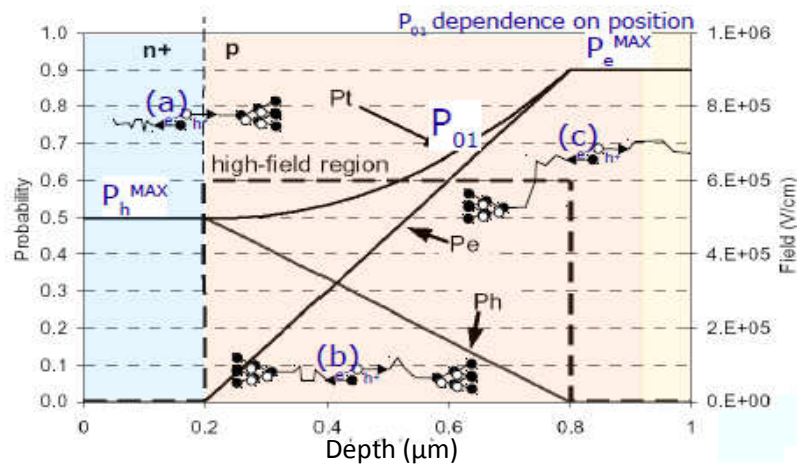


Fig. 3.26 Avalanche triggering probability as a function of the photo-generation position (Oldham, 1972)

The P_e and P_h can be calculated if the electrical field profile and the e-h ionization coefficients (α_n and α_p), are known. The parameters α_n and α_p are not well determined yet, and large discrepancies exist among the values extracted from the various models (Grant, 1973; Overstraten, 1970, see also Annex 2). Anyway, despite the differences in absolute values, some features are well established (Fig. 3.27 (a)): (i) both coefficients increase with the electric field, (ii) the electron has an ionization rate higher than the hole (e.g., at 5×10^5 V/cm, α_n is about twice α_p (Grant, 1973)), and (iii) their difference decreases with increasing fields. This behavior is reflected in the probabilities P_e and P_h (Fig. 3.27 (b)). Thus, 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. It must be noted that V_{BIAS} cannot, in any case, exceed the value for which the current flowing through the diode is higher than I_{latch} .

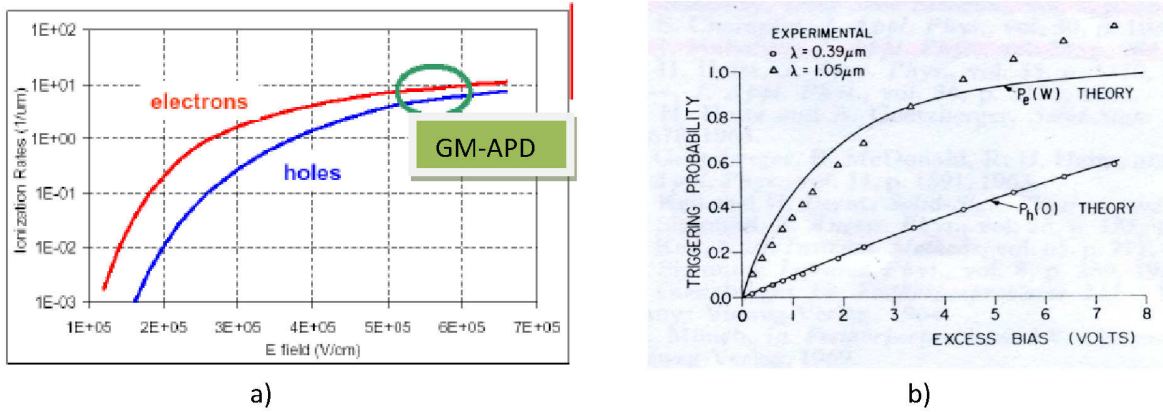


Fig. 3.27 (a) Electrons and holes ionization coefficients as a function of the electrical field; (b) experimental and theoretical electrons and holes triggering probabilities as a function of the excess bias voltage or overvoltage (Oldham, 1972)

For the case of GM-APD detectors produced at FBK-IRST, the solution of shallow top n+ doping layer does not maximize the breakdown initiation probability for short wavelengths. Indeed, most of the light is absorbed in the region where the P_t is low (Figs. 3.28 (a) and (b)). Anyway, this is the best one can do with this doping configuration, because the junction is extremely shallow, and the high field region is very thin.

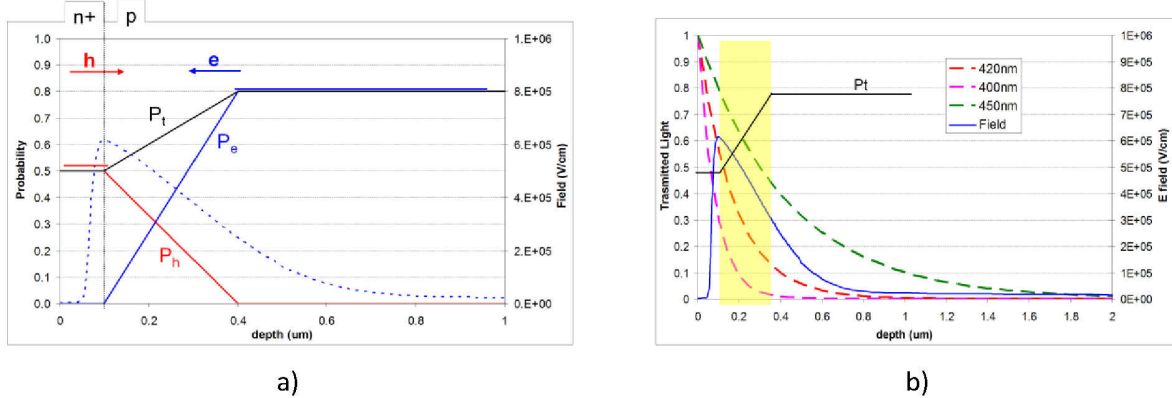


Fig. 3.28 (a) Triggering probabilities and the electric field profile for the GM-APD device produce at FBK-IRST; (b) simulation of short wavelengths light transmission in GM-APD produced at IRC-IRST; the blue light is almost absorbed in the region where the P_t is low

Geometrical efficiency

The geometrical efficiency or fill factor FF of a GM-APD represents the ratio of the detector sensitive area to the total surface. As we can see in the Fig. 3.29 (a) representing a GM-APD device produced in the 2006 first production run at FBK-IRST without any geometrical optimization, the FF is particularly determined by the dead area present at the borders of the diode (required to prevent the edge breakdown) with an additional contribution coming from the poly-silicon resistor and the external contacts. Neglecting for simplicity the last contributions, the maximum fill factor achievable for a dead border region respectively of 4 and $6 \mu\text{m}$ is shown in Fig. 3.29 (b) (Piemonte & al., 2006). We can observe that, for the first production run, the FF was not optimized: the size of the GM-APD prototype was $40 \mu\text{m}$ and, at that time we had two layout versions with fill factor of 20 and 30% respectively (a large fraction of the dead area is indeed given by the poly-silicon resistor). So it was room for an improvement and with the subsequent design (Piemonte & al., 2006) we reached a fill factor higher

than 40%. Notably, with larger pixels it is easier to reach a fill factor but at the cost of a reduction of the dynamic range.

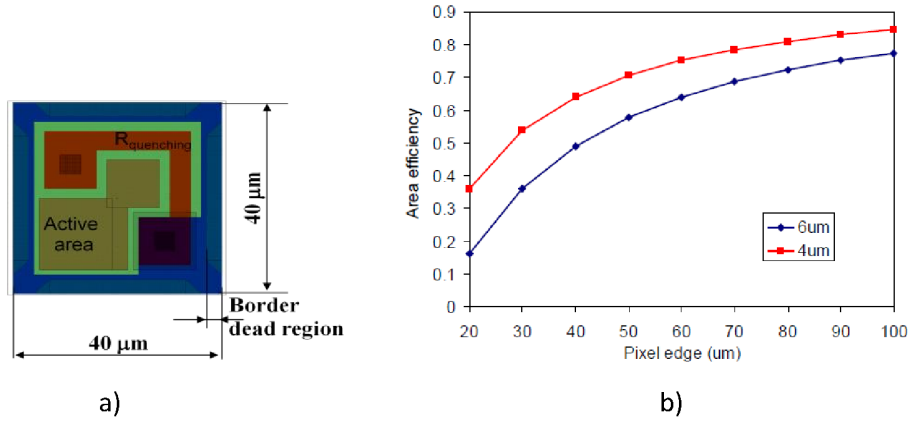


Fig. 3.29 (a) Image of a GM-APD device produced during the first run at FBK-IRST (not optimized for FF); **(b)** FF as a function of the edge size for two different technologies having a border region respectively of 4 and 6 μm

For the experimental measurement of the GM-APD PDE produced at FBK-irst as a function of wavelength, an optical set-up has been developed, equipped with a white light source, attenuation filters and a monochromator (Fig. 3.30). The calibration of every filter and of the light from the monochromator has been performed for a set of wavelengths ranging from 380nm to 700nm using a calibrated diode (i.e. UDT 221).

Since during my work at FBK-irst, the PDE measurements have been performed only on monolithic arrays of GM-APDs (i.e. SiPM), the PDE of a single GM-APD has not been done. Nevertheless, particular measurements of PDE have been done on some devices working in “photodiode mode”. They were extracted from the same fabrication wafers, which have the same anti-reflective coating and doping profiles as the GM-APD devices. These tests help us to measure the separate QE contribution (Fig. 3.31, Piemonte et al. 2006). The measurements have been performed at different bias voltages: from 0 to 5V. Up to 3V we obtained the same curve, whereas for higher biases the QE starts increasing because of impact ionization (note that diode begins to work as avalanche photodiode, even if the breakdown voltage is $\sim 30\text{V}$). Notably, due to FBK-irst technological choices, the QE is higher than 90% for wavelengths ranging from 380 to 530nm. TCAD simulations of the QE were carried out to tune the technology and actually the agreement is very good (red line in Fig. 3.31). The transmittance of the ARC is also shown in Fig. 3.31 as well (dashed blue line). It can be noticed that the decrease of QE at short λ is caused by the poor transmittance of the ARC while at long λ the reduced thickness of the epitaxial layer determines the slow QE reduction. As we will see later in the Section 4.2.3, the poor total PDE of the arrays of GM-APD in the blue light region is determined, besides of the small FF, by a poor triggering efficiency, since this parameter is not optimized for such kind of devices.

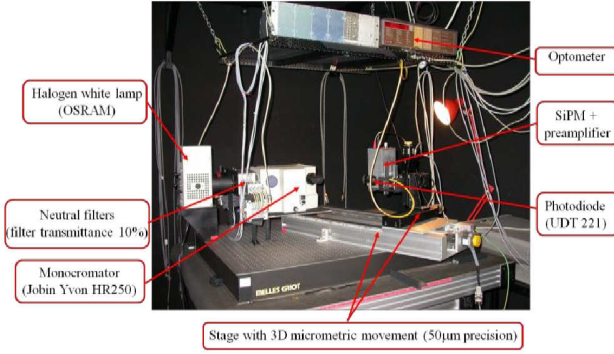


Fig. 3.30 Optical set-up developed at FBK-irst for the measurement of the optical characteristics of GM-APD devices

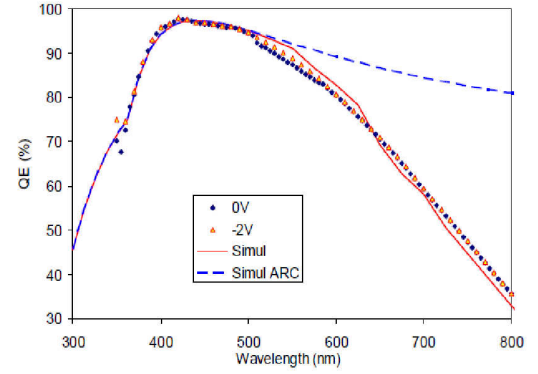


Fig. 3.31 QE vs wavelength of a 1 mm² diode having the same doping profiles and an ARC coating as the GM-APD

3.2.4. Timing resolution

The GM-APD timing resolution is defined as the time jitter (TJ) between the true arrival time of the photon at the sensor and the instant when the output current pulse is recorded.

The time jitter of a GM-APD is usually measured by sending picoseconds laser pulses directly to the detector, and measuring the statistical distribution of the time delay between the laser trigger and the measured pulse (Sciaccia, 2003).

A method based on the measurement of the fluctuations of the difference in time between successive pulses can be also used (see Collazuol, 2007). Of course, the method employed for detecting the onset of an avalanche event is of high importance, and the read-out circuit should be designed to minimize (or eliminate) the time walk effects (Leo, chapter 17, 1994).

Independent of the employed method, the timing resolution is usually quoted by the full-width at half maximum (FWHM) of the measured curve. During my work period on GM-APD and SiPM (i.e. monolithic arrays of GM-APD) either at Trento or at LAL, we did not perform direct measurements on single GM-APD devices. Therefore, in this chapter, only the physical mechanisms related to the timing resolution of the GM-APD will be discussed. The experimental results of timing resolution measurements on SiPM detectors will be reported later, in the section 4.2.4.

Fig. 3.32 presents the typical experimental histogram of a GM-APD time response to a short laser pulse characterized by a fast peak followed by a slow tail (Cova & al., 1987). The fast peak component has a Gaussian shape with a FWHM of the order of tens up to ~100ps. The main physics mechanisms contributing to this fast component are related to:

- the variation caused by the *generated carrier transit time from depletion layer to multiplication region*, which is dependent of the depth of the absorption of the incident photon (as a guideline, the transit time at carrier saturation velocity is 10ps per micron); this contribution is represented schematically in Fig. 3.33 (a), and is usually named *longitudinal build-up*. It follows that the thinner the depleted region, the better the ultimate time resolution. But, we can observe that this contribution is much lower than the measured one, therefore other jitter mechanisms presented in the following do not allow reaching this limit;

- the *statistical fluctuations of the avalanche build-up time*. Since the avalanche is triggered in a single seed point and, at steady-state, the multiplication occurs over the entire junction area, there should be a *transverse propagation of the avalanche activation* (Fig. 3.33 (b)). This transversal propagation determines a jitter of the order of 100 ps and represents the main contribution to the GM-APD timing jitter; it can be related to two physical phenomena, depending of the device structure: *transversal propagation via multiplication assisted diffusion*, dominating in few μm thin devices (Lacaita et al., 1990) or to a *transversal spreading via photon assisted propagation*, dominating in thick devices of $\sim 100 \mu\text{m}$ (Lacaita, 1993).

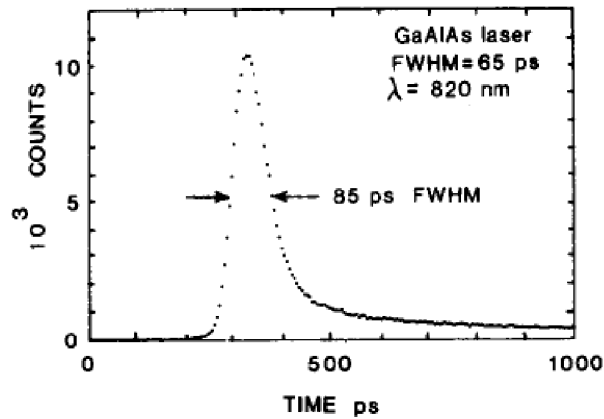


Fig. 3.32 Time response of a GM-APD to a short laser pulse at 820 nm wavelength

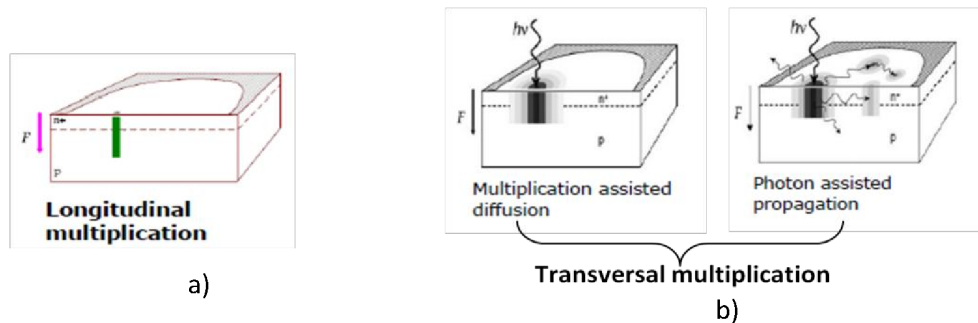


Fig. 3.33 Schematic representation of (a) longitudinal multiplication phenomenon; (b) transversal multiplication phenomena by multiplication assisted diffusion and photon-assisted propagation in GM-APD devices (Spinelli & Lacaita., 1997)

The slow tail of the time jitter distribution is related to the minority carriers, photo-generated in the neutral regions above or beneath the junction and reaching the high field region by diffusion (Fig. 3.34). For GM-APD devices build directly in the substrate, the shape of the tail depends on the incident radiation wavelength, because the absorption coefficient is wavelength dependent:

- neutral regions underneath the junction determine timing tails for long wavelengths ;
- neutral regions in the APD entrance determine timing tails for short wavelengths.

For GM-APD devices build into an epitaxial layer with a thinner junction, however, the tail is shorter (Ripamonti & Cova., 1985).

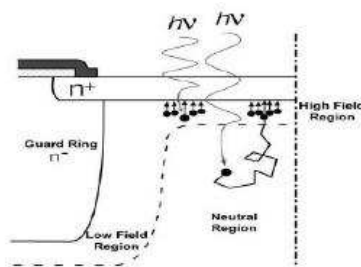


Fig. 3.34 Schematic representation of the photo-generated minority carriers above or beneath the junction and reaching the high field region by diffusion

3.3 Summary: advantages and drawbacks of GM-APD's

The GM-APD's discussed in this chapter have gain of the order of 10^6 comparable with that one of photomultiplier tubes (PMT). Therefore, they are well adapted detectors for measuring extremely low light intensities down to single photon, the reason for which they are also called Single Photon Avalanche Photodiodes (SPAD). In addition, the single photon-sensitivity of such devices is achieved along with a sub-nanosecond timing precision.

Because of good performances of GM-APD's in single photon detection with very good timing resolution, they have found diverse applications such as time-resolved spectroscopy; chemistry, physics and biology (Louis et al., 1990; Isoshima et al., 1995; Hungerford and Birch, 1996); laser ranging (Aull et al., 2002); optical time-domain reflectometry (Bethea et al., 1988); astronomy (Nightingale, 1991); investigations of quantum-mechanical phenomena (Zou et al., 1991; Kwiat et al., 1993) and studies of high field properties of semiconductors (Lacaita et al., 1993).

Moreover, they are limited in measuring high count rates, since high count rates determine an increase of the non-linearity of the device. To overcome this limitation, arrays of GM-APD's called Silicon Photomultiplier (SiPM) have been developed and they will be described in details in the next chapter.

References on GM-APD's:

- Aull, B.F.; Loomis, A.H.; Young D.J.; Heinrichs, R.M.; Felton, B.J.; Daniels, P.J.; Landers D.J.; (2002); Geiger-Mode Avalanche Photodiodes for three-dimensional imaging, *Lincoln Laboratory Journal*, Vol. 13, Issue 2, 335-350
- Bethea, C.G.; Levine, B.F.; Cova, S.; Ripamonti, G.; (1988); High resolution and high-sensitivity optical-time-domain reflectometer; *Optical Letters*, Vol. 13, Issue 3, 233-235
- Collazuol, G.; Ambrosi, G.; Boscardin, M.; Corsi, F.; Dalla Betta, G.F.; Del Guerra A.; **Dinu, N.**; Galiberti, M.; Giuliotti, D.; Gizzi, L.; Llosa, G.; Marcatili, S.; Morsani, F.; Piemonte, C.; Pozza, A.; Zaccarelli, L.; Zorzi, N.; (2007); Single photon timing resolution and detection efficiency of theIRST silicon photo-multipliers; , *Nuclear Instruments and Methods in Physics Research A*, Vol. 581, 461-464
- Corsi, F.; Marzocca, C.; Perrota, A.; Dragone, A.; Foresta, M.; Del Guerra, A.; Marcatili, S.; Llosa, G.; Collazuol, G.; Dalla Betta, G.F.; **Dinu, N.**; Piemonte, C.; Pignatelli, G.U.; Levi, G.; (2006), Electrical characterization of silicon photomultiplier detectors for optimal front-end design, *IEEE Nuclear Science Symposium Conference Record*, N30-222, 1276-1280
- Cova, S.; Ripamonti, G.; Lacaita A.; (1987); Avalanche semiconductor detector for single optical photons with a time resolution of 60ps; *Nuclear Instruments and Methods in Physics Research A*, Vol.253; 482-487

- Cova, S.; Lacaita, A.; Ripamonti, G. (1991); Trapping phenomena in avalanche photodiodes on nanosecond scale, *IEEE Transactions on Electron Devices*, Vol. 12, No. 12, 685-687
- Cova, S.; Ghioni, M.; Lacaita, A.; Samori, C.; Zappa, F. (1996), Avalanche photodiodes and quenching circuits for single-photon detection. *Applied Optics*, Vol. 35, No. 12, 1956-1976, ISSN 1943-8206
- Frenkel, J.; (1938); On pre-breakdown phenomena in insulators and electronic semi-conductors, *Physical Review*, Vol. 54, Issue 8, 647-648
- Ganichev, S.D.; Yassievich, I.N.; Prettl, W.; Diener, J.; Meyer, B.K.; Benz, K.W.; (1995); Tunneling ionization of auto-localized DX centers in terahertz fields; *Physical Review Letters*, Vol. 75, Issue 8, 1590
- Ganichev, S.D.; Ziemann, E. ; Prettl, W. ; Yassievich, I.N.; Istratov, A.A.; Weber, E.R.; (2000); Distinction between the Poole-Frenkel and tunneling models of electric-field stimulated carrier emission from deep levels in semiconductors, *Physical review*, Vol. 61, Issue 15, 10361-10365
- Golovin, V.M., et al, (1999) ; *Russian patent*, nr. 1644708
- Grant, W.N. ; (1973) ; Electron and hole ionization rates in epitaxial silicon at high electric fields, *Solid State Electronics*, Vol. 16, Issue 10, 1189-1203
- Grove, A.S. (1967); Physics and technology of semiconductor devices, *J. Wiley and Sons*
- Haitz, R.H. (1964); Model for the electrical behavior of the micro-plasma, *Journal of Applied Physics*, Vol. 35, No. 5, 1370-1376
- Haitz, R.H. (1965); Mechanisms contributing to the noise pulse rate of avalanche diodes, *Journal of Applied Physics*, Vol. 36, No. 10, 3123-3130
- Hungerford, G.; Birch, D.J.S.; (1996); Single-photon timing detectors for fluorescence lifetime spectroscopy, *Measurements Science and Technology*, Vol. 7, Issue 2, 121-135
- Hurkx, G.A.M.; Graaff, de H.C., Kloosterman W.J., Knuvers, M.P.G.; (1992); A new analytical diode model including tunneling and avalanche breakdown, *IEEE Transactions on Electron Devices*, Vol. 39, Issue 9, 2090-2098
- Isoshima, T. Isojima, Y.; Hakomori, K.; Kikuchi, K.; Nagai, K.; Nakagawa H.; (1995); Ultrahigh sensitivity single-photon detector using a Si avalanche photodiode for the measurement of ultra-weak bio-chemi-luminescence; *Journal of Scientific Instruments*, Vol. 66, Issue 4, 2922-2926
- Kwiat, P.G.; Steinberg, A.M.; Eberhard, P.H.; Chiao, R.Y.; Petroff, D.; (1993); High-efficiency single photon detectors, *Physical Review A*, Vol. 48, Issue 2, 867-870
- Lacaita, A.; Mastrapasqua, M.; Ghioni, M. ; Vanoli, S. ; (1990) ; Observation of avalanche propagation by multiplication assisted diffusion in p-n junctions; *Applied Physics Letters*, Vol. 57, Issue 5, 489-491
- Lacaita, A.; Cova, S.; Spinelli, A. ; Zappa, F. ; (1993) ; Photon-assisted avalanche spreading in reach-through photodiodes; ; *Applied Physics Letters*, Vol. 62, Issue 6, 606-608
- Lacaita, A.; Zappa, F.; Bigliardi, S.; Manfredi, M.; (1993) ; On the bremsstrahlung origin of hot-carrier-induced-photons in silicon devices; *IEEE Transaction on Electron devices*, Vol. 40, Issue 3, 577-582
- Leo, W.R.; (1994); Techniques for nuclear and particle physics experiments, Second Edition, *Springer-Verlag Berlin Heidelberg*
- Louis, T.A.; Ripamonti, G. ; Lacaita, A.; (1990) ; Photoluminescence lifetime microscope spectrometer based on time-correlated single-photon counting with an avalanche diode detector; *Review of Scientific Instruments*, Vol. 61, Issue 1, 11-22
- McIntyre, R.J. (1961); Theory of micro-plasma instability in silicon, *Journal of Applied Physics*, Vol. 32, No. 6, 983-995
- McIntyre, R.J. (1966); Multiplication noise in uniform avalanche diodes, *IEEE Transactions on Electron Devices*, Vol. 13, Issue 1, 164-168
- Nightingale, N.S.; (1991); A new silicon avalanche photodiode photon counting detector for astronomy, *Experimental Astronomy*, Vol. 1, Issue 6, 407-422
- Oldham, W. G.; Samuelson, R.R.; Antognetti, P. ; (1972) ; Triggering phenomena in avalanche diodes, *IEEE Transactions on Electron Devices*, Vol. 19, Issue 9, 1056-1060
- Overstraten, Van R.; Man, de H; (1970); Measurement of the ionization rates in diffused silicon p-n junctions, *Solid State Electronics*, Vol. 13, Issue 5, 583-608
- Piemonte, C. (2006); A new silicon photomultiplier structure for blue light detection, *Nuclear Instruments and Methods in Physics Research A* 568, 224-232
- Piemonte, C. ; Battiston, R. ; Boscardin, M. ; Dalla Betta, G.F.; Del Guerra, A.; **Dinu, N.**; Pozza, A.; Zorzi, N.; (2007); Characterization of the first prototypes of Silicon Photomultiplier fabricated at FBK-irst, *IEEE Transactions on Nuclear Sciences*, Vol. 54, Issue 1, Part 2, 236-244

Piemonte, C.; Battiston, R.; Collazuol, G. ; Dalla Betta G.F. ; Del Guerra, A. ; **Dinu, N.** ;, Levi, G. ; Llosa, G. ;, Marcatili, S.; Marzocca, C. ; Pozza, A. ; Zorzi, N. ; (2006) New results on the characterization of FBK-first silicon photomultipliers, *IEEE Nuclear Science Symposium Conference Record*, N42-4, 1566-1569

Renker, D.; (2005); Silicon Photomultipliers; *International Conference on New developments in photo-detection*; Beaune, France

Ripamonti, G.; Cova, S., (1985); Carrier diffusion effects in the time-response of a fast photodiode; *Solid state Electronics*, Vol. 28, 925-931.

Sadygov, Z.; et al.; (1997), *Russian Patent*, nr. 2086047 C1

Sciacca, E.; Giudice, A.; Sanfilippo D. ; Zappa, F. ; Lombardo, S. ; Consentino, R. ; Di Franco, C. ; Ghioni, M. ; Fallica, G. ; Bonanno, G.; Cova, S.; Rimini, E.; (2003); Silicon planar technology for single-photon optical detectors; *IEEE Transaction on Electron Devices*, Vol. 50, Issue 4, 918-925

Senitzky, B.; Moll, J.L.; (1958); Breakdown in silicon, *Physical Review*, Vol. 110, Issue 3.

Shockley, W.; Read, W.T.; (1952); Statistics of the recombinations of holes and electrons, *Physical Review*, Vol. 87, Issue 5, 835-842

Spinelli, A.; Lacaita, L.; (1997); Physics and numerical simulation of single photon avalanche diodes, *IEEE Transactions on Electron Devices*, Vol. 44, Issue 11, 1991-1943

Sze, S.M.; (1981); Physics of semiconductor devices, 2nd edition, *John Willey & Sons*

Sze, S.M.; Ng, K.K.; (2007); Physics of semiconductor devices – 3rd edition, *John Willey & Sons*, Hoboken, New Jersey

Tyagi, M.S.; (1991); Introduction to Semiconductor Materials and Devices, *Willey*, New York

Zappa, F.; Tisa, S.; Tosi, A. ; Coca, S. ; (2007), Principles and features of single-photon avalanche diode arrays, *Sensors and Actuators*, Vol. 140; 103-112

Zou, X.Y.; Wang, L.J.; Mandel, L. ; (1991) ; Induced coherence and indistinguishability in optical interference, *Physical Review Letters*, Vol. 67, Issue 3, 318-321

4. Silicon Photomultiplier (SiPM)

In the previous Section of this dissertation we have seen that a p-n junction working at reverse bias voltage exceeding the breakdown voltage represents a nice detector in silicon technology, having an internal gain of about 10^6 (the same order of magnitude as a photomultiplier tube) and allowing the detection of low light levels, down to single photon.

We will show in this Section how arrays of GM-APD can improve the performances of a single device, allowing the measurement of light intensity.

4.1. Physics principle

As mentioned in the previous Section, the GM-APD detector works as a binary device and it does not give information on light intensity: its signal is the same independent if one or more photons fire the detector at the same time. To overcome this drawback, a matrix structure has been proposed in the late '90-ies by Golovin and Sadygov (Golovin, 1999, Sadygov, 1997). Such structure, called Silicon Photomultiplier, is a photodiode with many micro-cells connected in parallel on a common silicon substrate. Each micro-cell is represented by a p/n junction working in Geiger-mode and connected in series with its integrated passive quenching resistance R_{quench} .

A schematic drawing of a SiPM device is presented in Fig. 4.1 (a) and pictures of SiPM's developed by FBK-irst and Hamamatsu HPK are presented in Figs. 4.2 (b) and (c). We can distinguish the micro-cells build in a thin epitaxial layer, where each micro-cell is represented by a pn junction and the quenching resistance. The micro-cells are connected in parallel through a common Si substrate on the rear side and through a metal grid on the front side. The antireflective coating covering the active surface and dedicated to improve the light transmission is also represented.

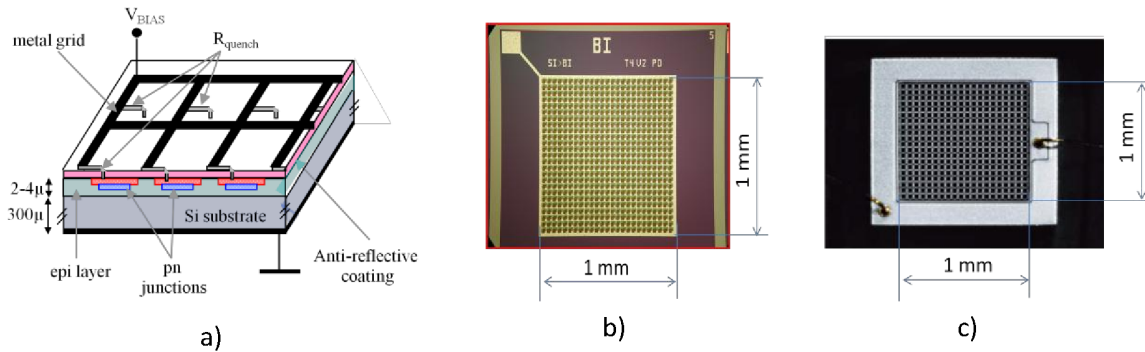


Fig. 4.1 (a) Schematic drawing of a SiPM device; (b) picture of a SiPM developed at FBK-irst; total area $1 \times 1 \text{ mm}^2$, 625 micro-cells, each micro-cell: $40 \times 40 \text{ } \mu\text{m}^2$; (c) picture of a SiPM developed by Hamamatsu HPK; total area $1 \times 1 \text{ mm}^2$, 400 micro-cells, each micro-cell $50 \times 50 \text{ } \mu\text{m}^2$

Actually, each SiPM micro-cell operates as an independent GM-APD, under reversed bias voltage V_{BIAS} of 10-20% more than breakdown voltage V_{BD} , as presented in the previous Section. Therefore, each micro-cell Geiger signal does not depend on the number of incident photons which fire the cell ("Geiger-mode"). The micro-cell gain is determined by the charge accumulated in the micro-cell capacity $C_{\text{micro-cell}}$ and the overvoltage $\Delta V = V_{\text{BIAS}} - V_{\text{BD}}$: $Q_{\text{micro-cell}} = C_{\text{micro-cell}} \times \Delta V$.

Typically, $C_{\text{micro-cell}} \cong 100\text{fF}$, and the $\Delta V \cong \text{few volts}$, so $Q_{\text{micro-cell}} \cong \text{few times } \times 100 \text{ fC}$ and therefore, the single micro-cell gain $G = Q_{\text{micro-cell}}/q \approx 10^6$ (q is the electron charge), i.e. the same order as a vacuum PMT. The micro-cell size is usually of 15 to 100 μm , and the total number of SiPM micro-cells is $\sim 100\text{-}4000/\text{mm}^2$.

Because all SiPM micro-cells work together in parallel on a common load, the output signal is the sum of the signals from all fired micro-cells. So such a device, where each micro-cell operates digitally as a binary element, works as an analogue device and it can measure the light intensity. The output charge of a SiPM is, in principle, proportional to the number of incident photons in different cells. The dynamic range is limited by the number of cells composing the device, and the probability that 2 or more photons hit the same micro-cell depends on the size of the micro-cell itself.

A schematic electrical circuit of a SiPM and an oscilloscope image showing the cumulative signals coming from different fired micro-cells are presented in Fig. 4.2 (a) and (b).

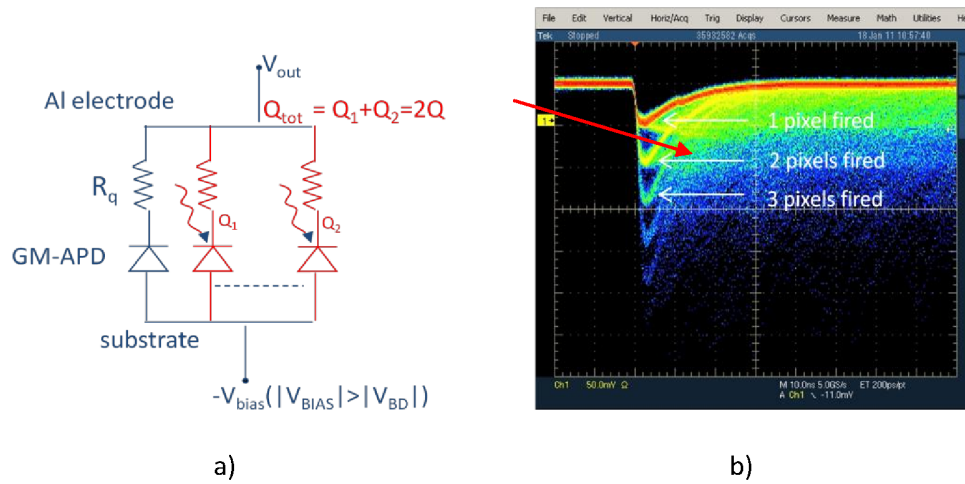


Fig. 4.2 The output signal of a SiPM is the sum of standard signals coming from different micro-cells fired in the same time; example showed on (a) simplified electrical circuit of a SiPM and (b) oscilloscope view of SiPM signals

4.2. Characteristics and performances

Technical supervision: Vincent Chaumat, Jean-François Vagnucci, Z. Amara, Cyril Bazin (LAL)

Internal collaboration: Veronique Puill, Christophe Sylvia, Chafik Cheikali (LAL)

External collaboration: Adam Para (Fermilab, USA)

After my fellowship at Trento, I continued my activities on SiPM at the Laboratory of Linear Accelerator (LAL-Orsay). As Research engineer at LAL since December 2006, my main research interest was related to the study of the physical phenomena of SiPM detectors and their applications in medical imaging. For such purposes, I have studied SiPM detectors fabricated in different technologies and their characteristics will be described in details in the following.

At the time when this work has been performed, only SiPM devices of total area of $1 \times 1 \text{ mm}^2$ were supplied by the foundries. Devices supplied by FBK-irst (Italy), SensL (Cork, Ireland), Hamamatsu HPK (Japan) during 2007 have been measured and compared (Dinu et al., 2009, a). The geometrical characteristics of these devices are presented in the Table 4.1. It can be noticed that, even if all devices cover an area of $1 \times 1 \text{ mm}^2$, the size of micro-

cells (implicitly the number of micro-cells) as well as the fill factor FF (here called geometrical efficiency ϵ_{geom}) is variable from one device to the other, depending on the producer design and technology.

The micro-cell of the SiPM's produced at FBK-irst is exactly the GM-APD device presented in the previous Section (see Section 3.2). Regarding the micro-cells of the other producers it is difficult to describe them in details, since their intrinsic characteristics are not published. However, following the polarity of the V_{BIAS} applied in between the substrate and the front side of the SiPM, we could distinguish two main categories: the n+/p micro-cells build on p-type substrate are characteristic to the FBK-irst and SensL devices, and the p+/n micro-cells build on n-type substrate are characteristic to the HPK devices. The schematic structure of these junctions as well as the profile of the electrical field is presented in Fig. 4.3.

Producer	SiPM ID	Nr. micro-cells	micro-cell size (μm)	ϵ_{geom} (%)
FBK-irst	W20-B10-T3V2PD/ I run	625	40 x 40	20
FBK-irst	W3-B3-T6V1PD/ II run	625	40 x 40	16
SensL	SPM-20	848	29 x 32	43
SensL	SPM-35	400	44 x 47	59
SensL	SPM-50	216	59x 62	68
HPK	S10362-11-25	1600	25 x 25	31
HPK	S10362-11-50	400	50 x 50	61.6
HPK	S10362-11-100	100	100x 100	78.5

Table 4.1 Geometrical characteristics of the SiPM prototypes

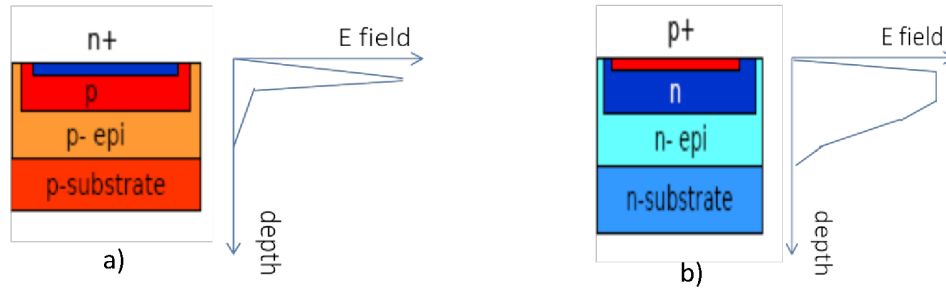


Fig.4.3 Schematic representation of two categories of SiPM micro-cells and the corresponding electrical field shape: **(a)** FBK-irst and SensL devices (micro-cells of n+/p/p-epi/p substrate); **(b)** HPK devices (micro-cells of p+/n/n-epi/n substrate)

4.2.1. Static characteristics

The schematic representation of the experimental set-up developed at LAL for the DC characterization of the SiPM devices is presented in Fig. 4.4. It consists of the SiPM connection to a Keithley source-meter (current resolution ~ 2 pA) through triaxial cables ($R_{\text{isolation}} \sim 10$ T Ω). A controlled temperature ($25 \pm 0.5^\circ\text{C}$) of the electrical tests has been assured by an LMS Fisher Bioblock climatic chamber. The temperature has been monitored by a Pt100 probe connected to a Keithley multimeter. LabView and C++ programs have been developed for the automatic data acquisition and analysis of the SiPM parameters, through a GPIB-USB connection.

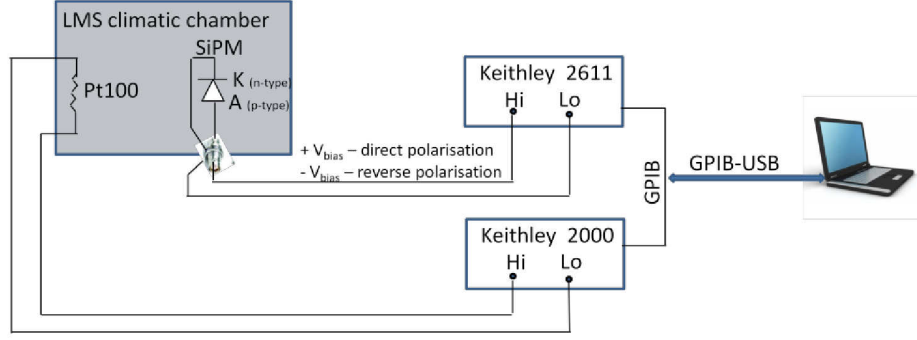


Fig. 4.4 Experimental set-up developed at LAL for the DC characterization of SiPM devices

The static characteristics (reverse and forward IV plots) of different SiPM prototypes are presented in Fig. 4.5 (a) and respectively (b). The IV plots are shown as a function of the Abs V_{BIAS} to simplify the graphical representation.

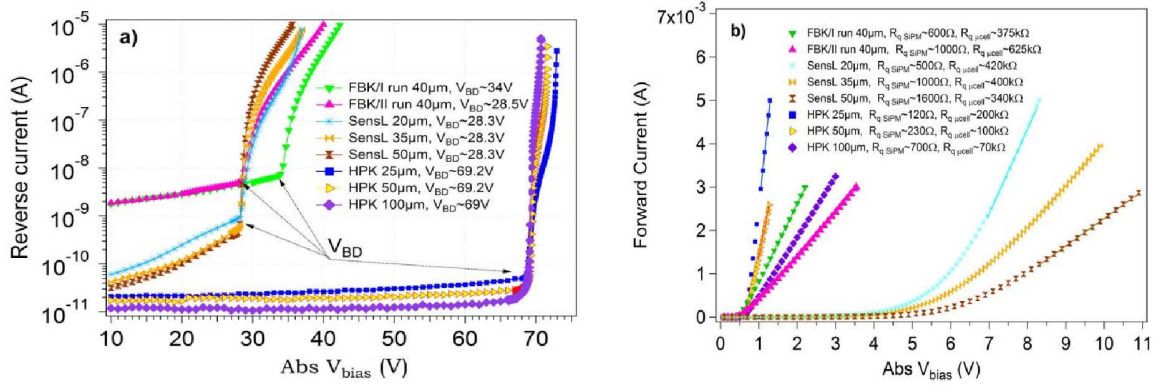


Fig. 4.5 DC characteristics of the SiPM prototypes (a) reverse; (b) forward

From Fig. 4.5 (a), the V_{BD} values @ 25°C have been extracted (indicated in the legend). It can be notice that a uniform V_{BD} has been measured over the devices coming from the same producer (e.g. $V_{BD \text{ SensL}} \sim 28 \text{ V}$, $V_{BD \text{ HPK}} \sim 69 \text{ V}$) independent of their geometrical characteristics, while different values have been measured over devices coming from different producers: $V_{BD \text{ SensL}}$ and $V_{BD \text{ FBK}} \sim 28\text{--}32 \text{ V}$ (i.e. micro-cells of n+/p/p-epi/p substrate), $V_{BD \text{ HPK}} \sim 69 \text{ V}$ (i.e. micro-cells of p+/n/n-epi/n substrate).

The dark current above V_{BD} represents the average current flowing through different SiPM micro-cells during 1 s and it depends on the charged released in one pulse (gain G) and the total number of dark pulses per second (dark count rate DCR). Assuming that a maximum value of $\sim 10 \mu\text{A}$ can represent a good compromise in between the G and the DCR at a certain V_{BIAS} (will be confirmed by the AC measurements), the value of the ratio $\Delta V/V_{BD}$ can be estimated. It is worth stressing that when comparing devices with different V_{BD} , the value $\Delta V/V_{BD}$ matter and not only the overvoltage ΔV , because the performances are related to the excess electrical field above the breakdown level. A $\Delta V/V_{BD}$ ranging from 0% to 10–15% has been found for FBK-irst and SensL devices (i.e. $V_{BD} \sim 28\text{--}30 \text{ V}$), while $\Delta V/V_{BD}$ ranging from 0 to 3–5% has been estimated for HPK devices (i.e. $V_{BD} \sim 69 \text{ V}$).

From the linear fit of the forward IV plots (Fig. 4.5 (b)), the values of micro-cell R_Q can be calculated as $N_{\text{micro-cells}} \times R_{\text{measured}}$ (R_Q values indicated in the legend). It can be noticed that hundreds of $k\Omega$ have been found for the micro-cell R_Q of all measured devices (ranging from 70 $k\Omega$ to 600 $k\Omega$).

4.2.2. Dynamic characteristics

As in the case of a GM-APD device, the SiPM signals generated by an absorbed photon or by a dark generated carrier are identical. Therefore, a basic characterization of a SiPM can be easily performed in the dark, without any light source.

We have developed at LAL a dedicated experimental set-up for the dynamic characterization of the SiPM devices in the dark (Fig. 4.6). This set-up allows the bias of the SiPM by the Keithley source-meter and its readout by an MITEQ wide-band (0.01–500 MHz) voltage amplifier connected to a Tektronix digital oscilloscope (500 MHz, 5 GS/s). The amplifier presents 50Ω input impedance, which acts as a current-to-voltage converter followed by an amplifying stage having a gain of 45 dB. The dark conditions and the protection against electromagnetic noise have been assured by mounting the SiPM and the amplifier in a metallic box. A controlled temperature ($25\pm0.5^\circ\text{C}$) has been assured by an LMS Fisher Bioblock climatic chamber. LabView programs have been developed for the automatic data acquisition and analysis of the SiPM dynamic parameters.

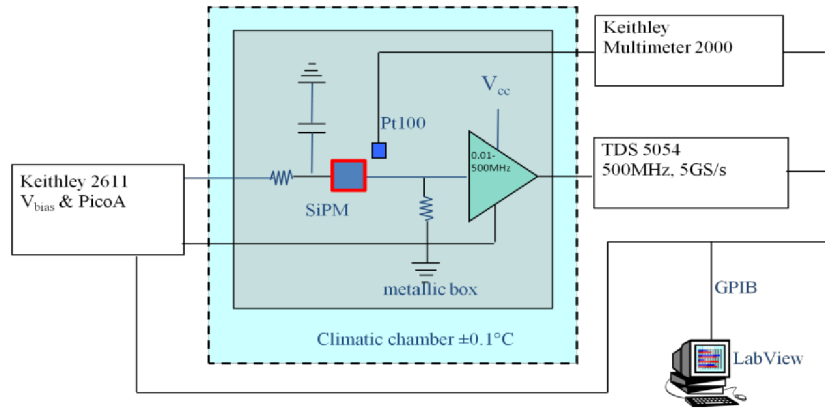


Fig. 4.6 Experimental set-up developed at LAL for the AC characterization of SiPM devices

a. Signal shape

In this Section we will study the SiPM signals coming from a single micro-cell and applied to the rest of the array.

Since each SiPM micro-cell is represented by a GM-APD connected in series with its quenching resistance, we expect that the SiPM single micro-cell signal alone has the same shape as described in the Section 3.2.2: a very fast (hundreds of ps) leading edge, with a time constant $\tau_{rise} = R_D \times (C_D + C_Q)$ and a slow exponential decay (tens to hundreds of ns), determined by the time constant $\tau_{fall\ slow} \sim R_Q \times (C_D + C_Q)$. The diode capacitance C_D is related to the geometry of the micro-cell and C_Q is the capacitance in parallel with R_Q .

The signal from a single micro-cell described above is applied to the load circuit composed by the other passive (not fired) micro-cells elements, a parasitic C_g (order of few pF) and R_{load} .

Fig. 4.7 (a) presents the equivalent circuit of the SiPM (Seifert et al. 2009), connected to a load resistance R_{load} (conversion $I \rightarrow V$), a voltage amplifier and read-out by an oscilloscope. In this figure, we can distinguish the equivalent circuit of the fired micro-cell (GM-APD diode in series with the R_Q and the parasitic C_Q as presented in the Section 3.2.2. a), the equivalent circuit of other passive micro-cells, and the parasitic grid capacitance C_g (due to the metal grid which spans over the entire surface of the detector).

A practical set-up of the circuit as used in our measurements is presented in Fig. 4.7 (b): the up figure shows the FBK and SensL SiPM structure and down figure shows the HPK SiPM structure; a positive V_{BIAS} is applied on the cathode of the diode in order to obtain a reverse bias voltage.

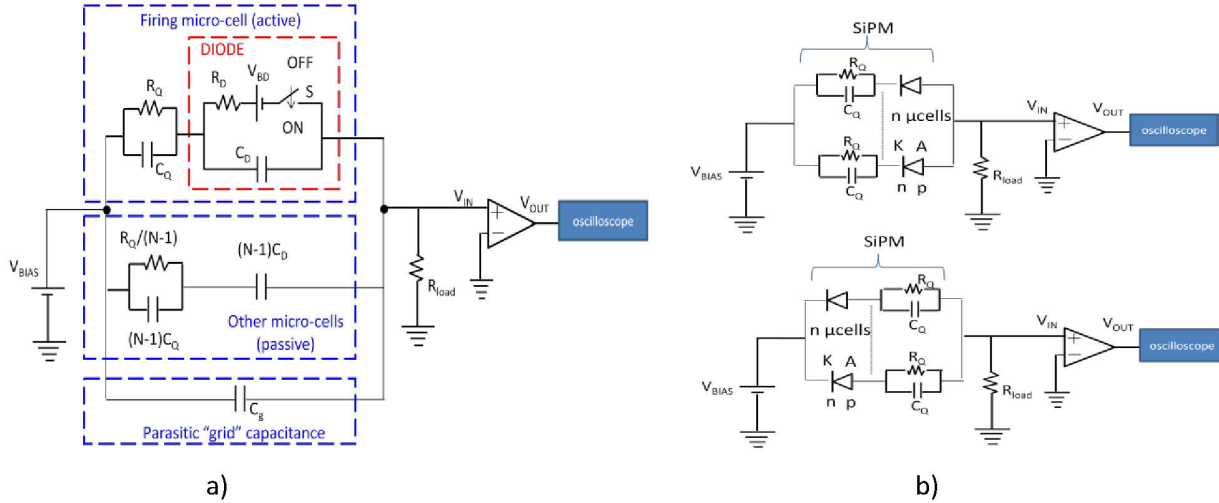


Fig. 4.7 (a) Equivalent circuit of a SiPM, showing the fired micro-cell, the passive micro-cells and the parasitic grid capacitance; **(b)** a practical circuit of the SiPM device (up figure – FBK and SensL SiPM structure, down figure – Hamamatsu HPK SiPM structure), connected to the voltage amplifier and the oscilloscope as used in our measurements (bandwidth of the amplifier and oscilloscope of 500 MHz)

The single micro-cell SiPM signals at the output of the amplifier for different devices are presented in Fig.4.8. Assuming that the time constant of the “load circuit” is much greater than the avalanche spread time, we can observe, as expected, that all devices present a fast rise time, characterized by a time constant $\tau_{rise} \sim 1.5\text{-}3\text{ ns}$ and a slow exponential decay, characterized by a time constant $\tau_{fall\ slow}$ going from $\sim 5\text{ns}$ (HPK device, $25\text{ }\mu\text{m}$ micro-cell size and $R_Q=200\text{ k}\Omega$) to 100ns (SensL device, $50\text{ }\mu\text{m}$ micro-cell size and $R_Q=340\text{ k}\Omega$). The total recovery time, defined as the time needed to recharge the diode from 0 to 99% of the bias voltage (5τ) is indicated in each plot of the Fig.4.8.

With respect to the other devices, the FBK device shows very well the additional fast component of the trailing edge as presented in the Section 3.2.2.a). This fast component is characterized by a time constant $\tau_{fall\ fast} \sim R_{load} \times (C_{tot} + C_g)$, assuming $R_{load} \ll R_Q/N$, where N is the total number of microcells. Typically R_{load} is of the order of 10 to 50Ω , $C_{tot} = C_D\ tot + C_Q\ tot \sim NC_D + NC_Q \sim N(C_D + C_Q)$, and $R_D \ll R_Q$.

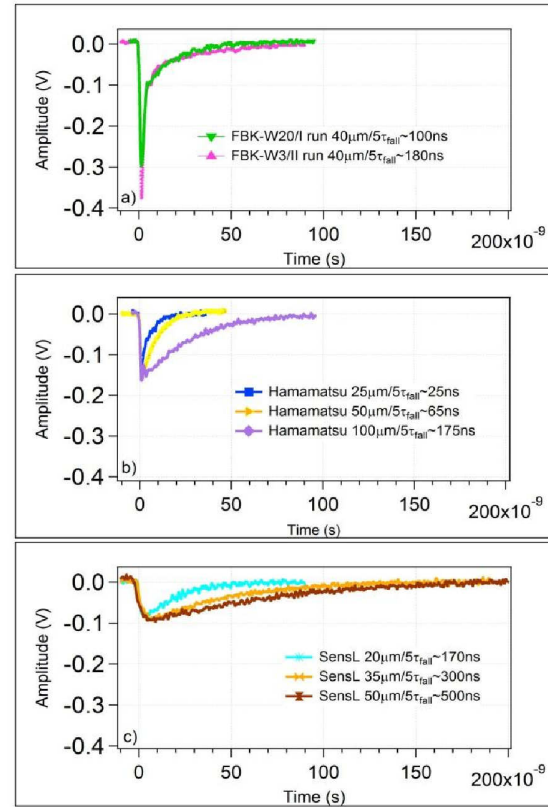


Fig. 4.8 Single micro-cell SiPM signals produced by **(a)** FBK, **(b)** HPK **(c)** SensL

Even if the fast component of the trailing edge cannot be distinguished in the shape of the signals from HPK and SensL devices, we expect that the parasitic capacitance C_Q is present in all measured devices (as shown in the electrical circuit), but its value is higher in the FBK devices with respect to the others: the devices present different metal and R_Q layouts as shown in Fig. 4.9 (a) and (b), determining different values of parasitic capacitance C_Q (see also Piazza et al. 2010; Yamamura et al., 2009). More details about the simulation of SiPM equivalent circuit and its corresponding signals can be found in the references: Corsi et al. 2006, Corsi et al. 2007, Seifert et al. 2009.

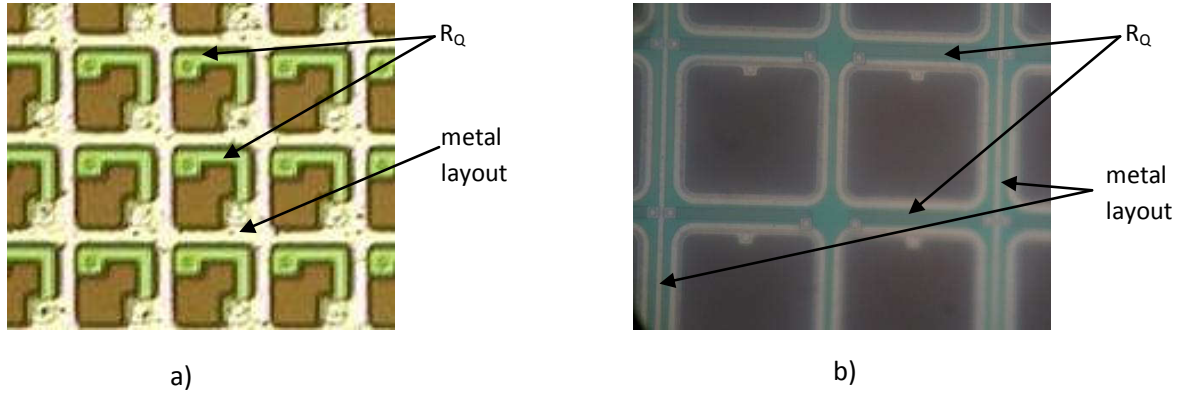


Fig. 4.9 Pictures of R_Q and metal layouts for SiPM devices from (a) FBK (micro-cell size $40 \times 40 \mu\text{m}^2$) and (b) HPK (micro-cell size $50 \times 50 \mu\text{m}^2$); the pictures have not the same scale

b. Signal gain

The gain G of the SiPM is defined as the number of carriers created during an avalanche discharge of a micro-cell diode. Therefore, it will be expressed by the same formula as the gain of a GM-APD:

$$G = \frac{Q}{q_e} = \frac{(C_D + C_Q)\Delta V}{q_e} = \frac{(C_D + C_Q)(V_{BIAS} - V_{BD})}{q_e} \quad 4.1$$

where q_e is the electron charge, C_D is the diode capacitance, C_Q is the parasitic capacitance in parallel to R_Q , V_{BIAS} is the applied bias voltage and the V_{BD} is the breakdown voltage.

To determine experimentally the SiPM gain, we integrate the SiPM output signals. If a fixed integration window is used ($5\tau_{fall}$, to record the full charge developed in a micro-cell), a second pulse (e.g. afterpulse) coming during this integration window is partially integrated. Therefore, a LabView acquisition program using a floating integration window has been developed. In this case, if a primary pulse is followed by a secondary one during its recharge (e.g. $5\tau_{fall}$), the integration window (which starts at the primary pulse) is increased with the time interval in between the pulses. Using this method, a complete spectrum of the charge distribution can be obtained (an example of such charge histogram for a SiPM detector produced by FBK is presented in Fig. 4.10). Such distribution has a dominant peak (fitted by a Gaussian function) corresponding to primary pulses, followed by a tail corresponding to primary pulses plus secondary effects (afterpulses or cross-talk). The centroid of the first Gaussian fit represents the most probable charge generated during one primary pulse and it allows determining the SiPM gain as this charge divided by the electron charge.

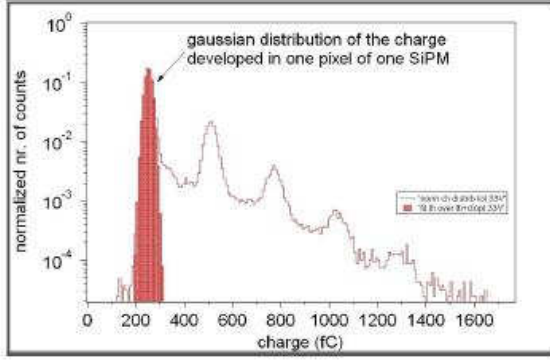


Fig. 4.10 Example of charge histogram allowing determining the gain of a SiPM

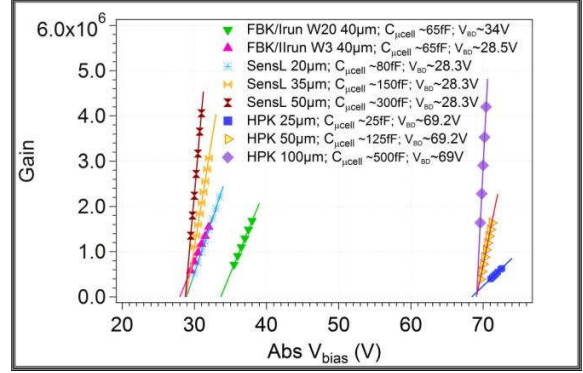


Fig. 4.11 Gain vs bias voltage for different SiPM's prototypes

The gain of different SiPM prototypes as a function of V_{BIAS} is shown in Fig. 4.11. As expected, a linear dependence of the G as a function of V_{BIAS} at a given V_{BD} has been found. The intercept of the linear fit with the x axis allows the calculation of the breakdown voltage V_{BD} . We can certainly recognize the values of $V_{BD} \sim 28-32$ V for SensL and FBK devices and the $V_{BD} \sim 69$ V for the HPK devices, as determined from the IV reverse characteristics.

Fig. 4.12 shows the gain G as a function of overvoltage ΔV for different measured prototypes: (a) FBK, (b) Hamamatsu HPK, (c) SensL devices. We can clearly observe the linear dependence of the G as a function of ΔV , but with different slopes, as a function of the micro-cell geometry (the slope increase with increasing the micro-cell area). The G ranges from 1×10^5 to 4×10^6 , once the $C_{micro-cell} = C_D + C_Q$, calculated from the slope of the linear fits, shows values ranging from 25 to 500 fF (values indicated in the Fig. 4.12).

The SiPM gain fluctuations are related to micro-cell to micro-cell uniformity over the device area (breakdown voltage as well as capacitances):

$$\frac{\partial G}{G} \approx \frac{\partial V_{BD}}{V_{BD}} \oplus \frac{\partial C_{D\&Q}}{C_{D\&Q}} \quad 4.2$$

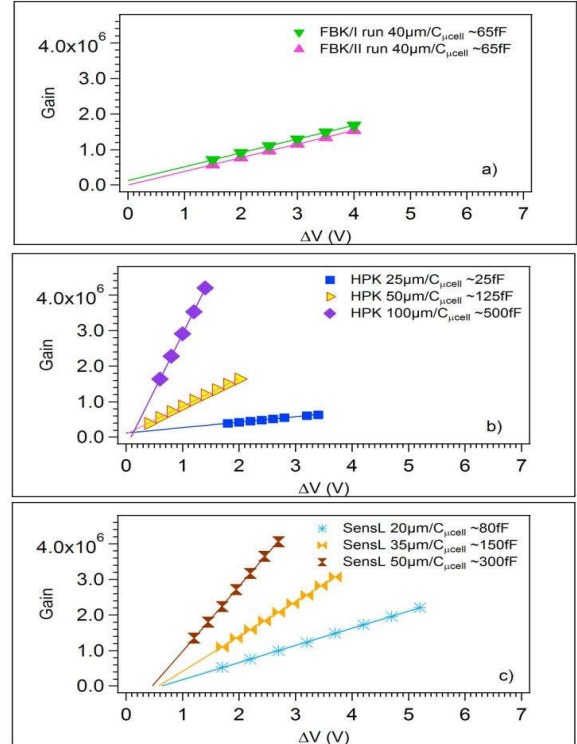


Fig. 4.12 Gain vs overvoltage for different SiPM's prototypes: (a) FBK, (b) Hamamatsu HPK, (c) SensL

c. Dark count rate (noise)

The dark count rate (DCR) of a SiPM represents the number of output pulses per second while the device is in the dark. This parameter is the intrinsic noise limit of the device, in particular when it has to be used for single photon detection.

As for the case of GM-APD, the main phenomena contributing to the dark pulses are: (1) thermal generated carriers, enhanced by trap-assisted tunneling (TAT) and (2) band-to-band tunneling. These phenomena have been described in detail in the Annex 1.

An oscilloscope image of the dark SiPM signals is presented in Fig. 4.13. We can observe, as in GM-APD case:

- signals coming from single micro-cells (i.e. primary uncorrelated pulses, determined by thermally generated carriers passing through the depleted region);
- primary signals followed by afterpulses (i.e. afterpulses are carriers captured by deep trap levels in the junction depletion layer and subsequently released, determining new pulses correlated to the primary ones).

Besides of the previously mentioned pulses, in the Fig. 4.13 we can distinguish a signal having double amplitude with respect to a signal coming from a single micro-cell. This kind of signal represents a second type of correlated pulses and it is related to the optical cross-talk phenomenon: a photon emitted in a primary avalanche can induce a second avalanche in a neighboring micro-cell, developed simultaneously as the primary one (i.e. photons are flying between micro-cells with the light speed). More details about the optical cross-talk phenomenon will be given in the following.

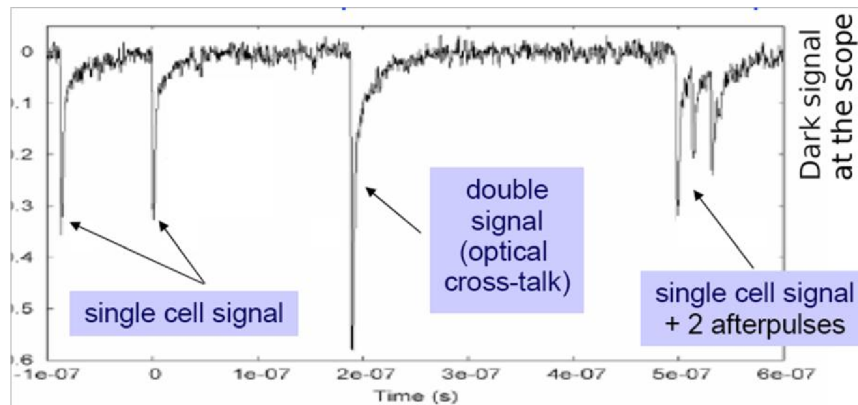


Fig. 4.13 Oscilloscope waveform of different types of SiPM dark signals (i.e. SiPM device produced by FBK)

Optical cross-talk

The optical cross-talk represents an additional contribution to the DCR of a SiPM detector and it is often observed when arrays of GM-APD devices are realized. When a SiPM micro-cell develops an avalanche determined by a photon or a thermally generated carrier, secondary photons are emitted by the micro-cell itself due to hot carriers relaxation (i.e. spontaneous direct relaxation of electrons between states of the conduction band). Schematic presentation of the optical cross-talk between two adjacent micro-cells is presented in Fig. 4.14.

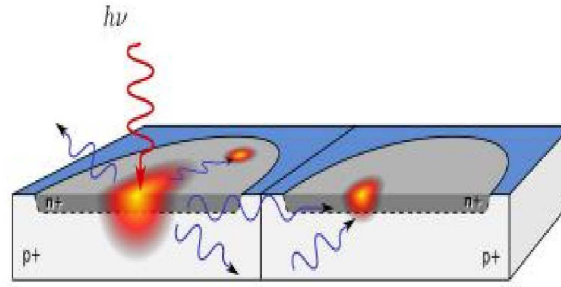


Fig. 4.14 Schematic representation of optical cross-talk between two adjacent micro-cells; when a primary avalanche is triggered in the first micro-cell (by a thermally generated carrier as well as by an incident photon), secondary photons are emitted by the micro-cell itself. These photons propagate through the bulk and they can be detected by an adjacent micro-cell

Lacaita et al. demonstrated that 2.9×10^5 photons can be emitted with energy higher than 1.14 eV per carrier crossing the junction (Lacaita, 1993 a). Therefore, for the case of SiPM where the number of carriers developed in one micro-cell avalanche is of the order of 10^6 (i.e. SiPM gain) there is a probability to have around 30 emitted photons which can determine optical cross-talk in the adjacent micro-cells.

The cross-talk intensity increases with reducing the distance between micro-cells; hence, this phenomenon sets a limit to the array density. Detailed analysis of this phenomenon has been reported (Haitz, 1965, Prochazka, 2005, Kindt, 1998) and all of them agreed that cross-talk depends only on direct optical paths between devices. For this reason some of these works (Prochazka, 2005, Kindt, 1998) proposed the introduction of trenches coated with metal between micro-cells in order to make cross-talk vanish (see Fig. 4.15), but experimental data showed just a decrease of its occurrence. More recent studies (Rech, 2008) showed that optical cross-talk depends strongly on indirect optical paths, i.e. on photons of near IR spectral components (between 1100 and 1200 nm) emitted by the micro-cell with an angle of incidence greater than the critical angle (since total internal reflection takes place) and reflecting off the bottom (substrate) of the chip (Fig. 4.16); these photons can by-pass trenches making them partially ineffective. Therefore, cross-talk of SiPM's can't be eliminated simply by means of trenches, but additional technological solutions are required to stop the contribution of bottom reflections.

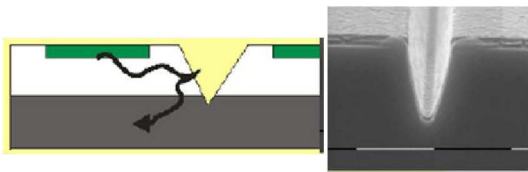


Fig. 4.15 Optical trenches coated with metal can decrease the magnitude of the cross-talk

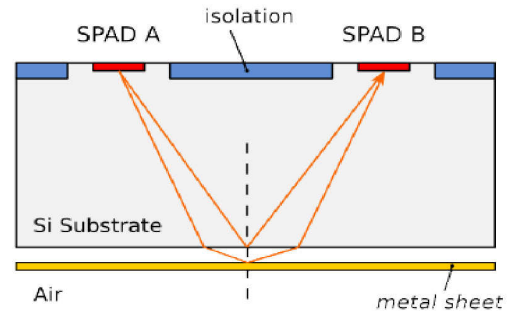


Fig. 4.16 Schematic representation of the indirect optical path contributing to the optical cross-talk: photons reflecting off the bottom (additional metal sheet to evidence the phenomenon (Rech, 2008)

Experimental results of SiPM DCR

We have measured the DCR for different SiPM prototypes by counting all dark pulses exceeding a threshold corresponding to 0.5 times one micro-cell pulse amplitude. The DCR as a function of overvoltage ΔV is presented in Fig. 4.17 (Dinu, 2009, a).

Fig. 4.17 (a), corresponding to FBK devices with the same micro-cell geometry but coming from two different production runs, shows that devices coming from the second run have improved DCR characteristics with respect of devices from the first production run: for a given overvoltage (i.e. 2V) the DCR has been reduced by about 4 times. This fact suggests that FBK succeed to improve the DCR parameter by modifying the technological conditions.

Fig. 4.17 (b) and (c), corresponding to HPK and respectively SensL devices with various micro-cell geometries, shows that DCR scales with the active surface: for a given technology and given overvoltage (i.e. 1V overvoltage) the DCR increases with increasing the micro-cell size.

Since the DCR should reflect the linear behavior of the thermal carriers triggering probability as a function of ΔV , we would expect a linear dependence of the DCR vs ΔV . We can see that SensL devices present a linear behavior of DCR vs ΔV , since FBK and HPK devices show a parabolic dependence of ΔV . We can expect that DCR of these last devices (FBK and HPK) is augmented by the contribution of afterpulsing and cross-talk.

As mentioned in the Section 4.2.1, when comparing devices with different V_{BD} , the value $\Delta V/V_{BD}$ matter and not only the overvoltage ΔV , because the performances are related to the excess electrical field above the breakdown level. Therefore, the DCR of different SiPM prototypes as a function of the ratio $\Delta V/V_{BD}$ has been represented in the Fig. 4.18. From this plot we can observe that DCR of HPK devices (p+/n micro-cells on n-type substrate technology) rise much faster as a function of $\Delta V/V_{BD}$ with respect of FBK and SensL devices (n+/p micro-cells on p-type substrate technology). This dependence is supposed to be related to different shapes of the electrical field profile corresponding to these devices (see Fig. 4.3).

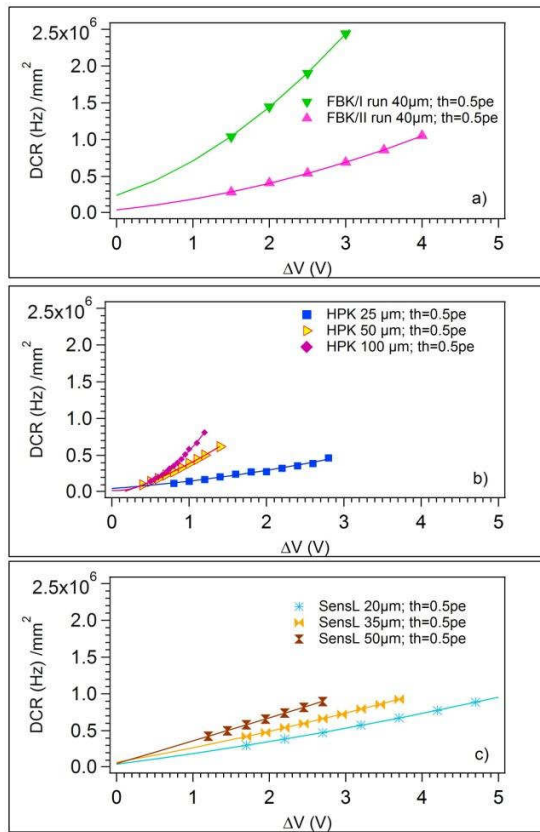


Fig. 4.17 DCR vs ΔV for different SiPM prototypes: (a) FBK devices; (b) Hamamatsu HPK devices; (c) SensL devices

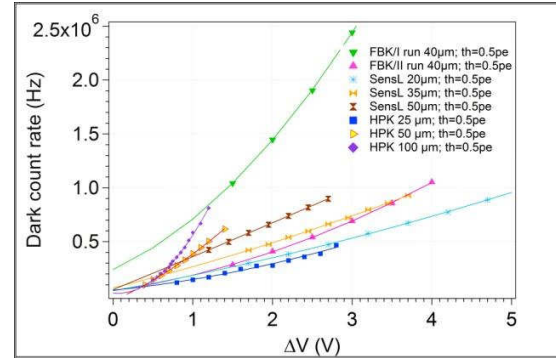


Fig. 4.18 DCR vs $\Delta V/V_{BD}$ for different SiPM prototypes

From the above plot we can estimate the optimal working range for different SiPM prototypes in terms of $\Delta V/V_{BD}$: ~2-5% for HPK devices and ~10-13% for FBK and SensL devices (the same ranges as found from IV reverse plots). If $\Delta V/V_{BD}$ exceeds the indicated values the DCR becomes too big (higher than ~1 MHz) and experimental difficulties can be found when devices are used for low light level detection and when no external trigger can be provided.

4.2.3. Photon detection efficiency

As for the case of a GM-APD, the photon detection efficiency (PDE) of a SiPM is defined as the ratio between the number of output pulses of the device exceeding its dark count rate and the number of photons impinging on the detector surface. Therefore, in order to be detected by a SiPM, a photon must be focalized in the active area of the detector, it should generate a primary carrier (more precisely an electron–hole pair) and the primary carrier should succeed in triggering an avalanche. Consequently, the PDE of a SiPM depends on three parameters: the quantum efficiency QE, the triggering probability P_{01} and the geometrical efficiency FF:

$$PDE = QE \cdot P_{01} \cdot FF \quad 4.3$$

A schematic representation of the three parameters influencing the PDE is presented in Fig. 4.19.

The quantum efficiency QE represents the probability for a photon to generate a carrier that reaches the high field region. It is the product of the transmission efficiency of the light through the dielectric layer covering the active area of the detector and the intrinsic quantum efficiency (the probability for a photon to generate an electron-hole pair). The QE is wavelength dependent.

The triggering probability P_{01} represents the probability for a carrier traversing the high-field region to generate an avalanche. This parameter is linked to impact ionization rates of electrons and holes (more details in the Section 3.2.3.) and therefore it is overvoltage dependent. It depends also on the wavelength, hence on the position of the generated carrier in the depth of the detector. For example, for the case of p+n junction on n-type substrate as presented in Fig. 4.3 (b), at short λ (high absorption and short penetration depth) only the electrons travelling from the p⁺ region to the n-type contact (positively biased) enter the junction and may trigger an avalanche, while at long λ only the holes drifting from the depletion region may do it.

The geometrical efficiency or fill factor FF represents the ratio of the sensitive area on the total detector surface. This dead area is usually due to structures between micro-cells (i.e. trenches, metal grid connections, quenching resistance etc.).

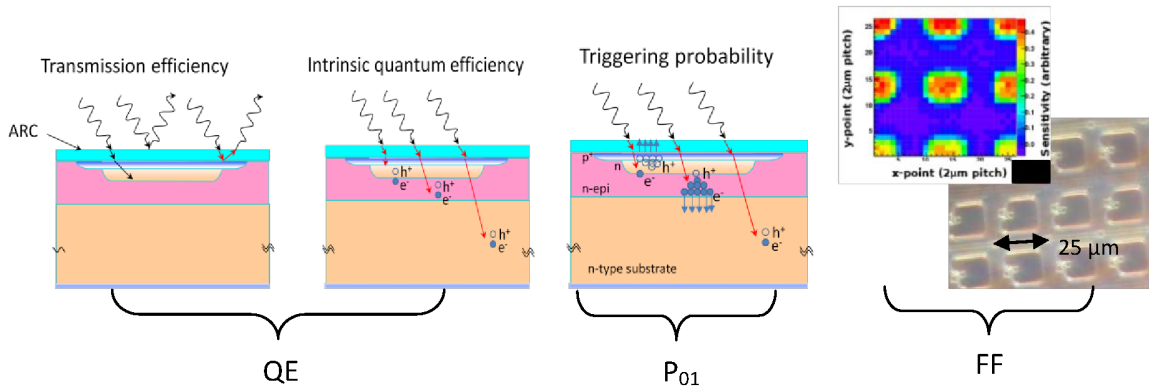


Fig. 4.19 Schematic presentation of the three parameters contributing to the PDE of a SiPM: quantum efficiency QE, triggering probability P_{01} and geometrical fill factor FF

From the experimental point of view, the PDE of a SiPM can be determined as:

$$PDE_{SiPM} = \frac{N_{photons\text{-}recorded\text{-}by\text{-}SiPM}}{N_{incident\text{-}photons\text{-}on\text{-}SiPM\text{-}surface}} \quad 4.4$$

To measure experimentally the PDE of different SiPM prototypes as a function of wavelength λ , we have built at LAL an optical set-up presented in Fig. 4.20. This set-up is composed of a halogen continuous light source (100W) followed by a grating monochromator, which separates the visible wavelengths (350–800 nm) with an accuracy of 2 nm. The number of incident photons on the detector surface has been determined as following:

$$N_{inc.ph. [s^{-1}mm^{-2}]} = \Phi_{[W/mm^2]} \cdot \frac{\lambda}{hc} \quad 4.5$$

where Φ represents the incident optical photon flux, λ is the incident wavelength, h the Plank constant and c the light speed. The incident optical photon flux Φ has been estimated by using two calibrated PIN photodiodes (HPK S3590-18, UDT Instrument 221) as following:

$$\Phi_{[W/mm^2]} = \frac{1}{A_{photodiode [mm^2]}} \cdot \frac{I_{photodiode [A]}}{R_{photodiode [A/W]}} \quad 4.6$$

where $A_{photodiode}$ is the calibrated photodiode area, $I_{photodiode}$ is the measured photodiode current at a given wavelength λ , and $R_{photodiode}$ is the photodiode responsivity (well known from fabricant calibrated curves). A good agreement has been found in between the incident optical photon flux Φ determined by using the two calibrated photodiodes. To avoid SiPM saturation, the incident optical flux was kept at low photon rate ($\sim 10^7$ /incident photons/s/mm²) by using neutral density filters. The precise replacement in front of the beam light in between the calibrated photodiodes and the SiPM has been assured by Polytec PI xyz translation stage (6 μ m accuracy on each direction). The optical tests have been performed at controlled room temperature (25 ± 1 °C).

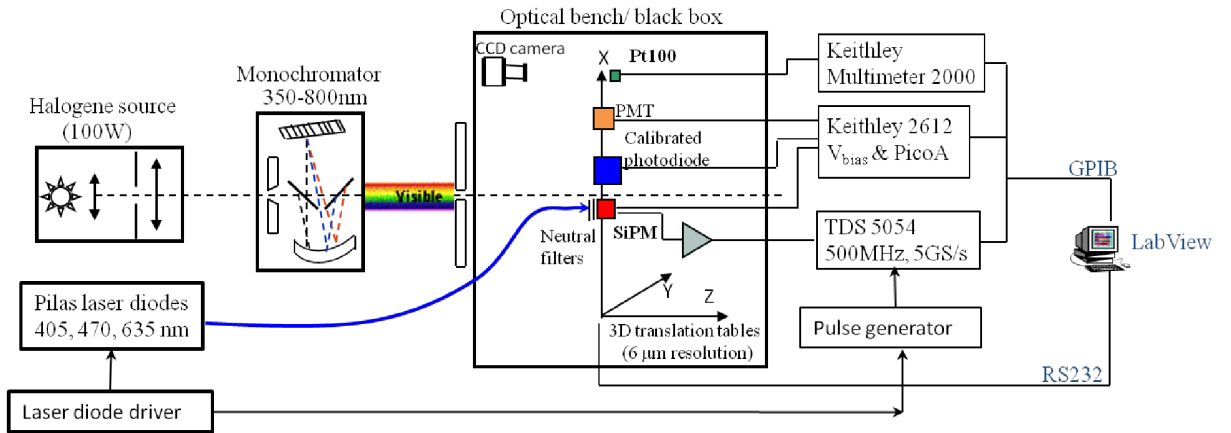


Fig. 4.20 Experimental set-up developed at LAL for the optical characterization of SiPM devices

The number of SiPM output pulses has been evaluated in two ways. In the first method we measured the DC current under illumination I_{light} . The difference between this value and the DC current in dark I_{dark} is proportional to the number of detected photons. The proportionality factor is the gain G of the SiPM which was previously determined (see Section 4.2.2.b). This is a very simple measurement but it requires the evaluation of the gain:

$$N_{\text{recorded-photon}}[s^{-1}mm^{-2}] = \frac{I_{\text{light}} - I_{\text{dark}}}{G_{\text{SiPM}} \cdot q} \quad 4.7$$

The gain G reported in the Section 4.2.2.b represents the charge developed only in one micro-cell and it does not take into consideration the charge carried by the afterpulses and cross-talk. Therefore, a corrected gain has been calculated using the average charge over the total spectrum of the charge distribution measured in the dark.

In order to confirm the validity of this method we performed a second test using the device as a photon counter, that is, we counted the pulses per second, with particular attention on the threshold value to eliminate the afterpulsing contribution (only pulses bigger than 0,85 of the primary pulses were counted). This number minus the dark count rate (at the same threshold) is the number of detected photons:

$$N_{\text{recorded-photon}}[s^{-1}mm^{-2}] = Rate_{\text{light}} - Rate_{\text{dark}} \quad 4.8$$

Such a method gives a direct measurement of the PDE but requires the presence of a more complicated read-out chain. We found a very good agreement between these two techniques, as presented in Fig. 4.21 (a) and (b) for the case of an FBK SiPM device, up to a level where the number of incident photons approaches the micro-cells number (i.e. non-linearity effects can be observed at high photon rate in the plot using the pulse counting method).

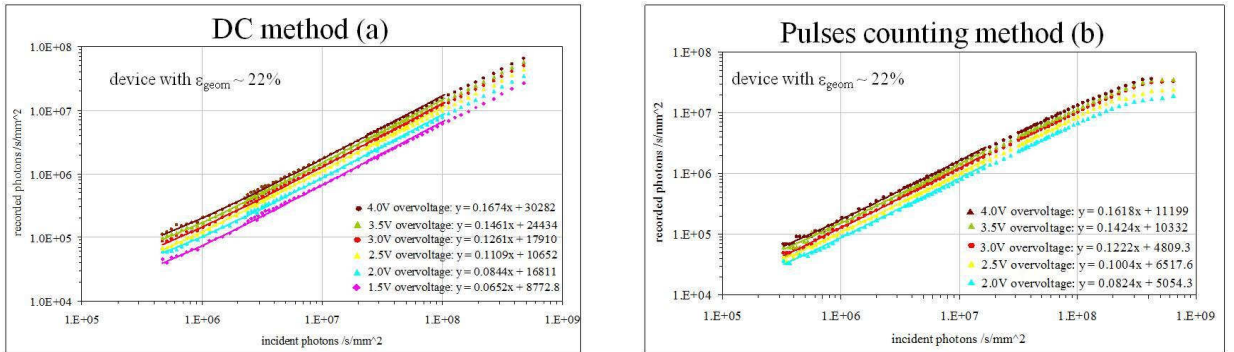


Fig. 4.21 Number of recorded photons as a function of the number of incident photons by using: **(a)** the DC method and **(b)** the pulses counting method, for a SiPM device fabricated at FBK; a nice agreement can be observed between two experimental methods at low photon rate

Since the two presented methods can still include errors determined by the after-pulses and cross-talk contributions to the calculation of the number of recorded photons, a third experimental method has been used by employing three pulsed Pilas laser diodes (405, 467 and 635 nm) driven at a repetition rate of 500 kHz (see Fig. xxx, exp set-up developed at LAL) (V. Chaumat, 2013). The spectral width of the laser pulses does not exceed 3 nm and, depending on the laser diode head, the pulse timing width was between 38 and 50 ps. The number of incident photons coming from the laser diodes has been determined by using a reference photomultiplier tube PMT (HAMAMATSU R7400U-01) with a gain and quantum efficiency known at $\pm 5\%$. The chosen PMT works in both continuous and pulsed mode allowing to measure the incident flux either from the mono-chromator or from the laser diodes. The error on the evaluation of the incident number of photons performed with this detector is therefore the same independent of the light source we have used.

The incident pulsed light coming from the Pilas laser diodes (405, 470 and 635 nm) is short enough (around 50 ps) to be considered as a Dirac pulse and as less than 10 photons/mm²/light pulse on average. Therefore, the number of recorded photons by SiPM can be assumed to follow the Poisson distribution:

$$N_{\text{recorded-photon}}[s^{-1}mm^{-2}] = -\ln(P(0)) \quad 4.9$$

where $P(0)$ is the probability of the non-converted events (pedestal) counted over the overall number of events. The number $N_{\text{recorded-photon}}$ calculated in this way is independent of afterpulsing and cross-talk. A good agreement has been found over the three experimental methods used for the PDE measurement, as will be shown in the following.

The PDE as a function of the wavelength λ (monochromator and pulsed lasers) for SiPM HPK with different geometries are presented in Fig. 4.22. We can observe that these devices present a peak of PDE in the blue region (420 nm), with a peak value increasing from 20% to 40% with increasing of the geometrical fill factor of the device. For comparison, the spectral sensitivity for the HPK devices of two different total areas: 1.3x1.3 mm² (T2K experiment) and 1x1 mm² (HPK catalogue) and 50x50µm² micro-cell size, published recently by Vacheret & al. (Vacheret, 2011) is presented in Fig. 4.23. We can observe a nice agreement between our measurements (i.e. the device of 50x50µm² micro-cell size) and those performed by the T2K group.

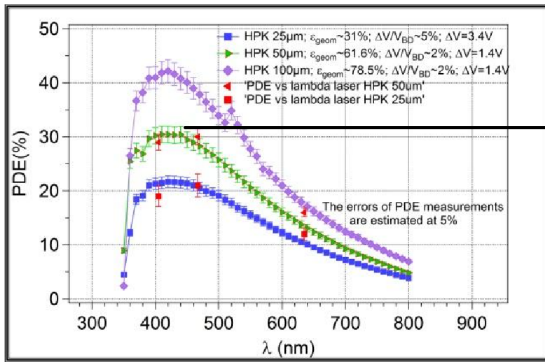


Fig. 4.22 PDE vs λ for HPK devices of 1x1mm² total area and different micro-cells geometries: 25x25µm², 50x50µm², 100x100µm² (Dinu et al., 2009, a; Chaumat et al., 2013)

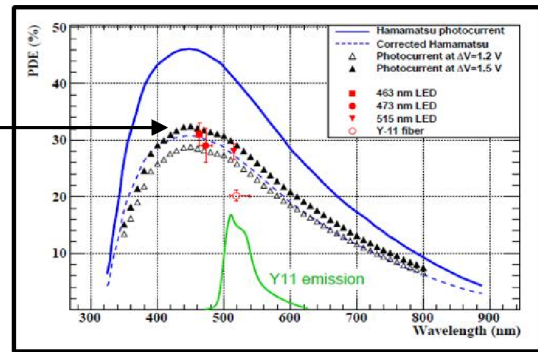


Fig. 4.23 PDE vs λ for HPK devices with total area of 1.3x1.3mm² (T2K experiment) and 1x1mm² (HPK catalogue) and micro-cell size of 50x50µm², as measured by the T2K group (Vacheret et al., 2011); the plot from the Hamamatsu catalogue (blue line) uses data not corrected for cross-talk and afterpulsing; the blue dashed line is the Hamamatsu plot scaled-down of 0.663 using the knowledge of the correlated noise contribution determined by T2K group

The PDE as a function of the wavelength λ for FBK and SensL SiPM prototypes are presented in Fig. 4.24. We can observe that the PDE peak is in the green region (500-600nm), with a maximum value of the order of 6-12%. The differences of the PDE values in between these devices and those from HPK could be explained by (1) higher FF of HPK devices with respect to the FF of the other devices (not yet optimized); and (2) higher $\epsilon_{\text{trigger}}$ of the HPK devices for blue incident light with respect to the FBK-irst and SensL devices: for the HPK devices (i.e. p+n junction on n-type substrate), the electrons trigger avalanches by traversing the high field region from the p⁺ region to the n-substrate contact, instead for the FBK and SensL devices (i.e. n+p junction on p-type substrate), the holes will trigger avalanches, since they are traversing the high field region by travelling from the n⁺ region to the p- substrate contact (see Fig. 4.25). Since the electrons have higher ionization coefficient than then holes, the $\epsilon_{\text{trigger}}$ for the blue light is higher for HPK devices with respect of FBK and SensL devices.

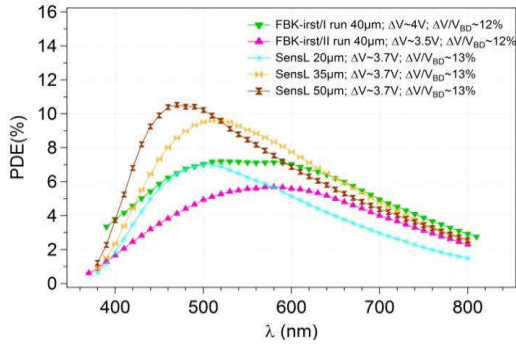


Fig. 4.24 PDE vs λ for FBK and SensL devices

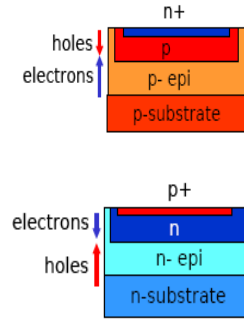
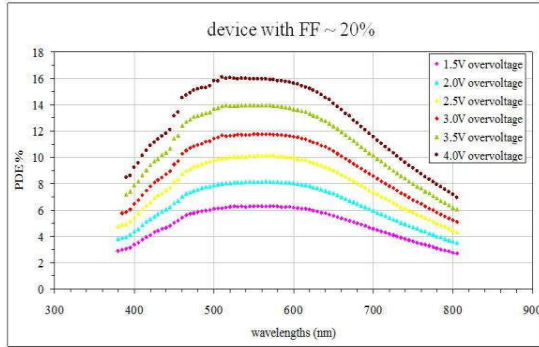
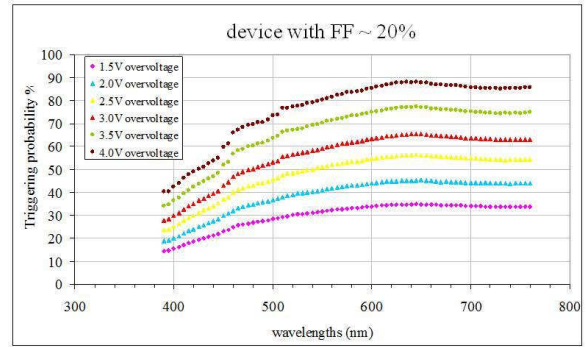


Fig. 4.25 Schematic representation of charge carriers travelling in two different types of SiPM junctions

In the Section 3.2.3. Fig. 3.31, we have presented the quantum efficiency of some diodes produced by FBK, extracted from the same wafers as the SiPM devices. In the Fig. 4.26 (a), the spectral sensitivity at different overvoltages for the FBK SiPM corresponding to the same production run is also presented. Knowing the total PDE of these devices, as well as the QE and the geometrical fill factor FF, we can extract the triggering probability $\epsilon_{\text{trigger}}$. The result of this calculation as a function of the wavelength, corresponding to FBK SiPM devices at different overvoltages is presented in Fig. 4.26 (b). The shape of the $\epsilon_{\text{trigger}}$ as a function of the λ can be explained as the following: at low wavelength of incident light ($\sim 400\text{-}450\text{nm}$), only h^+ cross the high electric field and trigger the avalanche, for medium wavelength of incident light ($500\text{-}600\text{ nm}$) both h^+ and e^- might trigger the avalanche (but cross only a fraction of high field region) and at high wavelengths of incident light (over 600 nm) only e^- cross the high field region and trigger the avalanche (i.e. the triggering probability attends its maximum). As expected, at a given wavelength, the $\epsilon_{\text{trigger}}$ increases with increasing the overvoltage.



a)



b)

Fig. 4.26 (a) PDE vs λ for an FBK device at different over-voltages (i.e. the same production run as the QE of the diodes presented in Fig. 3.31, Section 3.2.3.); **(b)** the calculated $\epsilon_{\text{trigger}}$ vs λ , at different over-voltages

4.2.4. Timing resolution

As for the case of a GM-APD device (Section 3.2.4), the SiPM timing resolution is defined as the time jitter (TJ) between the true arrival time of the photon at the sensor and the instant when the output current pulse is recorded.

The time resolution to single photons was studied on FBK SiPM devices by illuminating SiPM's with ultra-short laser pulses at fixed repetition rate and measuring the fluctuations of the difference in time between successive pulses (Collazuol et al., 2007). Measurements were carried out at the IPCF-CNR laboratory (Pisa) where a mode-locked Ti: sapphire laser, pumped by a continuous-wave solid-state green laser, provides red light ($\lambda = 800\text{nm}$) pulses $\sim 60\text{ fs}$ wide at a repetition rate of $\sim 80\text{ MHz}$, with jitter below 100 fs . Data were collected also at $\lambda = 400\text{nm}$ by exploiting a second harmonic generation crystal. After being filtered to the proper low intensity, the beam entered a black box, for rejecting environment light and electromagnetic noise, and the spot ($\sim 2\text{ cm}$ in diameter) impinged on the SiPM. The electronics to readout the SiPM current through a series resistor consisted in an AC coupled voltage amplifier (gain $\sim x50$) based on a low-noise and wide-band RF monolithic amplifier (gali-5, MiniCircuits) with low input impedance (50Ω). The amplifier output was sampled at 20 GSamples/s by a LeCroy SDA 6020 digital oscilloscope with 6GHz analog bandwidth (acknowledgments to E. Marcon, Lecroy). A schematic diagram of the set-up is presented in Fig. 4.27.

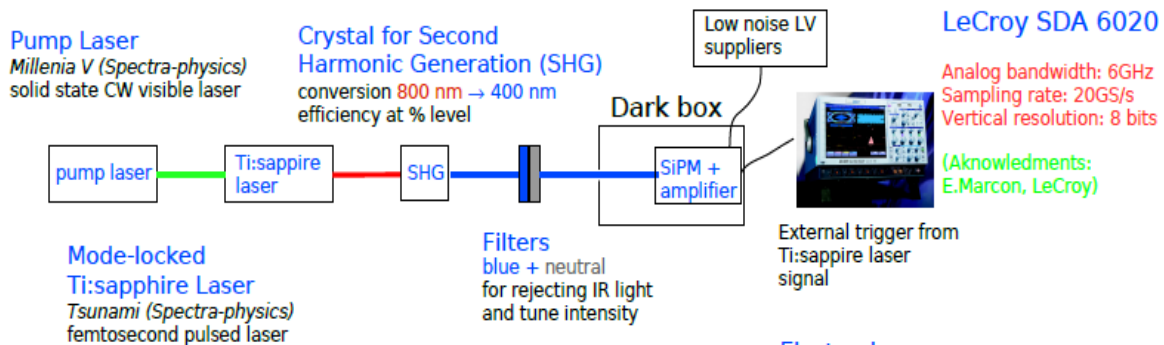


Fig. 4.27 Experimental set-up developed at IPCF-CNR laboratory (Pisa, Italy) for the SiPM timing resolution measurements

The SiPM signals have been sampled at high rate and the time of the pulses has been measured by waveform analysis as presented in the Fig. 4.28 and described below.

Different sets of data were collected for each device at $\lambda = 800$ and 400 nm , by varying the over-voltage (ΔV) and the light intensity conditions, but always keeping the rate on the single SiPM cell in the range $15\text{--}30\text{ kHz}$. The analysis of each set of data consists first in selecting single micro-cell signals called also single photo-electron signals (s.p.e. peaks) by requiring proper signal shape, pulse height and width. Typical rise time and FWHM width are ~ 1 and $\sim 3\text{ ns}$, respectively. Low instantaneous intensity and low noise are obtained by requiring absence of additional peaks and voltage fluctuations below 1 mV r.m.s. , respectively, within an interval of $\pm 50\text{ ns}$ around the selected s.p.e. peak.

To reconstruct the s.p.e. time with minimized effect from the electronic white noise a zero crossing filter (Wilmschurst, 1985) is applied as follows: first a “reference” s.p.e. signal $V_r(t - t_0)$ is built unaffected by the noise and depending on a reference time t_0 . Then $V_r(t - t_0)$ is slid along the time axis, by varying the reference time t_0 , until a best fit to the measured s.p.e. signal $V_s(t)$ is obtained. The condition of optimum timing is given by solving

in t_0 the following eq. $\int (\delta V_r(t - t_0)/\delta t) V_a(t) dt = 0$. Finally, the fluctuations of the SiPM response time are studied by taking the difference between successive s.p.e. peaks modulo the measured laser period ($T = 12.367$ ns).

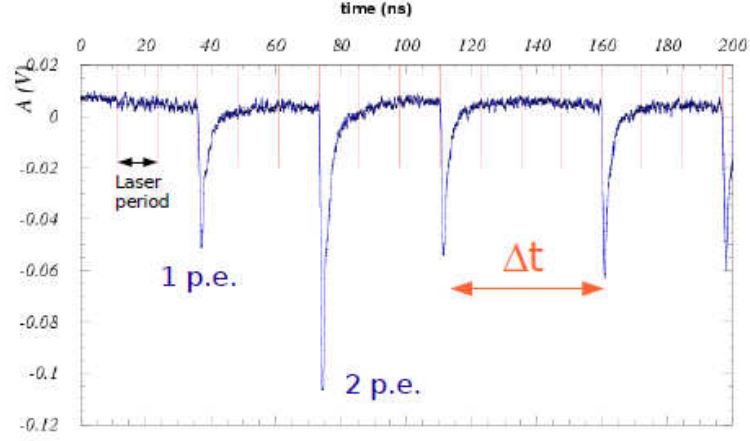


Fig. 4.28 Oscilloscope waveform of different signals provided by a SiPM illuminated with ultra-short laser pulses at fixed repetition rate; the time Δt between two consecutive single photon peaks has been used to determine the SiPM single photon timing resolution

In almost every set of measurements (Fig. 4.29 (a) and (b)) it is found that the distribution of time differences for data at $\lambda = 400$ nm fits (with good χ^2) to a gaussian plus a constant term, the latter being consistent with the contribution from the dark rate (% level). On the other hand data at $\lambda = 800$ nm need an additional exponential term, with typical time scale of 1 ns and integral contribution at the level of few 10%. Such slow component might be related to carriers generated by long wavelengths in the neutral regions beneath the depleted layer and reaching the latter by diffusion. The characteristic time is compatible with $\tau \sim L^2/\pi D$ (Cova et al., 2003) where L is the electron diffusion length and D is the electron diffusion coefficient in silicon. In the following we'll focus on the dominant contribution to the resolution, i.e. the gaussian part, and discuss the related sigma σ_t term extracted from the fit.

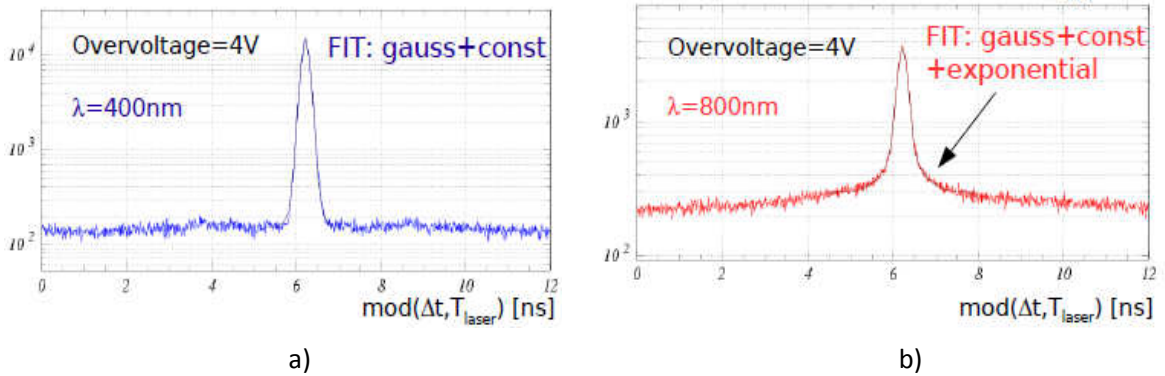


Fig. 4.29 Time distribution Δt between two consecutive single photon peaks at two wavelengths of incident light: (a) $\lambda = 400$ nm and (b) $\lambda = 800$ nm

Fig. 4.30 shows the resolution (σ_t) as a function of the overvoltage (ΔV) for one of the FBK devices. The resolution for the single photo-electron is at the level of 70 ps under standard working conditions ($\Delta V \sim 3$ -4V) and

reaches the level of 50 ps at $\Delta V \sim 6V$. The results for different measured FBK devices are in fair agreement. The shape of the resolution as a function of ΔV is in qualitative agreement with the results of timing resolution of GM-APD devices presented in the Section 3.2.4. (Spinelli, 1997). Statistical fluctuations in the avalanche development are due mainly to lateral propagation (Lacaita et al. 1990, Lacaita et al. 1993b). The worse resolution at long wavelengths (800 nm) may be explained by additional fluctuations on the longer drift time for carriers generated in depth by long wavelengths with respect to carriers generated near the high field region by short wavelengths.

The systematic uncertainties on the measurement are dominated by the contribution from the electronic noise which is directly measured as follows. Keeping the same laser and SiPM setup, after every standard measurement, data were collected with the electronics modified in order to evenly split the amplifier output signal, delay one component and recombine back the two components. As a result, for every s.p.e. peak from the SiPM two peaks at fixed delay were obtained at the amplifier output. The same optimum timing analysis described above is used for measuring the fixed delay and its fluctuation (green line in Fig. 4.30), which includes contribution of the electronic noise to the time resolution. This contribution is small compared (in quadrature) to the measured SiPM time resolution and is not corrected for. Only the uncertainty on the noise contribution is accounted in the measured SiPM time resolution. Residual systematic uncertainties (dependence on light intensity, sampling clock jitter, etc.) are estimated at a level below 10 ps.

Timing measurements with different beam spot size (circular with diameter in the range 10-200 μm) and illuminating different cells on the SiPM surface were performed by means of pinholes and careful positioning. No relevant spread in resolution was observed, indicating uniformity in rise-time for the avalanche signal among the different cells.

Data taking under higher light intensity conditions and special trigger were performed for investigating the timing resolution in presence of >1 simultaneous photo-electrons. The dependence of σ_t as a function of the number of photoelectrons is shown in Fig. 4.31 to be in fair agreement with Poisson statistics. The resolution with 15 photo-electrons, typical of applications where SiPM are coupled to small volume high yield scintillators, is better than 25 ps (important for medical imaging applications: Positron Emission Tomography).

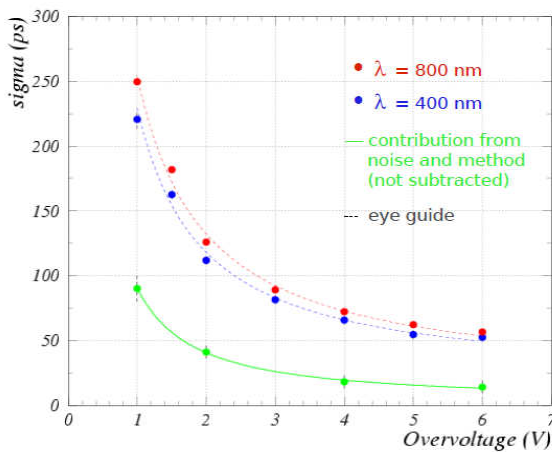


Fig. 4.30 Time resolution σ_t as a function of over-voltage at $\lambda=400$ nm (red circles) and at $\lambda=800$ nm (blue circles). Better resolution is reached at shorter wavelengths. The electronic noise contribution is directly measured (green circles) to be small compared to the device resolution. Lines are for eye guiding/

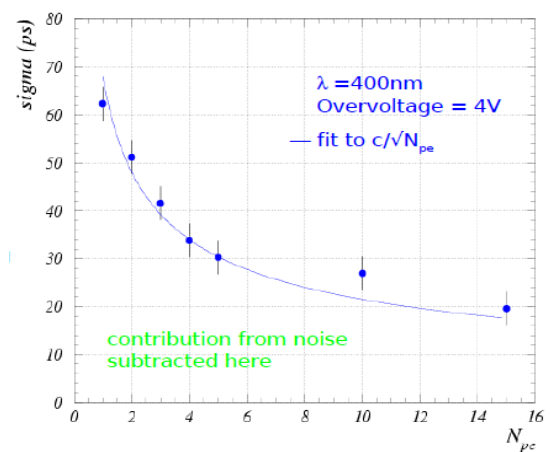


Fig. 4.31 Time resolution σ_t as a function of the number of simultaneous photoelectrons at $\lambda=400$ nm and over-voltage $\Delta V = 4V$ (circles). The best fit to the function $\sim 1/\sqrt{N_{p.e.}}$ (Poisson statistics) is also shown (line).

Single photon timing resolution of FBK devices, in typical working conditions has been shown to be at the level of <70 ps, with good cell-to-cell uniformity. The time resolution scales with the number of photo-electrons as expected from Poisson's statistics. Non gaussian tails in the resolution are found only for long wavelengths and not at $\lambda = 400$ nm.

4.2.5. Thermal effects – first part

Technical supervision: Cyril Bazin, Jean-François Vagnucci, (LAL)

Internal collaboration : Veronique Puill, Vincent Chaumat, C. Sylvia, C. Cheikali (LAL)

External collaboration: Adam Para (Fermilab, USA) – Memorandum of Understanding LAL/Fermilab

In the previous paragraphs of this Section we have learned about different characteristics of SiPM devices and how we can measure them experimentally. We could remark that each experimental set-up used for the measurement of SiPM parameters was equipped with controlled temperature climatic chambers, allowing the measurements in temperature controlled atmosphere. The main reason of this is that SiPM parameters are changing with temperature and therefore, the study of the SiPM behavior at different temperatures represents an important step in understanding of the physical phenomena staying behind of this temperature dependence.

Moreover, we have learned also that SiPM's are attractive alternative devices for many detectors that would use typically PMTs, in particular for applications at room temperature where multi-photon detection is involved. Actually, state of the art sensitivity in light detection can be reached by operating SiPM's at low temperature. There, a whole range of potential applications is opened, from particle detection with noble liquids, to Cherenkov light detection where the small photon yield, opposed to the huge dark noise of SiPM at room temperature makes cooling mandatory.

Therefore, following these two requirements, I have started two years ago a program of SiPMs analysis up to cryogenic temperatures and the results of this work will be presented in the following (Dinu et al. 2010).

For the characterization of SiPM devices as a function of temperature T , two experimental set-ups have been developed at LAL: the first one allowing temperature changes in the range from 0°C to 40°C by using a programmable climatic chamber (Fig. 4.32 (a)); a second one using a liquid nitrogen cryostat and covering a temperature range from -110°C to -50°C (Fig. 4.32 (b)).

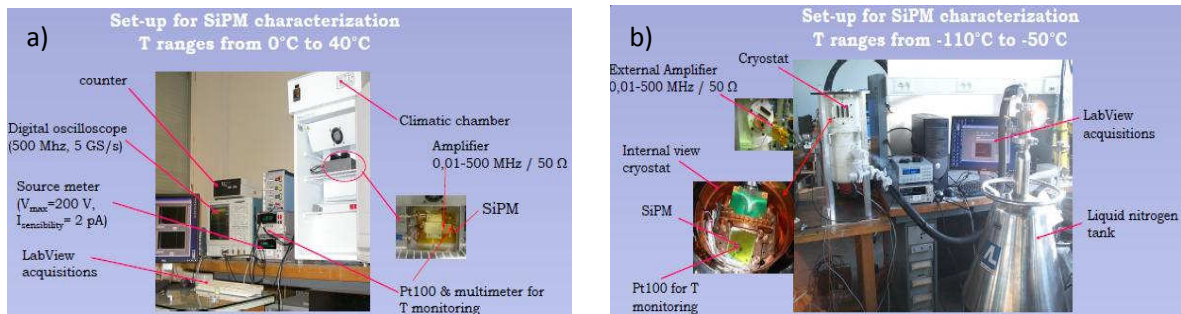


Fig. 4.32 Experimental set-ups developed at LAL for the thermal effects studies of SiPM devices

The T of the cryostat was controlled by a heater ($R \sim 20\Omega$) and stabilized by a cryogenic control system (Cryo.con model 22) for setting the heater current while the T of the climatic chamber was controlled by a ventilation system. In both set-ups, the T has been monitored by a Pt100 sensor mounted close to SiPM detector and read-out by a Keithley 2000 multimeter. To obtain reproducible results, the measurements have been carried out only when a good thermal equilibrium has been reached (e.g. $\pm 0.1^{\circ}\text{C}$ around assigned T).

The SiPM tests in the cryogenic set-up have been carried out under vacuum conditions at $P \sim 1.5 \times 10^{-3}$ mbar while the tests in the climatic chamber have been performed at atmospheric pressure. Equivalent data acquisition systems have been built to both set-ups:

- the IV static characteristics (e.g. R_Q) have been obtained by a direct connection of the SIPM detector to the Keithley 2611 source-meter;
- the AC characteristics (e.g. G , $C_{\text{micro-cell}}$, DCR) have been measured biasing the SIPM detector by a Keithley 2611 source-meter and reading-out by MITEQ wide bandwidth voltage amplifier (0.01 – 500 MHz) connected to a Tektronix digital oscilloscope (500 MHz, 5 GS/s). The amplifier presents $50\ \Omega$ input impedance, which acts as a current-to-voltage converter, followed by an amplifying stage having a gain of 45 dB. In particular for the DCR measurements at low T , when long counting time of 10 s was required to measure DCR of the order of Hz, the amplifier output signal has been connected to a counter.

To reduce the electromagnetic noise, the amplifier has been connected as close as possible to the detector. For the cryostat set-up, the amplifier has been connected outside of the cryostat to avoid the amplifier gain variation as a function of T . For the climatic chamber set-up, the amplifier gain variations have been evaluated to be less than 5% in the T range from 0° to 38°C and therefore the amplifier was mounted inside of the climatic chamber. For the automated data acquisition, the Keithley source meter and multimeter, the counter as well as the Tektronix oscilloscope have been run by LabView and C++ programs through GPIB and USB connections.

Based on the statement that primary signals generated by a thermally generated carrier or by an incident photon in a SiPM micro-cell are identical, all characteristics presented in the following have been obtained by measuring SiPM in dark conditions.

We analyzed two detectors provided by Hamamatsu HPK:

- MPPC S10362-11-050U-3, with a total area of $1 \times 1\ \text{mm}^2$ and a micro-cell size of $50 \times 50\ \mu\text{m}^2$;
- MPPC S10362-33-050U-329, with a total area of $3 \times 3\ \text{mm}^2$ and a micro-cell size of $50 \times 50\ \mu\text{m}^2$.

Therefore, a good uniformity as well as a very similar dependence of the measured parameters with respect to T is expected over analyzed detectors. Moreover, the only parameter which should present significant different values is the DCR since it is directly related to the total active area of the detector.

Experimental results

a) Gain versus bias voltage

The gain G of the SIPM detectors has been determined from time integration of the single micro-cell signal seen on the oscilloscope during an integration window adapted on the signal shape (e.g. to collect 99% of the charge). The results of these measurements as a function of V_{BIAS} at different T are presented in Figs. 4.33 (a) and (b). Independent on the measured detector, at a given T , the G increases linearly with V_{BIAS} as expected (e.g. $G \sim C_{\text{micro-cell}} \times (V_{\text{BIAS}} - V_{\text{BD}})$).

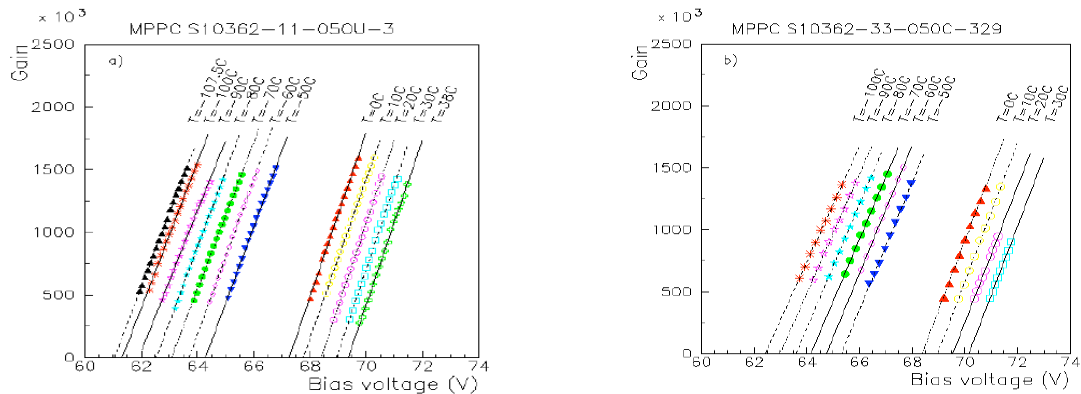


Fig. 4.33 Gain G vs V_{BIAS} for the SiPM of $1 \times 1\ \text{mm}^2$ (a) and $3 \times 3\ \text{mm}^2$ (b) at different temperatures

b) Breakdown voltage

The V_{BD} , determined from the intersection of the linear fits with abscise axis in the plot G vs V_{bias} , shows a linear increase when T increases from -110°C up to 38°C (Fig. 4.34), with a temperature coefficient of $\sim 58.5 \pm 0.5$ mV/ $^{\circ}\text{C}$ for both analyzed detectors. The increase in avalanche breakdown voltage with temperature results from a decrease in the ionization coefficients (see Annex 2) and it is to be understood as follows: as the temperature of the semiconductor increases, the optical phonon scattering of carriers in the depleted layer become more intense and the mean free path is reduced. Therefore, these carriers now lose more energy to the lattice and, to compensate for this energy loss, a higher voltage has to be applied across the junction to achieve the breakdown condition.

At a given T , a difference of $\sim 1.15\text{V}$ is observed between the V_{BD} of analyzed detectors. As presented in the Section 3.2.1 (equation 3.2) and from the plot of Crowell and Sze 1966 (see Annex 2), this difference is expected to be related to a difference in the background doping concentration N_b of analyzed devices.

c) Gain versus over-voltage

Independent on the measured detector, at a given V_{BIAS} , G decreases with T . Since this dependence is strongly related to the variations of the V_{BD} with T , uniform G values are expected if the overvoltage $\Delta V = V_{BIAS} - V_{BD}$ is maintained constant for different T . Figs. 4.35 (a) and (b) present the G dependence of ΔV . The G increases linearly with ΔV , with values ranging from 2.5×10^5 to 1.5×10^6 when ΔV varies from 0.5V to 2.5V . At a given ΔV , gain variations of $\sim 20\%$ are observed, which can be associated to some experimental systematic errors.

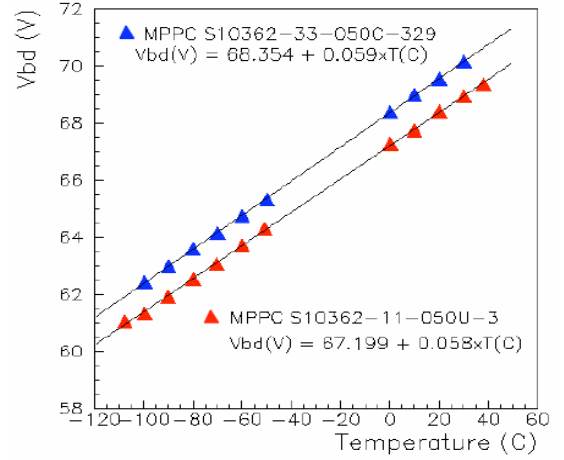


Fig. 4.34 Breakdown voltage V_{BD} versus T for the SiPM of 1×1 mm² and 3×3 mm² HPK detectors

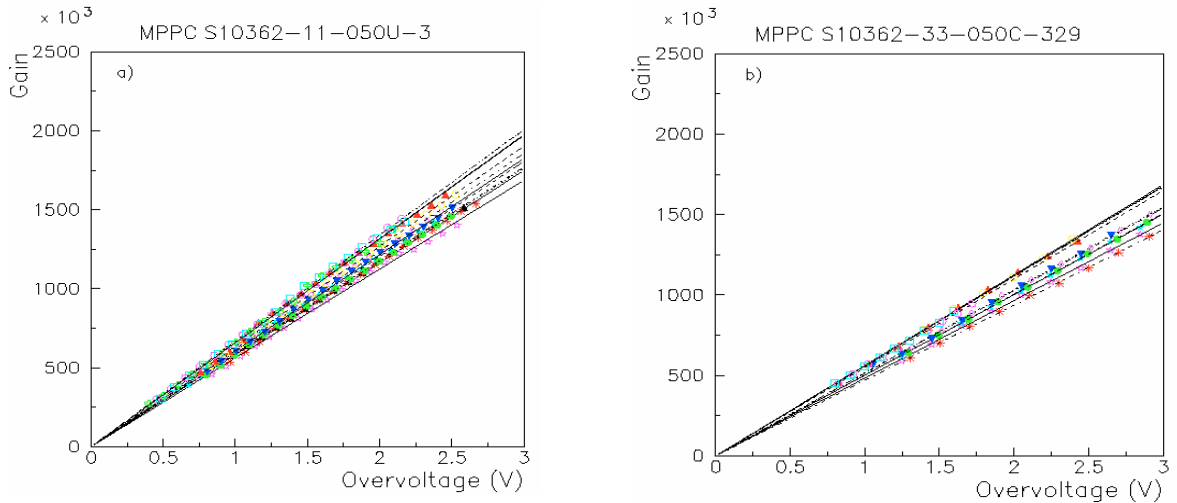


Fig. 4.35 Gain G vs overvoltage ΔV at different temperatures for the SiPM of (a) 1×1 mm² and (b) 3×3 mm²

d) Quenching resistance

Fig. 4.36 presents the dependence of quenching resistance R_Q as a function of T . The R_Q exhibits a quadratic variation, with decreasing values when T increases (negative temperature coefficient). The SiPM detector with an area of $1 \times 1 \text{ mm}^2$ shows R_Q values ranging from $\sim 240 \text{ k}\Omega$ to $\sim 90 \text{ k}\Omega$ when T varies from -100°C to 38°C and the detector with an area of $3 \times 3 \text{ mm}^2$ shows R_Q values from $\sim 460 \text{ k}\Omega$ to $\sim 150 \text{ k}\Omega$ for the same T range. For devices with the same micro-cell size we expected to found the same values of the R_Q , as can be observed in the Figs. 4.37 and 4.38, where the geometrical dimensions of the poly-silicon resistance (length L and width w) are identical for both types of measured detectors.

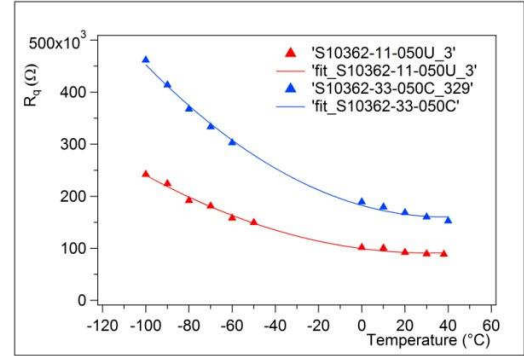


Fig. 4.36 Quenching resistance R_Q versus T for two HPK devices

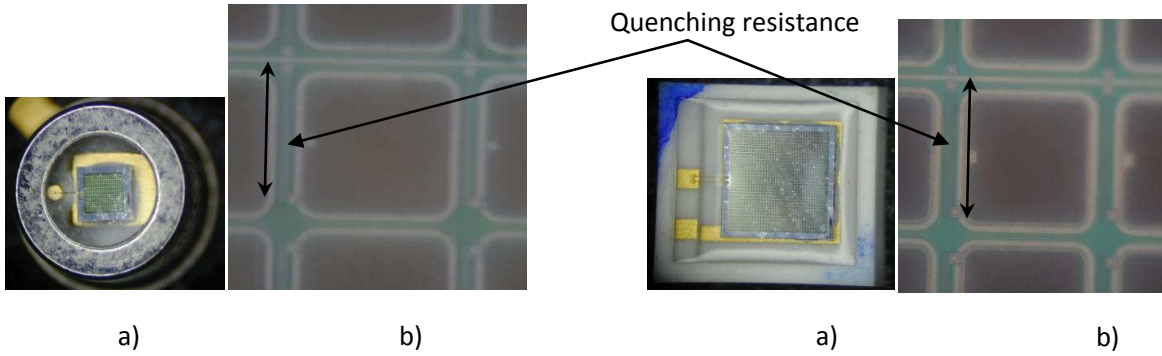


Fig. 4.37 Pictures of SiPM of $1 \times 1 \text{ mm}^2$ total area and $50 \times 50 \mu\text{m}$ micro-cell size: (a) overall view; (b) zoom over one micro-cell; the R_Q can be identified

Fig. 4.38 Pictures of SiPM of $3 \times 3 \text{ mm}^2$ total area and $50 \times 50 \mu\text{m}$ micro-cell size: (a) overall view; (b) zoom over one micro-cell; the R_Q can be identified

The higher R_Q values of $3 \times 3 \text{ mm}^2$ detector with respect to the $1 \times 1 \text{ mm}^2$ one are probably related to changes in the fabrication technological parameters of the detectors (i.e. doping concentration of poly-silicon). Note: for these devices R_Q is not on the active area but in between the micro-cells.

The variations of R_Q with temperature leads to significant variation of the micro-cells slow trailing edge characterized by the time constant $\tau_{\text{fall slow}} \sim R_Q \times (C_D + C_Q)$ (see Section 4.2.2). Figs. 4.39 (a) and (b) show normalized single micro-cell signal shapes for different temperatures. We can observe that, the total recovery time calculated as $5 \cdot \tau_{\text{fall slow}}$ for 99% recovery, decreases from 120 to 50 ns for SiPM's of $1 \times 1 \text{ mm}^2$ and from 300 to 160 ns for T increasing from -100°C to 38°C for the $3 \times 3 \text{ mm}^2$ area, because of the R_Q decreasing with T in the same range.

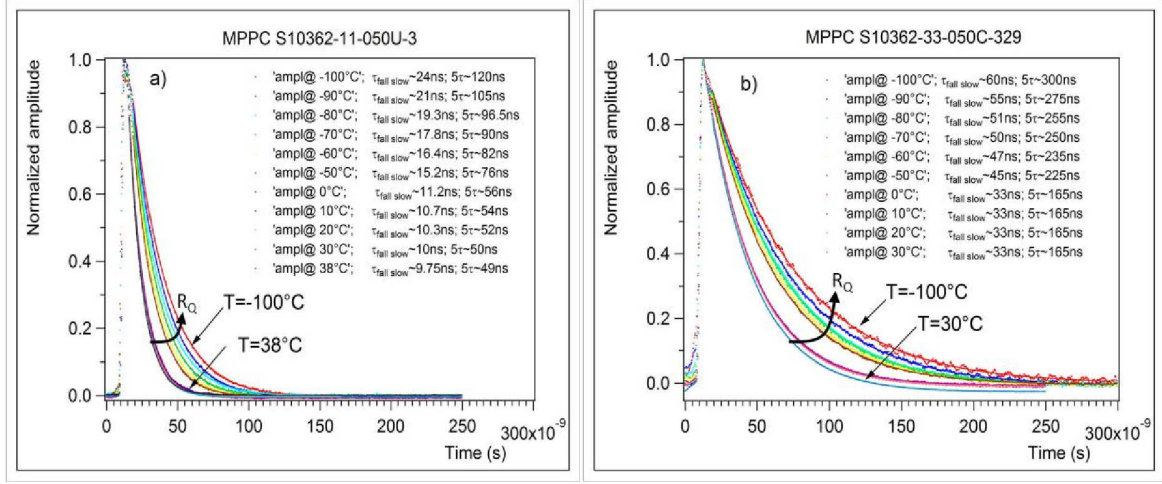


Fig. 4.39 Normalized signal shape at different temperatures for the SiPM of (a) 1x1 mm² and (b) 3x3 mm²

e) Dark count rate

The DCR versus ΔV has been also measured for $-100^\circ\text{C} < T < 38^\circ\text{C}$ (Figs. 4.40 (a) and (b)). At a given T , an increase of the DCR has been observed for ΔV increasing up to 2.5V. At a given ΔV , the DCR increases with T over many orders of magnitude.

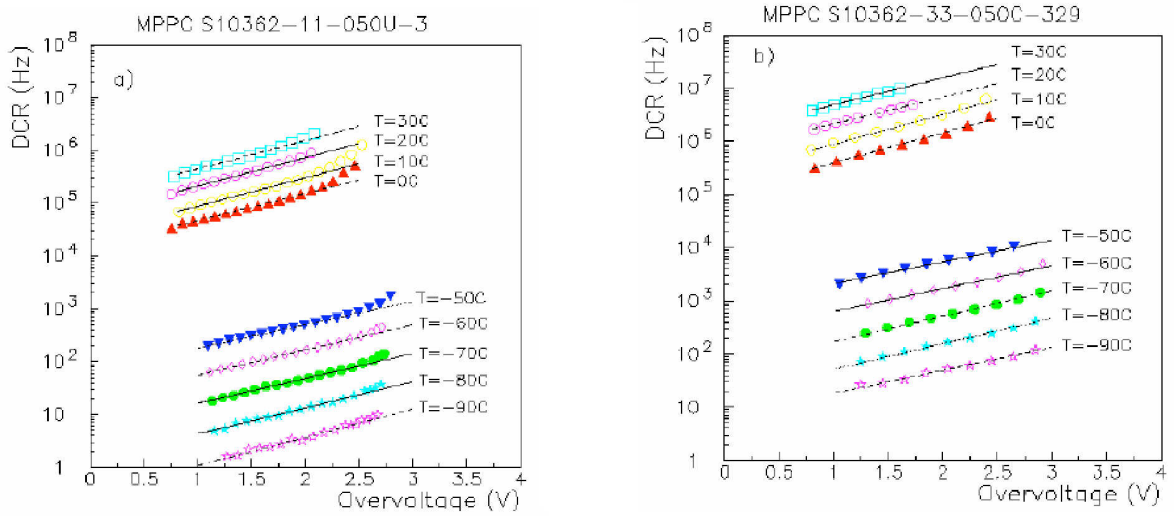


Fig. 4.40 Dark count rate DCR versus ΔV for different T for the SiPMs of (a) 1x1 mm² and (b) 3x3 mm² detector

As mentioned In the Annex 2, a temperature variation of the generation current $I_{\text{generation}}$ in the depleted region is expected to determine a temperature variation of the DCR. The temperature variation of $I_{\text{generation}}$ is dominated by the temperature variation of intrinsic carrier concentration (Annex 2), with a temperature coefficient given by:

$$\frac{1}{I_{\text{generation}}} \frac{dI_{\text{generation}}}{dT} = \frac{1}{2T} \left[3 + \frac{E_{g0}}{k_B T} \right] \quad 4.10$$

where T is the temperature, E_{g0} is the band gap energy at $T=0K$ (i.e. 1.166V), k_B is the Boltzmann constant. When generation is dominating, the plot $\log I_{generation}$ vs. $1/k_B T$ will yield a slope of $\sim E_{g0}/2$. Since the DCR is directly related to the $I_{generation}$, we expect a DCR dependence with temperature following a similar law as the $I_{generation}$.

Plotting the experimental values of the DCR (log scale) vs. $1/T$, in the temperature range $-60^\circ C$ to $38^\circ C$ the data can be well fitted with a form $A \cdot T^{3/2} \exp(-E/k_B T)$ determining a slope of ~ 0.54 eV (Fig. 4.41). This value estimates very well the expected value of $\sim E_{g0}/2$, demonstrating clearly that carrier generation is the dominating mechanisms in such devices, for temperatures ranging from $40^\circ C$ down to $-60^\circ C$. A deviation of the experimental points from the fitted line has been observed for T lower than $-60^\circ C$, for both investigated detectors and different over-voltages. This is probably an indication of a different mechanism of generation of free carriers in the conduction band with much weaker dependence on the temperature. Similar results were recently reported in Collazuol et al., 2011.

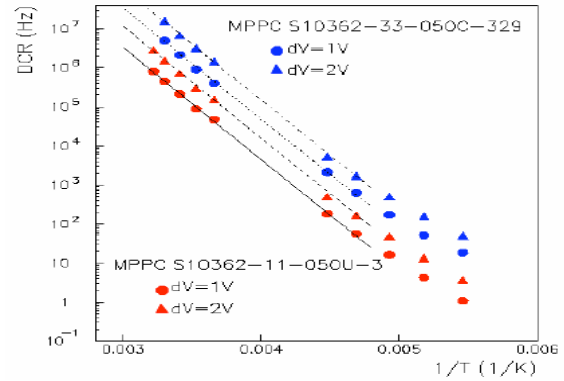


Fig. 4.41 Dark count rate DCR versus $1/T$ for two different over-voltages and SiPMs of $1 \times 1 \text{ mm}^2$ and $3 \times 3 \text{ mm}^2$

4.2.6. Thermal effects – second part

M1 & M2 pre-thesis internships supervision: Andrii Nagai (student from Kiev University, Physics Department)

External collaboration: Adam Para and technical staff at Fermilab, USA – MoU LAL/Fermilab

In the first part of this Section, we described the temperature variations of different SiPM parameters obtained with two different experimental set-ups developed at LAL and presented in Figs. 4.32 (a) and (b). The temperature range covered by these set-ups was ranging from $-100^\circ C$ to $-50^\circ C$ by using the cryogenic system and from $0^\circ C$ to $40^\circ C$ by using the climatic chamber. In order to fulfill the missing gap of temperature as well as to increase our studies in a larger temperature range, a new experimental set-up has been developed at Fermilab. A general view of this set-up is presented in Fig. 4.42 (a). A detailed zoom inside of the box (i.e. test-cube) where the SiPM's were located is presented in Fig. 4.42 (b).

The Fermilab system is equipped with a fully automatic system of nitrogen fulfilling and temperature control. The liquid nitrogen fill is interlocked to prevent fill until the vacuum, assured by a turbo vacuum pump, is below 2×10^{-4} Torr. An automatic heater controller via a PID control loop allowed covering a full temperature range from $-175^\circ C$ to $55^\circ C$. A copper cold finger, having one extremity in contact with the liquid nitrogen and the other one in contact with the SiPM mechanical support inside of the test-cube, has been used to assure the cold temperature transmission. A heater (based on a resistance) has been used to obtain higher temperatures with respect to liquid nitrogen.

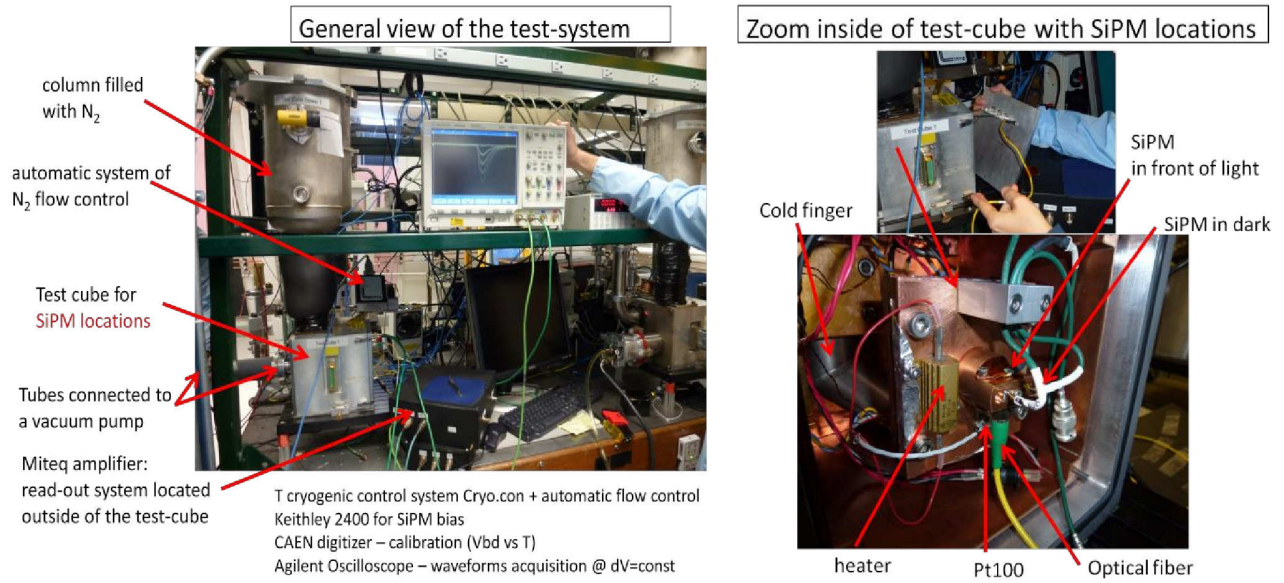


Fig. 4.42 Experimental set-up developed at Fermilab for the characterization of SiPM devices in a temperature range from -175°C to $+55^\circ\text{C}$

The mechanical system inside the test-cube have been optimized allowing to test three simultaneously SiPM detectors at a given temperature (see zoom in the Fig. 4.42). Two of the detectors are in the dark and the third one can be illuminated with light from a pulsed laser, brought inside of the test-cube and on the active surface of the detector by an optical fiber. The detectors in the dark are measured in DC conditions and the third one, light illuminated, is measured in AC conditions, using a readout system composed of a MITEQ voltage amplifier (500 MHz bandwidth and 60 db gain) and an Agilent oscilloscope. The amplifier is mounted outside of the test-cube, avoiding temperature variations of the amplifier characteristics. To reduce the electromagnetic noise on the SiPM signal shape, the oscilloscope bandwidth has been reduced at 200 MHz.

Two devices have been tested using the set-up from Fermilab:

- MPPC S10362-11-050U-3, with a total area of $1 \times 1 \text{ mm}^2$ and a micro-cell size of $50 \times 50 \mu\text{m}^2$ ($N_{\mu\text{cells}}=400$); this one has been already studied using the LAL set-up and it serves as reference for comparison calibration between LAL and Fermilab set-ups; we have to note that this device is coming from a production run of 2007;
- MPPC S10931-33-050P-63813, with a total area of $3 \times 3 \text{ mm}^2$ and a micro-cell size of $50 \times 50 \mu\text{m}^2$ ($N_{\mu\text{cells}}=3600$). This device is more recent (production run 2011) with respect to the $1 \times 1 \text{ mm}^2$ tested at LAL and Fermilab as well as to MPPC S10362-33-050U-329 of $3 \times 3 \text{ mm}^2$ device tested at LAL (production run 2007).

Besides of extension of the temperature range, the measurements performed at Fermilab allow us to compare the characteristics of devices coming from different production runs, performed at a time interval of 4 years, and to analyze the possible improvements or changes of device characteristics implemented by Hamamatsu.

The SiPM measurements have been performed in the temperature range from -175°C to $+55^\circ\text{C}$, in a step of 10°C . At each temperature T , twelve's bias voltages V_{bias} have been applied, in a step of 0.2V . The V_{bias} values have been chose in order to keep the same overvoltage with respect to the breakdown voltage V_{BD} , independent of T .

For the analysis of the huge amount of experimental data related to these measurements, an automatic procedure has been developed by A. Nagai during his M1&M2 internships. Details on the developed procedure can be found in his internship report (Nagai, 2013). In the following, only the physics results will be reported.

Experimental results

a) Gain versus bias voltage

The gain G as a function of V_{bias} for both tested detectors is presented in Figs. 4.43 (a) and (b). We can observe that for a given T , the G increases linearly with V_{BIAS} as expected (e.g. $G \sim C_{micro-cell} \times (V_{BIAS} - V_{BD})$) and this dependence is confirmed for both detectors and for the overall temperature range from -175°C to $+55^{\circ}\text{C}$.

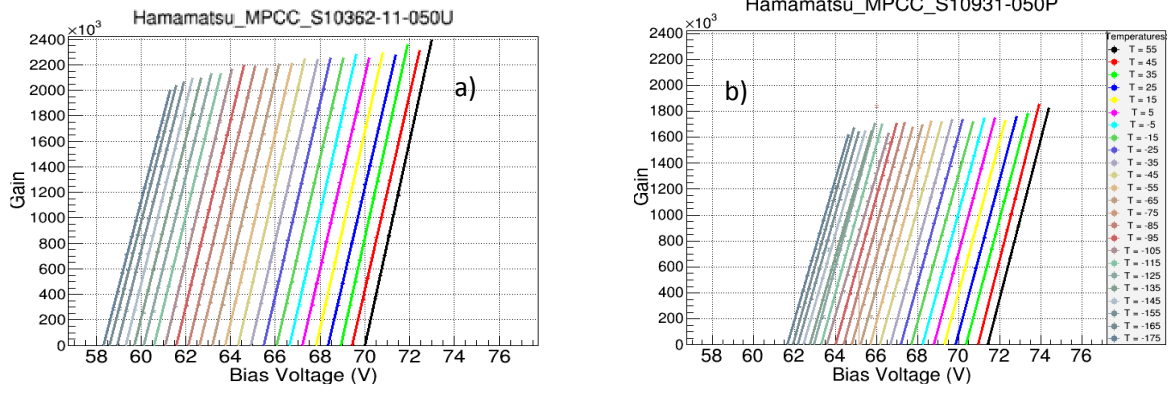


Fig. 4.43 Gain versus V_{bias} in the temperature range from -175°C to $+55^{\circ}\text{C}$ for (a) $1 \times 1 \text{ mm}^2$ and (b) $3 \times 3 \text{ mm}^2$ detector

b) Breakdown voltage

The V_{BD} versus T for both detectors is presented in Fig. 4.44. We can observe that V_{BD} decreases linearly with decreasing T up to $\sim -100^{\circ}\text{C}$ for the $1 \times 1 \text{ mm}^2$ detector (as already shown in the previous Section, Fig 4.34) and up to $\sim -50^{\circ}\text{C}$ for the $3 \times 3 \text{ mm}^2$ detector. Below these values of T , both detectors presents slower and non linear temperature dependence, showing quite different temperature coefficient. The normalized V_{BD} with respect to room temperature $V_{BD}(300\text{K})$ is presented in Fig. 4.45.

We can expect that V_{BD} of SiPM devices coming from different production runs can show certain differences, because of both process and wafer properties statistical variations. Therefore, the observed difference between the V_{BD} values of analyzed detectors at a given T is not very surprising. However, their different temperature dependence, in particular at low temperatures, could leave us imaging that the analyzed detectors have different structural or technological characteristics (Serra et al., 2011). More details on the physical phenomena involved in the temperature variation of V_{BD} as well as the influence of the technological parameters are given in the Annex 2.

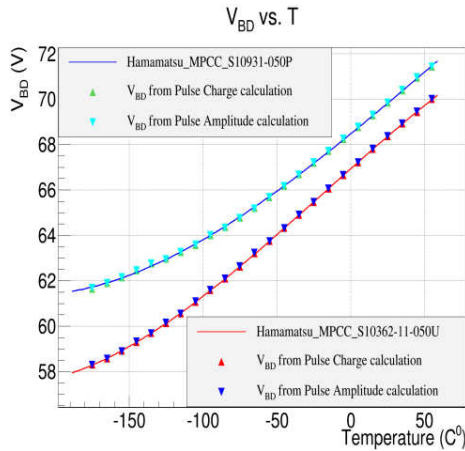


Fig. 4.44 Breakdown voltage versus T for $1 \times 1 \text{ mm}^2$ and $3 \times 3 \text{ mm}^2$ Hamamatsu detectors

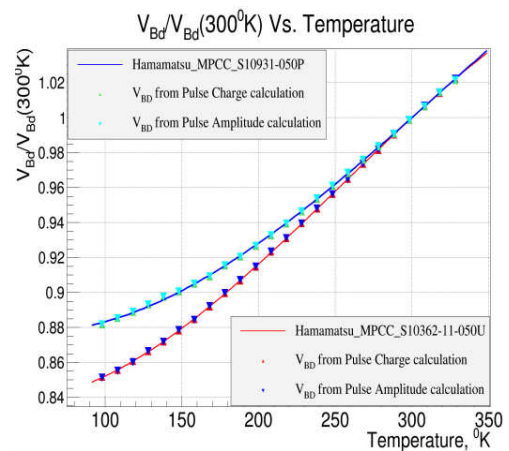


Fig. 4.45 Normalized breakdown voltage $V_{BD}/V_{BD}(300\text{K})$ versus T for $1 \times 1 \text{ mm}^2$ and $3 \times 3 \text{ mm}^2$ Hamamatsu detectors

c) Gain versus over-voltage

Figs. 4.46 present the gain G as a function of overvoltage ΔV for (a) $1 \times 1 \text{ mm}^2$ and (b) $3 \times 3 \text{ mm}^2$ detectors. At a given ΔV we can observe that the gain is almost constant, presenting a variation of only $\sim 8\%$ for $1 \times 1 \text{ mm}^2$ detector and $\sim 5\%$ for the second detector. These variations are very small if we keep into account that the measurements are performed over a very wide temperature range of 230°C (from -175°C to $+55^\circ\text{C}$).

The micro-cell capacitance has been determined from the slope of the linear fit. A value of $\sim 125 \text{ fF}$ has been found for $1 \times 1 \text{ mm}^2$ detector and $\sim 90 \text{ fF}$ for the $3 \times 3 \text{ mm}^2$ one. As mentioned previously, the micro-cell capacitance (see Section 3.2.2. a. or 4.2.2.b for more details) is given by the sum of the diode capacitance C_D and parasitic quencing capacitance C_Q . The micro-cell diode capacitance is related to its geometry: $C_D = \epsilon_{Si} \epsilon_0 A/d$, where $\epsilon_0 = 8.854 \times 10^{-14} \text{ F/cm}$ is the permittivity of the free space, $\epsilon_{Si} = 11.9$ is the silicon dielectric constant, A is the active area of the micro-cell and d is the depletion thickness. Geometrical measurements performed on the detectors surface showed that the overall micro-cell area (i.e. pitch) is $50 \mu\text{m} \times 50 \mu\text{m}$ for both detectors, as indicated by the Hamamatsu data sheet. However, the $3 \times 3 \text{ mm}^2$ detector shows an active window (black window in Figs. 4.48 (b) and 4.49 (b)) of $\sim 38 \times 38 \mu\text{m}^2$, smaller of $\sim 18\%$ than the one of the $1 \times 1 \text{ mm}^2$ detector which is $\sim 42 \times 42 \mu\text{m}^2$. Since the difference between measured micro-cells capacitances is $\sim 28\%$, the remaining difference can be imagined to be related to some structural or technological differences (i.e. influencing either the depletion thickness or parasitic quencing capacitance).

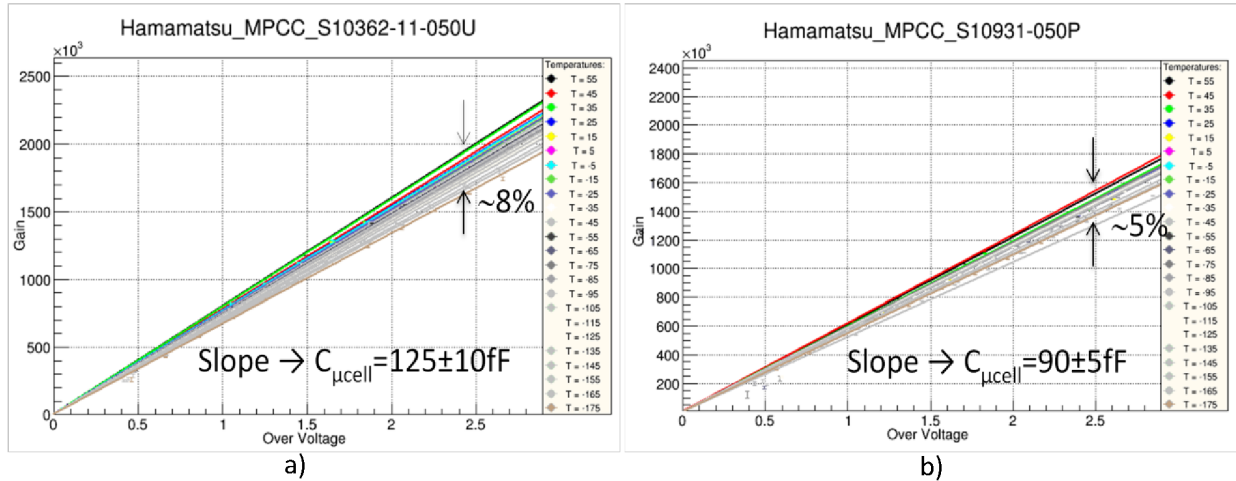


Fig. 4.46 Gain versus overvoltage for (a) $1 \times 1 \text{ mm}^2$ and (b) $3 \times 3 \text{ mm}^2$ detectors

d) Quenching resistance

Fig. 4.47 presents the quenching resistance R_Q temperature variation for (a) $1 \times 1 \text{ mm}^2$ and (b) $3 \times 3 \text{ mm}^2$ detectors. As reported in the previous Section 4.2.5.d. we can observe that R_Q exhibit a large variation with T , with decreasing values when T increases independent of measured detector (i.e. negative temperature coefficient). Their different temperature dependence determines us to imagine that different technological parameters have been used for the fabrication of the poly-silicon resistors (i.e. different doping concentrations of poly-silicon).

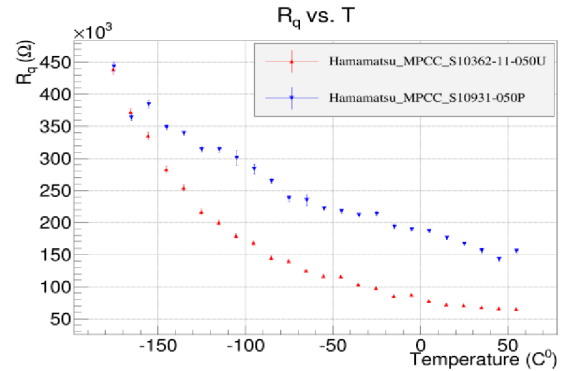


Fig. 4.47 Quenching resistance R_Q versus T for $1 \times 1 \text{ mm}^2$ and $3 \times 3 \text{ mm}^2$ HPK devices

This is confirmed also by the fact that the geometrical characteristics of the quenching resistor measured on the surface of the detectors (Figs. 4.48 and 4.49) are identical (i.e. length $\sim 45\mu\text{m}$; width $\sim 3\mu\text{m}$).

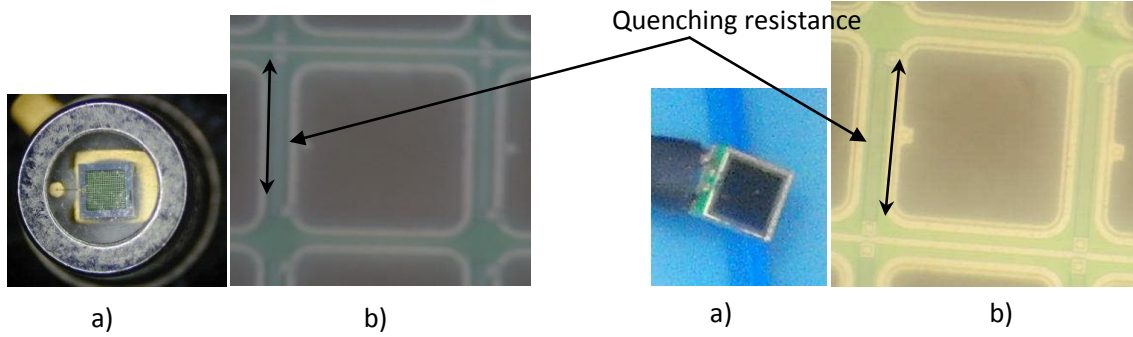


Fig. 4.48 Pictures of SiPM of $1 \times 1 \text{ mm}^2$ total area and $50 \times 50 \mu\text{m}$ micro-cell size: (a) overall view; (b) zoom over one micro-cell; the R_Q can be identified

Fig. 4.49 Pictures of SiPM of $3 \times 3 \text{ mm}^2$ total area and $50 \times 50 \mu\text{m}$ micro-cell size: (a) overall view; (b) zoom over one micro-cell; the R_Q can be identified

The variations of the R_Q with temperature lead as expected to variation of the single micro-cell SiPM shape (Fig. 4.50 (a) for $1 \times 1 \text{ mm}^2$ and (b) $3 \times 3 \text{ mm}^2$), determining a decreasing of the recovery time constant with increasing the temperature.

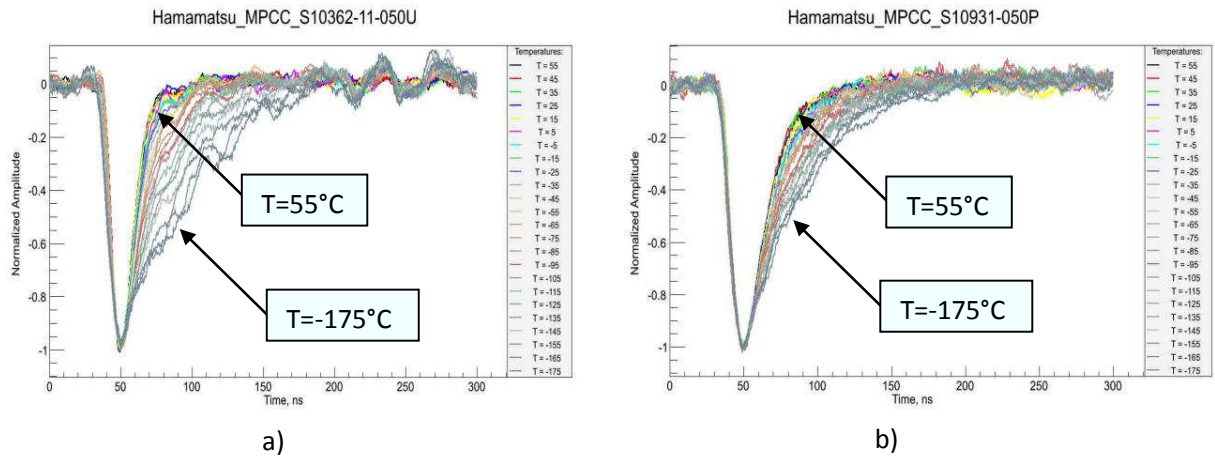


Fig. 4.50 Normalized signal shape for (a) $1 \times 1 \text{ mm}^2$ and (b) $3 \times 3 \text{ mm}^2$ detectors at different temperatures

e) Dark count rate

The dark count rate DCR versus overvoltage for different temperatures is presented in Fig. 4.51 (a) for $1 \times 1 \text{ mm}^2$ and (b) for $3 \times 3 \text{ mm}^2$ detector. Unfortunately the DCR could not be calculated for temperatures below -25°C because the DCR values are below 10^3 Hz and we have not enough statistics (i.e. $2.5 \mu\text{s}$ length of waveform has been used for DCR calculations).

As we already know, the DCR of a SiPM detector represents the sum of DCR signals coming from different micro-cells; therefore we expect that DCR increases with increasing the total number of micro-cells, therefore with the detector area. We can remark that at a given ΔV and T , both detectors exhibit quite equal DCR

(i.e. $\text{DCR} \sim 2 \times 10^6 \text{ Hz}$ @ $\Delta V = 1.5 \text{ V}$ and $T = 25^\circ \text{C}$ for both detectors), even if they have different total areas of $1 \times 1 \text{ mm}^2$ and respectively $3 \times 3 \text{ mm}^2$. This is clear indication showing that Hamamatsu modified the technological parameters, allowing a great reduction of this “non-desired” parameter (DCR/mm^2 has been reduced of about 10 times).

Independent of the analyzed detector, at a given T , the DCR increases with the ΔV as expected, since both triggering probability as well as after-pulse probability increases with increasing over-voltage. The values of DCR for temperatures below -25°C could not be calculated because our acquisition windows for dark signals was too short ($2.5 \mu\text{s}$ /acquired waveform) and not enough statistics was available for DCR below 1 kHz .

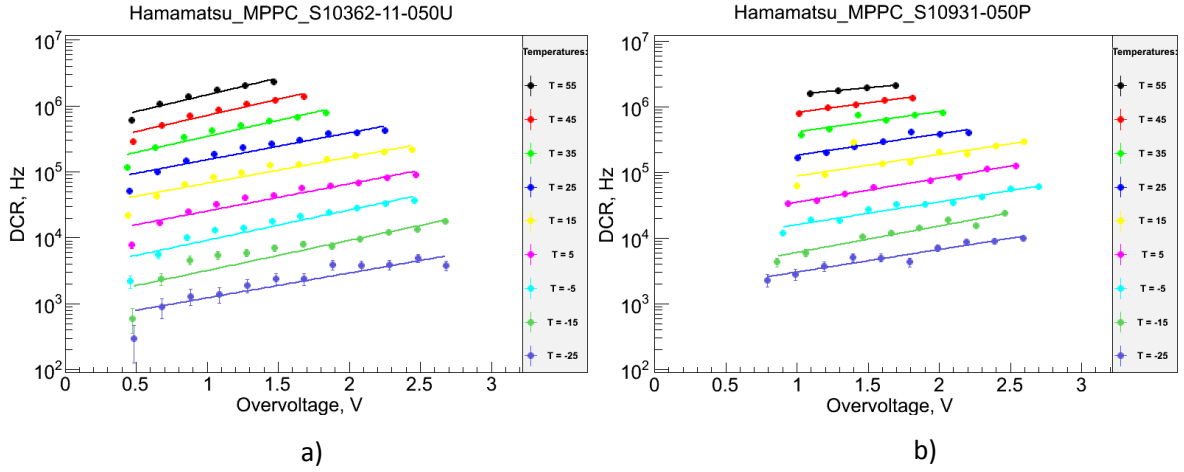


Fig. 4.51 Dark count rate versus overvoltage for (a) $1 \times 1 \text{ mm}^2$ and (b) $3 \times 3 \text{ mm}^2$ detector at different temperatures

As expected, at a given ΔV , the DCR decreases with decreasing T . Presently, detailed analysis of different DCR contributions (i.e. thermal generated carriers, after-pulses) and their temperature dependence is in progress. To improve the statistics of DCR calculation at low temperatures, new experimental measurements with longer acquisition time are planned to be done in the near future.

4.3. SiPM arrays

We have shown in the previous section that SiPM's present important advantages such as high internal gain ($\sim 10^5 - 10^6$), single photon timing resolution, low power consumption ($V < 100V$, $I \sim$ tens of μA), insensitivity to magnetic fields, stability and compactness. They have also an excellent signal to noise ratio, making unnecessary the use of sophisticated low-noise front-end electronics. These characteristics make these devices very attractive for physics and medical applications where multi-photon detection is involved. However, such applications require large detection areas, of the order of at least few cm^2 .

The SiPM's presented in the previous section were coming from the very early production runs of such devices and they had a total area of $1 \times 1 \text{ mm}^2$. Subsequently, SiPM's of $3 \times 3 \text{ mm}^2$, $4 \times 4 \text{ mm}^2$ and $6 \times 6 \text{ mm}^2$ have been also produced, but such surfaces cannot be increased more since the noise level scales with the area of the device (Piemonte, 2007). Therefore, an array of SiPM's represent the best solution for enlarging the active surface allowing to maintain, on each read-out channel, a noise level corresponding to a single SiPM. Moreover, the array can also give information on the 2D position with a spatial resolution equals or even less than the SiPM size.

There are two different approaches that can create a large-area of SiPMs: lining up discrete elements or designing a monolithic array.

- The advantage of discrete elements approach is that one can select and sort discrete SiPM's from multiple lots and therefore it is possible to fabricate an array that is more uniform in gain than would otherwise be possible if all the devices were from the same wafer. The main disadvantage of such array is the dead area between different SiPM's which could be quite significant.
- Monolithic arrays present the advantage of reduced dead area between different SiPM's but they are more risky from the yield point of view. If the variation in breakdown voltage is large, some channels may not operate in Geiger-mode, or the other extreme would be if the overvoltage value is too high for a channel. Even if the difference between breakdown voltages over different channels is not so significant, the uniformity of charge signals undoubtedly gets more problematic as the array size increases.

Another challenge behind making a monolithic array is the design of the trace layout. The impedance of the aluminum traces significantly affects the pulses shape so any variation in the length or width of these traces would change the pulse shape coming out of each SiPM channel in the array.

My activity on SiPM arrays started in 2007 on a monolithic array of 4×4 SiPM's, each SiPM of $1 \times 1 \text{ mm}^2$, fabricated by FBK-irst. This work has been reported to NDIP conference in 2008 (N. Dinu et al, 2009, b) and it represented the first publication on a characterization of a monolithic SiPM array. My second activity on SiPM arrays took place in 2011 and it is related to monolithic arrays of 4×4 SiPM's, each SiPM of $3 \times 3 \text{ mm}^2$, from Hamamatsu HPK. Since these activities were closely connected to particular projects (i.e. medical applications), the detailed characteristics of these devices will be presented in the next Section. In the following, the state of the art of SiPM discrete and monolithic arrays from main SiPM producers will be presented.

4.3.1. Discrete SiPM arrays

Table 4.2 presents the most recent SiPM arrays built from discrete elements, available today from main SiPM producers.

We can remark that the maximum number of channels obtained by assembling discrete elements reaches today the value of 256. Such system, based for example on the S11064-050P Hamamatsu discrete array, covers a total area of $\sim 72 \times 64.8 \text{ mm}^2$ (geometrical fill factor of $\sim 50\%$). This fill factor is determined by the SiPM's assembling. The geometrical fill factor of each channel (i.e. each SiPM) is not taken into account.

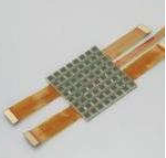
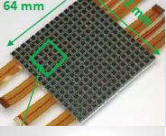
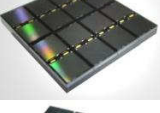
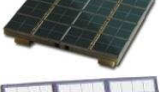
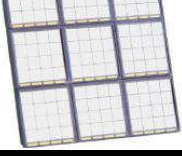
Producer	Device ID	Picture	Total area (mm^2)	SiPM area ($\text{mm}^2/\text{channel}$)	Nr. channels	μcell size
Hamamatsu	S11064-025P S11064-050P		18 x 16.2	3x3	16(4x4) ch	25x25 μm 50x50 μm
Hamamatsu	C11206-0404DF			3x3	64(8x8) ch	
Hamamatsu	S11834-3388DF		72x64.8	3x3	256(16x16) ch	
FBK AdvanSiD	ASD-SiPM4s-P-4x4T-50 ASD-SiPM4s-P-4x4T-69		8.2 x 8.2	4x4	16(4x4) ch	50x50 μm 69x69 μm
FBK AdvanSiD	SiPM tile		32.7x32.7	4x4	64(8x8) ch	
SensL	ArraySM-4P9 ArraySB-4P9 (blue sensitive)		46.3 x 47.8	3x3	144(12x12) ch (based on monolithic Array SM4)	35x35 μm

Table 4.2 Table of SiPM's arrays built using discrete elements

4.3.2. Monolithic SiPM arrays

Table 4.3 presents the most recent monolithic SiPM arrays, available today from main SiPM producers. We can remark that maximum number of channels reached on monolithic device is 64, proposed by FBK-AdvanSiD.

A simple calculation shows that using the monolithic arrays S11828-3344M from Hamamatsu, a system of 256 channels will cover $\sim 54 \times 56 \text{ mm}^2$ total area, determining a geometrical fill factor $\sim 75\%$, therefore 25% more in comparison with a photo-detection system having similar number of channels but built with discrete components (see previous page). This fill factor is determined by the arrays assembling. The geometrical fill factor of each channel (i.e. each SiPM) is not taken into account.

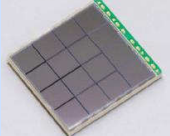
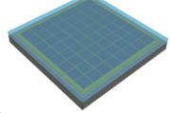
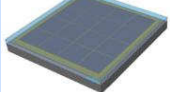
Producer	Device ID	Picture	Effective area (mm^2)	SiPM area/channel (mm^2)	Nr. channels	μcell size
Hamamatsu	S10984-025P S10984-050P S10984-100P		1 x 4	1x1	4(1x4) ch	25x25 μm 50x50 μm 100 x 100 μm
Hamamatsu	S10985-025C S10985-050C S10985-100C		6 x 6	3x3	4(2x2) ch	25x25 μm 50x50 μm 100 x 100 μm
Hamamatsu	S11828-3344M		12 x 12	3x3	16(4x4) ch	50x50 μm
FBK AdvanSiD	ASD-SiPM1.5s-P-8X8A		11.6 x 11.6	1.45x1.45	64(8x8) ch	50x50 μm
FBK AdvanSiD	ASD-SiPM3S-P-4X4A		11.8 x 11.8	2.95x2.95	16(4x4) ch	50x50 μm
SensL	Array SM-4 Array SB-4 (blue sensitive)		12 x 12	3x3	16(4x4) ch	35x35 μm

Table 4.3 Table of monolithic SiPM's arrays

4.4. Summary : advantages and drawbacks of SiPM's

We have seen that Silicon Photomultiplier detector can address the challenge of detecting, time tagging and quantifying light signals with a large dynamic range down to single-photon level. Such detector is represented by an array of parallel micro-cells, where each micro-cell is a Geiger-mode Avalanche Photodiode connected in series with integrated passive quenching resistance.

The elementary micro-cell of a SiPM operates as a photon-triggered switch: regardless of the number of photons interacting within a diode at the same time it will produce a signal of "1" photon. However, the SiPM overcomes the lack of proportionality of a single GM-APD, integrating a dense array of GM-APD's on the same substrate and giving information on the magnitude of an incident photon flux working as an "analog" device.

As a single GM-APD, the SiPM achieves high gain (10^6) with low bias operating voltage (from 30 to 75V, depending of the technology). Because of the high degree of uniformity between the micro-cells, the SiPM is capable of discriminating the precise number of photons detected as distinct, discrete levels at the output node. It presents a moderate PDE ($\sim 20\text{-}40\%$) with a peak in blue or green wavelength as a function device technology, as well as a very good timing resolution (single photon timing resolution less than 50 ps).

Moreover, it has the physical benefits of compactness, ruggedness and magnetic insensitivity as any solid-state photon detectors.

Despite of the fact that the SiPM is sensitive to single photons, its dark count rate of ~ 100 kcps / mm^2 at room temperature renders it unsuitable for use in applications at very low light levels if no external trigger (i.e. coincidence) can be applied. However, with the application of cooling, the contribution of the primary phenomenon (i.e. thermal generated carries) can be drastically reduced, determining a reduction in the dark count rate by orders of magnitude (less than $1\text{kcps}/\text{mm}^2$ at -25°C). The contribution of secondary effects (i.e. cross-talk and afterpulses) can be still at some % level for many devices and the related physical phenomena as well as technological solutions for their reduction are still opened research subjects.

Another two important effects of temperature on the SiPM characteristics are: (1) a change in the breakdown voltage and (2) a change of the signal shape.

Unless compensated by temperature regulation, the breakdown voltage variation will result in a change of the effective overvoltage that may affect the stable operation of the detector. However, knowing the temperature dependence of the breakdown voltage and adjusting the bias voltage in order to maintain a constant overvoltage independent of temperature, the main characteristics of the device will remain the same as at room temperature.

However, even if the overvoltage is kept constant, the temperature dependence of the quenching resistance determines different signal shapes for different temperatures. Therefore, the read-out electronics parameters must be adjusted (integration gate) as a function of signal shape variation.

The SiPM device achieved today a mature technology, arrays of such detectors being available either by lining up discrete elements or designing monolithic arrays. Due to the attractive perspective of using such arrays for building large detection areas with the great advantage of a high spatial resolution, this device is gaining an important place in many applications as medical imaging, high energy physics and LiDAR. The work I performed on analyzing the uniformity of SiPM arrays as well as their properties when they are used for two medical imaging applications will be presented in the next Section.

References on SiPM:

- Chaumat, V.; Bazin, C.; **Dinu, N.**; Puill, V.; Vagnucci, J.F.; (2013), Absolute photo-detection efficiency measurement of SiPM, *Proceedings of Science*, in press
- Collazuol, G.; Ambrosi, G.; Boscardin, M.; Corsi, F.; Dalla Betta, G.F.; Del Guerra, A.; **Dinu, N.**; Galiberti, M.; Giuliotti, D.; Gizzi, L.; Llosa, G.; Marcatili, S.; Morsani, F.; Piemonte, C.; Pozza, A.; Zaccarelli, L.; Zorzi, N.; (2007); Single photon timing resolution and detection efficiency of the IRST silicon photo-multipliers; , *Nuclear Instruments and Methods in Physics Research A*, Vol. 581, 461-464
- Collazuol, G.; Bisogni, M.G.; Marcatili, S.; Piemonte, C.; Del Guerra, A.; (2011); Studies of silicon photomultipliers at cryogenic temperatures, *Nuclear Instruments and Methods A*, Vol. 628, Issue 1, 389-392
- Corsi, F.; Marzocca, C.; Perrota, A.; Dragone, A.; Foresta, M.; Del Guerra, A.; Marcatili, S.; Llosa, G.; Collazuol, G.; Dalla Betta G.F.; **Dinu, N.**; Piemonte C.; Pignatel, G.U.; Levi, G.; (2006), Electrical characterization of Silicon Photomultipliers detectors for optimal front-end design, *IEEE Nuclear Science Symposium Conference Record*, N30-222, 1276-1280
- Corsi, F.; Dragone, A.; Marzocca, C.; Del Guerra, A.; Delizia P.; **Dinu N.**; Piemonte, C.; Boscardin M.; Dalla Betta G.F.; (2007), Modeling a silicon photomultiplier (SiPM) as a signal source for optimum front-end design, *Nuclear Instruments and Methods A*, Vol. 572, 416-418
- Cova, S.; et al.; (2003); NIST workshop on Single Photon Detectors
- Crowell, C.R.; Sze, S.M.; (1966); Temperature dependence of avalanche multiplication in semiconductors, *Applied Physics Letters*, Vol. 9, Issue 6, 242-244
- Dinu, N.**; Battiston, R.; Boscardin, M.; Collazuol, G.; Corsi, F.; Dalla Betta G.F.; Del Guerra, A.; Llosa, G.; Ionica, M.; Levi, G.; Marcatili, S.; Marzocca, C.; Piemonte, C.; Pignatel, G.; Pozza, A.; Quadroni, L.; Sbarra, C.; Zorzi, N.; (2007); Development of the first prototypes of silicon photomultiplier (SiPM) at FBK-irst, *Nuclear Instruments and Methods A*, Vol. 572, 422-424
- Dinu, N.**; Amara, Z.; Bazin, C.; Chaumat, V.; Cheikali, C.; Guilhem, G.; Puill, V.; Sylvia, C.; Vagnucci, J.F.; (2009, a); Electro-optical characterization of SiPM: a comparative study; *Nuclear Instruments and Methods A*, Vol. 610, 423-426
- Dinu, N.**; Barrillon, P.; Bazin, C.; Belcari, N.; Bisogni, M.G.; Bondil-Blin, S.; Boscardin, M.; Chaumat, V.; Collazuol, G.; De la Taille, C.; Del Guerra, A.; Llosa, G.; Marcatili, S.; Melchiorri, M.; Piemonte, C.; Puill, V.; Taroli, A.; Vagnucci, J.F.; Zorzi, N.; (2009, b); Characterization of a prototype matrix of SiPM, *Nuclear Instruments and Methods in Physics Research A*, Vol. 610, 101-104.
- Dinu, N.**; Bazin, C.; Chaumat, V.; Para, A.; Puill, V.; Sylvia, C.; Vagnucci, J.F. (2010); Temperature and bias voltage dependence of the MPPC detectors, *IEEE Nuclear Science Symposium Conference Record*, 215-219
- Dolgoshein, B.; Buzhan, P.; Ilyin, A.; Kantserov, V.; Kaplin, V.; Karakash, A.; Pleshko, A.; Popova, E.; Smirnov, S.; Volkov, Y.; (2001); An Advanced study of silicon photomultiplier, *ICFA Instrumentation Bulletin*
- Golovin, V.M., et al, (1999) ; *Russian patent*, nr. 1644708
- Haitz, R.H.; (1965), Studies on optical coupling between silicon p-n junctions, *Solid-State Electronics*, Vol. 8, 417-425
- Kindt, W.J.; Zeijl, van H.W.; Middelhoeck, S.; (1998); Optical cross-talk in Geiger mode avalanche photodiodes arrays: modeling, prevention and measurement, *Solid-State Device research Conference*, Proceedings of the 28th European
- Lacaita, A.; Mastrapasqua, M.; Ghioni, M.; Vanoli, S.; (1990); Observation of avalanche propagation by multiplication assisted diffusion in p-n junctions; *Applied Physics Letters*, Vol. 57, Issue 5, 489-491
- Lacaita, A.; Zappa, F.; Bigliradi, S.; Manfredi, M.; (1993 a); On the bremsstrahlung origin of hot-carrier-induced photons in silicon devices, *IEEE Transactions on Electron devices*, Vol. 40, Issue 3, 577-582
- Lacaita, A.; Cova, S.; Spinelli, A.; Zappa, F.; (1993 b); Photon-assisted avalanche spreading in reach-through photodiodes; ; *Applied Physics Letters*, Vol. 62, Issue 6, 606-608
- Nagai, A.; (2013); Temperature dependence of SiPM characteristics, Internship report supervised by **N. Dinu**, A. Para and O. Bezshyyko.
- Piazza, A.; Boscardin, M.; Dalla Betta G.F.; Del Guerra, A.; Melchiorri M.; Piemonte, C.; Tarolli, A.; Zorzi, N.; (2010); Characterization and simulation of different SiPM structures produced at FBK, *Nuclear Instruments and Methods A*, Vol. 617, 417-419
- Piemonte, C.; Boscardin, M.; Dalla Betta G.F.; Melchiorri, M.; Zorzi, N.; Battiston, R.; Del Guerra, A.; Llosa, G.; (2007); Recent developments on silicon photomultipliers at FBK-irst, *IEEE Nuclear Science Symposium Conference Record*, N41-2, 2089-2092.

Prochzka, I. ; Hamal, K. ; Kral, L. ; Blazej, J.; (2005); Silicon photon counting detector optical cross-talk effect, *5th International Conference on Photonics, Devices and Systems*, Prague

Rech, I.; Ingargiola, A.; Spinelli, R.; Labanca, I.; Marangoni, S.; Ghioni, M.; Cova, S.; (2008) ; Optical crosstalk in single photon avalanche diode arrays: a new complete model, *Optics Express*, Vol. 16, Issue 12, 8381-8394

Sadygov, Z.; et al.; (1997), *Russian Patent*, nr. 2086047 C1

Seifert, S.; Dam van H.T. ; Huizenga J. ; Vinke R. ; Dendooven P.; Lohner, H.; Schaart, D.R.; (2009), Simulation of silicon photomultiplier signals, *IEEE Transactions on Nuclear Science*, Vol. 56, Issue 6, 3726-3733

Serra, N.; Giacomini, G.; Piazza, A.; Piemonte, C.; Taroli, A.; Zorzi, N.; (2011); Experimental and TCAD study of breakdown voltage temperature behavior in n+/p SiPMs, *IEEE Transactions on Nuclear Science*, Vol. 58, Issue 3, 1233-1240.

Spinelli, A.; Lacaita, L.; (1997); Physics and numerical simulation of single photon avalanche diodes, *IEEE Transactions on Electron Devices*, Vol. 44, Issue 11, 1991-1943

Vacheret, A.; Barker, G.J.; Dziewiecki, M.; Guzowski, P.; Haigh, M.D.; Hartfiel, B.; Izmaylov, A.; Johnston, W.; Khabibullin, M.; Khotjantsev, A.; Kudenko, Yu.; Kurjata, R.; Kutter, T.; Lindner, T.; Masliah, P.; Marzec, J.; Mineev, O.; Musienko, Yu.; Oser, S.; Retiere, F.; Salih, R.O.; Shaikhiev, A.; Thompson, L.F.; Ward, M.A.; Wilson, R.J.; Yershov, N.; Zaremba, K.; Ziembicki, M.; (2011); Characterization and simulation of the response of multi pixel photon counters to low light levels, *Nuclear Instruments and Methods A*, Vol. 656, Issue 1, 2011, 69-83

Yamamura, K.; Sato, K.; Kamakura, S.; Ohsuka, S.; (2009); Products and Development status of MPPC, *PD09 International workshop on new photon-detectors*, Matsumoto.

Wilmshurst, T.H. (1985), Signal recovery from noise in electronic instrumentation, *A. Hilger Publ.*

5. Arrays of SiPMs for medical applications

As mentioned previously, my first work on SiPM's arrays started in 2007 through a collaboration between LAL, INFN & Pisa University and FBK-irst, of which I was the initiator. In the framework of this collaboration, we received a prototype array consisting of monolithic 4x4 SiPM's fabricated at FBK-irst and we connected it to a multi-channels integrated read-out electronics developed by the Omega micro-electronics group. Such array has been subsequently evaluated for a small animal PET scanner by the Pisa group. The characteristics of the SiPM's array produced by FBK-irst as well as their performances for the small animal PET scanner will be presented in the first part of this Section.

My second work on SiPM's array took place in 2011 as LAL responsible for a project called SIPMED (Silicon Photomultiplier for Medical Imaging) developed in collaboration with IMNC-Orsay and Hospital Lariboisiere and financed by l'Institut National de la Santé et de la Recherche Médicale (INSERM). In the framework of this project, we are using monolithic SiPM's arrays produced by Hamamatsu HPK, connected to a multi-channels read-out ASIC developed by the Omega micro-electronics group and test-boards developed by the electronic service at LAL. Such system is in progress to be build and evaluated as prototype of a very compact intra-operative gamma camera required in imaged guided surgery of tumors. The characteristics of the SiPM's arrays produced by HPK as well as the work under progress for the development of the miniaturized read-out electronics and calibration tests will be presented in the second part of this section.

5.1. Arrays of SiPM's for high resolution small animal PET scanner

5.1.1. The principle of PET

Technical supervision: Vincent Chaumat, Jean-François Vagnucci (LAL)
Internal collaboration : Pierre Barrillon, Sylvie Bondil-Blin, Christophe de la Taille, Veronique Puill (LAL)
External collaborations: Gabriela Llosá and Alberto del Guerra –Pisa University, Italy; Claudio Piemonte – FBK-irst, Italy (Memorandum of Understanding: LAL/INFN & Pisa University/FBK-irst)

Medical imaging is nowadays an increasing field of research, because of the many morphologic and functional information's that it's possible to obtain without almost any invasive action in the patient. Positron Emission Tomography (PET) is one of those imaging techniques, dedicated to functional studies (e.g. cancer diagnostic). The mechanism of PET is simple (Fig. 5.1): a β^+ emitter tracer is injected into the body of the patient. Emitted positrons interact with electrons, producing two almost anti-parallel photons which are detected in coincidence by opposite arrays of detectors. As a result, a complete projection of the activity in the patient for a complete set of angles is obtained. The image is reconstructed by back-projection algorithms.

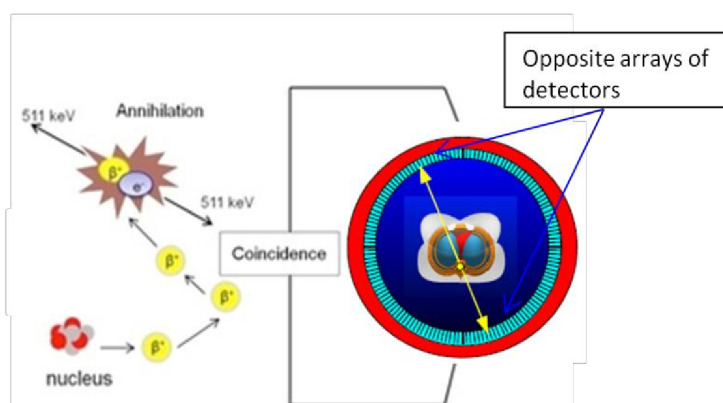


Fig. 5.1 Schematic principle of PET

Tracers (called also radio-pharmaceutical or radio-nuclide) are usually pharmaceutical organic molecules metabolized by the organs under study (e.g. glucose metabolism), coupled with β^+ emitters nuclides, mostly produced by dedicated medical low-energy proton and deuteron accelerators. The usual injected activity for a 70 kg patient is 5-15mCi. Positrons are emitted with an initial kinetic energy whose distribution is continuum up to a maximum value. When emitted in the medium, positrons start a process of slowing down and thermalization. They lose energy and change direction by elastic scattering with atomic electrons and nuclei. This process lasts until the positron reaches thermal equilibrium with the surrounding matter: at this point it interacts with an electron. Consequently, positron mean free path in matter depends on two factors: the initial energy of the emitted particle and the properties of the surrounding matter in which positron is thermalized. For PET the most important range is that in water, because water is the principal component of human body. In the Table 5.1 it's shown the mean range of positrons emitted by common nuclides used in PET (Valk et al, 2003).

Nuclide	$E_{\max}$ (MeV)	$t_{1/2}$ (mins)	Mean Range(mm)
^{11}C	0.959	20.4	1.1
^{13}N	1.197	9.96	1.5
^{15}O	1.738	2.03	2.5
^{18}F	0.633	109.8	0.6
^{68}Ga	1.898	68.3	2.9
^{82}Rb	3.40	1.25	5.9
$^{94\text{m}}\text{Tc}$	2.44	52	n.r.

Table 5.1 Maximum energy, the half-life and the range in water of positrons emitted by the more common nuclides used in PET

Positrons usually annihilate with electrons, resulting in the emission into opposite direction of two 511 keV photons. The annihilation photons have then to travel through the biological target for a distance depending on the shape of the target and on the point of the production of the photons. In human body scanners this distance can be tens of centimeters. As the linear attenuation coefficient of a 511 keV photons in water is 9.68 cm^{-1} , there's a finite probability that one of them undergoes Compton scatter before going out the object, losing energy and changing considerably its flight direction. The total radiation reaching the detector is then a sum of the photons directly coming from annihilation, Compton Scattered ones and background radiation coming from outside of the field of view (FOV).

The purpose of a PET system is to detect in coincidence pairs of annihilation photons generated in the biological target. As they are randomly and almost uniformly distributed in the solid angle around the annihilation region, a PET scanner needs detectors in opposite position and covering as much FOV as possible. Two steps are required for 511 keV photons detection:

- first an interface is needed, which converts 511 keV photons in visible radiation. This role is usually played by crystal scintillators commonly arranged in crystal blocks; in PET systems, the most commonly used are the inorganic scintillator crystals such as: (i) sodium iodide doped with thallium (NaI(Tl)), (ii) bismuth germanate $\text{Bi}_4\text{Ge}_3\text{O}_{12}$ (BGO), (iii) lutetium oxyorthosilicate doped with cerium $\text{Lu}_2\text{SiO}_5\text{:Ce}$ (LSO), (iv) yttrium oxyorthosilicate doped with cerium $\text{Y}_2\text{SiO}_5\text{:Ce}$ (YSO), (v) gadolinium oxyorthosilicate doped with cerium $\text{Gd}_2\text{SiO}_5\text{:Ce}$ (GSO), and (vi) barium fluoride (BaF_2).
- the second detection step is a photo-detector, which detect the light emitted by the scintillator block, and whose output is a measurable electrical signal that should be proportional to the energy of the photons; vacuum photon detectors such as photomultiplier tubes (PMT) as well as solid state detectors such as photodiodes, avalanche photodiodes have been commonly used up today.

More details about the detectors features required for PET systems will be presented in the Section 5.1.3.

Besides, reconstructing algorithms for PET scanners are based on the back-projection of the projected activity in a complete set of different azimuthal angles around the patient. Different configurations were implemented to comply with those two needs (Fig. 5.2). The most common configuration is a stationary full ring of detectors. Another stationary configuration is made by flat detectors (typically 6-8). Moving configurations are obtained by rotating two or four opposite partial-ring circular or flat detectors, resulting in a ring acquisition around the studied object (Fig. 5.3 (a)).

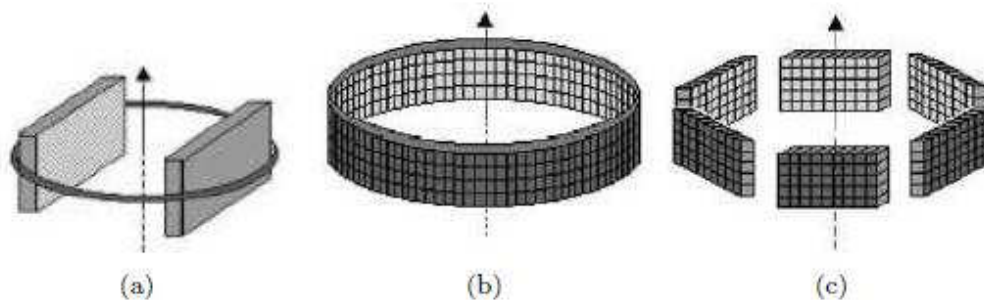


Fig. 5.2 Examples of PET detection systems: **(a)** two rotating heads, **(b)** ring geometry, **(c)** polygonal detector ring

PET systems can be designed to operate in 2D or 3D mode. In 2D mode, each ring of detectors is used to acquire activity projections coming from a single trans-axial slice. For this purpose, in a multi-ring geometry, each ring is physically decoupled by leads of tungsten shields (Fig. 5.3 (a)). In a 3D mode, the acquisition is not any more limited to a planar slice, but is performed on the whole solid angle between any detector pair, therefore achieving a fully 3D image (Fig. 5.3 (b)).

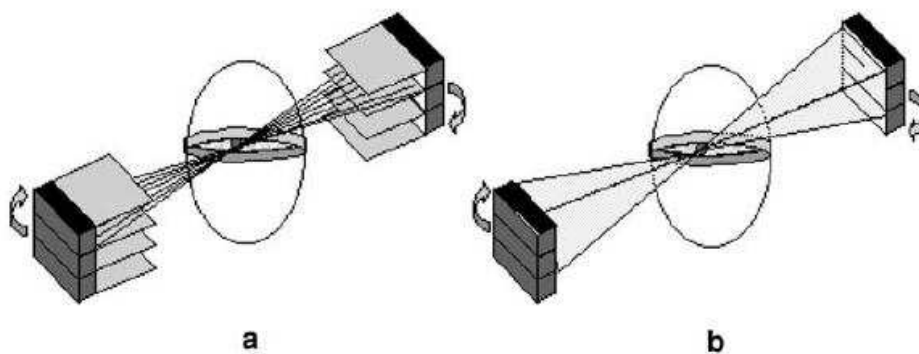


Fig. 5.3 **(a)** 2D and **(b)** 3D PET acquisition modes

The output signal from the read-out detector triggers the time gate generator and information about the energy of the incident photon is extracted. Dedicated electronics establish which detector was hit. The information is stored in a data block and it's then compared with the events registered in the opposite detectors during the generated time slot, to see if there is a coincidence or not. Two detectors hit in coincidence define a coincidence line of response (LOR). With several LORs scanning the target in different directions, it's possible to define projections profiles of the radioactivity distributed within the target. These projections will be used then by the reconstruction algorithms.

The acquisition for each projection can be made in histogram mode or list mode. The first one is a scalar increment for each channel when a coincidence event comes up. The second one is more complicated: coincidences are sorted out by the electronics, and then data characterizing one single event are stored in one data block. The difference with histogram mode is that channels are processed and stored into histograms after the acquisition and not on line. This second mode can be useful in modern PET systems, where an increase of coincidence channels should make difficult a real time handling of a huge amount of event data.

5.1.2. Detection features for PET systems

The imaging performances of a PET system can be described in terms of *spatial resolution* and *sensitivity*.

a) Spatial resolution

The physical phenomena involved in a PET system give a first limitation to the *image or spatial resolution*. The *positron range in the medium* and the *non-co-linearity of the emitted photons* are intrinsic non-avoidable and causes a *blurring* in the final image. An empirical formula (Lecomte, 2004) described the reconstructed full-width at half-maximum (FWHM) image resolution:

$$FWHM = 1.25 \sqrt{\left(\frac{d}{2}\right)^2 + b^2 + r^2 + (0.0022D)^2} \quad 5.1$$

where d is the crystal width, D is the diameter of the detector ring, r is the mean range of positron in tissue and b is an additional term which is required to explain poorer resolution of scanners with block scintillator detectors. The two latter terms are the contribution of the *range effect* and of the *2 γ non-co-linearity* that is proportional to the diameter of the scanner system D . This intrinsic effect (Figs. 5.4 (a) and (b)), taking for example ^{15}O for a 1m diameter scanner and using the positron mean range of Table 5.1, gives a FWHM of 4.16 mm, that can be reduced to 2.85 mm using ^{18}F , in the ideal condition when d and b are null.

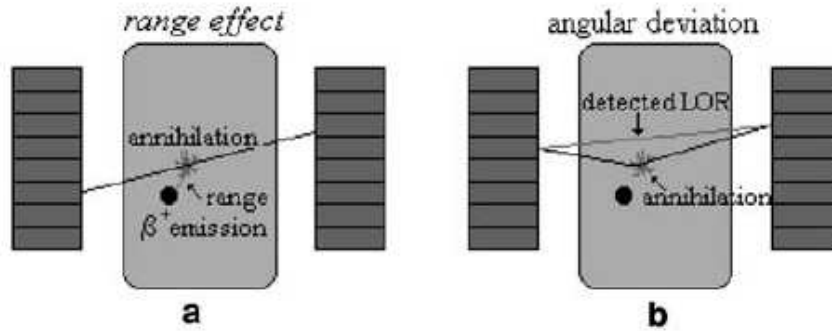


Fig. 5.4 Fundamental limits in PET: **(a)** *Range effect*: the detected line of response (LOR) contains the annihilation point, not the emission point; **(b)** *2 γ non-co-linearity*: the tomography measures an LOR on which the annihilation point does not lie (Humm & al, 2003)

The first two terms are linked to the properties of the scintillator and the photon detector, which is the core of a good PET detection system: the width of the scintillator crystal d and the resolution effects b , depending on both the scintillator and photon detector characteristics (e.g. crystal dimensions, and its readout) as well as on the direct coupling between them. It has been shown (Thompson et al. 2005) that neglecting the contribution of positron range and the non co-linearity, the spatial resolution FWHM is a linear function of the crystal width, with an offset depending on the coupling of the scintillators and the photon-detector. The algorithm used to reconstruct the interaction position is accounted by the multiplication factor 1.25. Eriksson et al, (2004), illustrates that in 85 cm diameter PET system using ^{18}F and LSO crystals, increasing the width of the crystals from 2 to 4 mm and the coupling factor from 0 to 0.5, this degrades significantly the image resolution from 3.04 mm to 4.56 mm.

b) Sensitivity

System sensitivity is considered the most important parameter since it determines the image quality per unit scan time or alternatively the time required to perform a whole-body PET scan. The physical definition is the ratio of the number of detected events to the number of decays within the scanner field of view (i.e. the probability of detecting two annihilation photons). On a clinical system, sensitivity depends on:

- the detectors characteristics and their readout electronics;
- the axial field of view of the camera, since it determines the length of the body segment that can be imaged per bed position; the axial field of view of modern PET scanners is 15-16 cm;
- upon whether the scans are acquired in 2D or 3D modes.

High sensitivity means high statistics, an important feature needed for a good image reconstruction. This property depends strongly on the detectors efficiency. If we take two opposite detectors each with efficiency η , the probability that these detectors detect two photons in coincidence is proportional to η^2 . As coincidences are detected in a total solid angle given by the field of view, the sensitivity will be given by:

$$\zeta \propto v\eta^2 \quad 5.2$$

where v is the fraction of solid angle covered by the detectors. More details about the detectors features required for high sensitivity PET systems will be presented in the Section 5.1.3.

The statistics in the image acquisition is limited by *random and scattered events*. Random coincidences are a direct consequence of having a large coincidence timing window. They arise when two unrelated photons enter the opposing detectors and are temporally close enough to be recorded within the coincidence timing window. For such events, the system produces a false coincident event. Due to the random nature of such events, they are labeled as random or accidental coincidences. The random rate is proportional to the coincidence timing window and to the activity detected by the detectors:

$$R_{1,2} = 2 \cdot \tau \cdot N_1 N_2 \quad 5.3$$

where N_1 and N_2 are the single event rate for the two detectors and τ is the timing window. Random coincidences add uncorrelated background counts to an acquired PET image and hence decrease image contrast if no corrections are applied to the acquired data.

Scattered events also give their contribution to lower the rate of the true coincidences. The density of tissue in human body is approximately the same as that of water, and so the mean free path of a 511 keV photon is about 7 cm in human tissue. Since the cross-section of a human body is much greater than 7 cm, many of the photons originating inside the human body are Compton scattered before they enter the PET detectors. Since scatter involves loss of energy, in principle some of these scattered coincidences can be rejected using an energy gating technique around the photopeak in the energy spectrum. Good energy resolution for the detector allows the application of a very narrow energy gate, and thus a more extensive and accurate rejection of scatter coincidences can be performed. However, some scattered events may be indistinguishable from true coincidences based upon the energy if they lie within the photopeak. Hence, additional scatter-correction techniques which estimate the distribution of scattered radiation are then employed in order to remove them from the image and improve image contrast.

A quantitative statistic measurement of the significance of the coincidence rate compared to the random and scattered events is the noise equivalent count (NECR):

$$NECR = \frac{T^2}{T + S + \alpha R} \quad 5.4$$

where T is the true coincidence rate, R is the random event rate, S is the scatter rate, α is a parameter depending on the correction algorithm used for random events.

The measured rate is limited by the recovery time of the detector and the read-out electronics. If the activity in the target is increased, there will be a further increase both of true and random coincidence rates. The later one increases faster. The consequence of this process is that there is a peak of the NECR, depending on the geometry of the system. Substituting the random event value R with equation 5.3, a relationship between NECR and time gate τ can be found:

$$NECR = \frac{T^2}{T + S + 2 \cdot \alpha \cdot \tau \cdot N^2} \quad 5.5$$

This relationship shows that the *peak* of NECR is also strongly connected to the time coincidence window, hence to the detector timing properties. It was shown (Moses, 2003) that a decrease of the timing coincidence window reduces the maximum event rate requirements on data processing electronics (usually set by the random events rate) and increases the NECR. Typical NECR values for 82 cm detector ring diameter and 15 cm axial extent, for a 10 ns timing window, are 100-150 kcps, when a specific 3 μ Ci/cc is injected in a 20 cm diameter cylinder phantom.

Other cause of blurring in the image is the so-called *parallax error*. In ring geometry, photons are supposed to hit the scintillator perpendicularly to the front face. This happens for photons emitted in the center of FOV. In peripheral zones it is possible that they will not interact perpendicularly: they can cross the detector, they enter and then be absorbed by the near detector. Consequently a count to a wrong coincidence channel is added. Parallax error is as higher as longer are the scintillators and smaller the diameter of the ring. If a detector configuration is able to give a Depth of Interaction (DOI) resolution, it solves the parallax problem.

The overall geometry of the PET system also contributes to the image definition and noise events. In a 2D PET the total scattered and background radiation detected is typically 10%, in a 3D PET it can go up to 40%. A 3D PET, on the other side, has a higher sensitivity, as compared to the performances of a 2D PET.

Therefore, the technological challenges in building high performance PET systems could be defined as following: it has to be able to detect photons, emitted simultaneously and almost collinearly, to distinguish them from a background radiation that is considerably not negligible if compared to good events. The earlier is the discrimination of good from bad events, the better the final image is resolved. *Therefore, the detector elements play a very important role in the final performance of a PET system.*

5.1.3. PET detector modules

The main requirements for PET detector modules can be found in Moses et al, 1994. A successful PET detector module must identify 511-keV photons with: high efficiency (> 85%), high spatial resolution (< 5 mm FWHM), low cost (less than \$600/inch²), low dead time (< 4 μ s inch²), good timing resolution (< 5 ns FWHM for conventional PET, < 200 ps FWHM for time of flight), and good energy resolution (< 20%), where these requirements are listed in order of decreasing importance. The high efficiency requirement also implies that the detector module (i.e. scintillator crystal plus the photo-detector) must pack together without inactive gaps. To

obtain such characteristics, each detector element should satisfy given requirements as presented in the following.

a) Scintillators crystals

There are four main properties of a scintillator which are crucial for its application in a PET detector: the stopping power for 511 keV photons, signal decay time, light output, and the intrinsic energy resolution.

The stopping power of a scintillator is characterized by the mean distance travelled by the photon before it deposits its energy within the crystal. For a PET scanner with high sensitivity, it is desirable to maximize the number of photons which interact and deposit energy in the detector. Thus, a scintillator with a short attenuation length will provide maximum efficiency in stopping the 511 keV photons. The attenuation length of a scintillator depends upon its density (ρ) and the effective atomic number (Z_{eff}). The important parameter determining the scintillator material's ability to stop 511 keV photons is ρZ_{eff}^4 .

The decay constant affects the timing characteristics of the scanner. A short decay time is desirable to process each pulse individually at high counting rates, as well as to reduce the number of random coincidence events occurring within the scanner geometry.

A high light-output scintillator influences a PET detector design in two ways: it helps achieve good spatial resolution with a high encoding ratio (ratio of number of resolution elements, or crystals, to number of photo-detectors) and attain good energy resolution. Good energy resolution is needed to efficiently reject events which may Compton scatter in the patient before entering the detector. The energy resolution ($\Delta E/E$) achieved by a PET detector is dependent not only upon the scintillator light output but also the intrinsic energy resolution of the scintillator. The intrinsic energy resolution of a scintillator arises due to inhomogeneities in the crystal growth process as well as non-uniform light output for interactions within it.

Scintillators are commonly arranged in crystal blocks. They consist in scintillator blocks partially sawed into multiple pseudo individual crystals. The use of crystal blocks clearly reduces the packing fraction and the cost of a PET system. Nowadays also pixelated matrices, made by several single optically insulated pixels glued together, are available.

The dimension of a single pixel in a block has to deal with the efficiency for 511 keV photons detection; hence it has to be longer than the attenuation length of these photons in the material. Furthermore, its section has to guarantee a good charge collection across the whole block. In fact, smaller crystals would provide a better spatial resolution, but their light output would be so low that the signal to noise ratio, due to statistical fluctuation in charge collection, would decrease. As an example, it was shown (Shao et al. 2005) that the light output from a LSO $1 \times 1 \times (10) \text{ mm}^3$ pixel is about 75% of the one from a $2 \times 2 \times (10) \text{ mm}^3$ pixel of the same material.

b) Photon detectors

To achieve high imaging performances in a PET system the photon-detector should present characteristics such as: high efficiency, fast response, short recovery time, low noise. These properties are so important in the performances of a PET system that a lot of research is done to improve them.

The first and most common used photon-detector is the photomultiplier tube (PMT). It consists in a vacuum tube with a light transmitting window covered on the inner side by a semitransparent photocathode, made by a thin layer of a semiconductor material. The detection process begins when the light passes through the input window and incidents the photocathode. Electrons in the photocathode are excited and emitted as photoelectrons into the vacuum tube, the phenomenon governing this process being the internal photoelectric effect. They are then accelerated by a strong electric field inside the device and a multiplication occurs by hitting

a series of electrodes called dynodes, placed inside of the tube. The multiplied secondary electrons emitted by the final dynode are collected by the anode, which output the electron current to the external circuit. These devices have a typical quantum efficiency of 25%, also due to the optical properties of the light transmitting window. A gain of up to 1×10^6 is achieved with a quite high bias ($\sim 1500\text{V}$). The PMT's are stable and no noise, but geometrically limited. They are rather bulky, with an active area in the range of 10-500 mm. Their shape is in particular determined by the technical requirements in the fabrication of the vacuum tubes.

A common and widely used phototube from Hamamatsu is the R1450 (19 mm diameter, 10 dynodes stages assuring a 1.7×10^6 gain, bialkali photocathode, 1500V bias voltage). In PET systems, the usual configuration of scintillator block readout with PMT consists in coupling each block with 4 PMT's (Fig. 5.5). A typical PET detector module, based on PMT couplet to BGO crystals has 5 mm FWHM spatial resolution, $4\mu\text{s}$ dead time, 2 ns FWHM timing resolution, 20% FWHM energy resolution (Moses et al. 1994).

The typical intrinsic spatial resolution of PET camera based on PMTs is 3-10 mm FWHM. This is limited by the scintillator PMT block design (PMT blocks have intrinsic problem of coupling with scintillators) and the low quantum efficiency of PMTs. To achieve a higher spatial resolution, a higher segmentation of scintillators is required with the ability to readout each segment independently.

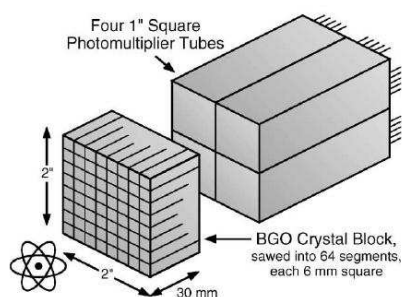


Fig. 5.5 Conventional PET detector module: scintillation light from gamma ray interactions in the crystal block is detected by multiple PMT's

Although no commercial clinical PET camera up to date uses semiconductor radiation detectors, these devices are promising as a mean to overcome the drawback of PMT-based PET instrumentation. Solid state photodiodes, in contrast to PMTs, are small, operate at much smaller voltage, and exhibit higher quantum efficiencies. This allows the possibility of easy read-out patterns of the sensitive area, especially through the introduction of monolithic pixelated photo-detectors, permitting advanced PET detector modules packaging.

A prototype PET system based on 64 LSO scintillator crystals matrix ($3 \times 3 \times 20$ mm) coupled on one end to a single PMT and on the opposite end to a 64 array of 3 mm^2 silicon PN photodiodes (typical parameters are 5 pF capacitance, 300 pA dark current, and 73% quantum efficiency at 415 nm) was developed in Berkley (Huber & al. 1997). The timing and the initial energy information are provided by the PMT and the photodiodes identifies the crystal of interaction. The PMT and photodiodes signals are combined to measure the depth of interaction and total energy. The system showed 24% energy resolution, 1 ns FWHM timing resolution, 8 mm DOI determination uncertainty. As the photodiodes do not provide internal gain, a readout charge amplifier has to be used, with a consequent increase of the noise of the system and decrease in timing resolution, as the short time response of the silicon device is completely dominated by the timing characteristics of the amplifier.

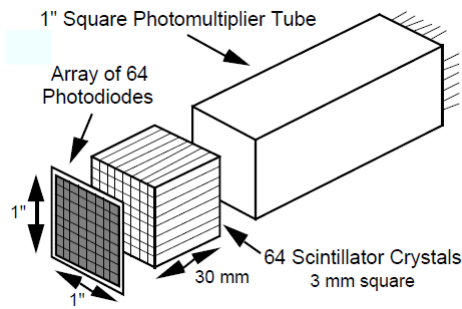


Fig. 5.6 Diagram of a prototype PET module, developed at Berkley, consisting of 64 LSO scintillator crystals coupled to one end to a single PMT and to the opposite end to an array of 64 silicon photodiodes (Huber & al. 1997)

As high signal to noise ratio is needed for a good statistics in image reconstruction, photo-detectors with internal gain are preferred. A small animal PET system has been developed in Munich, called MADPET II, totally based on APD's (McElroy et al. 2005). It consists on a ring of dual layer detector modules (71 mm diameter), each containing a 4x8 array of $2 \times 2 \times 6 \text{ mm}^3$ (front) and $2 \times 2 \times 8 \text{ mm}^3$ (back) LSO crystals that are each optically isolated and coupled one-to-one to a monolithic 4x8 APD array. It shows 12% FWHM energy resolution and 10.2 ns FWHM time resolution for the overall system, and 4 ns for individual LOR.

Tavernier et al., 2005, has designed other interesting application of APD's in PET, coupling them with a continuous LSO crystal. In this way it can be possible to avoid the parallax error and sensitivity reduction caused by eventual smaller crystals. Different reconstruction methods were tested and the FWHM average resolution was shown to go down to 1.86 mm for perpendicular gammas incidence. On the other side, since a quite high bias is needed and the gain is strongly dependent on it, stability problems on bias supplies can limit the performances of a complete PET system.

SiPMs present improved properties with respect to vacuum PMT's and other solid-state photo-detectors. As we have learned in the previous sections, they have high intrinsic gain (10^5 - 10^6), single photoelectron sensitivity, and excellent S/N ratio, making unnecessary the use of sophisticated low noise front-end electronics. They are extremely fast devices with a timing resolution of the order of tens of picoseconds, as well as insensitive to magnetic fields. Two others interesting characteristics are their low operating voltage ($<100\text{V}$), that results in reduced power consumption as well as the possibility to arrange them in arrays, that provides position information. In addition, they are compact, robust and their cost is reasonable, in continuous decrease. Therefore, they are presenting very "appetizing" characteristics to be used in PET systems. The work performed in collaboration with Pisa group for evaluating the SiPM arrays performances as photo-detectors for PET application will be presented in the following.

5.1.4. Could SiPM's array be implemented as photon detector for PET systems?

Even though SiPM's have showed their suitability and value for PET applications, certain aspects still need to be taken into account when addressing their system integration. In particular, the uniformity of characteristics such as breakdown voltage, gain and detection efficiency is crucial. If such characteristics are not uniform within different SiPM's used in a PET system, the position of the photo-peak may shift from device to device and the image resolution will degrade. For this reason, before trying to use arrays of SiPM's for PET system, a characterization of their uniformity is required.

In the framework of the LAL, Pisa-University and FBK collaboration, we obtained a 2D prototype array consisting of 4x4 SiPM's on the same substrate fabricated at FBK-irst, as presented in Fig. 5.7.

Each SiPM device, representing one read-out element of the array, has an area of $1 \times 1 \text{ mm}^2$ and it is composed of 625 microcells connected in parallel. Each microcell of the SiPM has an area of $40 \times 40 \mu\text{m}^2$ and it is represented by a GM-APD (n+/p junction on p+ substrate) connected in series with its integrated poly-silicon resistance R_D . The total fill factor of the array is $\sim 30\text{--}35\%$, calculated as a product of the array fill factor ($\sim 92\%$) and the SiPM fill factor ($\sim 35\text{--}40\%$). The signals of each SiPM are routed to the bonding pads of two opposite sides of the array, 8 on each side (Fig. xxx). The array has been wire bonded to an external PCB allowing its bias as well as the connection of each channel to the read-out electronics (Fig. 5.8)

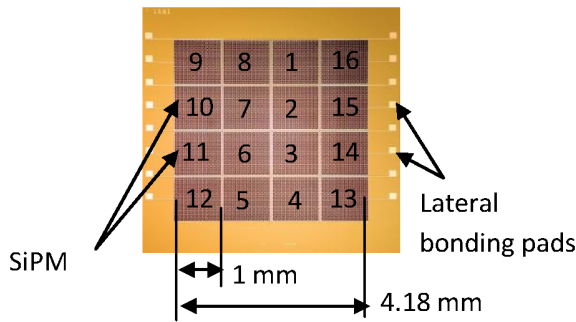


Fig. 5.7 Layout of the 2D prototype array of 4x4 SiPM's on the same substrate from FBK-irst

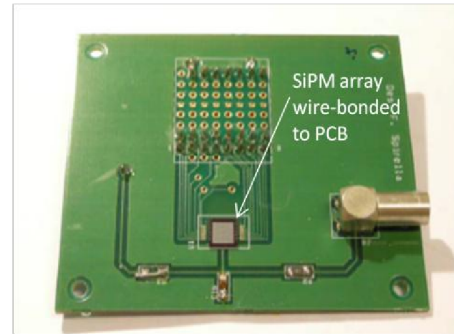


Fig. 5.8 2D prototype array of 4x4 SiPM's wire-bonded to a PCB

Static and dynamic electrical characterizations have been performed at LAL for each element of the array. The photon detection efficiency uniformity has been tested as well (Dinu et al., 2009 a, b). The results will be shown in the following.

a) Static characterization

The electrical measurement set-up used for static tests consists on the connection of each SiPM of the array to Keithley 2612 source-meter (Fig. 5.9).

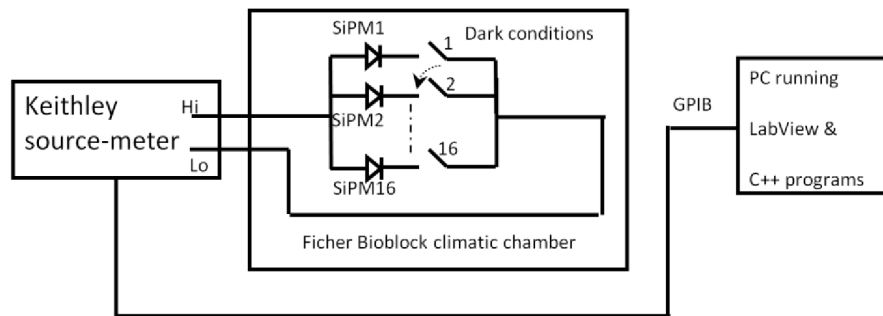


Fig. 5.9 Experimental set-up used for the static characterization of each SiPM of the 2D array

The static characteristics (reverse and forward IV plots) of the 16 SiPM's of the array are presented in Figs. 5.10 (a) and (b). From the reverse IV plots, a mean value of $V_{BD} \sim 30.5 \text{ V}$ has been determined, with a good uniformity over the 16 elements (RMS $\sim 0.15 \text{ V}$). The analysis of the $I_{post-BD}$ (see Section 3.2.1. and Annexe 1) has been also performed, showing that $I_{post-BD}$ lies in between two parabolic trends corresponding to the elements

SiPM13 (lowest) and SiPM16 (highest). The mean value of the $I_{post-BD}$ has been calculated, showing a value of $\sim 0.5\mu A$ at $V_{bias}=33V$ (the overvoltage $\Delta V = V_{BIAS} - V_{BD} = 2.5V$) with an RMS of $\sim 0.1\mu A$ ($\sim 12\%$). From the forward IV plots, a mean value of R_{qSiPM} of $\sim 520\Omega$ has been found over the 16 elements of the array, with a very good uniformity (RMS $\sim 5\Omega$). Considering that this represents the equivalent resistance of 625 micro-cells connected in parallel, a mean value of $R_{qcell} \sim 330k\Omega$ has been calculated.

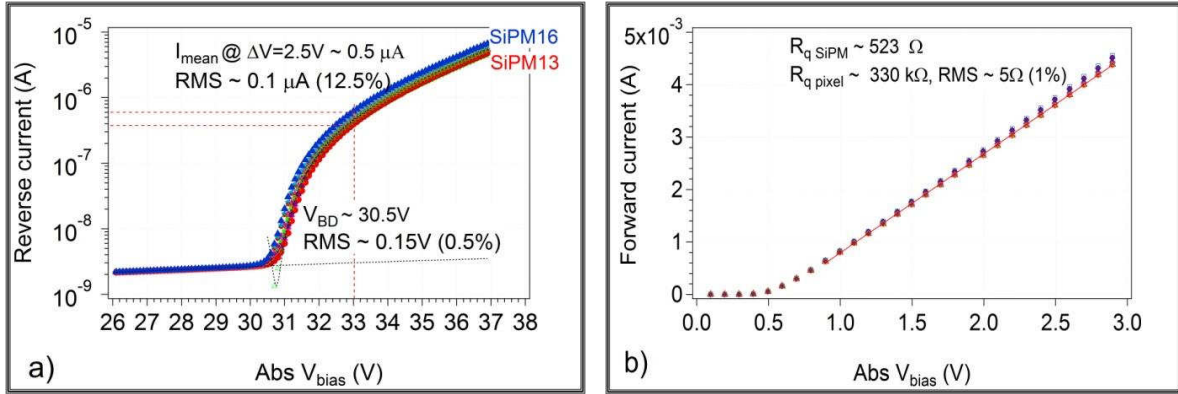


Fig. 5.10 (a) Reverse and (b) forward IV characteristics of the 16 SiPM elements of the 2D array

b) Dynamic characterization

Based on the observation that a signal triggered by an absorbed photon and by a thermally generated carrier are identical, the dynamic tests of different SiPM's of the 2D array have been performed in dark conditions. A controlled temperature of both static and dynamic tests ($25 \pm 0.5^\circ C$) have been assured by Fisher Bioblock LMS A49401 climatic chamber.

The dynamic characterization has been performed with two set-ups configurations.

The first one (Fig. 5.11) consists on one-by-one connection of each SiPM of the array to a MITEQ AU-1332 single-channel wide-band (0.01–500MHz) voltage amplifier read-out by a Tektronix TDS 5054 (500MHz, 5 GS/s) digital oscilloscope. The amplifier presents 50Ω input impedance, which acts as a current-to-voltage converter, followed by an amplifying stage having a gain of 45 dB. LabView and C++ programs have been used for the automatic data acquisition and analysis of the SiPM parameters.

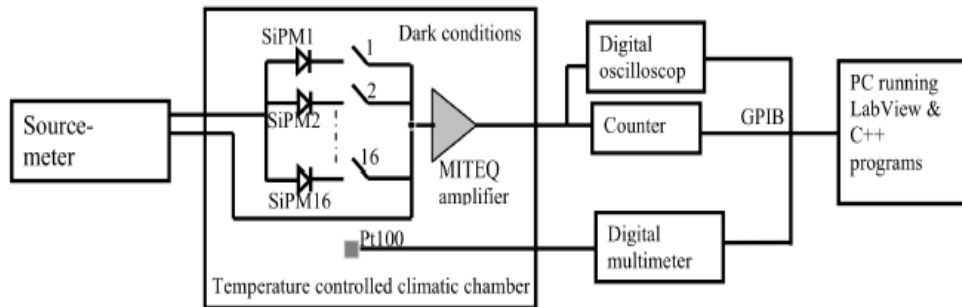


Fig. 5.11 Experimental set-up used for the one-by-one dynamic characterization of the SiPMs from the 2D array

Using the presented set-up, the oscilloscope analysis of any SiPM of the array in dark dynamic conditions, revealed different signals (Fig. 5.12) determined by different physical phenomena, as explained in the Section 4.2.2.c :

- signals coming from single micro-cells (e.g. primary pulses); they are generated by thermal carriers and represent the main source of the SiPM noise;
- signals following the primary pulses (e.g. afterpulses); they represent carriers trapped during one avalanche and when they released (during tens of ns), they trigger a new avalanche;
- signals coming from one or more neighbors micro-cells overlaid with the primary pulses (e.g. optical cross-talk); such signals are generated by photons created during the primary avalanche which can trigger avalanches in the adjacent pixels.

Counting all types of signals presented in Fig. 5.12, the dark count rate (DCR) has been determined. The DCR as a function of the V_{bias} is represented in Fig. 5.13 for the SiPM13 (having the lowest $I_{post-BD}$) and the SiPM16 (having the highest $I_{post-BD}$). The DCR of all the others SiPM's in the array are also represented, at V_{bias} of 33V.

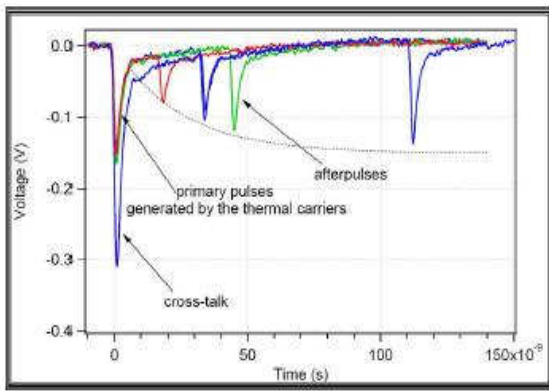


Fig. 5.12 Dark signals shapes for SiPM's in the 2D array

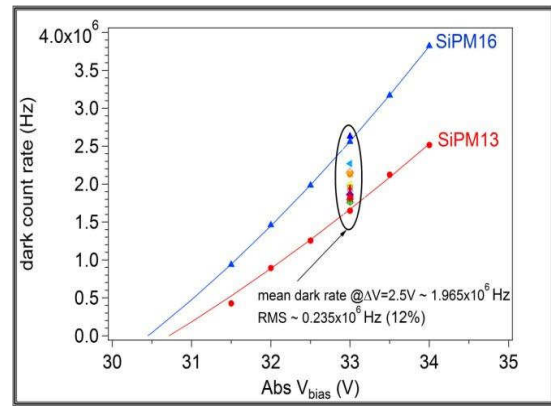


Fig. 5.13 Dark count rate vs bias voltage for different SiPM's of the 2D array

The DCR shows a direct correlation with the $I_{post-BD}$: the element SiPM13 having the lowest $I_{post-BD}$ has the lowest DCR while the element SiPM16 having the highest $I_{post-BD}$ has the highest DCR. The remaining 14 SiPM's show a DCR in between the elements SiPM13 and SiPM16, with a mean value ~ 2 MHz @ $V_{bias}=33V$ and $RMS=0.235 \times 10^6$ Hz ($\sim 12\%$, the same dispersion as for the $I_{post-BD}$). This confirm, as presented in the Section 3.2.1., that $I_{post-BD}$ of a given SiPM represents the average current of all dark pulses contributions (e.g. thermal generated carriers, afterpulses, cross-talk) coming from different micro-cells (see Annexe 1 for more details).

For each SiPM of the array, the charge distribution of dark signals has been obtained by integrating the dark SiPM output signals. An example of such distribution is presented in Fig. 5.14. The centroid of the first Gaussian fit represents the most probable charge generated during the Geiger avalanche in one SiPM micro-cell and it allows determining the SiPM gain as this charge divided by the electron charge. The gain measured using the described method is represented in Fig.5.15 as a function of the V_{bias} for the SiPM13 and the SiPM16. A linear increase of the G with V_{bias} has been observed for both elements. The G has been also evaluated for the other 14 elements of the matrix at V_{bias} of 33V. A mean value of $\sim 1.5 \times 10^6$ @ 33V has been determined, with a good uniformity ($RMS=0.057 \times 10^6 \sim 4\%$).

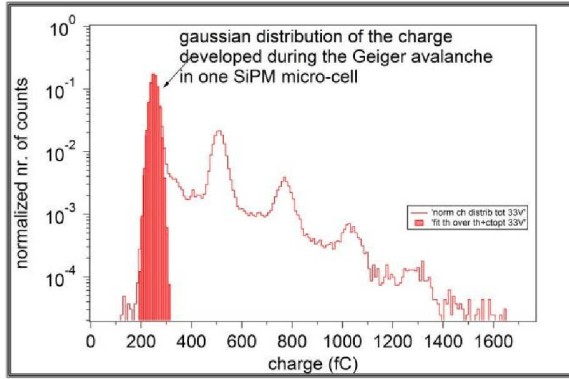


Fig. 5.14 Example of charge distribution of dark signals for one SiPM of the array

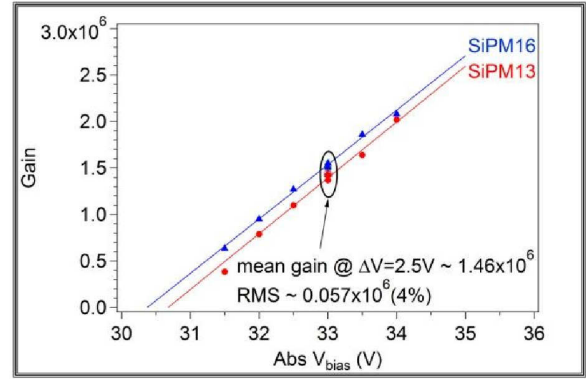


Fig. 5.15 Gain vs bias voltage for different SiPM's of the 2D array evaluated with the set-up employing the MITEQ amplifier

Since our final aim is the use of the SiPM's arrays in building PET system, our primary experimental set-up allowing one-by-one measurement of the SiPM's is very useful in laboratory tests but it cannot be implemented in a PET system. For a simultaneously measurement of different SiPM's, a multichannels read-out electronics is necessary. Therefore we developed at LAL a second readout set-up (Figs. 5.16 (a) and (b)), by implementing a simultaneously parallel connection of the 16 SiPM's of the array to the multi-channels read-out chip called MAROC2 (Barrillon et al. 2007). MAROC2 is a 64-channels ASIC based on AMS Si-Ge 0.35 μ m technology. Each channel features a low noise, variable gain preamplifier (6bits), and two (slow and fast) shapers. The slow shaper has an adjustable shaping time ranging from 50 to 150ns. The fast ($\tau=15$ ns) shaper (unipolar and bipolar) is followed by three discriminators. The main application of MAROC2 is the readout of the Hamamatsu H7546 multi-anode photomultiplier tubes used in the ATLAS luminometer (ATLAS collaboration, 2007) and therefore it is not optimized for the SiPM readout. In spite of some limitations, the MAROC2 ASIC allowed us to readout the SiPM array as well as to perform the first tests of SiPM array coupled to LYSO crystals as employed in PET scanners (more details will be presented in the next Section).

A board designed at LAL for testing the ASIC was utilized for data acquisition. The board hosts an ALTERA FPGA that controls the data acquisition and provides the OR signal of the trigger outputs of all channels in the ASIC. The digitized output signal is provided by a 12-bit ADC located on the board. A sample and hold signal triggered by the OR signal is sent to the ADC, synchronized with the maximum value of the shaped signal.

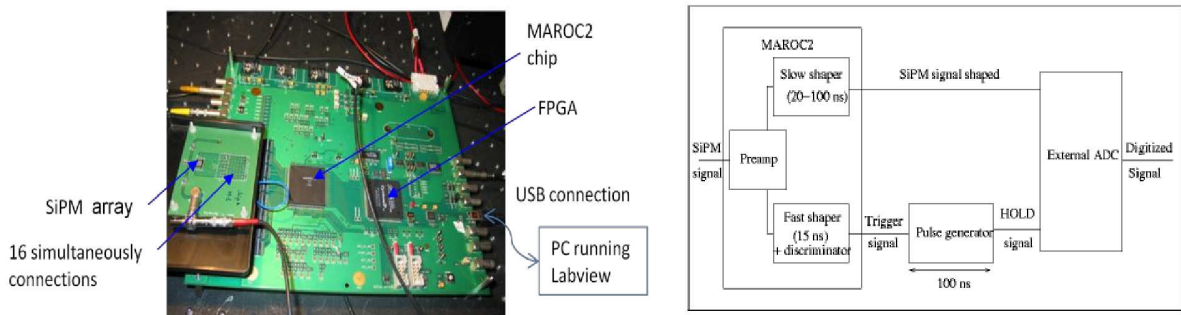


Fig. 5.16 (a) Picture of the experimental set-up used for the simultaneously dynamic characterization of the SiPMs from the 2D array; **(b)** Diagram of the SiPM's signals acquisition system implementing the MAROC2 ASIC and the external ADC

The PCB board hosting the MAROC2 ASIC allows us to perform the ASIC calibration. A calibration pulse is sent through a specific input with a variable capacitor to the channel selected by means of switches, and the digital charge output is provided by an on-board 12 bit ADC Wilkinson. Measurements have been performed for different injected charges as well as for different preamplifier gains. The response of one ASIC channel is presented in Fig. 5.17. For the highest preamplifier gain, a saturation at about 3-4 pC is observed. However, for the two lowest gains, the ASIC has a linear and uniform response up to 8-10 pC (RMS=0.9% at 7pC); for higher values of the charge, the response is still very close to the linearity, but a change in the slope is observed. For values above 60 pC (Fig. 5.18, for all channels) the response clearly saturates, and the uniformity between channels get worse (RMS=1.5% at 94 pC).

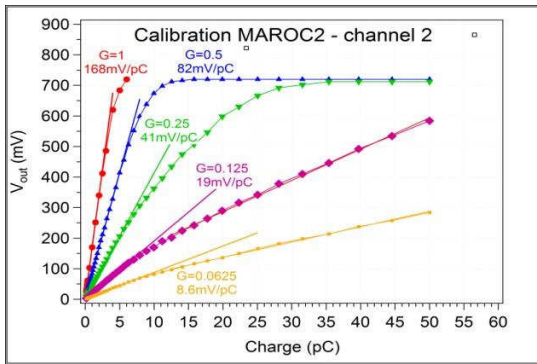


Fig. 5.17 Response of one MAROC2 channel for different injected charges and various preamplifier gains

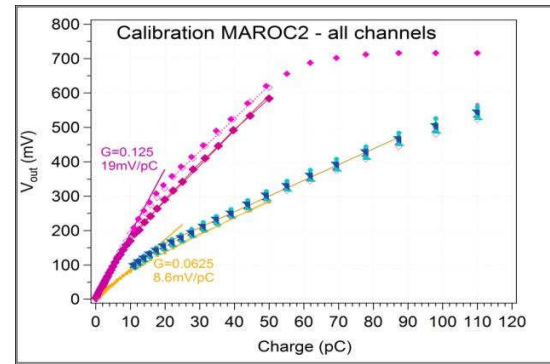


Fig. 5.18 Response of different MAROC2 channels for two lowest possible preamplifier gains

The signals waveforms of each SiPM's from the 2D array at $V_{BIAS} = 33V$ have been recorded using a preamplifier gain of 19mV/pC (Fig. 5.19). Following the calibration curve, the gain of different SiPM's in the array has been determined (Fig. 5.20). The mean value of the gain is $\sim 1.3 \times 10^6$, showing a good agreement in between the G values measured with both MAROC2 set-up as well as with the MITEQ one. The low dispersion ($\sigma \sim 5.5\%$) demonstrates not only the gain uniformity through different SiPM's of the array but also the uniformity of the channels through the MAROC2 ASIC.

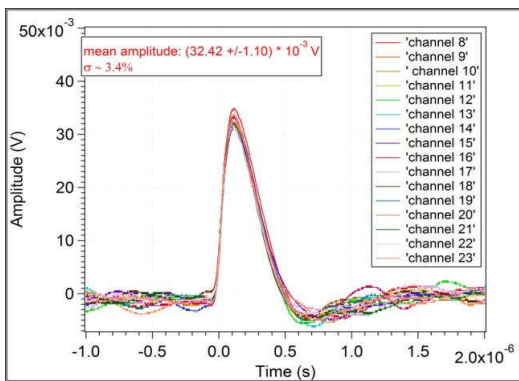


Fig. 5.19 Waveforms of different SiPM's of the 2D array recorded after the slow shaper of the MAROC2 ASIC

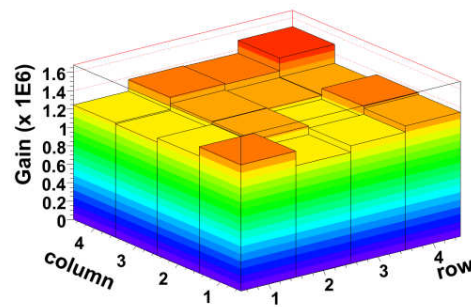


Fig. 5.20 Gain @ $V_{BIAS}=33V$ of 16 SiPM's of the array using the set-up employing the MAROC2 ASIC

c) Optical characterization

The optical test has been performed using a halogen continuous light source (100 W) followed by a LOT ORIEL 77250 grating monochromator which separates the visible wavelengths (350– 800 nm) with an accuracy of 2 nm (Fig. 5.21). The incident flux, evaluated by two calibrated photodiodes (HPK S3590-18, UDT Instrument 221) was kept low ($\sim 10^7$ /incident photons/s/mm²) to avoid the SiPM and the acquisition system saturation. A precise positioning of the SiPM array and the calibrated diode in the light beam was assured by Polytec PI M413 tri-dimensional translation stage. The optical test has been performed at room temperature ($25 \pm 1^\circ\text{C}$).

The optical characteristics of each SiPM in the array have been evaluated using two data acquisition systems: the DC set-up described in Fig. 5.9 (photocurrent technique) and the AC set-up involving the MITEQ amplifier described in Fig. 5.11 (photon counting technique). The two methods for PDE calculations have been described in details in the Section 4.2.3.

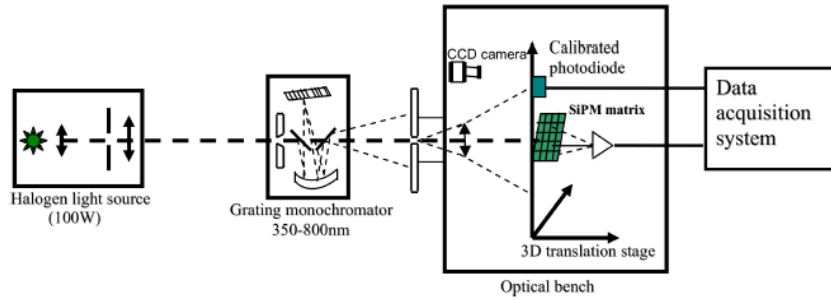


Fig. 5.21 Experimental optical set-up used for the optical characterization of the 2D SiPM's array

The PDE as a function of wavelength is represented in Fig. 5.22 for the SiPM13 and the SiPM16 a $V_{\text{BIAS}}=33$ V and 34V. The PDE has been also evaluated for the others 14 elements of the matrix at $V_{\text{BIAS}}=33$ V and $\lambda =580$ nm. A PDE value of ~ 8 – 10% @ $V_{\text{BIAS}}=34$ V has been found for a wide range of wavelengths (420–680 nm) as well as a very good uniformity over the elements of the array has been observed (RMS $\sim 5\%$).

The PDE of the SiPM13 and the SiPM16 has been evaluated as a function of V_{BIAS} for a given wavelength $\lambda =580$ nm (Fig. 5.23). We can observe that PDE increases quite linearly with V_{BIAS} . At the same wavelength and a given $V_{\text{BIAS}}=33$ V, the PDE has been measured for all SiPM's of the array. A good uniformity is observed (PDE_{mean}@33V $\sim 7.4\%$, RMS=5%).

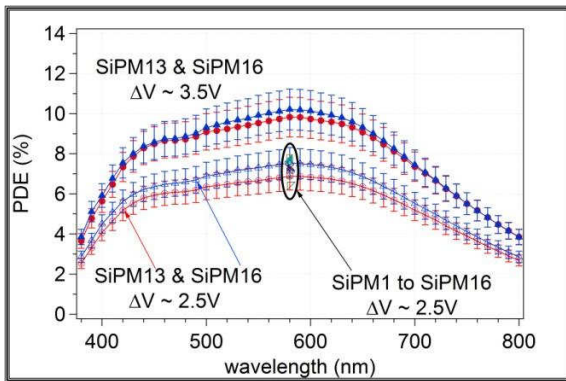


Fig. 5.22 PDE of SiPM13 & SiPM16 vs. λ @ $V_{\text{BIAS}}=33$ V and 34V; PDE of all SiPM's of the array @ $\lambda=580$ nm & $V_{\text{bias}}=33$ V

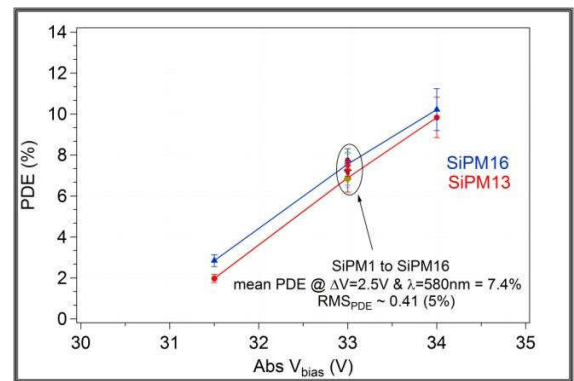


Fig. 5.23 PDE of SiPM13 & SiPM16 vs. V_{BIAS} @ $\lambda=580$ nm; PDE of all SiPM's of the array @ $\lambda=580$ nm & $V_{\text{bias}}=33$ V

Following the presented tests, the most important characteristics of the SiPM's of the 2D array such as breakdown voltage, gain and photon detection efficiency showed a very good uniformity (dispersion less than 5%). The only parameter showing higher dispersion (~12%) is the dark count rate, but it presents less importance in the case of using SiPM's for a PET system (i.e. the scintillator light signal is high enough to exceed the threshold of dark pulse signals). Therefore, the full characterization of the 2D SiPM's array allow us to have good confidence in the characteristics of these devices and to continue their characterization using scintillator crystals coupled to SiPM's array as required in a PET system. In the framework of LAL/FBK-irst/Pisa University, these tests have been performed by our collaborators in Pisa (Llosa et al., 2008 a, b and 2009 a, b) and selected results will be presented in the following Section.

5.1.5. Evaluation of SiPM's array for small animal PET scanner

Radioisotope imaging of small animals, such as mice and rats, using high-performance PET scanners, has become a valuable tool for studying animal models of human disease (Paigen, 1995; Chien, 1996; Phelps, 2000). The mouse has long been used by molecular biologists to study fundamental cellular events *in vivo*, but the relatively small size of the mouse (weight 20-30 g) makes it difficult to use imaging instruments developed for human subjects. In this context, dedicated rodent nuclear imaging techniques as the one developed at the Department of Physics of the Pisa University (del Guerra, et al. 2006) provide a powerful non-invasive method to follow the temporal kinetics of pharmaceuticals in the same animal, without the need to sacrifice, dissect, and count, as in traditional pharmacokinetic and bio-distribution studies.

A new design of a small animal PET tomograph is proposed by the Pisa group, employing LSO scintillator readout by SiPM's arrays (Fig. 5.24). The tomograph consists of four 4 cm x 4 cm detector heads that rotate around the object to be imaged. The opposed heads are separated either by 10 or 15 cm. A head is composed of three layers, each one made of a continuous LSO slab read out by a SiPM array. The use of monolithic scintillators slabs plus pixellated SiPM array can provide better performance than pixellated scintillator blocks: it allows very good spatial resolution with maximum likelihood algorithms as well as high efficiency (no light losses due to small pixel size). Additionally, the cost of the detector is reduced. The use of SiPMs allows stacking several detector layers, ensuring high detection efficiency, while providing depth-of-interaction (DOI) information that helps to reduce the parallax error. The tomograph will have a FOV of 4 cm axial, 4 cm transaxial. A spatial resolution better than 1 mm FWHM is expected, according to GEANT4 simulations.

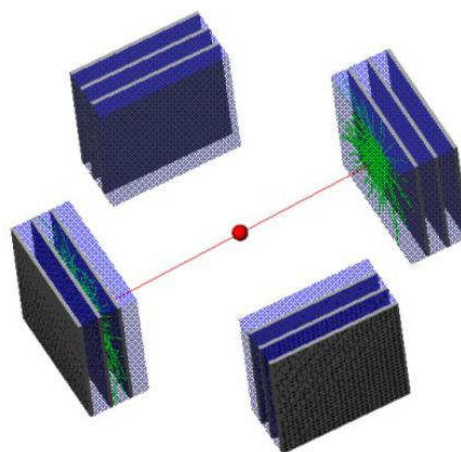


Fig. 5.24 Design of a small animal PET tomograph proposed by the group from Pisa University. It consists of four 4cm x 4cm detector head that rotate around the object. The detector heads are composed of three layers, each one consisting of a 5 mm thick LSO slab readout by a SiPM array.

The SiPM 2D array produced by FBK-irst and readout by MAROC2 ASIC (see description in the previous Section 5.1.4. b), have been used by Pisa group to perform specific tests required for small animal PET application: energy and timing resolution as well as position determination.

Tests have been carried out with both pixellated and continuous crystals (Figs. 5.25 (a) and (b)). The crystal array is coupled one-to-one to the SiPM array. It consists of 16 LYSO crystals of 0.96 mm x 0.96 mm x 10 mm, with 100 μ m of white epoxy resin between them (Fig. 5.25 (a)). The continuous crystals are blocks of 4 mm x 4 mm x 5 mm that cover all the SiPM's of the 2D array. Two samples have been used in the measurements shown here. The so-called *white slab* is wrapped in white Teflon tape and it has been utilized for the energy resolution tests. The *black slab* is painted black in five faces, and it has been used in the position determination tests (Fig. 5.25 (b)).

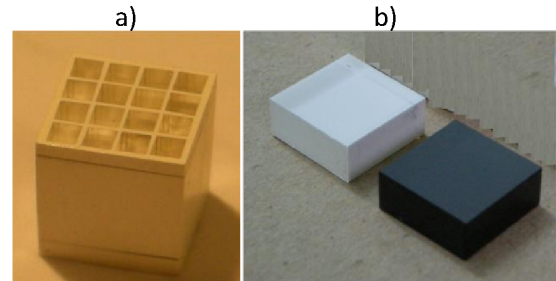


Fig. 5.25 (a) Crystal array composed of 16 LYSO crystals of 0.96 x 0.96 x 10 mm, with 100 μ m of white epoxy resin between them; **(b)** white and black slab continuous crystals of 4x4x5 mm

In order to have a reference for the results obtained with the SiPM array, tests have also been carried out with single SiPMs and crystals of similar dimensions in each case, to obtain similar conditions:

- a 1 mm x 1 mm single SiPM coupled to a 1 mm x 1 mm x 10 mm LYSO crystal is tested as a reference for the 16 SiPM array, coupled to the 16 LYSO crystal array;
- a 4 mm x 4 mm single SiPM coupled to the 4 mm x 4 mm x 5 mm *white slab* has also been studied for comparison with the SiPM array coupled to the same crystal. In this case, the microcell size is slightly different (40 μ m x 40 μ m for the pixels in the matrix, 50 μ m x 50 μ m in the single SiPM).

a) Set-up description

The readout of the single SiPM has been carried out either with a QDC CAEN V972 or directly with an oscilloscope LeCroy LC684DM, histogramming the signal area that is proportional to the charge. In the latter case, the signal is split with a fan-in-fan-out (FIFO) NIM module CAEN N401. One signal is sent directly to the QDC, while the other goes through a linear discriminator to generate the integration gate needed by the QDC. A picoammeter/voltage source Keithley 6487 is employed for biasing the SiPM photodetectors.

The readout of the SiPM array has been carried out with the MAROC2 ASIC, as presented in the Section 5.1.4. b. As mentioned previously, the MAROC2 is a 64 channel ASIC originally designed for multi-anode PMTs and it was not been optimized to readout the SiPMs coupled to scintillator crystals (i.e. the dynamic range is not large enough). In spite of that, it has allowed to make the first tests of SiPM array presented here satisfactorily.

For the coincidence timing measurements, the trigger signal from the SiPM array generated by the MAROC2 ASIC is acquired in coincidence with a single SiPM coupled to a 1 mm x 1 mm x 10 mm LYSO crystal (Fig. 5.26). For an improved measurement of the timing resolution the single SiPM signal is amplified with a custom made board with a Gali5 linear amplifier, and split with the FIFO (Llosa et al., 2008, b):

- one of the signals is sent to a high threshold discriminator that selects only the total absorption (511 keV photopeak of ^{22}Na) events to minimize the time-walk. The photopeak signals of the single SiPM are sent to a coincidence unit CAEN N405, to select the events in coincidence with the trigger signal of the array.

- the second signal of the SiPM is sent to a low threshold level discriminator in order to discriminate the signals at the lowest possible level.

For the SiPM array, the signal given by MAROC2 board is split to generate both the coincidence signal and the one employed to measure the time difference. In this case only one trigger level, either high or low, can be set. The signals from the two detectors are sent to the oscilloscope to histogram the time difference between them. The oscilloscope is triggered by the coincidence signal previously generated.

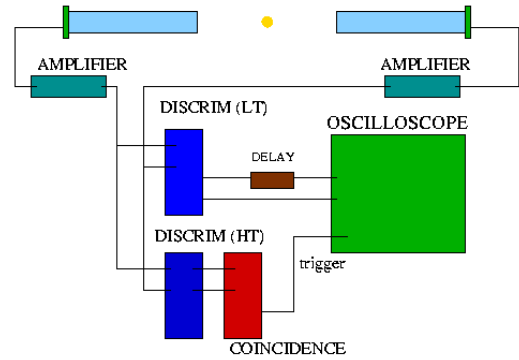


Fig. 5.26 Schematic drawing of the coincidence set-up employed for the timing measurements

A dedicated setup (Fig. 5.27 (a)) has been mounted for the position determination measurements with the SiPM array, which allows us to determine the interaction position of the photons. The board with the SiPM array attached is placed vertically, and operated in coincidence with a single SiPM coupled to a 1 mm x 1 mm x 10 mm crystal. The single SiPM is fixed to a set of three micropositioners in a 3D arrangement. A holder for the source keeps it attached to this second detector at a distance of 35 mm and at about 5 mm from the SiPM array (Fig. 5.27 (b)). The source and the second detector are moved together in 0.5 mm step. The time coincidence of the signals from both detectors is acquired, restricting the detected photons to those that hit the array in a precise position. The micropositioners allow us to make a 2D scan of the array surface.

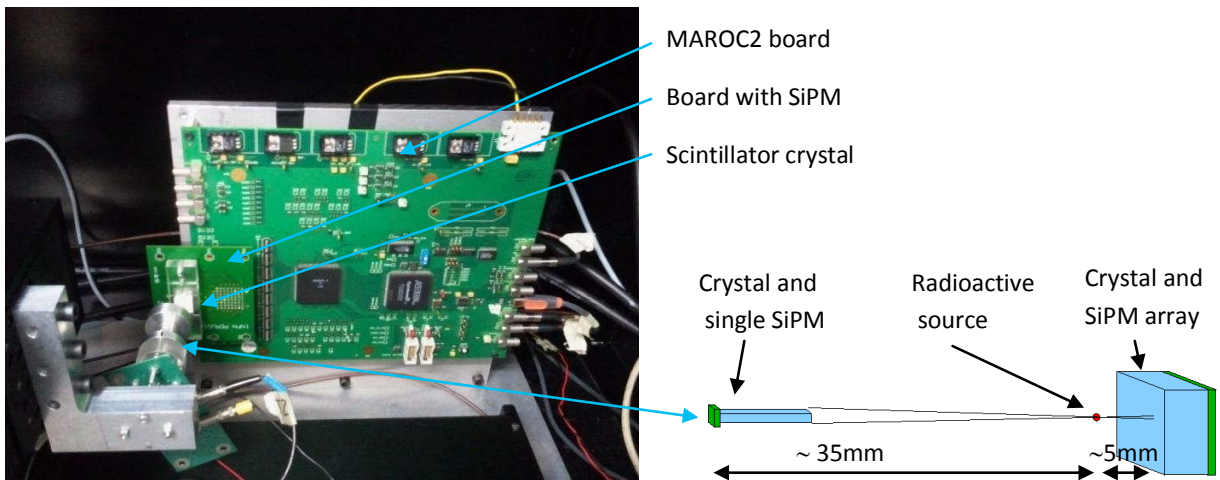


Fig. 5.27 Dedicated set-up for position determination measurements: (a) MAROC2 board with SiPM array attached is placed vertically and operated in coincidence with a single SiPM ; (b) schematics set-up for position determination measurements (pictures by courtesy of G. Llosá)

b) Results with pixellated crystals

Energy spectra

As a reference, tests have been carried out with a LYSO crystal of 1 mm x 1 mm x 10 mm coupled to a single SiPM of 1 mm x 1 mm of the same type as those that compose the array. Energy spectra of ^{57}Co (122.1 keV), ^{22}Na (511.0 keV and 1274.5 keV) and ^{137}Cs (661.7 keV) have been acquired with the QDC. The position of the photopeak in each case has been plotted versus the gamma ray energy (Fig. 5.28). The results show that the SiPM follows the theoretical curve, deviating from linearity when gamma energies exceeding 600 keV (i.e. high number of photons arriving on the SiPM surface as compared to the total number of microcells). For a precise characterization of the SiPM response, the non-linearity of LYSO, that is relevant at energies below 100 keV (Pidol & al., 2004), should be considered in the measurements and corrected for. When the same measurements were carried out coupling the same LYSO crystal to a PMT, the response is very close to a straight line (Fig. 5.29), showing that the deviation from linearity observed for the SiPM is due to the SiPM saturation. We have to keep in mind that the purpose of these tests was not to make a precise study, but to understand the response of single device, which will be comparable with the response of SiPM's from the array.

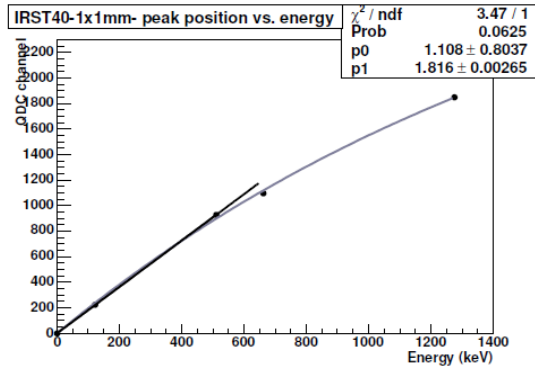


Fig. 5.28 Response curve of the SiPM FBK-irst 1 mm x 1 mm area and 40 μm microcell size coupled to a 1 mm x 1 mm x 10 mm LYSO crystal vs. energy. The SiPM saturates for a high number of photons and its response deviates from linearity.

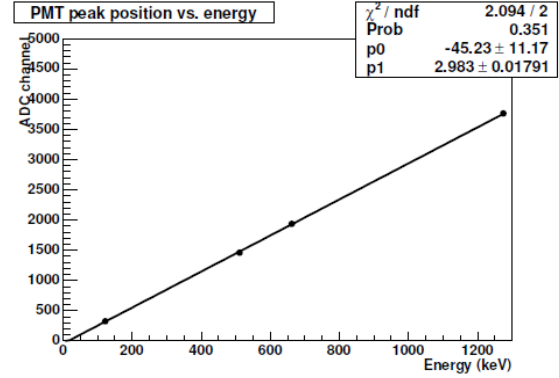


Fig. 5.29 Response of a PMT coupled to a 1 mm x 1 mm x 10 mm LYSO crystal vs. energy. In spite of the LYSO non-linearity at low energies, the curve is close to linearity.

Once the response of the single SiPM is understood, the crystal array has been coupled to the SiPM array, and data with ^{22}Na have been taken, requiring the coincidence with the single SiPM previously used. The bias voltage of the SiPM array was limited to an overvoltage of 2 V to avoid the saturation of the ASIC (i.e. to obtain a highest possible dynamic range, the MAROC2 ASIC preamplifier gain has been set at the lowest value, $G=0.625$; see the MAROC2 calibration in the Section 5.1.4.b). In order to obtain a precise value of the energy resolution, the deviations from linearity due to both the ASIC and the SiPM response must be corrected.

However, the relatively short dynamic range of the MAROC2 ASIC does not allow us to carry out the measurements with different sources, and therefore it is not required to make the necessary corrections.

Fig. 5.30 shows a typical ^{22}Na energy spectrum corresponding to one channel of the SiPM array, obtained with the MAROC2 ASIC. We can observe that 511 keV photopeak from ^{22}Na is clearly separate from the Compton continuum. The variations in the peak position for all measured channels are 4.4%, including variations in the gain and PDE of the SiPMs in the array, and of the electronics response.

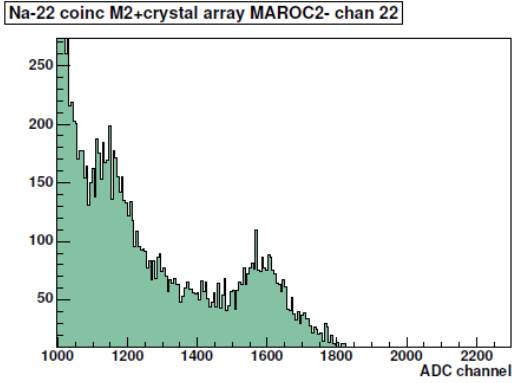


Fig. 5.30 ^{22}Na energy spectrum from one channel of the SiPM array coupled to the crystal array, employing MAROC2 ASIC for the read-out

Timing resolution

The coincidence timing resolution has also been measured, operating the SiPM array and the crystal array in coincidence with a single SiPM coupled to a 1 mm x 1 mm x 10 mm crystal. Measurements have been made by setting the MAROC2 discriminator high, to consider only the photopeak events, and low, to acquire also the Compton events. In addition, tests have been made with just one channel, setting to zero the preamplifier gain for all other channels, and with all channels enabled.

The results are summarized in Table 5.2. The measurements corresponding to one channel at low and high threshold are shown in Fig. 5.32 (a) and (b) respectively. In the case of low threshold we can see the tail due to the time-walk. When all channels are active, the results worsen with respect to the ones obtained with only one channel by about 35%.

	timing FWHM (ns)	
	one channel	all channels
low threshold	2.26	3.29
high threshold	2.10	3.01

Table 5.2 Timing resolution of the SiPM array coupled to the crystal array with the MAROC2 readout system

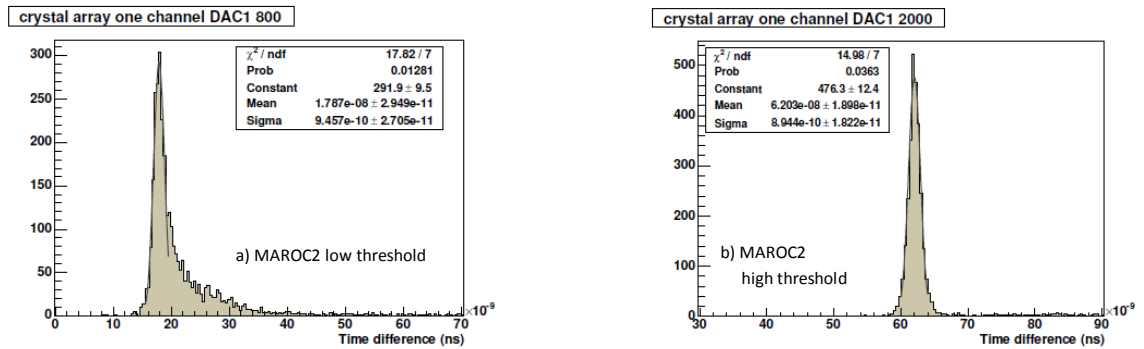


Fig. 5.32 Timing resolution distributions of one channel of the SiPM array coupled to the crystal array in coincidence with a single SiPM, setting the threshold level of the MAROC2 readout system low **(a)** and high **(b)**

Position determination

The interaction position in the case of pixellated crystals is given by the channel that triggers. Using the setup described previously for the determination of the interaction position, a linear scan across the device have been made, moving the source in 0.5 mm steps. In the initial position, with the source centered in one of the crystals on one side of the array we see more than 90% of the hits in the corresponding channel. If we shift the source by 0.5 mm it is positioned in between two adjacent crystals, we get hits in the two corresponding

channels. Shifting again the source by 0.5 mm, it is centered on the neighbor crystal, and again we get more than 90% hits in the corresponding channel (Fig. 5.33).

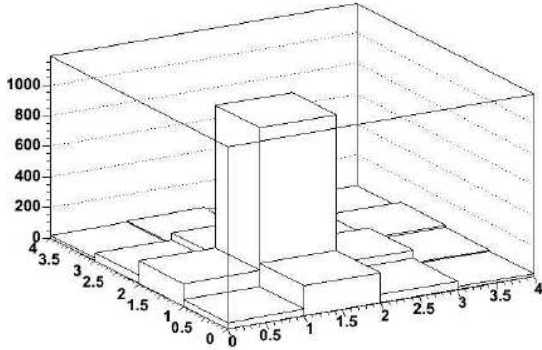


Fig. 5.33 Hit map of the SiPM array coupled to the crystal array when the source is placed at a given position. More than 85% of the hits are in the corresponding channel.

c) Results with continuous crystal

Measurements have been carried out with the *white and black slabs*. The light reflected in the sides of the *white slab* is detected by the photo-detectors, resulting in a larger amount of light collected and therefore in a better energy resolution. However, since all the SiPM's in the array are illuminated due to the reflections in the edges of the crystal, the position information can be distorted. Therefore, the position determination measurements have been carried out only with the *black slab*, in which the light that arrives to the sides and the back of the crystal is absorbed.

Energy spectra

First, the *white slab* is coupled to a single SiPM of 4 mm x 4 mm, with a microcell size of 50 μm x 50 μm (6400 microcells), and data with ^{22}Na , ^{57}Co and ^{137}Cs sources are taken at 2V, 3V and 4V overvoltage.

The results are shown in Fig. 5.34. In each case, the peak position is plotted versus the peak energy. The experimental points are fitted with a Poissonian distribution, a linear fit up to 511 keV (dashed line) and a linear fit up to 662 keV (solid straight line). At 2V overvoltage, a slight saturation effect is observed only at energies above 1 MeV. At 3V overvoltage, the difference between the SiPM response curve and the two linear fits up to 511 keV and 662 keV can hardly be appreciated at 511 keV. At 4V overvoltage this difference is still below 1% at 511 keV, so we can assume that the saturation of the SiPM at 511 keV in this configuration is negligible.

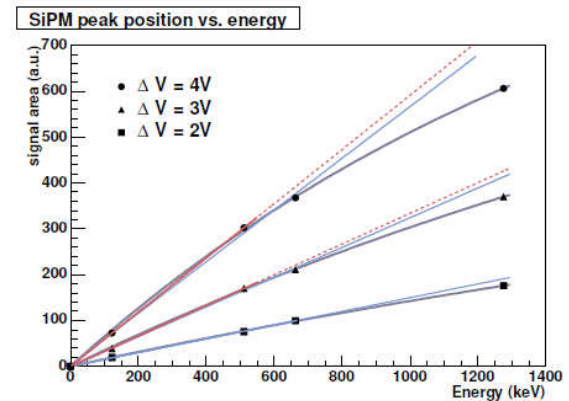


Fig. 5.34 Response curve of the SiPM FBK-irst of 4x4 mm active area coupled to the *white slab* as a function of the source energy for different values of the overvoltage ΔV .

The energy resolution at 511 keV has been obtained in each case by fitting the photopeak with a gaussian function. A value of $(14.06 \pm 0.18)\%$ FWHM has been obtained at 4V overvoltage, slightly better than the resolution obtained with the PMT. The results are shown in Table 5.3.

Overvoltage (V)	Resolution raw data (FWHM)
2	$17.2 \pm 0.3\%$
3	$15.4 \pm 0.3\%$
4	$14.06 \pm 0.18\%$

Table 5.3 Energy resolution for the SiPM FBK-irst of 4x4 mm active area at 511 keV coupled to the *white slab*

Next, the *white slab* is coupled to the SiPM array, and data are taken again with ^{22}Na , ^{57}Co and ^{137}Cs sources at 4V overvoltage (Fig. 5.35 (a)). The signals of all SiPMs in the array are summed for each event in order to obtain the total energy of the event. The peak position is plotted versus the peak energy (Fig. 5.35 (b)) and fitted with the response curve of the SiPM, showing that the array follows the same response curve as the single SiPM. The PDE of the array is lower than that of the single SiPM of 4 mm x 4 mm due to the smaller fill factor of the microcells, and to the loss of active area due to the gaps between the SiPM elements (8.75% lower active area). The saturation effects at 511 keV are also negligible in this case, and the energy resolution obtained is $(14.8 \pm 0.3)\%$ FWHM.

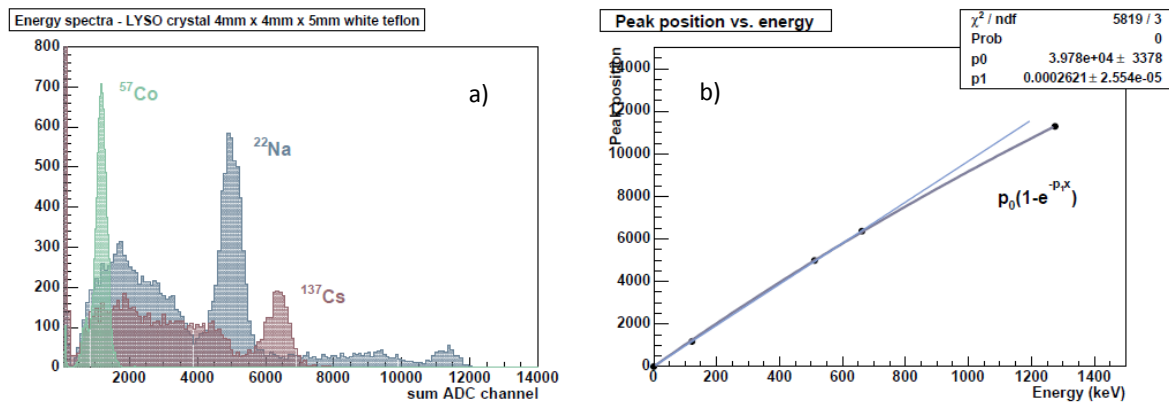


Fig. 5.35 Response curve of the SiPM array coupled to the white slab: (a) energy spectra obtained with different sources; (b) response curve as a function of the source energy

The response curve has been used to calibrate the ^{22}Na histogram. Fig. 5.36 shows the resulting spectrum with the two photo-peaks (i.e. 511 keV and 1274.5 keV) in the corresponding position. The energy resolution obtained from the fit of the calibrated photo-peak is $(15.5 \pm 3.2)\%$ FWHM.

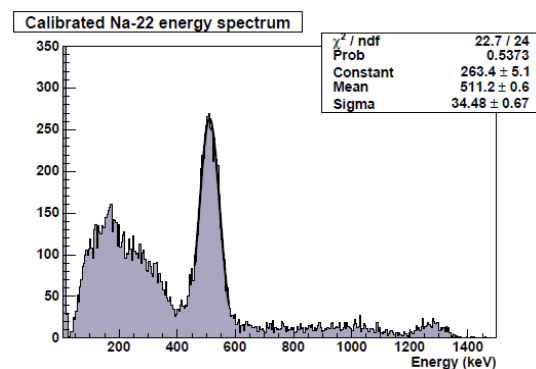


Fig. 5.36 ^{22}Na energy spectrum obtained with the SiPM array coupled to the white slab, calibrated employing the response curve

Timing resolution

The coincidence timing resolution has also been measured in this case with all channels enabled, obtaining similar results as those with pixellated crystals.

Position determination

The first tests on position determination with continuous crystals have also been carried out, using the SiPM array coupled to the *black slab*. Coincidence data have been acquired with the source in different positions, and the position has been reconstructed with the center-of-gravity algorithm, calculated as:

$$X = \frac{\sum X_i ADC_i}{\sum ADC_i}, \quad Y = \frac{\sum Y_i ADC_i}{\sum ADC_i} \quad 5.6$$

where X_i and Y_i are the X and Y coordinates of the SiPM's that fired, and ADC_i is the corresponding ADC value.

For a **given source position**, in Fig. 5.37 we can see the hitmap (a) and the reconstructed position (b). The resolution obtained by fitting the X and Y profiles of the reconstructed position with a Gaussian function at the maximum value is $\sigma = 0.24$ mm (Fig. 5.38).

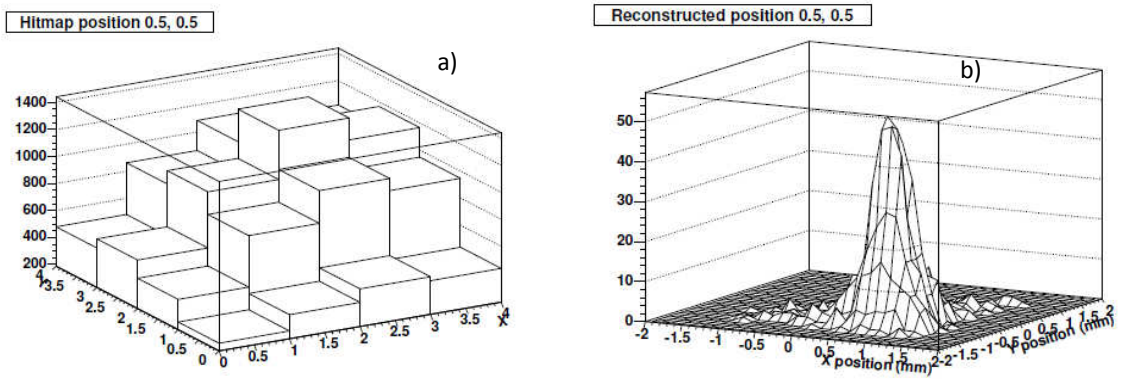


Fig. 5.37 Position determination with the SiPM array coupled to the black slab. (a) Hitmap; (b) reconstructed position employing the center of gravity algorithm

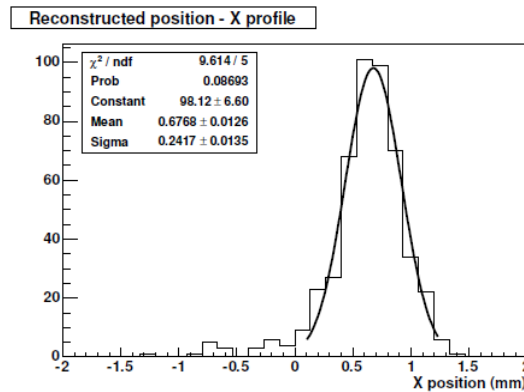


Fig. 5.38 Position determination with the SiPM array coupled to the black slab. X profile at the maximum of the Gaussian distribution of the position reconstructed with the center of gravity algorithm

A linear scan across the array, parallel to the X axis has been performed, taking data at steps of 0.5 mm. The real interaction position of the 511 keV photons cannot be determined since there is no reference point for the second detector (see Fig. 5.27). Once the initial position has been arbitrarily fixed, the distance from one position to the next is determined with a resolution of 0.02 mm. The initial position is about (-1.75mm, 1.25mm) from the geometrical center of the detector. No significant errors in the orthogonality of the X and Y axes have been observed. Fig. 5.39 shows the mean value of each reconstructed (x,y) position. The points are clearly separated, and the resolution of the reconstructed positions ranges from 0.56 to 0.63 mm FWHM, being the average value 0.61 mm FWHM. As expected when using the center-of-gravity algorithm, close to the edges of the crystal the calculated distance between the reconstructed positions does not correspond to the real distance between the interaction points. This effect can be observed for all the positions in the 16-SiPM array, given the reduced dimensions of the device. The reconstructed positions get closer to one another as we approach the edges of the detector. Fig. 5.40 shows the distance from each reconstructed position to the first reconstructed position versus the real distance between the different positions. The error bars represent the sigma of the gaussian distribution for each position. For the positions near to the center of the detector, the distance is closer to the real one, while for the positions near the edges of the detector, the reconstructed positions are very close to one another, and the difference between the real and the calculated position is larger.

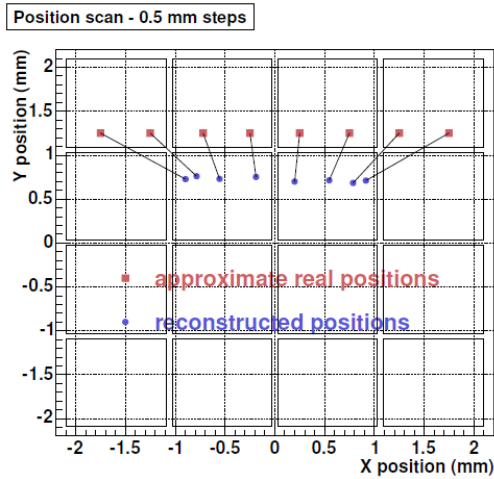


Fig. 5.39 Approximate real positions (red squares) and reconstructed positions (blue circles) in a linear scan parallel to X axis of 0.5 mm step. The distance between the reconstructed positions decreases as the positions approach the edges of the detector. The 4x4 squares represent the SiPM elements in the array.

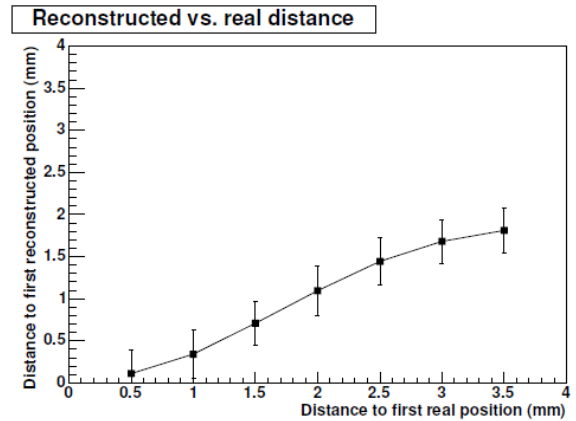


Fig. 5.40 Distance from each reconstructed position to the first reconstructed position versus the real distance between the different positions. The reconstructed positions get closer to one another as they approach the edges of the detector.

5.1.6. Summary on SiPM's arrays for small animal PET scanner

We have seen that the first monolithic SiPM array has been evaluated as photon detectors for PET systems. This work has been done in the framework of a Memorandum of Understanding in between LAL, FBK-irst and Pisa University.

The SiPM array, composed of 16 (4 x 4) SiPM elements of 1 mm x 1 mm size on a common substrate, has been produced at FBK-irst.

The device has been fully electrically and optically characterized at LAL, employing a single channel as well as multi-channels readout electronics. The main characteristics as breakdown voltage, gain, and photon detection efficiency showed an excellent uniformity (less than 5% dispersion) over different elements of the array, giving good confidence in using such detectors in PET systems.

Therefore, the devices have been subsequently tested in Pisa with both pixellated and continuous crystals employing the MAROC2 ASIC and its test board as readout system. Even if MAROC 2 ASIC was not perfectly adapted for the readout of the SiPM array coupled to scintillator crystals, at the time when these tests have been performed, it allowed for the first time the simultaneous connection of an array of SiPM to a multi-channel readout electronics as well as its characterization as photon detector for PET system.

The measurements with pixellated crystals could not be fully exploited because a high amount of light is arriving on each SiPM of the array determining a saturation of the MAROC2 ASIC. However, some measurements have been performed at low overvoltage (i.e. 2V) showing that timing resolution measured employing the MAROC2 ASIC is degraded as compared with the results obtained with single SiPMs. The coincidence timing resolution measured with pixellated crystals is about 2.1 ns FWHM if only one channel is enabled, and 3 ns FWHM if all channels are enabled.

The measurements with continuous crystals show that the SiPM array has the same response function as the single devices. The saturation effect at 511 keV is negligible, being the difference with the linear response below 1%. An energy resolution of 14.8% FWHM has been measured at 511 keV, close to the one obtained with a PMT XP2020Q. The coincidence timing resolution has also been measured in this case with all channels enabled, obtaining similar results as those with pixellated crystals.

In addition, a setup for position determination studies has been constructed, and the first tests have been performed. The source position reconstructed with the center-of-gravity algorithm is in accordance with the real source position. The X and Y projections of the reconstructed positions have a resolution of $\sigma=0.61$ mm FWHM.

Recently, a PET detector ring using SiPM-based block detector has been built by a Japanese group (Yamamoto et al. 2010). High-resolution mouse and rat images were successfully obtained, confirming that the development of a SiPM-based PET system is promising for small animal imaging. The insensitivity of the SiPM to the static magnetic field makes it suitable for future development of integrated PET/MRI system.

5.2. Arrays of SiPM's for radio-guided surgery of tumors

Technical supervision: Taibi Ait Imando (LAL); Andrii Nagai (M2 pre-thesis internships, LAL/Kiev University)

Internal collaboration: Stephane Callier & Ludovic Raux (Omega microelectronic Group)

Beng Yu, Veronique Puill, Dominique Breton, Jihane Maalmi, Michel Gaspard, Stephane Trochet, Dominique Cuisy, Regis Sliva, Patrik Fave, Jean-Christophe Hernandez, Pascal Rusquart, Debennerot Bernard (LAL)

External collaborations: Laurent Menard, Laurent Pinot, IMNC-Orsay (SIPMED project)

We demonstrated in the previous section that SiPM's array can provide uniform and well adapted characteristics for building high resolution PET systems used for diagnostic purposes. We will see in this section that SiPM's arrays can find a new interesting field of medical application, being employed as photon detector in a hand-held radiation detector and used for therapeutic purposes, by assisting the surgeon in locating and removing tumors.

5.2.1. The principle of radio-guided surgery

As mentioned in the Section 5.1, the PET system represents one of the nuclear medical imaging techniques dedicated to functional studies, in particular for cancer detection and diagnostic. For many situations when solid tumors were identified following a PET exam (e.g. breast cancer, different malignances as cutaneous, gastrointestinal, head and neck, gynecologic, urologic etc.), the surgery treatment represents the most commonly used therapeutic solution. In such cases, over many years, the surgeon relied on palpation and visual inspection to identify the tumors position before incision. For helping surgeons on location and real-time information on disease extension, the nuclear imaging techniques exceeded progressively their original diagnostic role and consequently, the *radio-guided surgery concept* was borne (Sweet, 1951). This concept involves the use of a radiation detection system for the intra-operative detection of a radio-pharmaceutical product. Additionally, such technique helps also the surgeon to minimize the surgical invasiveness, improving the quality of the patient life after the intervention.

The design of a radiation detection system used for the intra-operative detection should satisfy particular constrains of surgical procedures, in particular regarding size and shape, the versatility of associated electronics and the need to work with sterile instruments. In this context, numerous hand-held intra-operative radiation detection systems have been developed and have been made commercially available for use in radio-guided surgery.

Such intra-operative radiation detection systems are divided into two main categories: (1) *gamma detection systems* and (2) *beta detection systems*, based upon the specific type of radiation detected. Gamma systems detect photon radiation, consisting of either gamma rays or X-rays (Hoffman & al., 1999; Zanzonico & Heller, 2001; Sarikaya & al., 2008). Beta systems detect beta radiation, consisting of either positrons or electrons (Hoffman et al., 1999; Zanzonico & Heller, 2001; Daghighian & al. 1993; Raylman & Wahl, 1996; Raylman, 2000; Yamamoto & al., 2005; Yamamoto & al., 2006). This includes some beta detection probe systems that are reported to have gamma photon background rejection capabilities (Yamamoto & al., 2005; Yamamoto & al., 2006).

Our **SIPMED (Silicon Photomultiplier for Medical Imaging) project**, developed in collaboration with IMNC-Orsay and l'Hospital Lariboisiere, uses SiPM arrays for building a prototype of intra-operative gamma camera.

Therefore, the present Section will give a short review of basic principles of gamma detection systems used in radio-guided surgery as well as of desirable design features that are important for surgeon. More details will be given about two gamma detection systems called POCI and TReCam, previously developed by our colleagues from IMNC laboratory (S. Pitre, PhD thesis, 2002 and E. Netter, PhD Thesis 2011), and the main reasons for which a new intra-operative gamma system based on SiPM arrays is desired. The beta intra-operative detection will not be addressed in this dissertation. But I would like to mention that we have a new project called SONIM (New miniaturized probes for charges particles detection in molecular imaging) that is in a very early stage of development (it received his financial support in March 2013). This project proposes a technological development of SiPM arrays with characteristics adapted for a prototype of a beta imaging detection system and it will be also realized in collaboration with IMNC. In the framework of this new project, Andrii Nagai, who already performed his M1 and M2 internships under my supervision, will start his PhD thesis at the University Paris 11 from September 2013, on the subject: "Silicon photomultiplier for medical imaging".

5.2.2. Performance parameters of intra-operative detection systems

The basic principle of the intra-operative gamma detection system relies on a radiation detector (usually a scintillator coupled to a photon detector or a semiconductor detector) and associated electronics to provide a sound signal or visual image related to the amount of radiation detected. The detection system providing a sound signal is usually called *intra-operative probe* (counter) and the system giving a visual image is called *intra-operative imaging camera*.

Besides the handling as well as the volume and robustness constraints which are inherent to their exploitation in the operating room, both intra-operative probes and imaging camera should answer to performance criteria which involves a given choice of the detection components. Therefore, before passing to the description of different types of detection systems and their components, the most important performance parameters of an intra-operative gamma detection system are reviewed in the following:

- *Overall sensitivity (detection efficiency)* which represents the number of counts per unit of injected dose (cps/KBq). This parameter depends particularly on the source-detection geometry and media, therefore we can define two distinct components:
 - o *Geometric sensitivity or spatial selectivity* – the fraction of emitted radiations that intersect the detector (the fraction of total solid angle subtended by the detector); it is directly proportional to the radiation-sensitive detector area and, for a point source, inversely proportional to the square of the source-detector distance. This is the parameter which usually limits the location of tumor lesions situated in the depth of the human tissue. For a collimated detector, geometric efficiency is inversely related to the collimator length.
 - o *Intrinsic sensitivity, or efficiency* – the fraction of radiations intersecting the detector that is stopped within the detector; it is directly related to the detector thickness, effective atomic number, and mass density and decreases with increasing photon energy because higher-energy photons are more penetrating and are more likely to pass through a detector without interacting.

Overall sensitivity represents an important parameters since it influence the fastness of the tumor location as well as the capacity of the detection system to detect tumors (the time of the tumor location step should be as short as possible to not delay the surgical intervention).

- *Energy resolution* – is the capacity of the detection system to discriminate between emitted radiations of different energies. Energy resolution (ΔE) is generally defined as the full-width at half maximum

height (FWHM) expressed as a percentage of the photo-peak energy (E_γ) of the photo-peak in the energy distribution spectrum: $\text{FWHM (\%)} = (\Delta E / E_\gamma) \times 100\%$.

- *Spatial resolution* – it reflects the ability of the detector to accurately determine the location of the source and it represents a critical parameter for a detector. The detected count rate as a function of the distance from the detector's central axis (e.g. by moving the point source in front of the detector) can be used to generate the point spread function (PSF) by using the point source in air. The probe's spatial resolution can then be expressed as the FWHM of the PSF.

The spatial resolution deteriorates dramatically with increasing distance of the source to the detector, primarily because the probe's field of view increases with increasing distance. The degradation of spatial resolution is less dramatic in air owing to the absence of a scattering medium. The use of a collimator tends to limit the detector's field of view to tissue directly in front of the detector and improve its spatial resolution. However, geometrical sensitivity and spatial resolution are inversely related performance parameters: geometrical sensitivity deteriorates as spatial resolution improves and spatial resolution is degraded as geometrical sensitivity increases (higher field of view).

- *Contrast or spatial selectivity* – the ability of the detection system to distinguish activity from a target region (i.e. tumor lesion) from the lower activity in a surrounding background region. It may be expressed as the difference of the target (T) and background (B) region count rates divided by the target region count rate, $(T-B)/T$. The advantage of this particular definition of detector contrast is that it varies between 0 and 1, with a value of 0 corresponding to no contrast and a value of 1 corresponding to maximal contrast (i.e. no background). Contrast is directly related to all the above mentioned detector performances: the detector sensitivity, energy resolution and spatial resolution.

5.2.3. Review of intra-operative gamma detection systems

As mentioned previously, two types of intra-operative gamma detection systems exists today: (1) *the gamma counting probes*, employed to assist the surgeon in detection and removal of tumors for more than 50 years, and (2) *gamma imaging camera*, representing more recently developed intra-operative detection systems. In the following, more details about the components of these detection systems as well as their characteristics will be given.

5.2.3.1. Intra-operative gamma probes

The intra-operative gamma probes are usually single-pixel probes of cylindrical shape. They present small dimensions (1-2 cm diameter and 10-20 cm length), an ergonomic shape (e.g. pencil type) and weight well studied to allow the surgeon going with the probe in contact with the tumor lesions. To not detract the surgeon attention when he is exploring the tumor lesion, such probes give in real time the radioactivity level through a sound signal. The count rate is also shown on a count meter (Fig. 5.41).

The gamma probes are usually composed of a photo-detection system connected to a readout system and they are equipped with a shielding

(lead, tungsten, gold, platinum) on the lateral sides and a collimator (lead, tungsten) on the front side.



Fig. 5.41 Example of intra-operative gamma probe

In the following, we will describe the collimation head and the different types of detectors.

a) Collimation head

The collimation head is used to limit the field of view of the probe avoiding the entrance into the detector of the diffused events coming from the surrounding background region with respect to the target region. Unfortunately this background rejection is done with the detriment of the probe size increase. In particular, the detectors dedicated to study the gammas of high energies (^{131}I : $E_\gamma = 364 \text{ keV}$, ^{124}I : $E_\gamma = 511 \text{ keV}$) need thicker shielding and collimator, consequently they are heavier and less compact. Moreover, even if the collimator improves the lateral spatial selectivity of the probe, which represents the real limitation of gamma probes, in the same time it degrades its efficiency. This is the main reason for which they are removable or retractable.

b) Detectors

Two general categories of gamma detection probe systems exist: the one based on scintillator detector connected to a visible light detector and the second one, based on semiconductor ionization detector.

Scintillation detectors

The typical configuration of a scintillator detector-based probe is presented in Fig. 5.42. This is basically the configuration used by Harris et al. in 1956. In this type of detector, the scintillator absorbs the radiation and emits visible light in proportion to the absorbed energy. The visible light must be measured by a photon detector, usually a photomultiplier tube (PMT) or a silicon photodiode PIN to assure the light conversion in electrical signal. The most used scintillator crystals are NaI(Tl) and CsI(Tl) and they present thicknesses up to 1 cm to assure a detection efficiency higher than 90% at 140 keV (the most used radio-isotope in radio-guided surgery is $^{99\text{m}}\text{Tc}$, emitting gamma rays of 140 keV). Other crystals presenting higher density and less hygroscopic such as YAP for the detection of 140 keV gamma or LSO and LYSO for the detection of 511 KeV gamma can also be used. Even if such scintillators allow to reduce the size of the detection system, they present lower light yield which degrades the energy resolution, and short time constants which require the development of a fast readout electronics.

A recent developed intra-operative probe called CarolIRES (Mathelin & al.; 2006) uses a YAP scintillator with 4x4 mm lateral dimensions and 12 mm thickness, determining a detection efficiency of 87% at 140 keV. The crystal is coupled to a Hamamatsu Photomultiplier tube H3164-10 and a dedicated electronics. Despite of these instrumental efforts, the scintillation detectors are generally less compact and they present lower performances

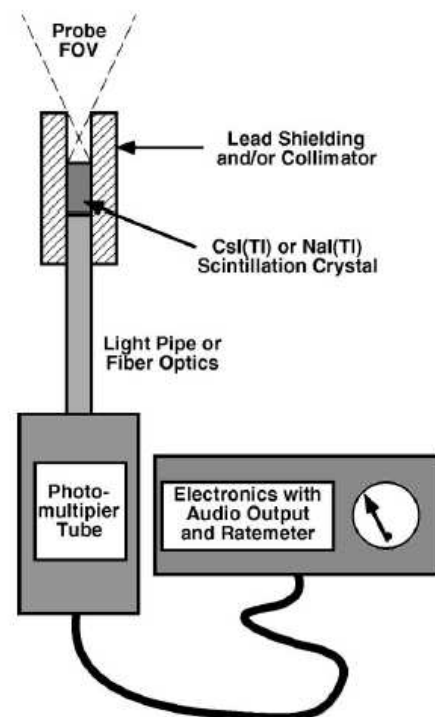


Fig. 5.42 The basic configuration of scintillation detector-based intra-operative probe (Harris & al. 1956)

in energy resolution than semiconductor detectors (for example, the CarolIRES system present an energy resolution of 56%). Instead of this, they present better detection efficiency.

Semiconductor detectors

This kind of detectors uses semiconductors working at room temperature and presenting high atomic number and density like cadmium telluride (CdTe), cadmium zinc telluride (CdZnTe), and mercuric iodide (HgI₂). These materials allow a direct conversion of the gamma interacting in the depleted region of the semiconductor into the electron-holes pairs. The charges are then collected to the anode under the action of the electric field. Such kind of detection confer to these detectors the advantage of working without additional device for gamma conversion, therefore they are more compact and in the same time they present a very good energy resolution of ~5% which allows a good efficiency of background rejection. However, to avoid the recombination phenomena in the depletion region (i.e. loss of signal to the anode level), and the noise related to the leakage current (direct proportional to the geometrical parameters of the detector, in particular its thickness), the thickness of the detector is limited to maximum 5 mm. Therefore, the detection of gamma with energies higher than 200 KeV becomes more difficult and the entire sensitivity of the system is limited.

5.2.3.2. Intra-operative gamma imaging camera

In addition to intra-operative counting probes, compact imaging gamma cameras are very attractive to provide more efficient tumor localization during radio-guided cancer surgery. These devices add the ability to see details of the detected activity, giving the potential of using the technique in a low-contrast environment. They can cover a larger area (field of view from 20 to 100 cm²) than a non-imaging probe and still pinpoint the location of the activity. In cases of relatively poor signal to noise ratio, the image can show the distribution of the activity to pinpoint the hot area, whereas the non-imaging probe simply gives a tone or an activity reading representing the average activity in the region.

Many miniaturized gamma cameras are currently under evaluation, especially for the sentinel lymphatic node mapping in melanoma and breast cancer (i.e. the sentinel lymphatic node is the first ganglion draining a tumor). Inspiring from the technology of “high field of view” gamma camera, the intra-operative gamma camera are composed from a collimator associated to a detection system which allows determining the position of gamma emission. The collimator geometry provides essentially the spatial resolution and the device efficiency. Regarding the detection system, it is belong the same two categories like the counting probes: scintillation crystal connected to a visible light detector and semiconductor ionization detectors. A nice review of different types of intra-operative gamma imaging camera and their principle is given by E. Netter in her PhD thesis, 2011. In this dissertation, to understand the improvements bring by using SiPM arrays for building intra-operative gamma camera, we will concentrate particularly on the review of the two devices called POCI and TReCam developed by our colleagues from IMNC.

POCI – Per-Operative Compact Imager

The POCI system (Fig. 5.43) consists of a head module composed of a interchangeable lead collimator and a CsI(Na) scintillator crystal optically coupled to a position sensitive diode (PSD) through an image intensifier tube. When a gamma ray is emitted by a tumor lesion, it is selected according to a preferential axis of collimator. The energy deposited in the scintillator creates a local light spot which is collected, by proximity, by the photocathode of the image intensifier tube. This last element amplifies the light intensity, keeping in the same

time the spatial and temporal information, and brings the light spot up to the PSD photodiode. The centroid of the position and the total energy of the light photons interacting with the photodiode are given by a charge division. Therefore, each gamma event is characterized by its position as well as by its energy left to the scintillator. The signals are read and amplified by a 4 channels readout system and they are finally transferred to a computer to be digitized and seen on a screen in real time.

A general picture of the POCl intra-operative gamma camera is presented in Fig. 5.44 (a). This system has been used in the operating room for the sentinel lymph node localization in the breast cancer (Fig. 5.44 (b)).

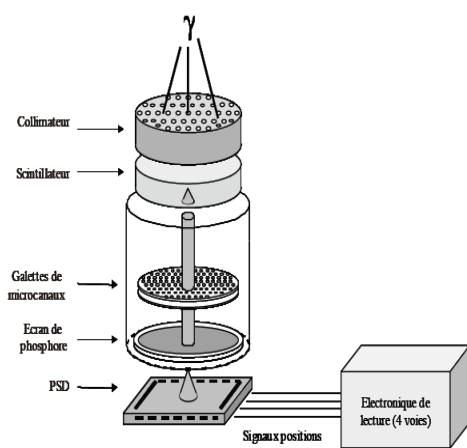


Fig. 5.43 Sketch of POCl camera

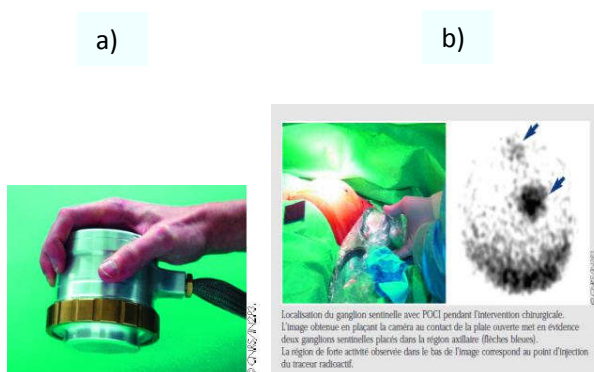


Fig. 5.44 (a) Picture of POCl camera; (b) POCl camera in the operating room used for lymph node localization (S. Pitre, 2002)

TReCam – Tumor Resection Camera

The more recently developed intra-operative gamma camera called TReCam takes the advantage of using new pixellated detectors and in the same time increases the field of view of 50% with respect to POCl camera (from 12.5 cm^2 corresponding to POCl to 25 cm^2 corresponding to TReCam).

The TReCam system (Fig. 5.45) is composed of a 15 mm thick parallel hole collimator, a 5 mm Saint Gobain $\text{LaBr}_3:\text{Ce}$ scintillating crystal optically coupled to a Hamamatsu H9500 flat panel multi-anode photomultiplier tube (MAPMT). The MAPMT presents 256 anodes and an active area of $49 \times 49 \text{ mm}^2$ (total dimensions: $L \times l \times h = 52 \times 52 \times 33 \text{ mm}^3$). The readout of the MAPMT is ensured by a specific integrated circuit called HARDROC2 (Seguin-Moreau et al. 2007).

The principle of TReCam system is very similar to POCl one. The gamma ray coming from the tumor lesion is selected, following its direction, by the collimator and interacts in the scintillator. The energy deposited by the gamma ray in the scintillator is converted in a light spot which is collected by the photocathode of the MAPMT. For each gamma event, the charge, collected on each of 256 anodes, is read by the corresponding readout electronics channel and then transferred to a computer through an USB connection, allowing for data analyzing, reconstruction and visualization of the signal (Fig 5.45).

The TReCam camera has been build in such way that it answers to electrical, electro-magnetical and mechanical constrains allowing to be approved by the Laboratoire National de Métrologie et d'essais (LNE) and to be used in the operating room. A general picture of the TReCAM intra-operative gamma camera is presented in Fig. 5.46 (a). This system has been evaluated in the clinical procedure of the sentinel node (Fig. 5.46 (b)).

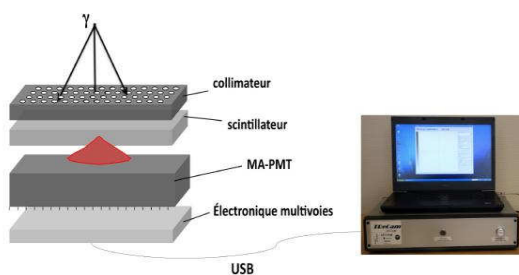


Fig. 5.45 Sketch of the TReCAM camera

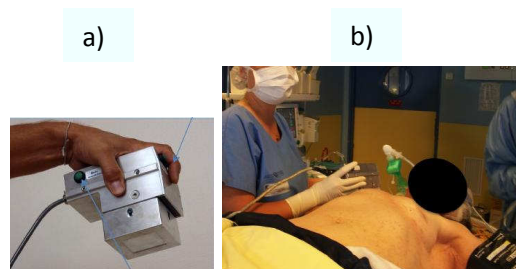


Fig. 5.46 (a) Picture of TReCAM camera; (b) TReCAM camera in the operating room used for lymph node localization and control of surgeon movement (E. Netter, 2011)

The main characteristics of the POCl and TReCAM intra-operatives gamma cameras for gamma rays of 140 keV are summarized in the Table 5.4.

	Spatial resolution	Efficiency	Energy resolution	Active surface	Dimension	Weight
POCl	3.2 mm (contact)	290 cps/MBq	32%	12.5 cm ²	h = 90 mm φ95 mm	1.2 kg
TReCAM	1.8 mm (contact)	300 cps/MBq	11.3%	25 cm ²	h = 117 mm 140 x 83 mm ²	2 kg

Table 5.4 The main characteristics of POCl and TReCAM for gamma rays of 140 keV

5.2.4. SIPMED project – Silicon Photomultiplier for medical imaging

As we have observed in the previous chapter, the use of MAPMT opened the way forward to build intra-operative gamma camera using pixellated photon detectors. The major limitation of the presented system generally concerns the compactness and the weight of the device. In this context, the arrays of SiPM detectors, presenting geometrical characteristics much thinner than MAPMT and providing very similar optical characteristics, became very good candidate for building intra-operative systems.

In this context, around two years ago we started a project called SIPMED (Silicon Photomultiplier for medical imaging) aimed to develop a prototype of intra-operative gamma imaging camera based on SiPM arrays connected to a miniaturized readout electronics. In the following, the general description of the SIPMED camera and the actual status of this development will be presented (the project is in progress).

The SIPMED system (Fig. 5.47) is composed of 49x49x5 mm³ LaBr₃(Ce) Saint Gobain scintillator, covered on its top with Teflon optical coating and on its edges with absorbing paint. It is optically coupled to 16(4x4) SiPM monolithic arrays from Hamamatsu HPK, each of them consisting of 16(4x4) SiPM's connected to a miniaturized readout electronics based on a dedicated ASIC. By using the mentioned components we keep the same field of view as TReCAM system (25 cm²), but we expect to reduce substantially the geometrical dimensions (less than LxLxh= 60x60x50 mm²) as well as the weight (< 1 kg). *Using the SiPM technology we expect also an improvement of spatial and energy resolution with respect of TReCAM system based on a multi-anode PMT.* The detection performances of the full system will be also evaluated in the framework of SIPMED project.

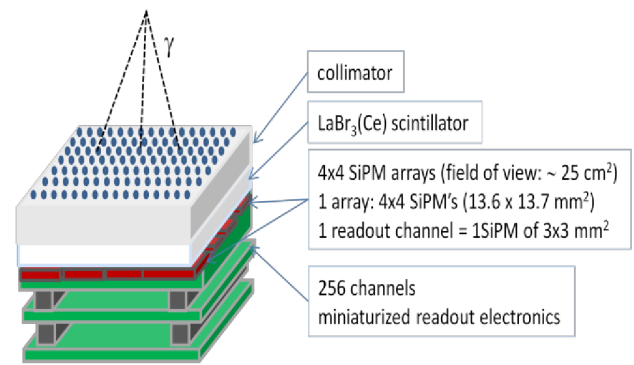


Fig. 5.47 Schematic representation of the SIPMED camera

5.2.4.1. Characteristics of the SiPM's arrays for SIPMED camera

The SiPM arrays used for building the SIPMED intra-operative gamma camera are monolithic arrays called S11828-3344M produced by Hamamatsu HPK (Fig. 5.48). The S11828-3344M is a 16-channels (4x4) monolithic SiPM array¹ mounted on a surface-mount package (SMD). Each channel is made up of 3600 micro-cells (i.e. each micro-cell is a GM-APD in series with the quenching resistance) at a 50 μm pitch and has a 3x3 mm² photosensitive area. The gap between each channel is minimal (200 μm) due to the monolithic chip structure which is highly efficient for collecting light when coupled to a scintillator. The distance from the photosensitive area to the package edge is 500μm, assuring a minimal loss of active surface when the arrays are arranged against each other in a buttable format. The dimensional outline of the array (unit=mm) is presented in Figs. 5.49 (a): the front side, (b): the back side, (c): the cross-section. We can notice the smaller thickness of the

¹ Hamamatsu HPK company names the SiPM detector as MPPC (Multi-Pixel Photon Counter)

SiPM array with respect to the one of flat panel MAPMT (33mm): the thickness of the SiPM chip is 300 μm and the overall thickness of the SMD mounted device is 1.45 mm.

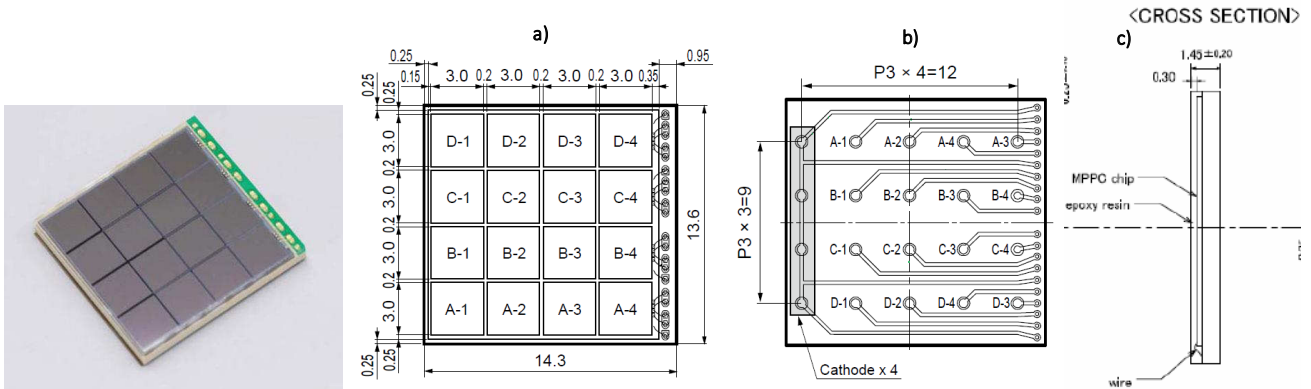


Fig. 5.48 Picture of S11828-3344M array from Hamamatsu HPK

Fig. 5.49 Dimensional outline (in mm) of S11828-3344M array: (a) front side, (b) back side, (c) cross-section

As already mentioned in the case of SiPM arrays used for the PET system, the uniformity of the SiPM characteristics such as breakdown voltage, gain, detection efficiency is extremely important. If such characteristics are not uniform within different SiPM's used to build the overall gamma camera, the position of the photo-peak may shift from channel to channel and the image resolution will degrade. For this reason, the first test to check the devices uniformity is by measuring the reverse IV characteristic, which will be presented in the following.

For performing IV characteristics, the SMD mounting of the SiPM arrays requires careful connections between the measurement instrument (Keithley 2612 Source-meter) and each SiPM. These connections have been assured by using thin probes provided by the KarlSuss test-station mounted in the ATLAS pixel clean-room at LAL (Fig. 5.50). More details about the characteristics of the KarlSuss test-station will be given in the Section relating the activity I developed for the ATLAS pixel detectors (Part B, Section 3.3.1.).

A zoom of the needle contacts provided by the Karl-Suss probes is presented in Fig. 5.51. For the measurement of each SiPM IV reverse characteristic, one probe is contacting the anode of corresponding SiPM and the other one is connecting the cathode, common for all SiPM's. The two probes are connected to the Keithley source meter.

Since the SiPM characteristics can change with the temperature, careful attention has been paid on temperature control. The ATLAS pixel clean room has a temperature and humidity control system and therefore the measurements have been carried at $20^\circ\text{C} \pm 0.5^\circ\text{C}$.

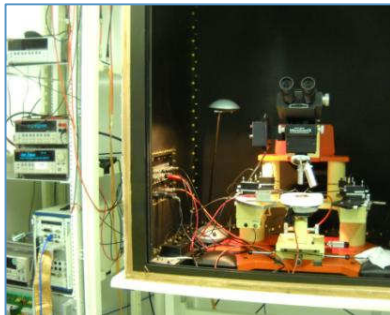


Fig. 5.50 General view of the KarlSuss probe-station (ATLAS pixel clean-room at LAL) used for IV characterization of the SiPM arrays

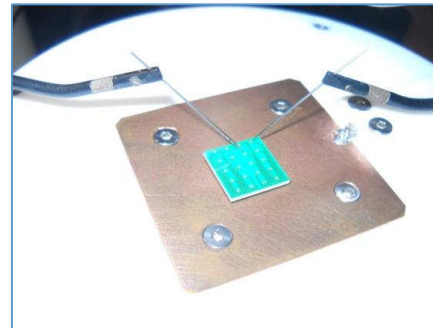


Fig. 5.51 Zoom of the SiPM array back side and the two probes used to contact each anode and the cathode of the array

We characterized a total number of 23 arrays, each of them having 16 SiPM's, totaling 368 IV characteristics. The cumulative plot of these characteristics is given in the Fig. 5.52. We can clearly observe a distribution of the breakdown knee from ~ 70.2 V to ~ 71.6 V, representing ~ 1.4 V range. Sixteen arrays having the closest breakdown voltage have been choised for building SIPMED camera. Their IV characteristics are presented in Fig. 5.53.

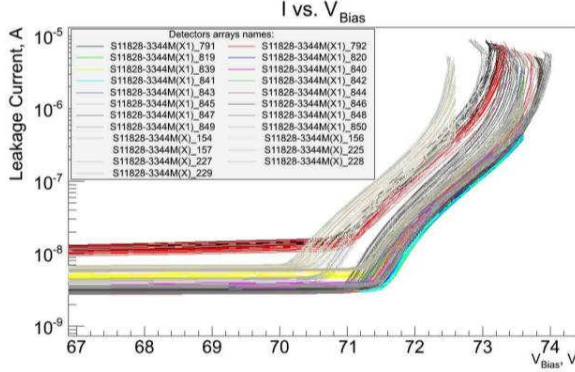


Fig. 5.52 Zoom around the breakdown region of the cumulative IV characteristics

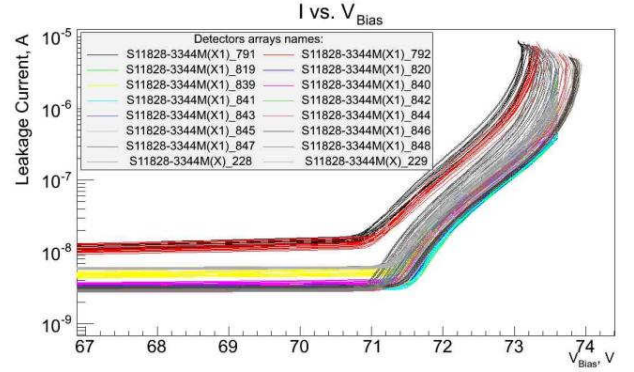


Fig. 5.53 Zoom around the breakdown region of the IV characteristics corresponding to 16 SiPM arrays selected for SIPMED camera

We can remark that the breakdown voltage range of the selected arrays is reduced to ~ 800 mV. This is confirmed also by the V_{BD} distributions presented in the Figs. 5.54 and 5.55. The V_{BD} values have been defined as the bias voltage on the IV characteristic at which the second derivative of the logarithmic of the current has its maximum.

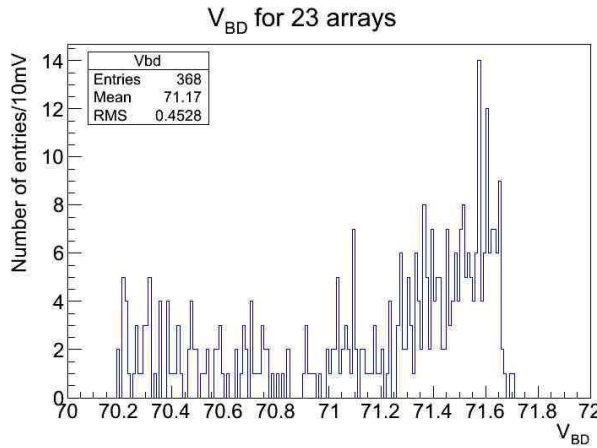


Fig. 5.54 Histogram of the V_{BD} for all tested arrays

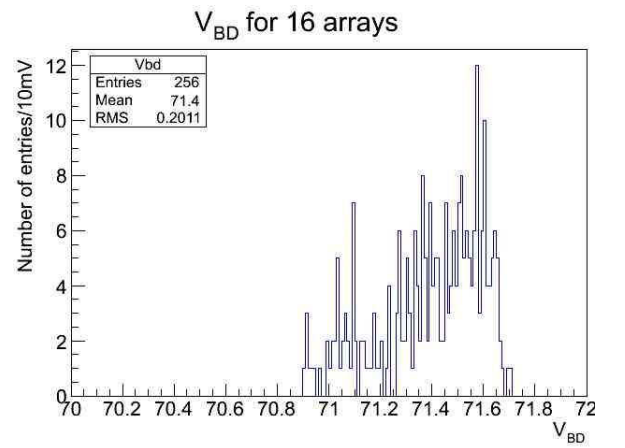


Fig. 5.55 Histogram of the V_{BD} for 16 arrays selected for SIPMED camera

In order to be easily connected to the readout electronics, the selected 16 SiPM arrays have been assembled on four test-boards or elementary modules (Fig. 5.56), each of them having 4(2x2) arrays (more details on the elementary modules assembly will be given in the next section;(note: the inactive green parts of the test-boards will be removed on final system). The arrays of each elementary module have been selected to have as closest as possible V_{BD} values. A 3D plot of the V_{BD} values of overall 256 channels as a function of their corresponding xy positions on different elementary modules are presented in Fig. 5.57. We can observe a V_{BD} range of about ~ 200 mV on each elementary module, but an overall ~ 800 mV over all 256 channels.

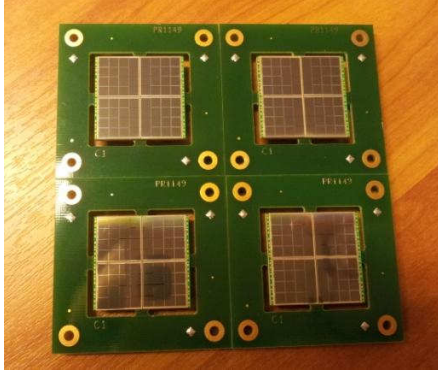


Fig. 5.56 Picture of 4(2x2) elementary modules, each of 4(2x2) arrays mounted on PCB's, building the full SIPMED camera

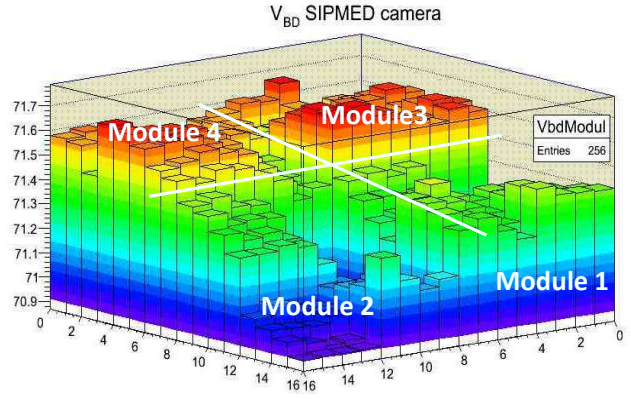


Fig. 5.57 3D plot of breakdown voltage of 256 SiPM's used for the SIPMED camera

If a common V_{Bias} of 72.5V is applied over all devices, the leakage current presents a variation of more than one order of magnitude. This can be observed in Fig. 5.58 showing the leakage currents distribution at $V_{bias} = 72.5V$. The same values of the leakage currents are represented in 3D plot as a function of xy position of the SiPM's in the assembly of the elementary modules (Fig. 5.59). We can clearly observe that the leakage current is much uniform for a given array, but a higher variation is seen on overall 256 channels.

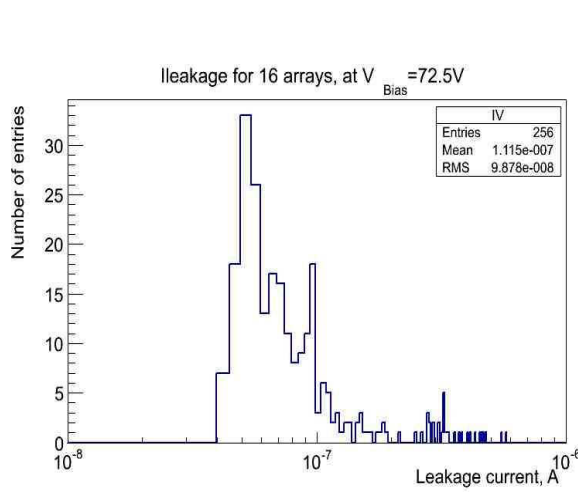


Fig. 5.58 Leakage current histograms of the 256 SiPM's of the SIPMED camera at a common bias voltage of 72.5V

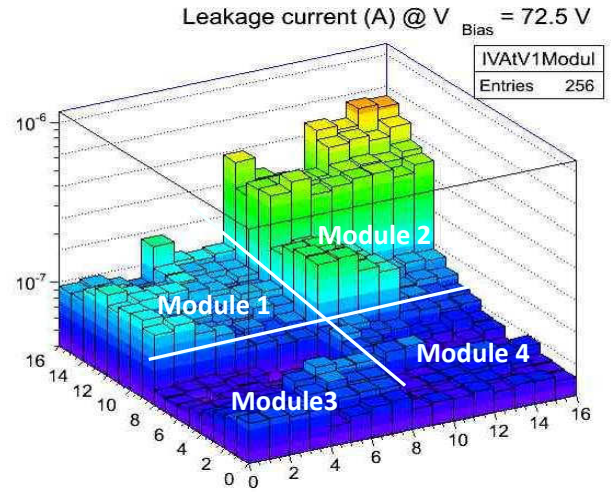


Fig. 5.59 3D plot of leakage current at $V_{bias}=72.5V$ as a function of xy position of 256 SiPM's used for the SIPMED camera

If we apply a different V_{bias} on each channel in such way that the overvoltage $\Delta V = V_{bias} - V_{BD}$ is kept constant (i.e. V_{BD} compensation), a much better uniformity is obtained as can be seen in Fig. 5.60. The same values of the leakage currents are represented in 3D plot as a function of xy position of the SiPM's in the assembly of the elementary modules (Fig. 5.61). We can clearly see that the leakage current is much more uniform over all 256 channels; two arrays can still be distinguished to have higher values with respect to the others.

As we will see in the next section, the dedicated ASIC used for SiPM's readout features an input DAC allowing to adjust the SiPM bias voltage channel by channel and to improve the uniformity of the system. We have to keep into account that presented leakage current distributions are determined by both gain and dark count rate

contributions (for more details see Section 3.2.1). Therefore, the final gain variation from channel to channel is expected to be lower than the overall leakage current variation.

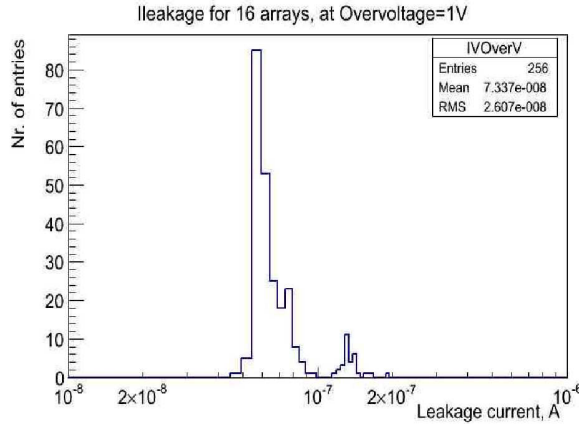


Fig. 5.60 Leakage current histogram of 256 SiPM's of SIPMED camera at $\Delta V=1V$ (bias voltage is adjusted on each channel in order to obtain the same overvoltage)

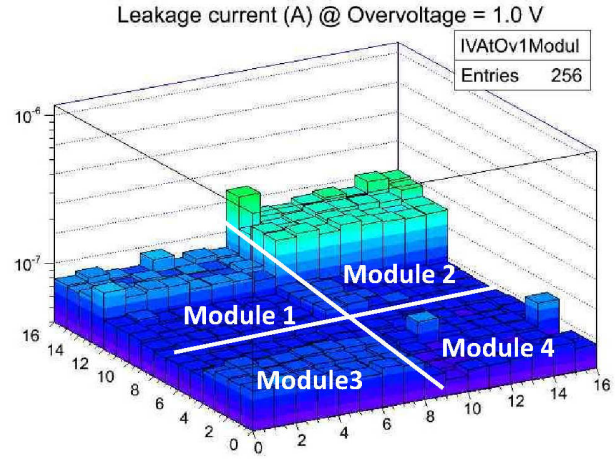


Fig. 5.61 3D plot of leakage current at $\Delta V=1V$ as a function of xy position of 256 SiPM's used for the SIPMED camera

5.2.4.2. SiPM's array implementation for SIPMED camera

The full field of view of the SIPMED camera is assured by mounting together, against each other, four elementary modules. Each elementary module covers a quarter of the full camera surface and it is based on 2x2 SiPM monolithic arrays, coupled to a miniaturized readout electronics based on a dedicated ASIC. The bloc diagram of an elementary module is presented in Fig. 5.62.

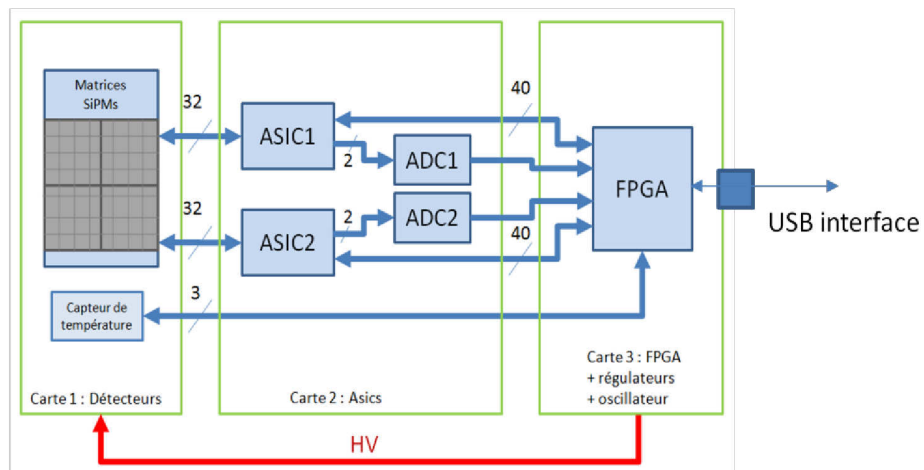


Fig. 5.62 Bloc diagram of the SIPMED camera elementary module

The elementary modules have been already designed and produced at LAL. Each elementary module is composed by a compact stack of three PCBs, having an overall dimension of $28.6 \times 27.2 \text{ mm}^2$ and a thickness of 30 mm (Fig. 5.63).

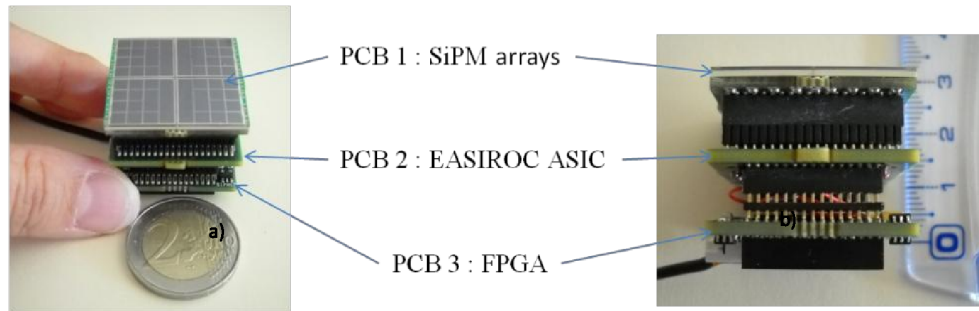


Fig. 5.63 Pictures of a SIPMED elementary module on (a) front side and (b) lateral side

The first board (Figs. 64 (a1) and (a2)) contains four monolithic arrays S11828-3344M on the front side and passive elements (i.e. resistors, capacitors) and a digital thermal sensor on the back side. The passive elements are aiming to adapt the SiPM signals to the input of the readout electronics. The temperature sensor monitors the temperature and allows, through a DAC embedded at the ASIC input, to adjust individually the SiPM bias voltage and to maintain a constant overvoltage (i.e. constant gain) on each channel during the system operation, independent of temperature variation.

The second board (Fig. 5.64 (b1) and (b2)) contains two ASIC's named EASIROC, standing for Extended Analogue Silicon pm Integrated Read Out Chip (Callier et al. 2011). Each ASIC (simplified bloc diagram presented in Fig. 5.65) has 32 channels fully-analogue front-end electronics dedicated to gain trimming and read-out of SiPM detectors. Each of the 64 inputs features a low power 0-2.5V range 8-bit DAC to adjust the SiPM bias voltage channel by channel and to improve the gain uniformity of the system. Each SiPM output signal is sent to two voltage sensitive preamplifiers with low (LG) and high gain (HG) to perform the requested dynamic range from 160 fC to 320 pC (i.e. 160 fC corresponds to a SiPM signal of one micro-cell, with a gain of 10^5). The charge is measured at the maximum amplitude of the two tunable slow shapers (peaking time from 50 to 175 ns) by two Track and Hold blocks. A common fast trigger line made of a very fast shaper and a discriminator provides the hold signal. The analog data are digitized using an external two-channels ADC (12 bits and 2Msample/s, ADS7947, Texas Instruments). The response linearity and uniformity over the 64 channels is better than 1% and 2%, respectively.

The third board (Fig. 5.64 (c1) and (c2)) combines an Altera Cyclone III FPGA, a FTDI FT232RL (USB protocol 2.0 Hi-Speed, 480Mbit/s) and a USB mini-connector for the communication with a PC, and a DC/DC converter to deliver the high voltage to SiPM array. The FPGA allows easy programming of all parameters of the ASIC using free software produced by ALTERA, to sample and to save digital data in a built-in FIFO before to be transferred to USB port. The power consumption of the electronic module is lower than 5mW/channel.

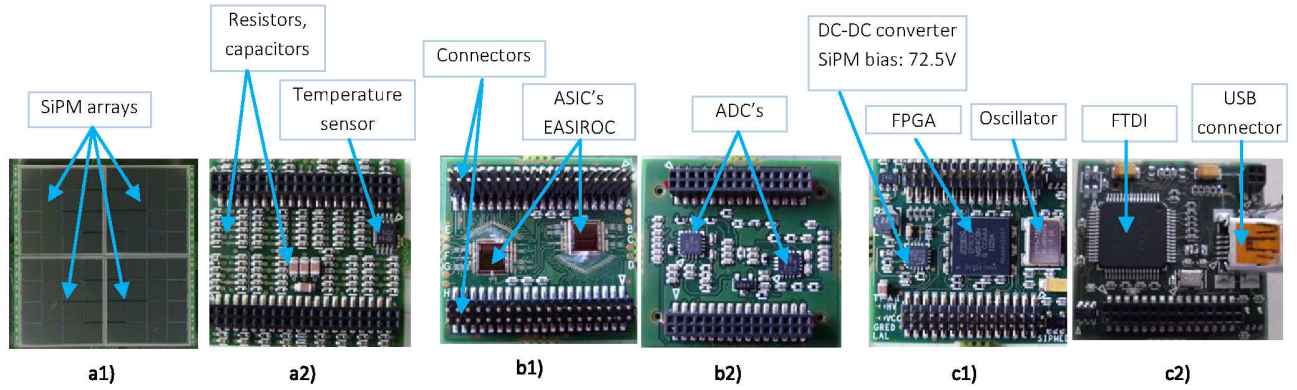


Fig. 5.64 Pictures of SiPM board on front side (a1) and back side (a2); ASIC board on front side (b1), back side (b2); FPGA board on front side (c1), back side (c2)

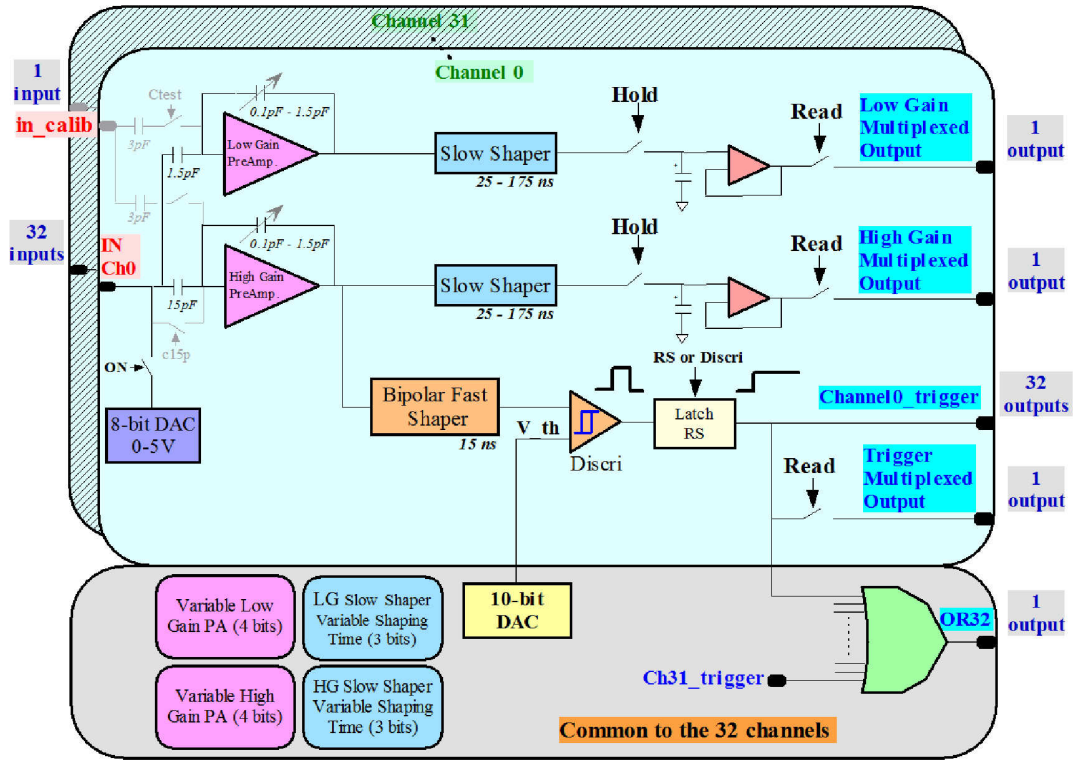


Fig. 5.65 Simplified bloc diagram of the EASIROC chip

5.2.4.3. Preliminary evaluation for gamma imaging

The photo-detection elementary module (Fig. 5.63) was coupled to a $49 \times 49 \text{ mm}^2$ and 5mm thick $\text{LaBr}_3(\text{Ce})$ crystal plate (Saint Gobain), covered on its top with teflon optical coating and on its edges with absorbing paint. Optical coupling grease layer is applied on the boundary (BC-630, Bicon). Because the view of

the elementary module does not match that of the scintillator (i.e. the scintillator has the dimensions of full gamma camera, including 4 elementary modules), only suboptimal performances can be achieved. However, this preliminary evaluation allows one to validate the good operation of the electronic read-out for gamma imaging.

Measurements have been made at 20°C and an optimal overvoltage of 1.2V. A preliminary gain calibration was made in order to correct the heterogeneity of the SiPM response. The bias voltage is adjusted on each SiPM with the 8-bit input DAC depending on its breakdown voltage obtained from the IV characteristics (see Section 5.2.4.1). The gain uniformity over the 64 SiPM's improves from 13% to 2% after gain corrections.

Energy resolution was evaluated by using a 2 mm collimated ^{57}Co point source placed at the centre of the crystal plate. The FWHM energy resolution at 122 keV full energy peak was found to be 13.9% (Fig. 5.66 (a)).

The spatial resolution and the position linearity were investigated by moving a 0.5 mm collimated ^{57}Co source along the x- and y- axes of the scintillating crystal in 3 mm scanning steps (Fig. 5.66 (b)). The reconstruction algorithm was based on the squared method (centroid of the charge distribution raised to the power two). The FWHM spatial resolution, calculated from the profile of the point-spread function of the pinpoint ^{57}Co source, ranges from 1.3 mm in the center of the detector to 1.95 mm at the edges (Fig. 5.66 (c)). The maximum error on position remains inferior to 0.25 mm in the center region of the field of view (central SiPM) and increases up to 1.5 mm at the edges.

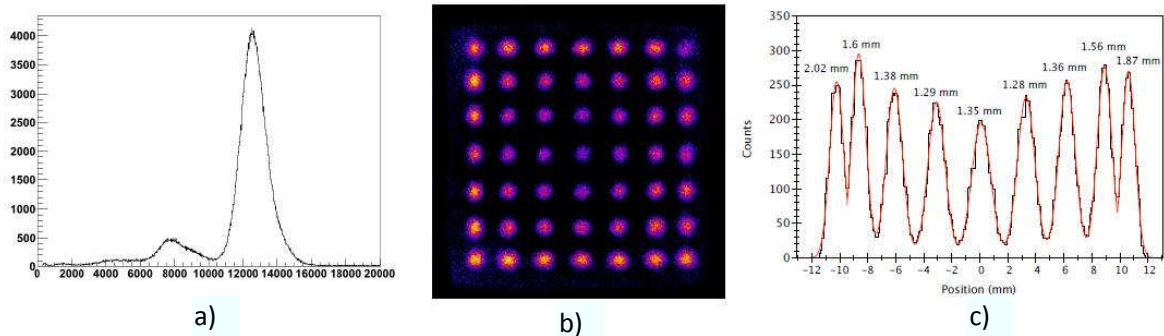


Fig. 5.66 (a) Energy spectrum of a 2 mm collimated ^{57}Co source; (b) Image of a matrix of 0.5 mm collimated ^{57}Co irradiation spots spaced 3 mm apart; (c) Profile of a line of 9 pinpoints ^{57}Co sources spaced 3 mm apart (Vandenbussche & al., 2013)

5.2.5. Summary on SiPM's for radio guided surgery

The SiPM arrays present important advantages as compactness and versatile design flexibility making them very attractive for building pixellated photon detectors required in high resolution imaging systems. In this section, monolithic arrays of 16(4x4) channels, each channel representing a SiPM of 3x3 mm² produced by Hamamatsu, have been evaluated for building an intra-operative hand-held gamma camera. The full system will integrate 256 channels (16 monolithic arrays), covering a field of view of 25 cm².

The project is still under development. However, preliminary evaluation of these devices shows very suitable characteristics for such kind of application.

An IV characterization carried out on 23 arrays (i.e. 368 SiPM's) allowed a very fast and easy analysis of the breakdown voltage uniformity. The breakdown voltage spreading has been found to be ~1.5V, showing that devices coming from different production runs can provide quite large non-uniformities (i.e. 1.5V difference in the operating voltage is expected to imply more than 50% of gain variation). To assure a good uniformity of our detection system, a preliminary selection of the arrays has been done, reducing the breakdown voltage spreading down to ~800 mV. The remaining difference will be compensated by a 8-bit digital-to-analog converter embedded in each channel input of

the ASIC. The uniformity of other electrical and optical characteristics of the selected arrays is presently under evaluation.

In the framework of this project, miniaturized read-out electronics based on a dedicated ASIC and test-boards with a high density of components has been developed. The full field of view of the camera is assured by 4 elementary modules, each of them composed of 4(2x2) arrays. Therefore, our goal in terms of miniaturization of the readout electronics was to match the size of the 2x2 SiPM arrays (28.6x27.2 mm²). This compact package enables to assemble side-by-side several elementary modules with no dead-zone in order to obtain versatile photo-detection system with various fields of view, while optimizing their compactness.

The preliminary evaluation of the elementary module for gamma imaging shows very promising results. Even if the full detection system was not yet not optimized (the scintillating crystal plate is four times larger than the surface of the elementary photo-detection module) the performances are very close to that achieved by coupling a Hamamatsu H9500 flat panel to the same type of scintillating crystal (energy resolution of 12.9% and spatial resolution of ~1 mm).

Further investigations will include the evaluation of the temperature-compensated circuit allowing to assure a constant overvoltage, therefore an uniformity of overall SiPM parameters over different SiPM devices building the camera. The overall performances of the full gamma system imaging of 25 cm² in terms of energy and spatial resolution will be also evaluated.

References on SiPM arrays and medical applications:

- ATLAS collaboration; (2007); ATLAS forward detectors for measurement of elastic scattering and luminosity determination, *Technical design report*, CERN/LHC/2007-004.
- Barrillon, P.; Blin, S.; Caceres, T.; Heller, M.; De La Taille, C.; Puzo, P.; Seguin-Moreau, N.; (2007); MAROC: Multi-Anode ReadOut Chip, *Topical Workshop on Electronics for Particle Physics*, Praga Czech Republic.
- Callier, S.; De la Taille, C.; Martin-Chassard, G.; Raux, L.; (2011), EASIROC, an easy and versatile ReadOut device for SiPM, *Physics Procedia*, TIPP 2011.
- Chien KR.; (1996); Genes and physiology: molecular physiology in genetically engineered animals, *Journal of Clinical Investigation*; Vol. 97, Issue 4, 901-909.
- Daghighian, F.; Mazziotta, J.C.; Hoffman, E.; Shenderov, P.; Eshaghian, B.; Siegel, S.; Phelps M.E.; (1993); Intraoperative beta probe: a device for detecting tissue labeled with positron or electron emitting isotopes during surgery, *Medical physics*, Vol. 21, Issue 1, 153-157.
- Dinu, N.**; Barrillon, P.; Bazin, C.; Belcari, N.; Bisogni, M.G.; Bondil-Blin, S.; Boscardin, M.; Chaumat, V.; Collazuol, G.; De la Taille, C.; Del Guerra, A.; Llosa, G.; Marcatili, S.; Melchiorri, M.; Piemonte, C.; Puill, V.; Taroli, A.; Vagnucci, J.F.; Zorzi, N.; (2009, a); Characterization of a prototype matrix of SiPM, *Nuclear Instruments and Methods in Physics Research A*, Vol. 610, 101-104.
- Dinu, N.**; Barrillon, P.; Bazin, C.; Belcari, N.; Bisogni, M.G.; Bondil-Blin, S.; Boscardin, M.; Chaumat, V.; Collazuol, G.; De la Taille, C.; Del Guerra, A.; Llosa, G.; Marcatili, S.; Melchiorri, M.; Piemonte, C.; Puill, V.; Taroli, A.; Vagnucci, J.F.; Zorzi, N.; (2009, b); Characteristics of a prototype matrix of silicon photomultipliers (SiPM), Pixel 2008 international workshop, Fermilab, *Journal of Instrumentation*, Vol. 4, P03016.
- Eriksson, L.; Townsend, D.; Eriksson, M.; Melcher, C.; Schmand, M.; Bendriem, B.; Nutt, R.; (2004); Experience with Scintillators for PET: towards the fifth generation of PET scanners, *Nuclear Instruments and Methods A*, Vol. 525, 242-248.
- Del Guerra, A.; Bartoli, A.; Belcari, N.; Herbert, D.; Motta, A.; Vaiano, A.; Di Domenico, G.; Sabba, N.; Moretti, E.; Zavattini, G.; Lazzarotti, M.; Sensi, L.; Larobina, M.; Uccelli, L.; (2006); Performance evaluation of the fully engineered YAP-(S)PET scanner for small animal imaging, *IEEE Transactions on Nuclear Science*, Vol. 53, Issue 3, 1078-1083.
- Harris, C.C.; Bigelow, R.R.; Francis, J.E.; Kelley, G.G.; Bell, P.R.; (1956); A Cs(Tl)-crystal surgical scintillation probe, *Nucleonics*, Vol. 14, 102-108.
- Hoffman, E.J.; Tornai, M.P.; Janecek, M.; Patt, B.E.; Iwanczyk, J.S.; (1999); Intraoperative probes and imaging probes, *European Journal of Nuclear Medicine*, Vol. 26, Issue 8, 913-935.

- Huber, J.S.; Moses, W.W.; Derenzo, S.E.; Ho, M.H.; Andreaco, M.S.; Paulus, M.J.; Nutt, R.; (1997); Characterisation of a 64 channel PET detector using photodiodes for crystal identification, *IEEE Transactions on Nuclear Science* NS-44, 1197-1201.
- Humm, L.J.; Rosenfeld, A.; del Guerra A.; (2003); From PET detectors to PET scanners; *European Journal of Nuclear Medicine and Molecular Imaging*, Vol. 30, 1574-1597.
- Llosá, G.; Battiston R.; Belcari, N.; Boscardin, M.; Collazuol, G.; Corsi, F.; Dalla Betta, G.-F.; Del Guerra, A.; **Dinu, N.**; Levi, G.; Marcatili, S.; Moehrs, S.; Marzocca, C.; Piemonte, C.; Pozza, A.; (2008, a); Novel Silicon Photomultipliers for PET applications, *IEEE Transactions on Nuclear Science*, Vol. 55, Issue 3, 877-881.
- Llosá, G.; Belcari, N.; Bisogni, M.G.; Del Guerra, A.; Marcatili, S.; Barrillon, P.; De la Taille, C.; Bondil-Blin, S.; **Dinu, N.**; Melchiorri, M.; Piemonte, C.; Zorzi, N.; (2008, b); Evaluation of the first silicon photomultiplier matrices for a small animal PET scanner, *IEEE Nuclear Science Symposium Conference Record NSS '08*, 3574-3580.
- Llosá, G.; Belcari, N.; Bisogni, M.G.; Collazuol, G.; Marcatili, S.; Boscardin, M.; Melchiorri, M.; Taroli, A.; Piemonte, C.; Zorzi, N.; Barrillon, P.; Bondil-Blin, S.; Chaumat, V.; De la Taille, C.; **Dinu, N.**; Puill, V.; Vagnucci, J.-F.; Del Guerra, A.; (2009, a); First results in the application of silicon photomultiplier matrices to small animal PET, *Nuclear Instruments and Methods*, Vol. 610, 196-199.
- Llosá, G.; Belcari, N.; Bisogni, M.G.; Collazuol, G.; Marcatili, S.; Barrillon, P.; De la Taille, C.; Bondil-Blin, S.; **Dinu, N.**; Melchiorri, M.; Taroli, A.; Piemonte, C.; Del Guerra, A.; (2009, b); Energy, timing and position resolution studies with 16-pixel Silicon Photomultiplier matrices for small animal PET; *IEEE Transactions on Nuclear Science*, Vol. 56, Issue 5, 2586-2593.
- Lecomte, R.; (2004); Technology challenges in small animal PET imaging, *Nuclear Instruments and Methods A*, Vol. 527, 157-165.
- Mathelin, C.; Piqueras, I.; Guyonnet, J.L.; (2006); Performances of a new prototype handheld scintillating probe for the search of sentinel lymph nodes in breast cancer, *European Journal of Surgical Oncology*, Vol. 32, Issue 1, 24-28.
- McElroy, D.P.; Pimpl, W.; Pichler, B.J.; Rafecas, M.; Schuler, T.; Ziegler, S.I.; (2005); Characterization and readout of MADPET-II detector modules: validation of a unique design concept for high resolution small animal PET, *IEEE Transactions on Nuclear Science*, Vol. 52, Issue 1, 199-204.
- Moses, W.W.; Derenzo, S.; Budinger, T.; (1994); PET detector modules based on novel detector technologies, *Nuclear Instruments and methods A*, Vol. 353, Issues 1-3, 189-194.
- Moses, W.W.; (2003); Time of flight in PET revised, *IEEE Transactions on Nuclear Science*, Vol. 50, Issue 5, 1325-1330.
- Paigen K.; (1995); A miracle enough: the power of mice. *Nature Medicine*, Vol. 1, Issue 3, 215-220.
- Phelps ME.; (2000); PET: the merging of biology and imaging into molecular imaging; *The Journal of Nuclear Medicine*; Vol. 41; Issue 4; 661-681.
- Pidol, L.; Kahr-Harari, A.; Viana, B.; Virey, E.; Ferrand, B.; Dorenbos, P.; de Haas, J.T.M.; van Eijk, C.W.E.; (2004); High efficiency of lutetium silicate scintillators, Ce-doped LPS, and LYSO crystals, *IEEE Transactions on Nuclear Science*, Vol. 51, Issue 3, 1084-1087.
- Piemonte, C.; Boscardin, M.; Dalla Betta G.F.; Melchiorri, M.; Zorzi, N.; Battiston, R.; Del Guerra, A.; Llosa, G.; (2007); Recent developments on silicon photomultipliers at FBK-irst, *IEEE Nuclear Science Symposium Conference Record*, N41-2, 2089-2092.
- Raylman, R.R.; Wahl, R.; (1996); Evaluation of ion-implanted-silicon detectors for use in intra-operative positron-sensitive probes, *Medical Physics*, Vol. 23, Issue 11, 1889 - 1895.
- Raylman, R.R.; (2000); A solid-state intra-operative beta probe system, *IEEE Transactions on Nuclear Science*, Vol. 47, Issue 4, 1696 - 1703.
- Sarikaya, I.; Sarikaya, A.; Reba, RC.; (2008); Gamma probes and their use in tumor detection in colorectal cancer, *International Seminars in Surgical Oncology*, Vol. 5:25, 1-8.
- Seguin-Moreau, N.; Bouchel, M.; De la Taille, C.; Martin-Chassard, G.; Gaglione, R.; Laktineh, I.; Mathez, H.; Brient, J.C.; Jauffret, C.; (2007); HARDROC, HAdronic Rpc Detector Read-Out Chip, TWEPP-07 *Topical Workshop on electronics for Particle Physics Search*, Prague.
- Shao, Y.; Cherry, S.; Chatzioannou, A.; (2005); Design and development of 1mm resolution PET detectors with position sensitive PMT's, *Nuclear Instruments and Methods A*, Vol. 447, 486-490.
- Sweet, W.H.; (1951); The uses of nuclear disintegration in the diagnosis and treatment of brain tumor, *The New England journal of medicine*, Vol. 245, 875-887.
- Tavernier, S.; Bruyndonckx, P.; Leonard, S.; Devroede, O.; (2005); A high-resolution PET detector based on continuous scintillators, *Nuclear Instruments and Methods A*, Vol. 537, Issues 1-2, 321-325.

- Thompson, C.J.; James, S.; Tomic, N.; (2005); Under-sampling in PET scanners as a source of image blurring; *Nuclear Instruments and Methods A*, Vol. 545, 436-445.
- Valk, P.E.; Bailey, D.L.; Townsend, D.W.; Maisey, M.N.; (2003); Positron Emission Tomography, *Springer*.
- Zanzonico, P.; Heller, S.; (2000); The intraoperative gamma probe: basic principles and choices available, *Seminars in Nuclear Medicine*, Vol. XXX, Issue 1, 33-48.
- Vandenbussche, V.; Imando, T.; **Dinu, N.**; Pinot, L.; Callier, S.; Janvier, B.; Cuisy, D.; Dinaucourt, P.; Gaspard, M.; Puill, V.; Raux, L.; Sliwa, R.; Trochet, S.; Charon, Y. ; Duval, M-A. ; Hudin, N.; Verdier, M-A.; Menard, L. ; (2013) ; Miniaturized multi-channels SiPM read-out electronics for the development of compact gamma cameras, *Abstract submitted to IEEE NSS 2013*.
- Yamamoto, S.; Matsumoto, K.; Sakamoto, S.; Tarutani, K.; Minato, K.; Senda, M.; (2005); An intra-operative positron probe with background rejection capability for FDG-guided surgery, *Annals of Nuclear Medicine*, Vol. 19, Issue 1, 23-28.
- Yamamoto, S.; Higashi, T.; Matsumoto, K.; Senda, M.; (2006); Development of a positron-imaging detector with background rejection capability, *Annals of Nuclear Medicine*, Vol. 20, Issue 10, 655-662.
- Yamamoto, S.; Imaizumi, M.; Warabe, T.; Watabe, H.; Kanai, Y.; Shimosegawa, E.; Hatazawa, J.; (2010); Development of a SiPM-based high-resolution PET system for small animals, *Physics in Medicine and Biology*, Vol. 55, 5817-5831.

6. Summary and perspectives of silicon light detectors for medical imaging

Based on the p - n junction properties, silicon detectors for visible light detection constitute a replacement technology having many practical advantages over the vacuum tube devices.

The primary developed detectors were PN and PIN photodiodes and they are still standing today as the simplest and cheapest solid state visible light detector. Since their appearance, they have been used in different applications where photomultiplier tubes were too bulky or delicate, or where high voltages were not possible or where the insensitivity to magnetic fields was required. However, PIN diodes are severely limited by their complete lack of internal gain. Even with state of the art very low noise amplifiers, the noise is at the level of several hundred electrons and consequently the smallest detectable light flash needs to consist of few hundred photons.

The avalanche photodiode (APD) represents a more recent technology where the reverse bias voltage is raised to a point where impact ionization allows for some internal multiplication. In this way, a gain of around a 100 is achieved for a bias of 100-200V. With special manufacture, it is possible to reach gains of several thousand using a HV bias > 1500V. The internal gain of APD's improves the signal-to-noise ratio, but still about tens of photons are needed for a detectable light pulse.

Some APD's devices have been tested as photon detectors in medical applications (i.e. Positron Emission Tomography). But since a quite high bias voltage is needed and the gain is strongly dependent on it, stability problems on bias supply can limit the performances of a complete PET system.

Moreover, arrays of such devices are still not available on the market, limiting the uses of these detectors in applications where large detection area is required.

Pioneering work in the development of silicon devices detecting single photons was done in the years 1960s by R.J. McIntyre in the RCA Company and by R.H. Haitz in the Shockley Research Laboratory.

Such devices work in Geiger-mode (GM-APD), employing the breakdown mechanism to achieve a high gain. The p - n junction is designed in such way that it can sustain a reverse bias beyond its nominal breakdown voltage, creating the necessary high field gradients across the junction. In this way, the electron-hole pair created by an incident photon in the high field region can trigger a self-sustaining ionization cascade that will spread throughout the silicon volume subjected to the field. The silicon will breakdown and become conductive, effectively amplifying the original charge into a macroscopic current flow. To be able to detect a new photon, the flow of the current should be stopped or 'quenched'. Passive quenching circuit, achieved through the means of a series resistor, is the most simple and frequently used quenching mechanism.

Therefore, the GM-APD's is the first detector in solid-state technology providing a high gain of the order of 10^6 comparable to the one of the photomultiplier tube build in vacuum technology. It functions as a binary device, in either an "on" or "off" state, allowing for single photon detection, the reason for which it is also called Single Photon Avalanche Diode (SPAD). However, it has a limited dynamic range and it cannot provide proportional information regarding the magnitude of an instantaneous photon flux.

To overcome the lack of the proportionality, Silicon Photomultiplier (SiPM) has been proposed in the years 1980 by Russian scientists V.M. Golovin and Z. Sadygov. Such device integrates a dense array of small micro-cells, where each micro-cell is a GM-APD connected in series with its integrated quenching resistor. In this way, each

micro-cell detects photons identically and independently and the signals of all micro-cells are summed to form the output of the SiPM, being therefore capable of giving information on the magnitude of an incident photon flux.

Such device presents important advantages such as low operating voltage ($< 75\text{V}$), low power consumption ($< 75\text{ }\mu\text{W}/\text{mm}^2$), insensitive to magnetic fields up to 15T, high gain ($\sim 10^6$), large dynamic range, good photon detection efficiency ($\sim 40\%$) and it is intrinsically very fast (single photon timing resolution $< 50\text{ ps}$). It is mechanically robust, presenting compactness and ruggedness and allowing for a flexible design in applications where large detection area is required.

Around ten years ago, when I have read for the first time the article of B. Dolgoshein “An advanced study of Silicon Photomultiplier” (ICFA Instrumentation Bulletin, 2001) I was very impressed by the great perspectives opened by the SiPM detectors in different field of applications. At that time, the SiPM was like a “mirage” detector, since only few Russian devices were available over the entire world and it was quite impossible to obtain a unique device. Today, only ten years later, these devices are produced by most important silicon detectors companies, in different shapes and sizes. The application area of these detectors was extensively enlarged, covering today the field of medical imaging, high energy physics, hazard and threat detection, biophotonics and LiDAR.

I hope that my modest work in this field, performed during last ten years and presented in this dissertation, contributed to the understanding of physics mechanisms of this device, its main characteristics and properties. Our future projects involving these detectors in medical imaging applications will allow me to use the knowledge acquired over many years of laboratory and instrumentation work and to bring further their development.

I tried to be as exhaustive as possible in this dissertation, showing not only my personal work but giving in the same time an overview of the technology and physics state-of-the art. However, the research field of SiPM detector is still largely open. Detailed physical models of avalanche multiplication are still missing. Such models may be of help in ultra-fast timing applications where there is still room for device improvement. Techniques for reducing long timing tails might be also exploited. The physical models might be also of help in further reducing dark count rate and afterpulsing contributions. The photon detection efficiency is another important parameter still open for improvements in UV, VUV as well as NIR and IR regions. Temperature dependence of different parameters and the way in which they can be tuned for stable operation should still be exploited.

When used in applications where broad detection field of view is required, large monolithic devices with uniform electrical and optical characteristics and very low dead area are desired. In this case, the development of dedicated ASIC's involving multi-channel readout electronics becomes an important task. Moreover, to reduce the dead-area between different channels and build compact detection systems, 3D interconnection technology would be very useful.

Digital Silicon Photomultiplier (dSiPM), recently developed by Philips Company, open a new perspective of building large areas of GM-APD or SPAD arrays, implying a digital reading of each element. However, these devices have still a high level cost and are distributed in compact close systems dedicated to a particular application, limiting its accessibility for research laboratories development.

PART B

Experimental techniques for doping profiles measurements in pixels detectors

Table of contents: PART B

1. Introduction	139
2. Experimental techniques for doping profiles measurements	141
2.1 Secondary Ion Mass Spectrometry (SIMS)	141
2.2 Spreading Resistance Profiling (SRP)	144
3. From experimental data to TCAD simulations of ATLAS pixels silicon detectors	147
3.1 ATLAS pixel silicon sensors	147
3.2 Doping profiles of ATLAS planar pixel sensors	153
3.2.1 Experimental results	153
3.2.2 Comparison with TCAD simulations	159
3.3 Static characterization	160
3.3.1 Set-up calibration	160
3.3.2 Guard-ring potential: measurement and comparison with TCAD simulations	162
3.4 Optimization of guard-ring region by TCAD simulations	163
4. Summary and perspectives of Part B	167
References on Part B	168

1. Introduction

In the Part A of my dissertation (Sections 3.2 & 3.2.1., 3.2.3., 4.2 & 4.2.1, Annexes 1 & 2) we could observe the important relationship between the semiconductor physics (i.e. electrical field, charge carrier distribution, the density of defects or impurities) and the electrical characteristics of Geiger-mode avalanche detectors as well as of the Silicon Photomultiplier detectors (i.e. avalanche breakdown voltage, gain and dark count rate, triggering probabilities, temperature dependence). Therefore, it becomes evident that an improvement of the final characteristics of such detectors can be obtained with a better understanding of the underlying physics.

The main tool for understanding the relationship between semiconductor physics and the electrical behavior of semiconductor detectors is the Computer-aided design (CAD) for semiconductor devices, referred as Technology CAD (TCAD). The origin of the modern TCAD programs can be found in the work of Prof. R. Dutton and his research group at Stanford University (Stanford Technology). More developed and modern versions exist today such as SILVACO (Silvaco Inc.) and SYNOPSIS (Synopsys Inc.). They are widely used in the semiconductor industry to understand the underlying semiconductor physics but also to reduce the device development time and cost.

Independent of the TCAD tool, such programs consist of two different simulators: one for the semiconductor manufacturing process and the other for the device behavior. The process simulator simulates the semiconductor processing steps such as: oxidation, deposition, etching, ion implantation and annealing and generates inputs to the device simulator as realistically as possible. The device simulator calculates the macroscopic electrical parameters by solving different equations such as: the Poisson equation of the electrical potential, the current density continuity equations of electrons and holes and the drift diffusion and heat conduction equations as a function of lattice temperature. The device simulators incorporate semiconductor physics models for transport, mobility, generation-recombination (Shockley-Read-Hall), Auger, impact ionization, trap, surface, etc., using default models and their parameters. The core of the simulation is the finite-element method, representing the cutting-edge advantage of the TCAD.

Besides of their many advantages, the TCAD programs (in particular the process simulator) are intended to be used by experts who have access to the process details and its parameters. Such technological parameters are usually not published or they are known with large uncertainties (few orders of magnitude). Therefore, realistic TCAD simulations are very difficult to be realized by physicists or engineers working in research laboratories. However, in the last years, an increase interest has been shown from physicist's side to use TCAD tools for simulation of silicon detectors (i.e. Journée de la simulation, IN2P3-IRFU, <https://indico.in2p3.fr/conferenceDisplay.py?confId=8337#>). Such interest is related to the fact that they are representing the main users of such devices and they need to improve the electrical or geometrical characteristics of the detectors to achieve the best physics performances of their experiments. *Therefore, the knowledge of the most important technological parameters has a critical importance when TCAD simulation of silicon detectors is performed* by "non-expert" people in semiconductor technology.

The ion implantation represents one of the key technological steps used in silicon detectors manufacturing, in particular for selective doping of the p and n-type geometrical defined regions in mono-crystalline substrate silicon. Boron (for the p-type regions) and arsenic, phosphorus and antimony (for n-type regions) are the preferred dopant atoms that are currently used. The ion implantation is usually used for pn junction realization of the visible light silicon detectors (i.e. APD, GM-APD, SiPM) as well as for other types of silicon detectors using the pn junction properties for detection of charged particles (i.e. micro-strip silicon detectors, pixel detectors).

To restore the mono-crystalline state and to locate the dopant atoms at lattice sites (electrical activation), the ion implantation is followed by a thermal step called annealing. Both ion implantation and annealing steps define the pn junctions doping profile, directly related to the underlying semiconductor device physics and its electrical characteristics. Therefore, approaching such technological parameters as close as possible to the real values allows calibrating the TCAD tool and provides silicon detector models as close as to the reality.

The natural way to characterize the ion implantation and annealing steps is by experimental measurement of the doping profiles. Two different types of dopant profiles can be defined: (1) the total *atomic* dopant profile of the overall implanted ions and (2) the *active* dopant profile of implanted ions located in the lattice sites and participating to the electrical properties of the device. Usually, a difference between total and active dopant profiles is observed, in particular at both surface and far tail region (i.e. damage of the silicon lattice during implantation, channeling). Therefore, both profiles should be measured to provide a full understanding of the device characteristics.

Accurate and reliable one-dimensional experimental techniques used for dopant profile measurement are: Secondary Ion Mass Spectrometry (SIMS) used for total *atomic* dopant profile characterization and Spreading Resistance Profiling (SRP) used for *active* dopant profile characterization. Presently, new 2D techniques are under development as for example Scanning Spreading Resistance Microscopy (SSRM) or Scanning Capacitance Microscopy (SCM) to provide a better spatial resolution (less than 1 nm), in particular required for *active* dopant profile characterization of very small devices (i.e. highly segmented silicon detectors, CMOS devices, superconducting junctions).

The Part B of my dissertation will show how these experimental techniques can be successfully applied for the characterization of doping profiles of silicon detectors and provide very useful information for the calibration of the TCAD simulation of such devices.

In particular, I will describe my personal work performed at LAL in the framework of the ATLAS physics group, in parallel with the SiPM activity. This is related to the use of previously mentioned experimental techniques (SIMS, SRP) for the doping profiles measurement of the ATLAS planar silicon pixel detectors. The extracted technological parameters have been successfully used to calibrate the TCAD simulations and to predict planar pixel detectors with improved geometrical and radiation hardness characteristics for the upgrade of the ATLAS inner detector at high-luminosity LHC. This work has been particularly addressed in the PhD thesis of M. Benoit (defended at LAL in 2011), PhD student that I co-supervised together with Dr. A. Lounis, the LAL-ATLAS pixel group coordinator.

2. Experimental techniques for doping profiles measurements

The one-dimensional experimental techniques mostly used in the semiconductor industry for the characterization of doping profiles are Secondary Ion Mass Spectrometry (SIMS) and Spreading Resistance Profiling (SRP). The SIMS and SRP methods are based on quite expensive facilities. Therefore, to apply these methods to the doping profiles characterization of ATLAS planar pixel detectors we relied on existing facilities, outside of our laboratory:

- SIMS instrument from the laboratory Groupe d'Etude de la Matière Condense (GEMaC) (CNRS, Meudon);
- SRP instrument from Evans Analytical Group (EAG, United Kingdom), a laboratory specialized in materials characterization.

The physics principle of these methods will be presented in the following.

2.1 Secondary Ion Mass Spectrometry (SIMS)

SIMS is a vacuum based technique which relies on the bombardment of a sample surface with a primary ion beam followed by mass spectrometry of the emitted ionized secondary ions. Details on the basic concepts and instrumental aspects can be found in the reference Benninghoven & al., 1987. In this document, only a general description of the SIMS system and the analysis method will be given. The main elements of a SIMS facility are presented in Fig. 2.1 and they are described as:

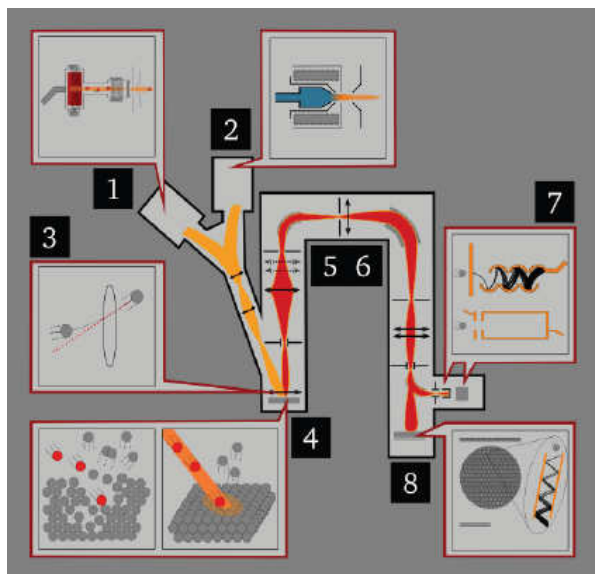


Fig.2.1. Schematic principle of SIMS technique

- *Primary ion source (1&2)*, represented by a duoplasmatron (1) or a surface ionization primary (2) ion source; the duoplasmatron is used to produce O_2^+ primary ions, dedicated particularly to the analysis of the electropositive elements (i.e. boron and aluminum in silicon); the surface ionization source is used to produce Cs^+ primary ions, dedicated particularly to the analysis of the electronegative elements (i.e. phosphorus, antimony, arsenic in silicon); the extracted energies of the primary beam are between 1 and 30 keV (i.e. usually 10 keV);
- *Primary ion column (3)*, containing a primary beam mass filter (i.e. electrostatic lenses and apertures) that transmits only ions with a specified mass-to-charge (m/z) ratio and eliminates impurities species as Cr, Fe, Ni etc.; beam impact energy on the sample is in the range 5 to 15 keV and various incident angles are available (i.e. 27° or 46°);
- *Secondary ion extraction and transfer (4)* is carried out once the primary beam is impinging on the sample surface; the mass spectrometer components are held at ground potential and the sample must be held at high voltage (i.e. typical accelerating potential is ± 4.5 kV);

- *Ion energy analyzer (5)* based on an electrostatic analyzer, which bends the lower energy ions more strongly than higher energy ions, and an energy slit that can be set to intercept the high energy ions;
- *Mass analyzer (6)* based on a magnetic sector or quadrupole (i.e. magnetic sector is most commonly used); secondary ions with a given mass-to-charge (m/z) ratio are selected at the exit slit from the magnetic field;
- *Secondary ion detectors (7&8)* including ion counting electron multiplier, a Faraday cup (7) and two ion image detectors (8). Ion counting electron multipliers (i.e. series of dynodes connected along a resistor string) are the most effective detectors, they operating with sufficiently high gain to produce a detectable pulse for every ion arrival; the Faraday cup is just an electrode from which electrical current is measured while a charge particle beam (electrons or ions) impinges on it. Ion image detectors are based on arrays of micro-channel plate electron multipliers; two combined micro-channel plates assuring a high gain (i.e. $1e6$) followed by a phosphor screen for visualizing the electron cascade provides an easy way to monitor the secondary ion beam.

Monitoring the secondary ion count rate of selected elements as a function of time leads to depth profiles. The depth profile can be seen as a plan by plan cutting out over the sample volume coupled with a measurement in each plan of the desired element concentration.

To convert the time into depth, a profilometer for measuring the sputter crater depth can be used. A profilometer is a separate instrument (see Fig. 2.2.) that determines depth by dragging a stylus across the crater and noting the vertical deflections. Fig. 2.3 presents a profilometer picture of the crater shape after a SIMS measurement. Total crater depth divided by total sputter time provides the average sputter rate. Then, the depth of each measurement point is calculated as average sputter rate multiplied by the time corresponding to the measured point.

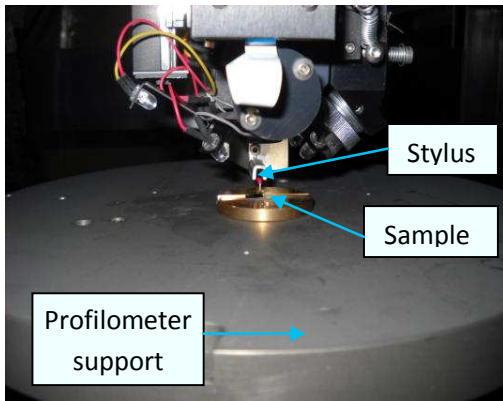


Fig.2.2. Zoom of the head and the support of a mechanical profilometer

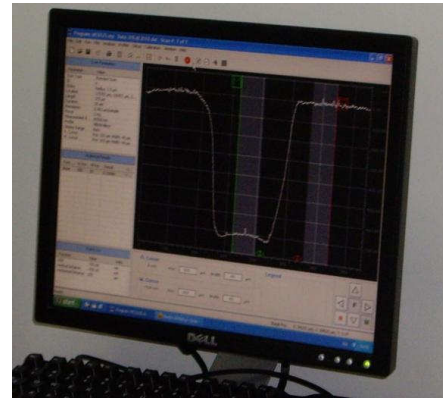


Fig.2.3. Crater shape of a SIMS measurement measured by a profilometer

The ion counts are converted in dopant concentration with the help of relative sensitivity factors (RSF) by using the following expression:

$$C_E = RSF \cdot \frac{I_E}{I_M} \quad 2.1$$

where C_E is the concentration of the element E of interest, I_E is the secondary ion intensity for the element E and I_M is the secondary ion intensity for the matrix element (the major or matrix element is usually chosen as reference; in our samples the matrix element is Si).

The RSF is a function of the element of interest (i.e. B or P) and the sample matrix (i.e. Si). Therefore, the RSF factors have to be determined by measuring calibrated samples (Si samples implanted with a well known dose of the analyzed element) and then calculated as following:

$$RSF = C_I \cdot \frac{I_M}{I_{Eref}} = \frac{dose}{crater \ depth} \cdot \frac{I_M}{I_{Eref}} \quad 2.2$$

where all elements are referred on the measurement performed on the calibrated sample: I_M is the secondary ion intensity for the matrix element, I_{Eref} is the secondary ion intensity for the element E, the crater depth is measured by a profilometer and the implanted dose is known. Low values of RSF and high ion yield or secondary ion intensity determines good detection sensitivity (i.e. $RSF_B^{Si} \sim 6 \times 10^{22} \text{ at.cm}^{-3}$; $RSF_P^{Si} \sim 3 \times 10^{21} \text{ at.cm}^{-3}$).

An example of SIMS row data (i.e. secondary ions vs. time) for the measurement of phosphorus in a silicon matrix is presented in Fig. 2.4 a) (calibrated sample measured with Cameca IMS 4F SIMS system at CNRS Meudon). The sample was prepared by ion implantation of P with energy of 150 keV and dose of $1 \times 10^{14} \text{ cm}^{-2}$. The SIMS measurement has been carried out with Cs+ primary ions and negative secondary ions have been analysed.

The time axis has been converted into depth by profilometer measurement of sputter crater depth. At the end of the above P depth profile, profilometer gives 853 nm, determining an average sputter rate of $\sim 4.8 \text{ \AA/s}$. The measured matrix current was of $\sim 1.272 \times 10^5$ silicon ion counts/s, determining a RSF factor of $2.963 \times 10^{21} \text{ at/cm}^3$. Consequently, the corresponding P depth profile plotted on depth and concentration axes is presented in Fig. 2.4 b).

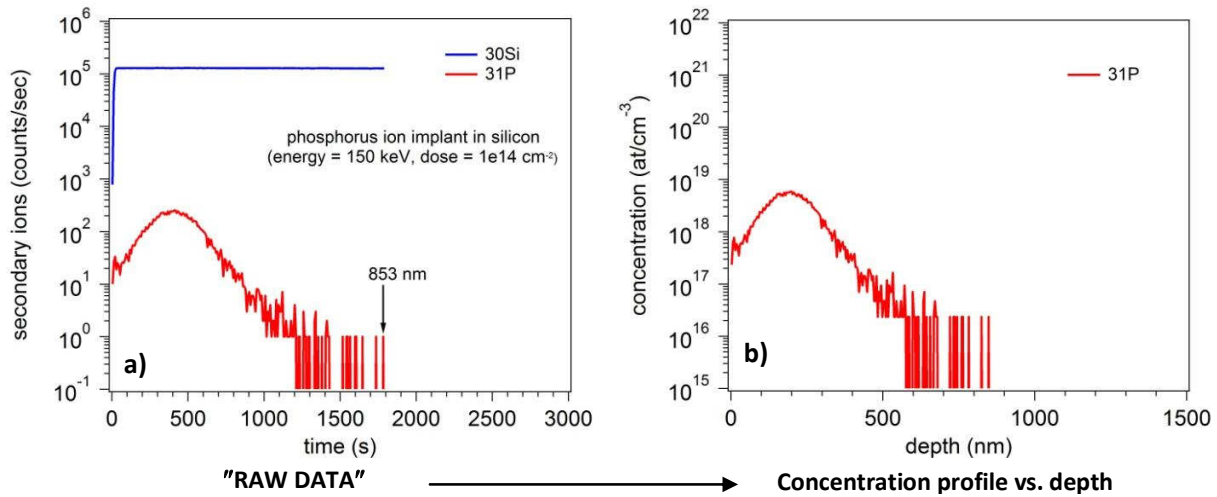


Fig.2.4. a) Row data of a SIMS profile corresponding to a phosphorus implant in silicon; **b)** the corresponding dopant concentration profile vs. depth

The SIMS method has good detection sensitivity for many elements (i.e. B, P, Al, As, Ni, Cu, Ge, Si, O, etc.), detecting dopant concentrations as low as 10^{15} at/cm^3 . It allows simultaneous detection of different elements (isotopes can be distinguished), in the surface and near surface regions (up to $30 \mu\text{m}$ depth) and it can provide a depth resolution of 2-5 nm (10-20 nm typical) and a lateral resolution of $20 \text{ nm} - 1 \mu\text{m}$, depending on primary source.

During SIMS measurement a crater is leaved on the sample because of the removing material by sputtering. Therefore, the SIMS method is destructive and usually test samples and not final sample detectors should be used for such characterization. The SIMS measurement determines the total *atomic* dopant profile and not the *active* dopant profile of implanted ions. Therefore for the characterization of the active dopant profiles of implanted ions located in the lattice sites and participating to the electrical properties of the device an electrical method should be used and one of such methods is described in the following section.

2.2 Spreading Resistance Profiling (SRP)

The technique of using spreading resistance to measure the thickness of diffused layers and impurity profiles was originally proposed by Mazur and Dickey (Mazur, 1966). The SRP method relies on using a pair of specially conditioned point contact probes (osmium and/or tungsten carbide) which are stepped across the bevel surface of a semiconductor sample (Fig. 2.5). The probe tips are shaped such that they can be positioned very close together, often with less than 20 μm separation (sometimes much less). Five mV are applied across the probes at each step and the resistance (called spreading resistance) is measured. If the contact area of the probes is made small enough (usually probe contact size is $\sim 2 \mu\text{m}$), the most important component of the measured resistance arises from the current flowing in the immediate vicinity of the probe tip. In this case, a very simplified (approximate) relationship between the measured resistance and the local resistivity can be written as following:

$$R_s = \frac{\rho}{2a} \quad 2.3$$

where R_s is the measured spreading resistance (ohms), ρ is the local resistivity (ohm-cm), a is the radius of the contact area of the probe (cm). In practice, calibrated samples with well known resistivity are measured and the contact radius “ a ” is an empirical quantity calculated for given experimental conditions (i.e. given pair of probe tips, probe loading). For a contact radius of 1 μm , equation 2.3 predicts $R_s \cong 2500 \cdot \rho$. The higher value of the spreading resistance with respect to ρ ($\sim 10^4$ times higher) is determined by the fact that R_s dominates over probe resistance and contact resistance.

Unless the resistivity is extremely high, the majority carrier concentration in a doped silicon sample will exceed minority carrier concentration by orders of magnitude. Therefore, an intimate relationship exists between measured resistivity (i.e. conductivity) and the majority carrier’s concentration:

$$\sigma = \frac{1}{\rho} = Nq\mu \quad 2.4$$

where σ is the conductivity ($\text{ohm}^{-1}\text{cm}^{-1}$), ρ is the resistivity (ohm-cm), N is the majority carriers concentration (cm^{-3}), q is the electron charge ($1.6\text{e-}19 \text{ C}$) and μ is the majority carrier mobility (electrons mobility $\sim 1500 \text{ cm}^2/\text{V.s}$; holes mobility $\sim 450 \text{ cm}^2/\text{V.s}$). Therefore, using the published values of the electrons and holes mobility’s, the carrier concentration can be calculated at each measured step.

The SRP does not sense doping type by itself. As a separate operation, one of the probe tips can be heated and the polarity of the Seebeck voltage can be sensed to determine conductivity type.

As illustrated in the Fig. 2.5, the depth profile resistivity is obtained by using a beveled sample. The sample is prepared by mounting it on a bevel block with melted wax (Fig. 2.6). Blocks with different bevel angles are usually available and are chosen to provide a desired resolution profiling: typically 1° - 5° for junction depths of 1-2 μm and $\leq 0.5^\circ$ for junction depths less than 0.5 μm . The bevel block is inserted into a well-fitted cylinder, and the sample is glided over a plate containing a polishing compound. Next, the sample is positioned in the measurement apparatus with the bevel edge perpendicular to the probe stepping direction.

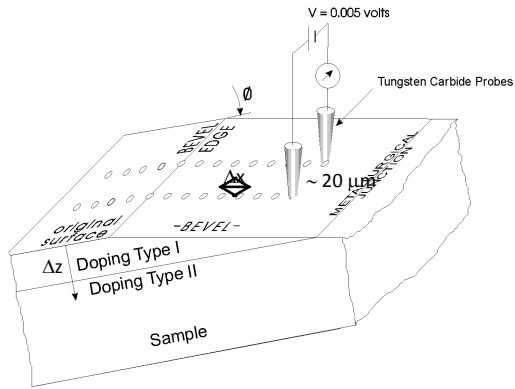


Fig.2.5. Illustration of a Spreading Resistance measurement on a bevel sample

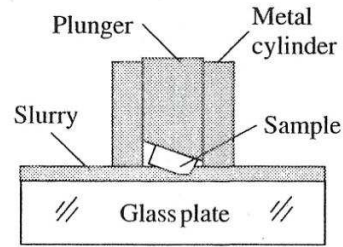


Fig.2.6 Schematic presentation of the bevel sample preparation

Clear bevel edge is required to identify the starting profile. Therefore, it is useful to provide samples with an insulating (oxide or nitride) layer on the surface. The oxide provides a sharp corner at the bevel and also clearly defines the start of the bevel surface because the spreading resistance of the insulator is very high. In the absence of the top oxide, a high quality optical microscope with dark field illumination can be used to identify the bevel edge. The bevel angle should be measured with a well-calibrated mechanical profilometer (as the one presented in Fig. 2.2). Then, the equivalent depth, Δz , at each Δx step along the surface beveled at angle θ is given by:

$$\Delta z = \Delta x \cdot \sin \theta \quad 2.5$$

An example of SRP profile corresponding to a high resistivity implanted layer on a low resistivity substrate (same doping type) is presented in Fig. 2.7 a) (Brennan, Solecon Labs Seminar). Using calibration charts of spreading resistance vs. resistivity (NIST traceable Si bulk standards), the resistivity plot presented in Fig. 2.7 b) is obtained. Then, using the published carrier mobility values, the final carrier concentration profile presented in Fig. 2.7 c) is obtained.

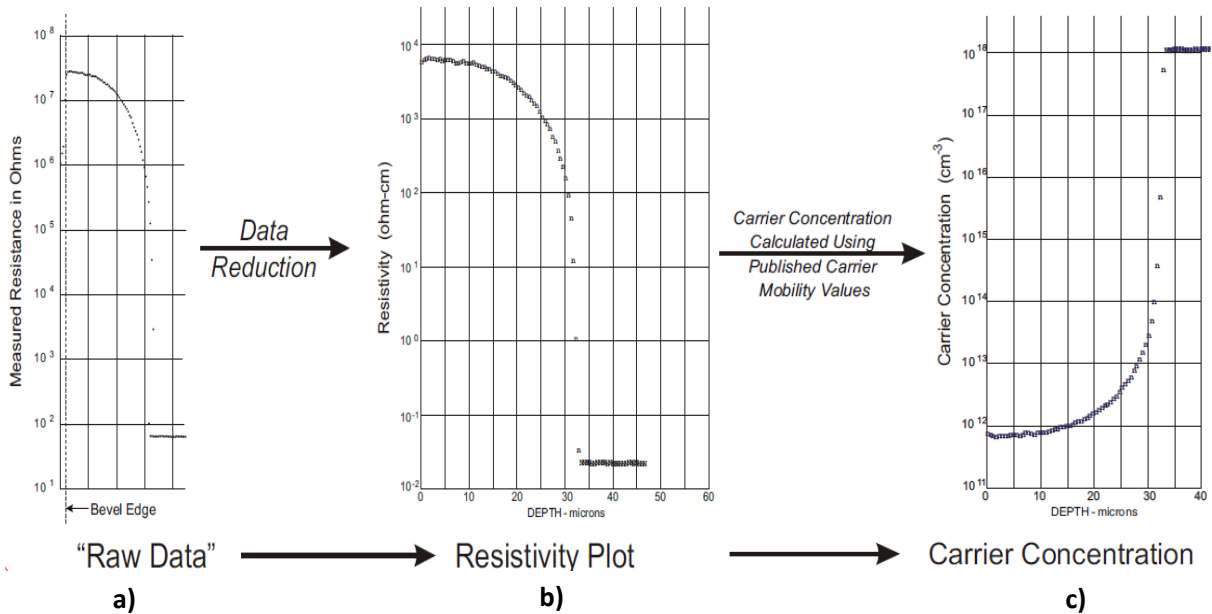


Fig.2.7 Example of **a)** measured spreading resistance profile; **b)** the corresponding resistivity plot and **c)** the final carrier concentration profile calculated using the mobilities values (Brennan, Solecon Labs Seminar)

The spreading resistance analysis provides general purpose and relatively inexpensive depth profiling with excellent depth accuracy. It has high dynamic range ($10^{12} - 10^{21} \text{ cm}^{-3}$) and is capable of profiling shallow junctions into the nm range. The procedure is destructive and technique dependent (i.e. relatively high measurement noise) and as a consequence it implies a given uncertainty in resistivity measurements. Substantial progress has been made in the last years for data collection and treatment. The later relates to improved sample preparation and probes conditioning procedures, universally applicable correction factors with appropriate radius calibration procedures, and the development of physically based Poisson schemes for the correction of the carrier diffusion (spilling) phenomena. Therefore, reproducibility of 10% can be obtained routinely by “qualified” SRP systems.

3. From experimental data to TCAD simulations of ATLAS pixel silicon detectors

As mentioned in the introduction of the Part B, the semiconductor physics of any silicon detector can be studied by using modern technology computer-aided design (TCAD) tools. In this way alternative detector layouts can be investigated, such studies being much cheaper compared with expensive and time consuming productions of real devices. The difficulty of this approach is the abundance of unknown parameters; in particular, the crucial parameters of the doping implants (profile shape, peak local concentration) are fundamental for realistic simulations of the devices under studies, but they are usually not available or known with large uncertainties (few orders of magnitude).

During 2007, a TCAD simulation activity started at LAL (M. Benoit's PhD thesis) for the design of new silicon pixels detectors with improved geometry and radiation hardness for the upgrade of the ATLAS inner detector at high-luminosity LHC. To calibrate these simulations by using realistic technological parameters, I proposed the design of dedicated test-structures for the measurement of doping profiles. These test-structures have been implemented in the mask layout of the ATLAS planar pixels sensors produced between 2009 and 2011 by CiS Institut für Mikrosensorik gGmbH Erfurt, Germany. The SIMS and SRP experimental techniques, described in the Section 2, have been used to measure the doping profiles corresponding to these detectors.

The results of experimental doping profiles measurements will be presented in the Section 3.2. Selected simulations of the ATLAS silicon planar pixel detectors based on the extracted technological parameters will be shown in the Section 3.4. The nice agreement between simulations and electrical parameters of real detectors will be also presented (Section 3.3) For a better understanding of the presented results and their relationship with the physics requirements of the ATLAS experiment, a short introduction on ATLAS experiment and characteristics of its pixel detectors (called in the following **pixel sensors** to distinguish them from the overall ATLAS pixel detector) will be given (Section 3.1).

3.1 ATLAS pixel silicon sensors

The Large Hadron Collider (LHC) at CERN is a proton-proton collider designed to reach an unprecedented high center-to-mass energy (14 TeV) and luminosity ($10^{34} \text{ cm}^{-2} \text{ s}^{-1}$). It operated smoothly from spring 2010 till February 2013 at a center-of-mass energy of 8 TeV and a luminosity which has constantly increased up to $\sim 5\text{-}7 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$ (the corresponding integrated luminosity was more than $\sim 20 \text{ fb}^{-1}$ at the end of 2012).

ATLAS (A Toroidal LHC Apparatus) is a general purpose detector built around one LHC interaction point in order to reconstruct the products of the highly energetic pp collisions (ATLAS collaboration, 2008, a). Its primary goals are to probe the Standard Model predictions at high energy, to discover the Higgs boson and to search for beyond-Standard Model evidences, like super-symmetric particles or other exotic processes. The announcement given by CERN on 4 July 2012 about the discovery of new boson at 125-126 GeV, confirmed today to be the long awaited Higgs particle, was the first fundamental discovery delivered by LHC nowadays.

The Inner Detector (ID) is the ATLAS sub-detector closer to the interaction point. It consists of three nested sub-systems kept under 2T axial magnetic field: a silicon pixel detector (ATLAS Collaboration, 2008, b) at small radial distance r from the beam axis ($50.5 < r < 122.5 \text{ mm}$), double layers of single-sided silicon micro-strip detectors

($299 < r < 560$ mm), and a straw tracker with transition radiation detection capabilities ($563 < r < 1066$ mm). The ID allows an accurate reconstruction of tracks from the primary pp collision region and from secondary vertices due to photon conversion or heavy flavor decays.

To exploit the physics potential of LHC, it is foreseen to upgrade the accelerator chain in three dedicated long shutdowns (LS1, 2 & 3), followed by longer data-taking phases, called phase I, II and III. The LHC baseline plan for the next ten years is given in the Fig. 3.1 (Rossi, 2012). After the accelerator upgrades performed during the LS1 (i.e. consolidate the splice at the magnet interconnects), the maximum centre-of-mass energy of 14 TeV will be reached by the end of 2014 and the nominal design luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ is expected to be reached in 2015. It is estimated that LHC running at nominal luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ will produce about 40 fb^{-1} of integrated luminosity per year. After more accelerator upgrades during LS2 (2018), two times nominal luminosity ($2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$) is expected to be obtained by the end of 2019. The last step upgrade LS3 will require a major effort (new final focusing design) and it will raise the luminosity further up to 5-10 times nominal one, allowing an integrated luminosity of 250 fb^{-1} per year.

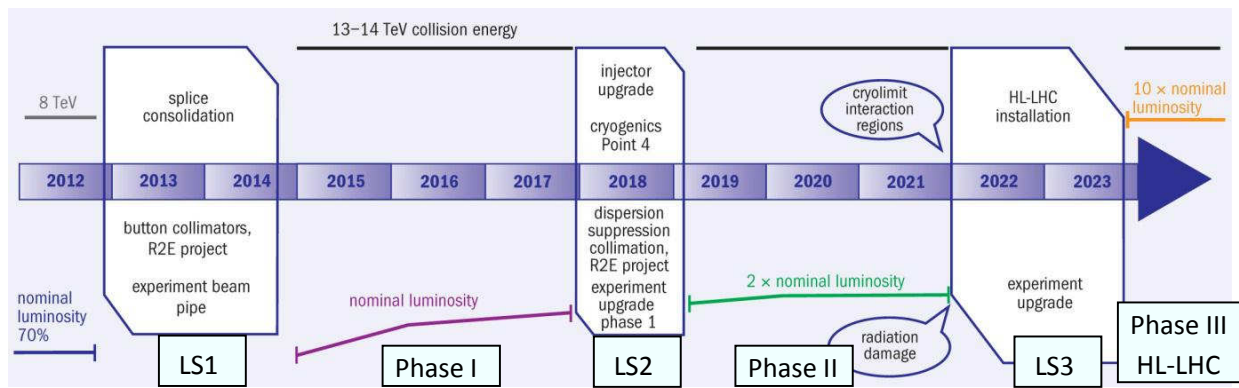


Fig.3.1 LHC baseline plan for the next ten years (Rossi, 2012)

Before the first long shutdown, the LHC run stable collisions at 4 TeV beam energy by using intense bunches with $1.5\text{-}1.6 \times 10^{11}$ protons/bunch spaced at 50 ns (instead of nominal 1.15×10^{11} protons/bunch spaced by 25 ns). The nominal luminosity of $1 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ expected by the end of 2015 associated with a bunch spacing of 25 ns (2808 bunches per beam) at nominal 1.15×10^{11} protons/bunch will give an average value of 27 event/crossing, or pile-up. An increasing of the peak luminosity of $5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ expected after the last long shutdown, based on the hypothesis of 25 ns bunch spacing and increasing the bunch population up to 2×10^{11} protons/bunch, will determine an mean pile-up of ~ 128 , and up to one third of all bunches will have a pileup of at least 140 (Allport & al. 2013).

Therefore, the LHC upgrade poses several challenges to the actual ATLAS silicon pixel detector:

- Data rate: the higher interaction rate requires smaller pixels to keep occupancy under control and to avoid the readout inefficiencies;
- Radiation damage: the actual pixel detector is designed to withstand a fluence of $10^{15} \text{ n}_{\text{eq}}/\text{cm}^2$; the inner most layer is expected to be exposed to a total fluence of $2 \times 10^{15} \text{ n}_{\text{eq}}/\text{cm}^2$ ($5 \times 10^{15} \text{ n}_{\text{eq}}/\text{cm}^2$ including safety factors) during data-taking phase II and to a fluence of $2 \times 10^{16} \text{ n}_{\text{eq}}/\text{cm}^2$ during HL-LHC period (phase III). Therefore, the much higher fluences imply increased leakage currents resulting in an increase of noise in the analog front-end, a higher bias voltage needed to (at least partially) deplete the devices, and larger charge trapping (lower charge collection efficiency). Signal-to-noise ratio will decrease.

Since the actual ATLAS detector was designed to operate at the nominal luminosity, modifications of the system are required to cope with the increased pile-up of interactions at each bunch crossing and to assure stable

detection efficiency, track pattern recognition capability and accuracy of reconstructed tracks, independent of luminosity. Two main upgrade projects of the ATLAS inner detector are planned:

- a fourth pixel layer (called Insertable B-Layer - IBL) will be added to the present pixel detector, between a new vacuum pipe ($r \sim 23$ mm) and the current inner pixel layer (called B-layer), during the long shutdown LS1 (2013-2014);
- the entire inner detector is supposed to be replaced during LS3 (2022-2023).

The experimental results and simulations presented in this Section are related to the R&D efforts we have done at LAL for the design of new pixel silicon sensors dedicated to IBL upgrade project, with improved geometrical and radiation hardness characteristics with respect to the actual pixel sensors.

The main motivations behind IBL upgrade are:

- Tracking robustness: failure of actual modules, as they age and suffer from radiation damage, will affect tracking efficiency and precision. An additional space point will provide a more robust tracking by compensating for the failed sensors and by giving information about tracks from a distance closer to the interaction point;
- Luminosity effects: increase in luminosity will produce more tracks per bunch crossing to be reconstructed by the pixel detector. Read-out inefficiencies on the current sensor will determine the overall detector inefficiency and the IBL will provide a system with an increased occupancy capacity to compensate for this effect;
- Tracking precision: the location of the IBL, closer to the interaction point, will allow for more accurate reconstruction of the primary vertex and b-tagging. The insertion of the IBL will result in an increase in sensitivity to physics channel using these parameters;
- Radiation damage: the actual inner layer is foreseen to resist to radiation fluences of $1 \times 10^{15} \text{ n}_{\text{eq}}/\text{cm}^2$. The increases of the luminosity will inflict radiation damage and will eventually lead to a failure of the system. The insertion of the IBL will provide an insurance policy against such problems if they occur.

The drawing of the IBL pixel layer is presented in Fig. 3.2. It is located at a mean active geometric radius r of 33.4 mm and it consists of 14 staves tilted by 14° and arranged in a barrel configuration.

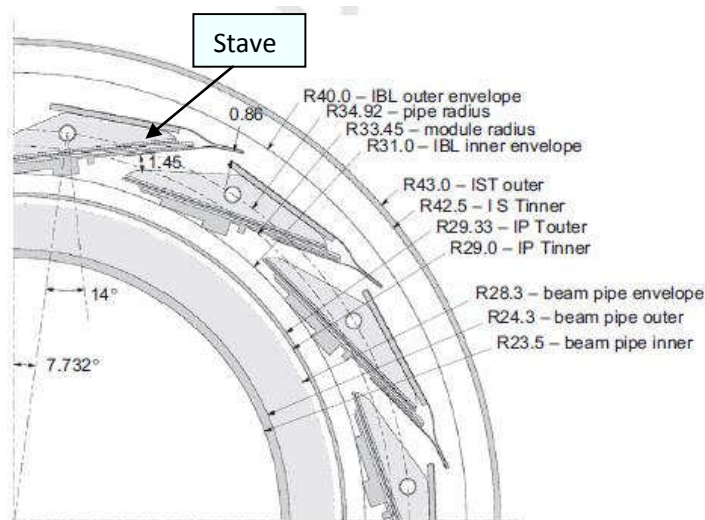


Fig.3.2 One quadrant cross-section of the IBL layer

Each stave of IBL layer has a total length of 724 mm and width of ~20 mm and it is loaded with 20 modules (Fig. 3.3), each module consisting on silicon sensors bump-bonded to the newly developed front-end electronics integrated circuit called FE-I4 built in 130 nm CMOS technology (details on the FE-I4 readout chip can be found in the reference Garcia-Sciveres & al., 2011). The 75% central part of the stave (12 modules, each module with two chips FE-I4) is built with planar pixel silicon sensors, while the remaining 25% (each extremity with 4 modules of one chip FE-I4) is based on 3D sensors.

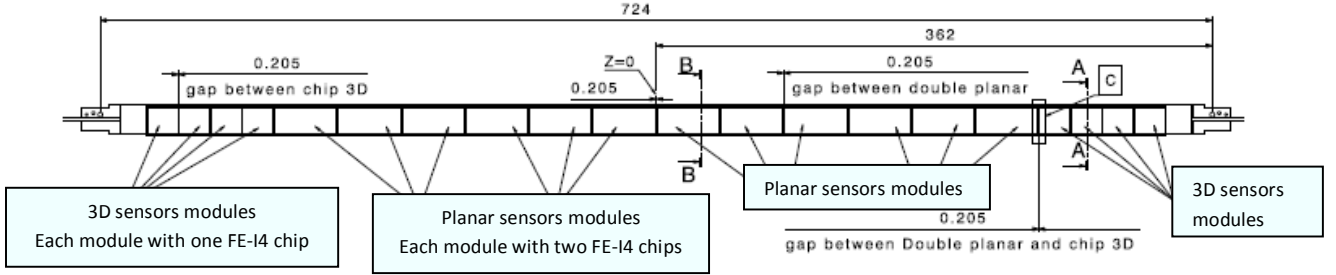


Fig.3.3 Stave layout with a mixed sensor scenario: 3D sensor modules populate the two extremities; the planar pixel sensors populate the central part; the gap between neighboring modules is fixed at 0.205 mm.

The planar pixel sensors (more details on their characteristics will be given later in this section) have been already widely used in present high energy physics (HEP) experiments and have very well established manufacturing process with very high yield (~90%) and low costs. However, at the irradiation levels expected in the inner pixel layer at the Phase II, they require high bias voltage and pose stronger constraints on the cooling system, especially when comparing with emerging sensor technologies like 3D sensors (Parker & al., 1997), which are still affected by lower yield (40-60%) and higher production costs, but made significant progress over the last few years (Grenier & al, 2011; Da Via & al., 2012).

It has been shown by very recent tests (ATLAS IBL Collaboration, 2012) that both sensors technologies fulfill the IBL physics requirements (i.e. radiation hardness up to $6 \times 10^{15} \text{ n}_{eq}/\text{cm}^2$ and 250 Mrad), therefore the staves will be build with both types of sensors. The vertical electrode of 3D sensors can offer the advantage of a better z-resolution after heavy irradiation; therefore, the 3D modules are used at the extremities of each stave.

The hermiticity in the transaxial ($r\phi$) plane of the IBL barrel is assured by about 20% stave-to-stave overlap (Fig. 3.2). However, the small radius of IBL and the limitation in the material budget inhibits overlapping modules in z direction; therefore, both planar and 3D sensors should present either active-edge or slim-edge guard-rings to reduce the geometrical inefficiencies along the beam axis.

The LAL ATLAS group is member of the ATLAS Planar Pixel Sensors international collaboration established in 2008 (Goessling & Muenstermann, 2008). In this context, our work was dedicated to the design of new planar silicon pixel sensors and in the following I will particularly describe these detectors.

The current ATLAS pixel sensors are represented by so-called n^+ -in-n planar pixel sensors design (Fig. 3.4): the substrate is n-type silicon while the pixels consist on n^+ implants on the front side (i.e. readout side). The junction is obtained by a large pad p^+ implantation located on the back side. Consequently, the planar pixel sensors are represented by arrays of pixels, where each pixel is a pn junction. The mechanical damage caused at the crystal edge, by sawing the wafer and extracting the sensor, makes the edge conductive and it would short the junction. Therefore, a multi guard-ring (GR) structure is surrounding the p^+ backside implant, allowing a smooth transition of the high voltage potential to ground while approaching the outer edge of the device. To avoid short-circuits between the neighboring n^+ pixels due to electron accumulation layer at the interface between silicon and silicon oxide, a p-spray isolation technique has been used (p-implant with a moderated dose adjusted to the saturation value of oxide

charge of $\sim 3 \times 10^{12} \text{ cm}^{-2}$). The p-spray isolation technique offers the advantage of the absence of a specific photolithographic step, allowing narrow spacing between neighboring n^+ -implants and the introduction of a punch-through biasing structure for testing the entire pixels array prior to bump-bonding. For more details about planar pixel sensors concepts, isolation techniques and biasing structures, the reader is required to refer to the reference Rossi & al. 2006.

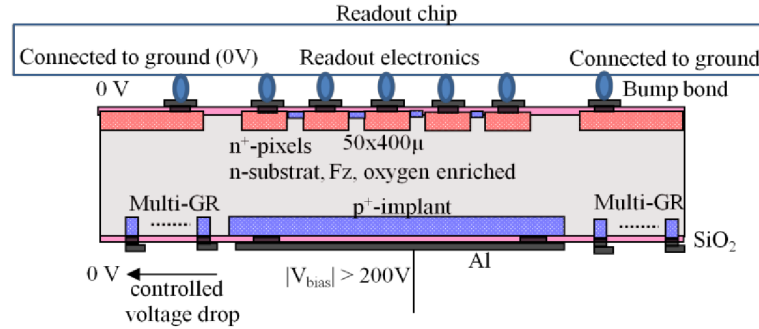


Fig.3.4 Cross-section of n^+ -in- n planar pixel sensor design connected to a front-end chip as used in the current ATLAS pixel detector

As presented in Fig.3.4, each pixel is connected to a read-out electronics channel by a vertical interconnection called bump-bonding, obtaining in such way a hybrid pixel detector (hybrid because electronics and sensors are fabricated separately and then assembled). Monolithic arrays of such pixels are built in such way that the size of the pixel matches the size of the front-end electronics channel.

The read-out electronics used in the current ATLAS pixel detector is called FE-I3 chip built in $0.25 \mu\text{m}$ CMOS technology (ATLAS collaboration 2008, b) and it consists of 2880 readout channels (18 columns x 164 rows), with a pixel size of $50 \times 400 \mu\text{m}$ (the small size in $r-\phi$ direction and long size in the z direction). Therefore, to match the size of the readout channels, the pixel sensor is represented by an array of 2880 pixels, arranged in 18 columns and 164 rows, with a pixel size of $50 \times 400 \mu\text{m}$. This monolithic array is named FE-I3 single-chip sensor and it has a full area of $\sim 10 \times 10 \text{ mm}^2$. The large surface of the ATLAS pixel detector ($\sim 1.7 \text{ m}^2$) has been build putting together many of these FE-I3 single-chip sensors, assembled in 1744 identical modules (i.e. one module has an active surface of $6.08 \times 1.64 \text{ cm}^2$ and it consists of 16 FE-I3 single-chip sensors built on the same substrate). The main characteristics of ATLAS FE-I3 single chip sensor are presented in the Table 3.1.

Parameters	FE-I3 single-chip planar pixel sensor
Full dimension	$\sim 10 \times 10 \text{ mm}^2$
Substrate	n-type silicon, FZ, oxygen enriched <111> orientation, $250 \mu\text{m}$ thickness, $2-5 \text{ k}\Omega\text{cm}$ resistivity
Pixels	n^+ implants of $30 \mu\text{m}$ wide and $380 \mu\text{m}$ long $50 \times 400 \mu\text{m}$ pitch 2880 pixels arranged in 164 rows x 18 columns
Pixels isolation	moderated p-spray technique
Guard-rings	16 GR's, each of $10 \mu\text{m}$ p-type implant width total width of $650 \mu\text{m}$
Electrodes covering the GR	16 to $22 \mu\text{m}$ wide, the largest at the outer side of the structure overshoot the implant by $2 \mu\text{m}$ in the edge direction and 1 to $40 \mu\text{m}$ in the active area direction; the distance between electrodes varies from 15 to $8 \mu\text{m}$
Dead zone along the long pixel size (z direction)	$1100 \mu\text{m} = 450 \mu\text{m}$ from cutting edge to outer GR + $650 \mu\text{m}$ of GR's zone width
Junction	p-type implant, back side

Table.3.1 Characteristics of the current ATLAS planar pixel sensors (FE-I3 single chip sensor)

The planar pixel sensors with the characteristics presented in the Table 3.1 have been successfully exploited by the ATLAS experience before LHC LS1 shutdown. To investigate the suitability of this technology for the upgraded tracker, the ATLAS Planar Pixel Sensors (PPS) international collaboration started intense R&D efforts since 2008 (Goessling & Muenstermann, 2008).

We could remark in the Table 2.1 that between the cutting edge of the sensor and the first pixel implant there is a width of $\sim 1100 \mu\text{m}$. A zoom over a corner including this zone is presented in Fig. 3.5. We can distinguish the multi GR structure composed of 16 GR's, with a total width of $650 \mu\text{m}$. Moreover, the distance from the last GR to the cutting edge has a width of $450 \mu\text{m}$. Therefore, a total width of $1100 \mu\text{m}$ from the last pixel edge to the cutting edge determines a complete inefficient zone of the sensor in the z direction.

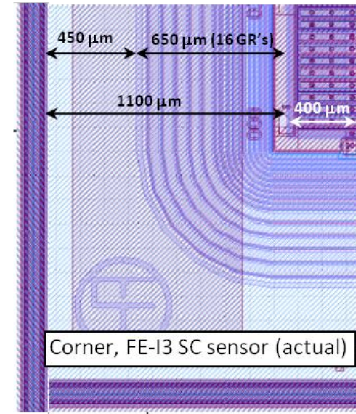


Fig.3.5 Corner view of a FE-I3 single chip planar pixel sensor showing the GR's structure and the first pixel column

To use the same design of planar pixel sensors for the IBL project upgrade, an overlapping of the sensors is necessary to avoid detection gap between sensors. Such overlap increases the amount of material present in the tracker and should be avoided to reduce the material budget of the inner detector and increases the simplicity of its configuration. Therefore, *reduction of the GR's zone for the sensors to be used in IBL was a key to minimize its material budget and its inactive zones.*

Consequently, the goal of TCAD simulations performed at LAL since 2008 (Benoit, 2011) was to evaluate the effects of modifying the geometrical GR configuration (number of GR's, widths, distances) for reducing the dead zone of planar pixel sensors while maintaining their adequate operation conditions and to exploit the effect of radiation damage on the efficiency of GR's structure.

To compare the TCAD simulations with real structures characteristics, since 2008 I coordinated two experimental activities:

- SIMS and SRP measurements for doping profiles characterization; SIMS tests were carried out at GEMaC laboratory (CNRS-Meudon, France) and SRP tests at EAG lab (Reading, United Kingdom); these experimental results and their comparison with TCAD simulations will be presented in Section 3.2;
- DC electrical measurements (i.e. IV, CV, GR potential) of planar pixel sensors; a manual Karl-Suss test-station has been installed in a clean-room at LAL and adapted for such kind of measurements; the experimental results and their comparison with TCAD simulations will be presented in the Section 3.3.

3.2 Doping profiles of ATLAS planar pixel sensors

3.2.1 Experimental results

Supervision: Mathieu Benoit (PhD Student, LAL ATLAS group)
Internal collaboration: Abdenour Lounis (LAL, ATLAS Pixel group coordinator);
 Idarraga John (LAL, post-doc ATLAS)
External collaboration: François Jomard (GEMaC laboratory, CNRS)
 ATLAS PPS collaboration: A. Macchiolo (MPI, Munich) & D. Muestermann (TU Dortmund, Germany)

As discussed previously, the doping profiles techniques are usually destructive. Therefore, special test-structures for the doping profiles measurements have been designed at LAL and implemented in the mask layout of the fabrication runs of planar pixel sensors produced by CiS Institut für Mikrosensorik gGmbH (Erfurt, Germany). The doping profile test-structure (Fig. 3.6) covers a full area of $\sim 7 \times 7 \text{ mm}^2$; on its front side we can observe three zones of $\sim 2 \times 7 \text{ mm}^2$, each corresponding to a given implant used in the fabrication of pixel sensors:

- Right zone: phosphorus implant corresponding to the n^+ implant of pixels cells;
- Middle zone: boron implant corresponding to the standard p-spray isolation implant;
- Left zone: boron implant corresponding to the low dose area of the moderated p-spray isolation.

A boron implant exists on the whole backside of the sample, corresponding to p^+ implant of junction side as well as of GR's implants.



Fig.3.6 Picture of test-structure dedicated to doping profiles measurements; the three zones have different types of implants: moderated p-spray implant (left zone), standard p-spray implant (middle zone), n^+ implant (right zone); a boron implant exists on the whole backside of the sample

For a better understanding of the correspondence between different implantation zones of our test-samples and the implants used for pixels fabrication and the isolation techniques, a cross-section of two pixels with standard and moderated p-spray isolation methods are presented in Fig. 3.7:

- standard p-spray isolation technique (Fig. 3.7 a) involves application of a medium-dose p-type implant to the whole n-side; some regions of p-spray are then overcompensated by the n^+ pixel implants themselves;
- moderated p-spray isolation technique (Fig. 3.7 b) implies a smaller p-type concentration at the interfaces with the n^+ implants and a higher concentration in the center of the gaps between pixels.

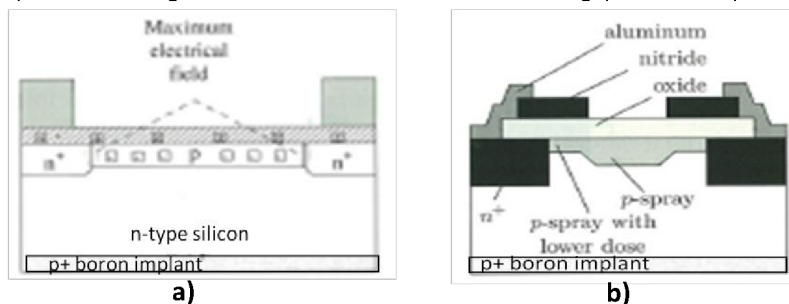


Fig.3.7 Cross-section of two pixels involving **a)** the standard p-spray technique; **b)** the moderated p-spray technique

Since 2008, different runs of planar pixel sensors have been produced by CiS: two pre-production runs during 2009 and 2010 (called here PPS09 and PPS10) and a production run for the IBL construction during 2011 (called here IBL2011). Firstly, we performed SIMS and SPR measurements on test structures coming from few wafers of PPS09 run. Subsequently, we carried out the measurements on wafers from following runs (PPS10 and IBL2011). For

simplicity of the presentation, but also to allow a comparison between profiles coming from different runs, cumulative plots of SIMS and SRP profiles from all runs will be presented in the following. However, the reader has to keep into account that we have used progressively the results coming out from these measurements, allowing us to calibrate the TCAD simulation code. Moreover, feedbacks with the producer have been done after each series of measurements and technological improvements took place in the more recent runs as we will see in the following.

The SIMS profiles, expressed in dopant concentration vs depth, are presented in Fig. 3.8 a) for n^+ (phosphorus) implant and Fig. 3.8 b) for the p^+ (boron) backside implant. I would like to mention that before performing SIMS measurements, we did not know exactly which type of dopant have been used by the producer to obtain the n^+ and p^+ implants (i.e. phosphorus and arsenic can be used as n-type dopants, while boron, aluminum, gallium can be used as p-type dopant). SIMS measurement allowed us to identify each dopant, namely phosphorus for the n-type implant and boron for the p-type implant.

We can observe that both n^+ and p^+ profiles of two PPS09 samples show a surface peak concentration of $\sim 2 \times 10^{19}$ at/cm³ with a depth of ~ 700 nm (corresponding to the SIMS detection limit of $\sim 5 \times 10^{15}$ at/cm³) while the profiles of PPS10 and IBL2011 samples present quite the same surface peak concentration of $\sim 2 \times 10^{19}$ at/cm³, but a deeper junction depth at ~ 1000 nm.

As mentioned previously, SIMS measurement provides a total *atomic* dopant profile of the overall implanted ions. However, not all implanted ions are usually electrically active and therefore electrical SRP measurements have been performed (Fig. 3.9 a) and b)). We can observe a good agreement between two types of measurements, demonstrating that the PPS09 samples have been produced in different technological conditions (i.e. lower annealing time) with respect to PPS10 and IBL2011 runs.

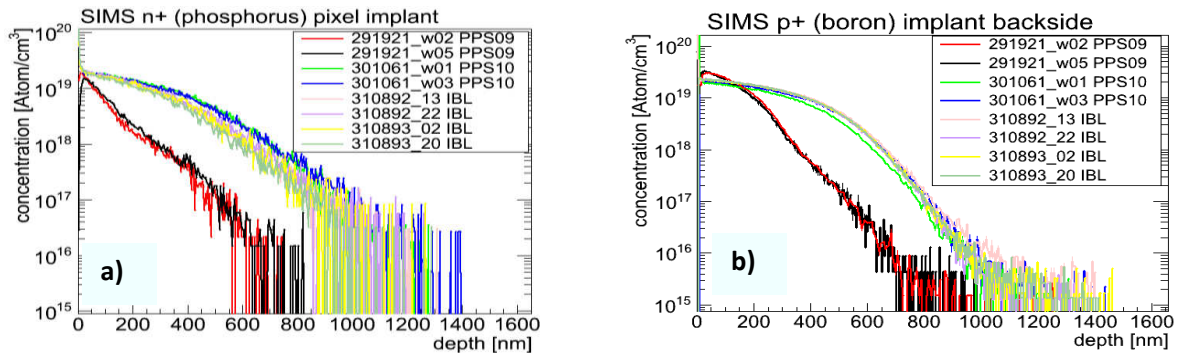


Fig.3.8 SIMS profiles of **a)** n^+ (phosphorus) implant; **b)** p^+ (boron) backside implant corresponding to three different production runs of n^+ -in-n ATLAS planar pixel sensors

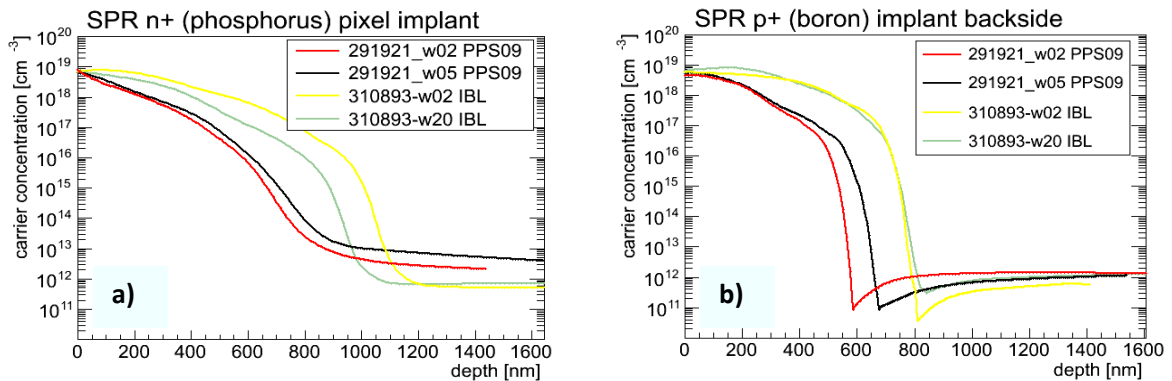


Fig.3.9 SRP profiles of **a)** n^+ (phosphorus) implant; **b)** p^+ (boron) backside implant corresponding to different production runs of n^+ -in-n ATLAS planar pixel sensors

The SRP measurement allows us to identify the conductivity type of each implanted layer. We already knew from the producer that n-type silicon wafers have been used for the mentioned fabrications runs. The SPR profile of the n^+ implant (Fig. 3.9 a) shows a smooth transition from high concentration implant to the low concentration of silicon substrate, confirming that both the implant and the substrate have the same type of conductivity (identified by SRP method to be n-type). However, the SPR profile of the p^+ implant (Fig. 3.9 b) shows a sharp transition from high concentration implant to the low concentration of silicon substrate, confirming that two types of conductivity are present (i.e. the point of minimum concentration in the profiles corresponds to metallurgical p^+/n junction). By using the SRP method, the high concentration implant has been identified of p-type conductivity and the substrate was confirmed to be n-type.

The SRP method allows also to determine the substrate resistivity by measurement of active dopant concentration far from the junction. The dopant profiles presented in Fig. 3.9 a) are not measured far enough in the depth to determine the correct bulk concentration. Therefore, from the SRP profiles presented in Fig. 3.9 b) we determined a donor bulk concentration of $N_{eff} = 1 \times 10^{12} \text{ cm}^{-3}$, corresponding to a resistivity of $4.4 \text{ k}\Omega \cdot \text{cm}$ ($\rho = 1/eN_{eff}\mu_e$, where ρ is the resistivity, e is the electron charge, N_{eff} is the effective doping concentration, μ_e is the electron mobility in silicon). This high value of the resistivity is in agreement with the requirements for manufacturing of silicon particle sensors used for high energy physics applications: a high resistivity is needed in order to fully deplete the sensor bulk with a thickness of $\sim 200 - 300 \text{ }\mu\text{m}$ by an adequate voltage below 300 V ($V_{dep} = eN_{eff}d^2/2\varepsilon$ where V_{dep} is the depletion voltage, e is the electron charge, N_{eff} is the effective doping concentration, d is the sensor thickness, ε is the silicon electrical permittivity). In the case of current ATLAS planar pixel detectors, the substrate thickness is $\sim 250 \text{ }\mu\text{m}$, determining a fully depletion voltage of $\sim 50 \text{ V}$ (confirmed also by a capacitance-voltage measurement presented in the Section 3.3.1).

The SIMS profiles of standard and moderated p-spray implants have been also measured (Figs. 3.10 a) and b)). As presented previously, moderated p-spray design presents two layers (oxide and nitride) on top of the implanted zone, while the standard p-spray design presents only an oxide layer. Therefore, the SIMS moderated p-spray profile shows a lower depth of $\sim 300 \text{ nm}$ with respect of $400-500 \text{ nm}$ for the case of SIMS standard p-spray profile. Beside of the different depth, we can observe also that the shape of the standard and moderated p-spray profiles is quite different: (1) the standard p-spray presents a Gaussian shape, with a surface concentration of $\sim 2-5 \times 10^{16} \text{ cm}^{-3}$ and a peak of $\sim 2 \times 10^{17} \text{ cm}^{-3}$ at a depth of $\sim 100-200 \text{ nm}$ (depending of the sample); (2) the moderated p-spray presents just a simple decreasing profile, with a concentration peak of $\sim 1 \times 10^{17} \text{ cm}^{-3}$ at the surface.

This difference is as expected, if we suppose that the same implantation step has been used to produce both implants (i.e. the same energy and dose) and consequently, the additional nitride layer filters part of the implanted boron ions for the case of moderated p-spray implant.

The SRP measurements of standard and moderated p-spray implants have been also performed (Figs. 3.11 a) and b)). The SRP profiles of PPS09 samples showed unexpected profiles, where no sign of a junction can be observed. However, the SRP measurements of the IBL2011 structures show expected profiles (i.e. presence of a p/n junction), well correlated with the SIMS ones. Since boron presence has been well confirmed by SIMS measurements in all measured samples, it is highly probable that the activation step of both standard and moderated p-spray implant of the two measured wafers of the PPS09 run was not fully completed, technological step that has been corrected for the successive run of the IBL2011 production.

The problem of the incomplete activation of p-spray implants detected during our SRP measurements demonstrates that this test is important not only for the TCAD calibration (which was our first goal), but it provides also a very useful test for process control: it can detect if the technological steps related to implantation and annealing are correctly performed, and it assures therefore a good functionality of the final sensors to be used in the experience.

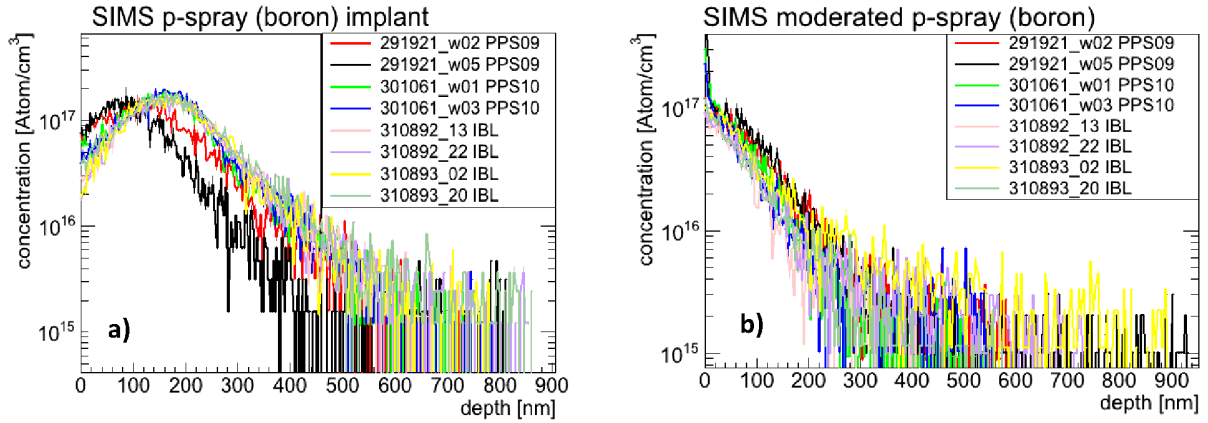


Fig.3.10 SIMS profiles of **a)** standard p-spray implant and **b)** moderated p-spray implant corresponding to three different production runs of n⁺-in-n ATLAS planar pixel sensors

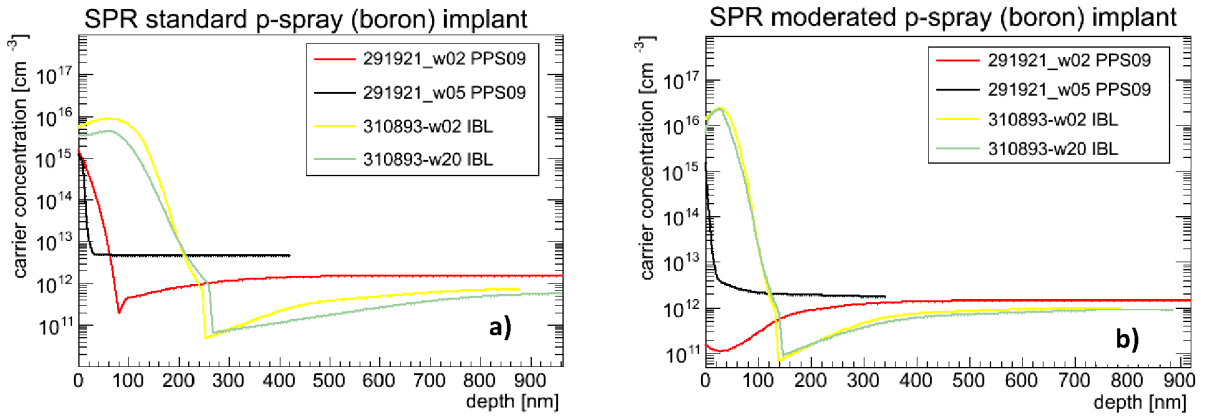


Fig.3.11 SRP profiles of **a)** standard p-spray implant and **b)** moderated p-spray implant corresponding to different production runs of n⁺-in-n ATLAS planar pixel sensors

In recent years, the interest of p-bulk devices as possible sensors for HEP experiments has rapidly increased as high-resistivity p-type wafers have become available and since they do not undergo type inversion after heavy irradiation. Therefore, in parallel with the runs of n⁺-in-n planar pixels sensors produced by CiS and already mentioned in this Section, runs of n⁺-in-p designs of planar pixels sensors have been also produced.

For a better understanding of the main differences of the n⁺-in-n and n⁺-in-p designs of planar pixels sensors, a cross section of both sensors are presented in Fig. 3.12 a) and b).

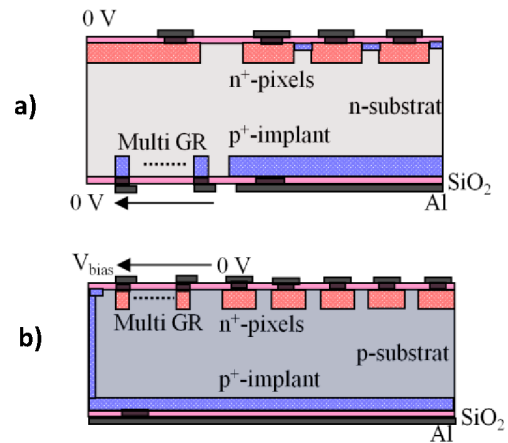


Fig.3.12 Cross-section of n⁺-in-n (a) and n⁺-in-p (b) planar pixels design

For the case of n^+ -in-p design, the p - n junction is located on the pixel implantation side. Therefore, the GR's structure can be placed on the front-side as well, and thus patterned processing is only needed on a single side. Consequently, no masks and no alignment for backside processing are needed, lowering the cost and enabling the use of more foundries for processing. Furthermore, the lack of patterned backside implantation eases subsequent handling and testing.

The special test-structure presented in Fig. 3.6, dedicated to doping profiles measurements, was implemented in the mask layout of the n^+ -in-p production run performed at CiS factory during 2009 (called here PPS09, p-type wafer). As for the n^+ -in-n samples, the SIMS and SRP measurements have been performed on different zones: n^+ corresponding to pixels implant and the GR's, p^+ corresponding to p^+ back side implantation, and p-sprays corresponding to both standard and moderated p-sprays implants.

The SIMS measurement confirmed the presence of phosphorus element in the n^+ implanted zone (red line, Fig. 3.13 a1) as well as of boron element corresponding to the back-side implant and moderated and standard p-sprays implants (red line, Figs. 3.13 a2), a3) and a4)). Both n^+ and p^+ backside SIMS profiles are similar with SIMS profiles of n^+ -in-n design, showing a surface peak concentration of $\sim 1\text{-}2 \times 10^{19}$ at/cm³ with a depth of ~ 900 nm for the n^+ profile and ~ 1400 nm for the p^+ backside implant. Regarding the SIMS profiles of standard and moderated p-spray, we can observe that they present similar implant depths as measured in the n^+ -in-n samples (i.e. 400-500 nm for the standard p-spray and 300 for the moderated p-spray), while the surface and peak concentrations are much lower (i.e. $1\text{-}2 \times 10^{16}$ cm⁻³ at the surface and 3×10^{16} cm⁻³ at the peak of the standard p-spray). For a better understanding of these profiles, the SRP electrical measurements have been performed.

The SRP profiles presented in Figs. 3.13 b1), b2), b3) and b4) (red line) are well correlated with the SIMS ones. However, the SRP measurements revealed that the silicon bulk, close to front and back surface up to the depth we have performed the measurements, is of n-type conductivity and not p-type:

- the SRP profile of n^+ implant on p-type layer should show the presence of a junction (i.e. sharp transition with a point of minimum concentration); however, an n-type conductivity has been detected on both n^+ implanted layer as well as on the substrate, up to a depth of 15 μ m (not shown in the figure as the steps were too large);
- the SRP profiles of different p-type implants (backside or p-sprays) should show no junction presence (i.e. smooth transition from the p-type implanted layer to the p-type substrate); however, the p-type conductivity has been measured on the implanted layers and, contrary of what was expected, the n-type conductivity was measured on the substrate too. Therefore, the sensor junction was found on the backside of the sample, incorrectly located since the bulk should be p-type.

These measurements lead us to suspect the presence of an additional donor in the bulk of the sensor that inverts the type conductivity of the bulk as for radiation damage. It was found that (A. Macchiolo, MPI, Munich) these wafers were subjected, at the very early stage of the fabrication, to an oxygenation step to increase their radiation hardness. The heavily oxygenated p-type bulk can develop thermal donors during process steps between 400 and 600°C (Hull, 1999). These thermal donors can be sufficiently abundant to invert the type conductivity of the silicon bulk from p-type to n-type. An attempt has been done to measure the oxygen depth profile by SIMS method, but unfortunately the lowest oxygen detection limit of our SIMS installation was $\sim 1 \times 10^{18}$ at/cm³ (with a vacuum of 10^{-9} bar), not enough to detect the oxygen concentration in our samples (the highest oxygen concentration in CZ silicon is expected to be of the order of 9×10^{17} cm⁻³ (Moll, 1999)). A much modern SIMS installation, with ultra-vacuum, is required to perform such measurement.

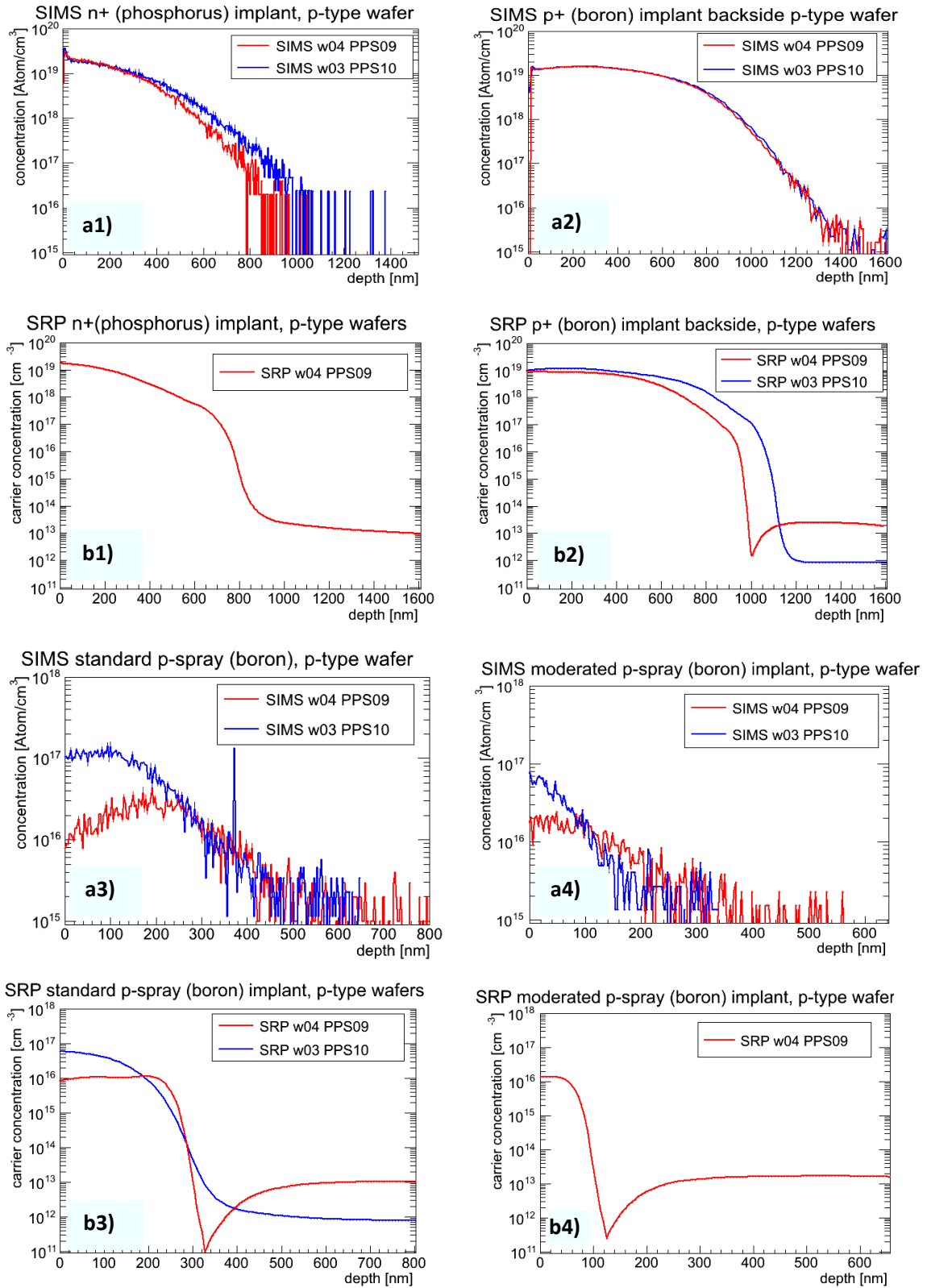


Fig.3.13 SIMS profiles **a1)**, **a2)**, **a3)**, **a4)** and SRP profiles **b1)**, **b2)**, **b3)**, **b4)** corresponding to PPS09 (including oxygenation step) and PPS10 (without oxygenation step) of n⁺-in-p ATLAS planar pixel sensors

Taking into account the results from our measurements, a new production run of n⁺-in-p design, without the oxygenation step, has been performed during 2010 (called here PPS2010, p-type wafer). The results of corresponding SIMS and SRP measurements performed on our dedicated test-structures are presented also in Figs. 3.13 b2) and b3), with blue color. We can observe that the problem of the bulk inversion type has been resolved, SRP measurements showing a silicon bulk of p-type conductivity in comparison with n-type previously detected. Moreover, the effective donor concentration has been changed from $N_{eff}=1 \times 10^{13} \text{ cm}^{-3}$, corresponding to a n-type bulk resistivity of $\sim 440 \text{ } \Omega \cdot \text{cm}$ to an effective acceptor concentration of $N_{eff}=1 \times 10^{12} \text{ cm}^{-3}$, corresponding to a p-type high resistivity of $\sim 13 \text{ k}\Omega \cdot \text{cm}$ (the value specified by the producer).

The inversion of the bulk conductivity detected by our SRP measurements in the n⁺-in-p runs demonstrates that such kind of test, as for the case of n⁺-in-n design, is extremely useful for process control during silicon sensors production runs, assuring a good functionality of our final detectors.

3.2.2 Comparison with TCAD simulations

The experimental data of the doping profiles structures presented in the previous Section allowed us to determine some of the main missing parameters needed for accurate TCAD simulation of the implantation step. In this Section, I will show how such parameters have been determined using the SIMS profiles of n⁺ and p⁺ implants of the n⁺-in-n sensors.

The parameters needed to reproduce the process of implantation are the energy and dose of the implanted ions, the thickness of the oxide or nitride layers used as a screen and the temperature and duration of the thermal annealing used to activate the dopants. The Modified Lavenberg-Marquart algorithm (Silvaco Inc., 2008), built-in SILVACO software, was used (Benoit, 2011) to simulate doping profiles most closely fitting the experimental SIMS and SRP measurements, allowing in such way to determine the needed parameters.

Two implantation models are available in SILVACO TCAD software: (1) a simple model which consists of a parameterization of implant profiles using various SIMS measurements; (2) a Monte-Carlo model, using the binary collision approximation for a more accurate description of the implant. Two models are also available to simulate the diffusion of the dopants during annealing step: the Fermi diffusion model and the fully coupled diffusion model. The fully coupled model includes the physics of the previous one and adds new phenomena needed for more precise simulation of small details of the doping profile. However, the needed computing time increases by an order of magnitude when using more accurate models.

Optimization of the implantation parameters was performed using the simple diffusion and the implantation models. The advanced models were then activated to compare the obtained profile to the SIMS measurements. Fig. 3.14 shows the comparison between the simulated doping profile and the experimental SIMS profiles corresponding to n⁺ and p⁺ implants of the n⁺-in-n sensors. It can be observed that simulated profiles (black curves) reproduce well the shapes of the experimental data. If more complex models are used, even more accurate description of the profile can be obtained. For confidentiality reasons, *the technological parameters extracted from these simulations will not be presented here. However, they have been successfully used to calibrate our TCAD simulation code and to simulate final characteristics of pixel sensors as will be shown in the next Section, allowing us to predict a new design of n⁺-in-n planar pixel sensors that have been selected for the IBL construction.*

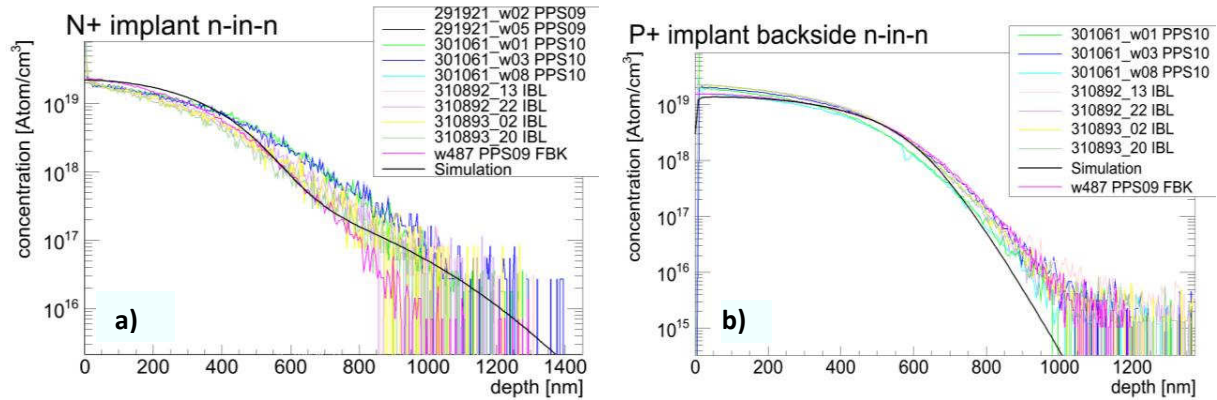


Fig.3.14 Comparison of TCAD simulation profiles (after optimization of implantation and diffusion parameters) and experimental SIMS profiles corresponding to n⁺ (a) and p⁺ (b) implants of n⁺-in-n planar pixel sensors.

3.3 Static characterization

3.3.1 Set-up calibration

Supervision: Jean-François Vagnucci (LAL), Mathieu Benoit (PhD Student, LAL ATLAS group), Ahmed Bassalat & Ali Harb (Master II students, LAL ATLAS group)

Internal collaboration: Abdenour Lounis (LAL), Gerard Guilhem (LAL), Tristan Vanderberge (LAL)

External collaboration: ATLAS PPS collaboration: A. Macchiolo (MPI, Munich) & Tobias Witting (TU Dortmund, Germany)

As mentioned previously, our main goal for the IBL project was to reduce the dead zone of the planar pixel sensors (i.e. GR's geometrical configuration, the zone between the last GR and the cutting edge) while maintaining their adequate operation conditions. To predict how we can reduce this dead zone, the TCAD simulation tool was used; in particular, the bias voltage taken by the GR's structure when the sensor is biased was required to be simulated (more results on TCAD simulations of GR potential will be given in the next Section).

It has been observed that simulated GR bias potential was very dependent on the doping profiles characteristics of the n⁺, p⁺ and p-spray implants (Benoit, 2011). Therefore, the use of the doping profiles measurements presented in the previous section was very useful to reduce the number of free parameters needed for the simulation.

To gain more confidence in our TCAD simulations and to perform reliable predictions of a new sensors design, electrical measurements on real planar pixel sensors have been also performed at LAL. To carry out such measurements, I coordinated the building of a clean-room set-up for the DC characterization of pixel sensors. A manual Karl-Suss test-station has been enclosed in a home-made Faraday cage as can be observed in the Fig. 3.15. A mechanical support (chuck), holding the sensors during measurements by vacuum has been also designed and built at LAL. Since the electrical characterization of pixel sensors requires stringent conditions on measurement equipment (i.e. 10% accuracy for current levels ranging from ~10 pA to ~1 mA), special attention has been paid by using triaxial connections as well as well-isolated panel-troughs in between contacting probes (i.e. needles) and the measurement instruments. A Keithley electrometer with a high input impedance (10²⁰ Ω) was used to measure the

GR potential. As GR's are floating structures unconnected to any bias source, using a standard voltmeter with lower impedance would draw current from the GR modifying its electrical potential.

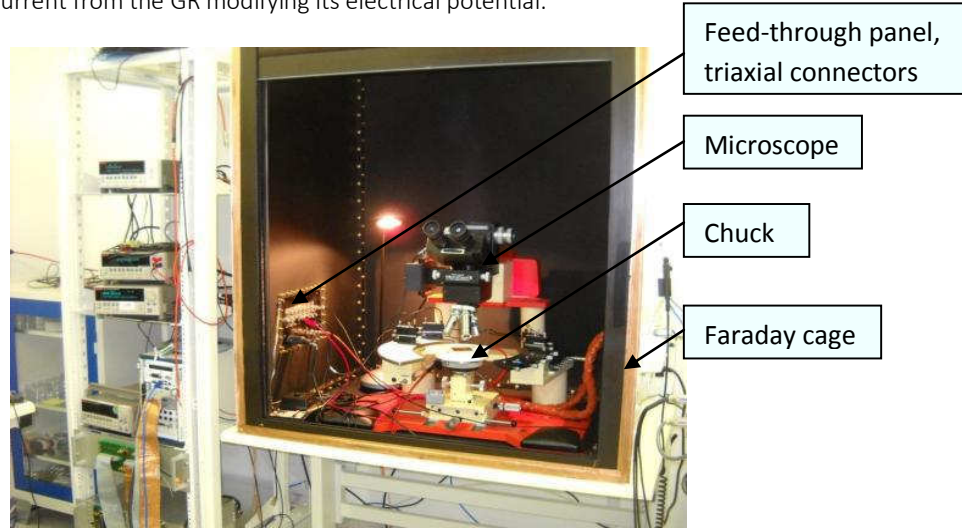


Fig.3.15 Set-up built at LAL for the DC characterization of ATLAS planar pixel sensors

To calibrate out experimental set-up, we performed firstly current-voltage (IV) and capacitance-voltage (CV) characteristics on different configurations of planar pixel sensors (these sensors have been produced by CiS company, PPS09 pre-production run, using the doping profiles characteristics presented in the previous Section). In particular, we have used FE-I3 single chip sensors with n^+ -in-p design (i.e. 2880 pixels and 15 GR's, compatible with FE-I3 single ship sensors of n^+ -in-n design) and we compared our measurements with those performed by our colleagues at MPI, Munich.

The IV characteristics are presented in Fig. 3.16 a) for sensors with standard p-spray design and Fig. 3.16 b) for moderated p-spray design (red color represents the measurements performed at MPI and those in blue performed at LAL). We can observe a good agreement between our measurements. However, at a given V_{bias} we can see that current values measured at MPI are about 8-10% higher than LAL current values. This difference can be explained by the fact that MPI measurements have been performed at 20°C, while the LAL measurements have been performed at ~19°C. Moreover, we can observe that sensors with moderated p-spray design present a breakdown voltage V_{BD} of about 300-400V, while the sensors with standard p-spray show lower $V_{BD} \sim 100V$. This comportment is as expected, since moderated p-spray technique determines lower electric field strength at the edges of the p-spray implants close to the pixels implants, allowing in such way higher V_{BD} values.

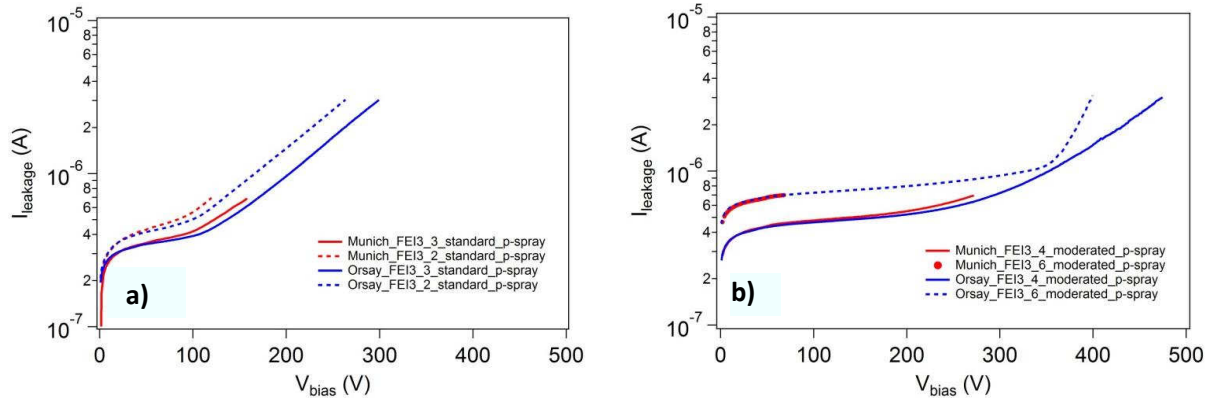


Fig.3.16 IV characteristics of n^+ -in-p FE-I3 single chip sensors produced by CiS: **a)** standard p-spray design; **b)** moderated p-spray design

The CV characteristic has been also measured by using Karl-Suss test-station and an Agilent CV meter. To avoid interferences from complex structure of pixels arrays, this measurement has been performed on a dedicated diodes designed at LAL (the one presenting 4 GR's is presented in Fig. 3.17), diodes which have been implemented in the mask of both n^+ -in-n and n^+ -in-p pixels design produced at CiS. Another advantage of such diodes is that they have large and contactable GR's allowing us to perform electrical measurements, while real pixel sensors presents very small GR's with very small openings where needle contacts are much more difficult to realize.



Fig.3.17 Picture of a diode with total area of $6 \times 6 \text{ mm}^2$ and 4 contactable GR's structure (LAL design)

The CV characteristic of such diode is presented in Fig. 3.18. a). We can observe that capacitance decreases progressively while the voltage increases, reaching a constant value when the diode is completely depleted. To determine the depletion voltage V_{dep} , the $1/C^2$ as a function of V_{bias} is presented in Fig. 3.18 b). The intersection of two linear fits allows us to determine a $V_{\text{dep}} \sim 50\text{V}$, in very good agreement with the one expected from our SRP measurements as presented in the previous Section 3.2.1.

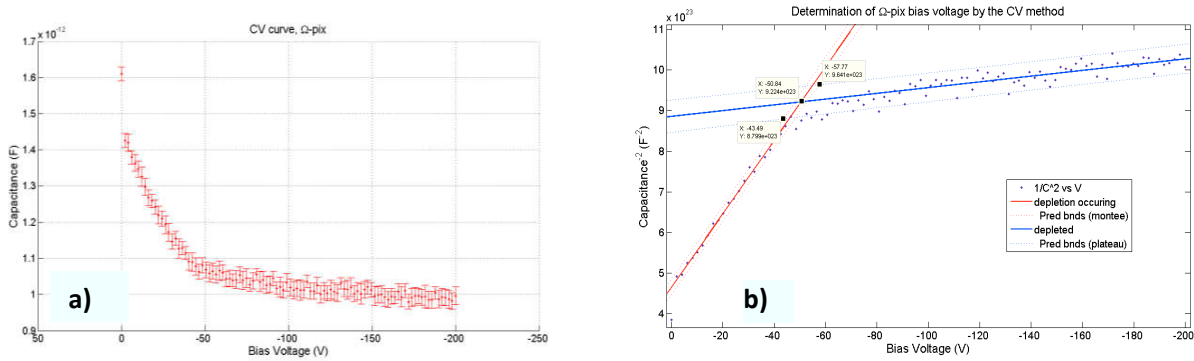


Fig.3.18 a) CV characteristic of a diode with the total area of $6 \times 6 \text{ mm}^2$ and 4 contactable GR's structure ; **b)** $1/C^2$ vs V_{bias} of the same diode

3.3.2 Guard-ring potential: measurement and comparison with TCAD simulations

As mentioned previously, the TCAD simulation represents a suitable tool to study the potential distribution of GR's geometries and to optimize them to improve the performances of the pixel sensors. The input parameters required to simulate the electrical GR's structure are: the doping profile of the GR implant, the metal overhang length and the distance between GR's. The doping profiles presented in Section 3.2.1 were extremely useful to fix one of the input parameters. The confidence on the values of other parameters has been obtained by comparing the TCAD simulations with experimental measurements of GR's potential.

The potential of different GR's have been measured on n^+ -in-n FE-I3 single chip sensor (i.e. 16 GR's). The experimental results (dots), compared with TCAD simulations (continuous lines) are presented in Fig. 3.19. The rather good consistency between simulations and experimental measurements gives us confidence in the

correctness of the simulation code and allowed us to use this code to predict new GR's structure design with different geometry to be used for the IBL pixel sensors design as will be presented in the next Section.

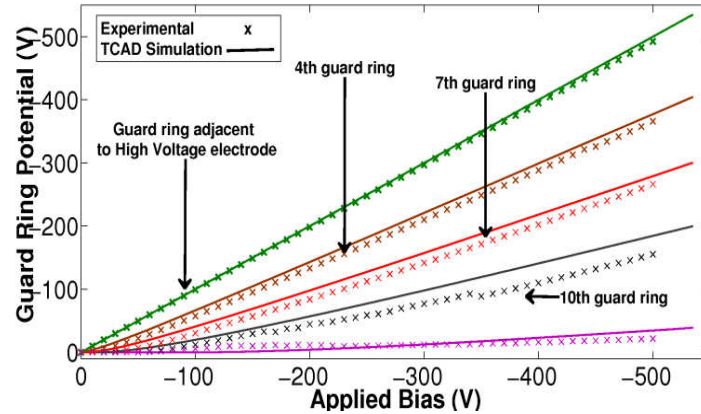


Fig.3.19 Comparison of simulated and measured GR potential for the n^+ -in-n FE-13 single chip sensor (PPS09 CiS production) after optimization of implantation and diffusion parameters

3.4 Optimization of guard-ring region by TCAD simulations

As described in the Section 3.1 (i.e. Fig. 3.5), the actual ATLAS pixels sensors design presents 16 GR's, determining a dead zone of $\sim 650 \mu\text{m}$ along the z-axis of the ATLAS detector. Since pixels sensors with as low as possible dead zone were required to be used for the IBL project, TCAD simulation tool was used to predict an optimization of the GR's structure. In particular, starting input parameters fixed by experimental parameters of doping profiles and GR potential have been used and different parameters as electrical potential, electrical field and electron concentration have been simulated by modifying the geometry of the GR's.

To emphasize the importance of the experimental work I performed in the framework of the ATLAS group and presented in the previous Sections (Dinu & al. 2011, a, b), selected results of the TCAD simulations (Benoit & al. 2009 a, b; Lounis & al. 2009, Benoit & al. 2010, Lounis & al. 2012) will be presented in the following.

The behavior of the electrical potential near the edges after gradual removal of some of the 16 GR's has been simulated when a reverse biased of 500 V is applied to the sensor. In Fig.3.20, the GR's potential as a function of GR position from the outer edge is illustrated. Since the potential value reaches an almost negligible value when up to 6 GR's are removed, one can predict that there is no real change, in this case to the sensor performance. For the same bias conditions, this result can be verified as shown in the electric field distribution at $0.1 \mu\text{m}$ under the GR's versus the distance from the outer edge (Fig. 3.21). One can see that after 4 GR's removal, the remaining electric field ($\sim 40 \text{ kV/cm}$) is extremely weak as compared to the few hundreds of kV/cm required to cause any breakdown situation or generate a noticeable leakage current.

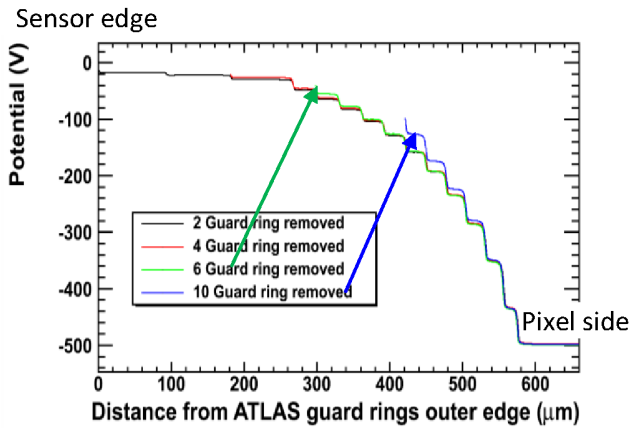


Fig.3.20 Simulated GR potential distribution as a function of distance from the sensor edge for the actual ATLAS sensors with 2, 4, 6 and 10 of the outer GR's removed

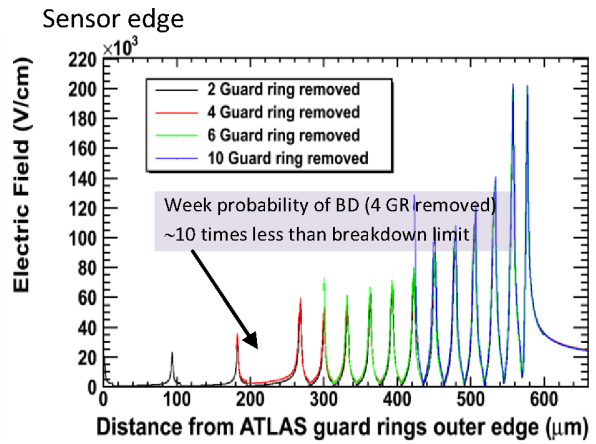


Fig.3.21 Simulated electrical field distribution as a function of distance from the sensor edge for the actual ATLAS sensors with 2, 4, 6 and 10 of the outer GR's removed

In the Fig. 3.5 (Section 3.1) it has shown that the overall dead zone of the actual ATLAS sensor is determined not only by the GR's structure but also by the dead zone between the outer GR and the cutting line, with a width of 450 μm . This is a safety insensitive area used to keep the sensor active area isolated from any border, because possible crystal damages or micro-cracks created during sensor dicing from the wafer can involves high leakage currents and consequently early breakdown. Reducing this zone to a width of 100 μm , no influence of the GR behavior to the electron concentration distribution as a function of the distance from the edge has been observed in our simulations (Fig. 3.22). One can clearly see that the lateral depletion boundary area stays sufficiently far away from the borders of the sensor and no malfunction is thus foreseen if we decrease the border size region from 450 μm (Fig. 3.22 a)) to 100 μm (Fig. 3.22 b)) which represents without any penalty, a gain of the sensitive area.

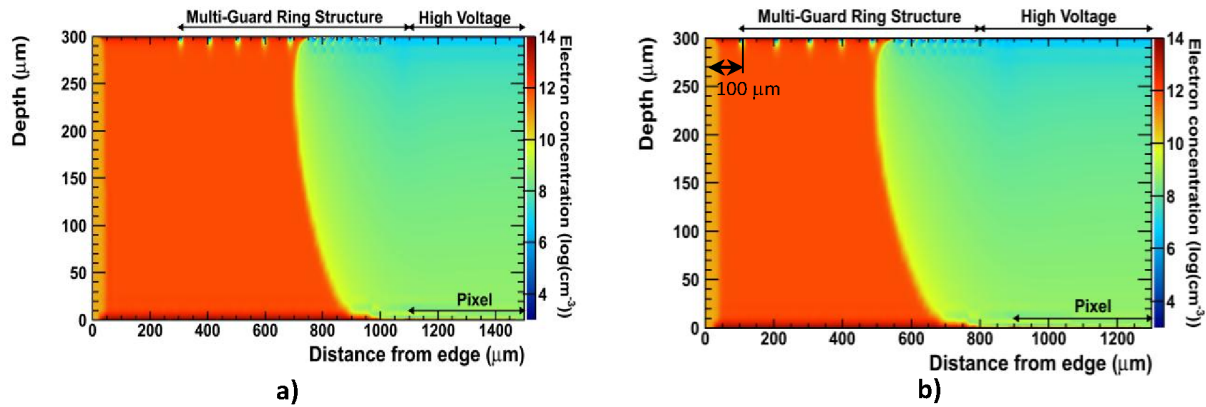


Fig.3.22 Simulated electron concentration ($\log(\text{cm}^{-3})$) as a function of the distance from the edge for ATLAS n^+ -in- n pixel sensors with: a) 450 μm and b) 100 μm dead zone from the outer GR to the cutting edge

Another idea, proposed by the ATLAS Planar Pixel Group, was to reduce further the inactive area of n^+ -in- n sensors by shifting the GR structure under the pixel area (the so-called "slim-edge" design). This geometrical structure is shown in Fig. 3.23 and it was also simulated to understand the effects on the sensor performances.

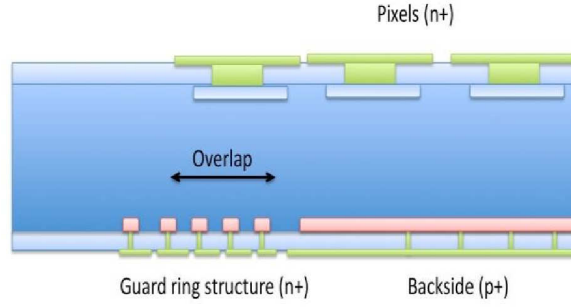


Fig.3.23 Geometrical structure of the slim-edge design

The model used was actual ATLAS n^+ -in- n design for which GR's were shifted successively under the pixel region by 100, 200 and 400 μm . In the simulation, the sensor thickness was 300 μm and a V_{bias} of 150 V was used for the case before irradiation. The edge was 500 μm , but reduction of the dead edge does not affect the GR behavior and the result can be valid as long as the width of this edge stays within the 100 μm previously mentioned.

Fig. 3.24 shows the electrical field magnitude in the region of the edge, where pixel overlaps by 200 μm the GR's structure. We can observe that GR presence affects the distribution of the field under the pixel and the field is weaker in the section of the pixel overlapping the GR. Fig. 3.25 shows the potential distribution in the case of 200 μm shift. We can observe that the potential gradient in the overlap region will drag charges away from the pixel as they drift closer to it. This is because this part of the pixel is in the lateral depletion zone which is normally located outside the volume of the edge pixel, determining charge collection losses. This can still be a beneficial method to reduce the edge of the sensor while keeping larger GR's structure if the signal collected is still sufficient to obtain good trigger efficiency.

The big advantage of the "slim edge" structure comes after irradiation. In this case, space charge sign inversion (SCSI) occurs and makes the sensor depletion region starting from the pixel region rather than from the backside region. The electric field distribution at the edge pixel is then changed (Fig. 3.26, for a sensor irradiated at $1 \times 10^{15} \text{ n}_{\text{eq}}/\text{cm}^2$ at $V_{\text{bias}} = 1000\text{V}$) and the problem of the lateral depletion zone disappears. Fig. 3.27 shows the corresponding potential distribution for the irradiated case. No lateral depletion appears near the pixel implant thus the full charge yield will be collected at the pixel electrode. The conclusion one can draw from these simulations is that slim edge design turns out to be a good solution to improve the active area in n -in- n sensor, especially after heavy irradiation exposure.

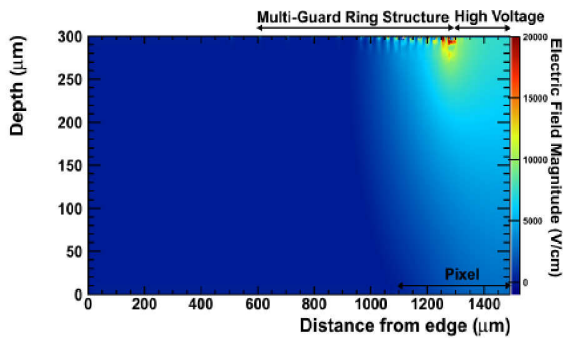


Fig.3.24 Electric field distribution for a slim edge design with 200 μm shift before irradiation

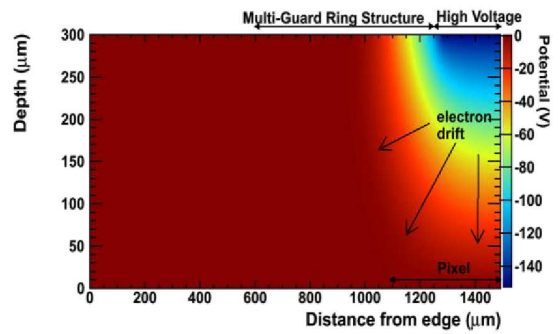


Fig.3.25 Potential distribution of a slim edge design with 200 μm shift before irradiation

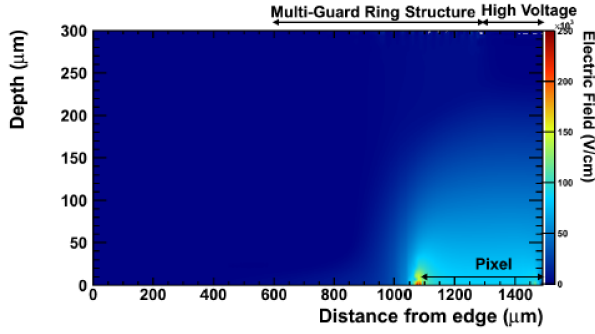


Fig.3.26 Electric field distribution for a slim edge design with 200 μm shift after irradiation of $1 \times 10^{15} \text{ n}_{\text{eq}}/\text{cm}^2$

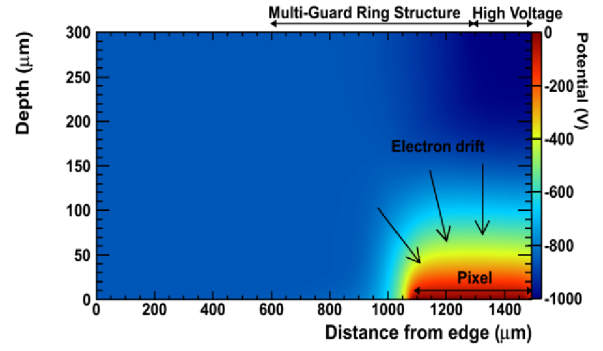


Fig.3.27 Potential distribution of for a slim edge design with 200 μm shift after irradiation of $1 \times 10^{15} \text{ n}_{\text{eq}}/\text{cm}^2$

Based on the simulations performed by our group, it has been shown that reduction of the inactive zone of the actual ATLAS planar pixel sensors of n^+ -in- n design can be safely done, meeting in the same time the IBL radiation hardness requirements. In particular, a reduction of the GR's number as well as a reduction of the dead zone between last GR and the cutting edge can be performed. Moreover, the "slim-edge" structure can be also safely implemented. Consequently, a n^+ -in- n pixel sensor design, characterized by 13 GR's, 100 μm dead edge and slim-edge with the outermost column pixels overlapping the GR structure by 250 μm (Fig.3.28) has been produced by CiS. In such way, the lateral dead zone in z direction has been reduced with respect to actual ATLAS sensors from 1100 μm to only 200 μm .

This design was chosen by IBL Sensor Review Panel (July 2011) to build the 75% of the IBL barrel.

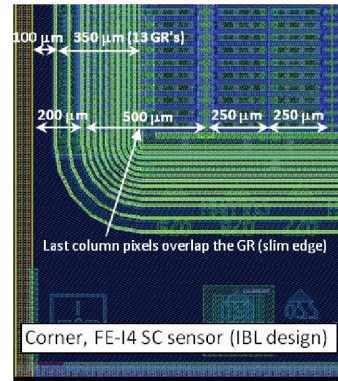


Fig.3.28 Corner view of a FE-I4 single chip planar pixel sensor used for the IBL construction: 13 GR's, 100 μm dead edge as well as slim-edge design can be observed. To fit the newly developed FE-I4 readout electronics, this sensor is represented by an array of $80 \times 336 = 26880$ pixels, with a pitch of $50 \times 250 \mu\text{m}$. The last pixels column is 500 μm length, overlapping the GR structure. Lateral dead zone remaining is only 200 μm .

4 Summary and perspectives of PART B

Experimental techniques as Secondary Ion Mass Spectrometry (SIMS) and Spreading Resistance Profiling (SRP) are today well known analysis methods used for the measurement of one-dimensional doping profiles in various silicon devices.

In this second part of my dissertation, I have shown how these methods can be successfully applied for the characterization of different technological steps as implantation and annealing used in the fabrication of silicon pixels sensors. In particular, planar pixels sensors candidates for the upgrade of the ATLAS inner detector at high-luminosity LHC have been measured and corresponding implants used to define the pixels regions, the guard-ring (GR) structures as well as the isolation regions between pixels have been analyzed.

For the first upgrade of ATLAS inner detector (i.e. Insertable BL project) new planar pixel sensors, with improved geometrical and radiation hardness characteristics with respect to actual design, were required to be investigated. In this context, TCAD simulation tool has been used and the experimental doping profiles obtained by SIMS and SRP methods were extremely useful to give input parameters for such kind of simulations. Moreover, DC electrical measurements performed on real devices and their comparison with TCAD simulations increased the confidence in our simulation code. Consequently, our TCAD simulations led to following quantitative results:

- actual GR's structure composed of 16 GR's can be safely reduced by at least 4 GR's;
- the actual dead-edge (i.e. from cutting edge to the outer GR) can be safely reduced from 450 μm to 100 μm ;
- the shifting of the last pixel column under the GR's region ("slim-edge" design) can subsequently reduce the inactive zone of the sensor, with even better characteristics after irradiations up to fluence of $1 \times 10^{15} \text{ n}_{\text{eq}}/\text{cm}^2$.

Such predictions lead to the fabrication of new planar pixel sensors based on n^+in-n design with only 200 μm dead zone in z-direction with respect to actual one of 1100 μm .

This design was chosen by IBL Sensor Review Panel (July 2011) to build the 75% of the IBL barrel.

Besides of the great role played by the experimental doping profiles in performing quantitative TCAD simulations, in this section we have seen also that such methods (in particular the SRP which measures the active doping density) are very useful tools for process control. Nevertheless, the most recent devices (highly segmented pixel sensors, CMOS devices, and superconducting junctions) have dimensions small enough to be sensitive to lateral diffusion mechanisms which escape the one-dimensional techniques.

A two-dimensional dopant profiling technique becomes therefore indispensable to provide a spatial resolution of the order of tenths nanometers covering a wide range of dopant concentrations (10^{15} to $5 \times 10^{21} \text{ at}/\text{cm}^3$). In this context, I am the initiator of a new project for the implementation, at the Centrale Technologique Universitaire Minerve (CTU Minerve) situated very close to our laboratory, of **a two-dimensional method** called Scanning Spreading Resistance Microscopy (SSRM). This project has already obtained a financial support from the NanoSaclay Labex and it is on the way to be developed in collaboration with our colleagues from CTU Minerve (thanks to Mr. Jean-Luc Perrossier) and l'Institut d'Electronique Fondamentale (IEF) d'Orsay (Mr. Dominique Débarre). This new method as well as the SIMS and SRP can be used in the future for the experimental investigations of new planar pixel sensors (i.e. edgeless design, small pixels) as well as to study the changes of doping profiles in irradiated sensors.

References:

- Allport, P.; di Girolamo, B.; Heinemann, B.; Torrence, E.; Wells, P.; (2013); Expected pileup values at the HL-LHC, *ATLAS-Communication-UPGRADE-2013-036*
- ATLAS Collaboration; (2008, a); The ATLAS experiment at the CERN Large Hadron Collider, *Journal of Instrumentation*, Vol. 3, S08003
- ATLAS Collaboration; (2008, b); ATLAS pixel detector electronics and sensors, *Journal of Instrumentation*, Vol. 3, P07007
- ATLAS IBL Collaboration; (2012); Prototype ATLAS IBL modules using the FE-I4A front-end readout chip, *Journal of Instrumentation*, Vol. 7, P11010
- Benoit, M; Lounis, A; **Dinu, N**; (2009, a); Simulation of guard-ring influence on the performance of ATLAS pixel detectors for inner layer replacement, *Journal of Instrumentation*, Vol. 4, P03025, Presented at Pixel 2008 International workshop, Fermilab, Batavia, IL, USA, September 2008.
- Benoit, M; Lounis, A; **Dinu, N**; (2009, b); Simulation of radiation damage effects on planar pixel guard-ring structure for ATLAS inner detector upgrade, *IEEE Transaction on Nuclear Science*, Vol. 56, Issue 6, 3236-3243.
- Benoit, M; Lounis, A; **Dinu, N**; (2010); Simulation of charge multiplication and trap-assisted tunneling in irradiated planar pixel sensors, *IEEE Nuclear Science Symposium Conference Record*, 612-616.
- Benoit, M.; (2011); Etude des détecteurs pixels planaires durcis aux radiations pour la mise à jour du détecteur de vertex d'ATLAS, *PhD Thesis*, LAL.
- Benninghoven, A.; Rydenauer, F.G.; Werner, H.W.; (1987); Secondary Ion Mass Spectrometry: Basic Concepts, Instrumental Aspects, Applications, and Trends, Wiley, New York.
- Brennan, R.; An introduction to spreading resistance analysis and its application in the semiconductor industry; *A Solecon Labs Seminar*, www.solecon.com
- Da Via, C. & al.; (2012); 3D silicon sensors: Design, large area production and quality assurance for the ATLAS IBL pixel detector upgrade, *Nuclear Instruments and Methods in Physics Research A*, Vol. 694, 321-330.
- Dinu, N**; Lounis, A.; Benoit, M.; Calderinu, G.; Idarraga, J.; Jomard, F.; Marchiori, G.; (2011, a); Dopant profiles of planar pixel sensors for the upgrade of the ATLAS inner detector, *IEEE Nuclear Science Symposium Conference, Valencia, Spain*.
- Dinu, N**; on behalf of ATLAS upgrade PPS collaboration; (2011, b); Recent progress of the ATLAS upgrade planar pixel sensors R&D project; Presented at *IEEE Nuclear Science Symposium Conference, Valencia, Spain*.
- Garcia-Sciveres, M.; & al.; (2011); The FE-I4 pixel readout integrated circuit, *Nuclear Instruments and Methods in Physics Research A*, Vol. 636, S155-S159.
- Goessling, C.; Muenstermann, D.; (2008); R&D on planar pixel sensor technology for the ATLAS inner detector upgrade (PPS proposal), CERN, <https://edms.cern.ch/document/966140>
- Grenier, P.; & al.; (2011); Test beam results of 3D silicon pixel sensors for the ATLAS upgrade, *Nuclear Instruments and Methods in Physics Research A*, Vol. 638, 33-40.
- Hull, R.; (1999); Properties of crystalline silicon, *The Institution of Engineering and Technology*, 129.
- Lounis, A.; Martinot, D.; Calderini, Marchiori, G.; Benoit, M.; **Dinu, N**; (2009); TCAD simulation of ATLAS pixel guard-ring and edge structure for Super-LHC upgrade, *ATLAS Note*, ATL-UPGRADE-PUB-2010-001.
- Lounis, A.; Benoit, M.; **Dinu, N**; (2012) ; TCAD simulations of planar pixel sensors, *Proceedings of Science (Vertex 2011)* 030.
- Mazur, R.G., Dickey, P.H.; (1966); A spreading resistance technique for resistivity measurements in silicon, *Journal of The Electrochemical Society*, Vol. 113, Issue 3, 255-259.
- Moll, M.; (1999); Radiation damage in silicon particle detectors, *PhD Thesis*, Hamburg
- Parker, S. I.; Kenney, C. J.; Segal, J.; (1997); 3D – A proposed new architecture for solid-state radiation detectors, *Nuclear Instruments and Methods in Physics Research A*, Vol. 395, 328-343.
- Rossi, L. Fischer, P.; Rohe, T. ; Wermes, N.; (2006); Pixel detectors: from fundamentals to applications, *Springer-Verlag Berlin Heidelberg*,
- Rossi, L.; Brüning, O.; (2012); High Luminosity Large Hadron Collider – A description for the European Strategy Preparatory Group.
- Silvaco Inc., <http://www.silvaco.fr/products/tcad/index.html>
- Silvaco Inc.; (2008); ATLAS User's Manual, Device Simulation Software, V 5.14.0.R 81, 139.
- Stanford Technology CAD home page, Stanford University, <http://www-tcad.stanford.edu/>
- Synopsys Inc., <http://www.synopsys.com/Tools/TCAD/Pages/default.aspx>

Annex 1

Reverse I-V characteristic for ideal and practical diode; relationship with the GM-APD characteristics

As well known from the literature, the ideal current-voltage static characteristic of a p-n diode is expressed by the Shockley ideal diode equation or diode law:

$$I_{diode} = I_{sat} \left(e^{\frac{qV_{BIAS}}{nk_B T}} - 1 \right) = I_{sat} \left(e^{\frac{V_{BIAS}}{nV_T}} - 1 \right) \quad A1.1$$

where I_{sat} is the reverse bias saturation current, V_{BIAS} is the voltage across the diode, V_T is the thermal voltage, n is the ideality factor, also known as the quality factor or sometimes emission coefficient. The factor n typically varies from 1 to 2, depending on the fabrication process and semiconductor material and in many cases is assumed to be approximately equal to 1. The thermal voltage V_T is approximately 25.85 at 300K. At any temperature it is known constant defined by: $V_T = k_B T / q$, where $k_B = 1.38 \times 10^{-23}$ J/K is the Boltzmann constant, T is the absolute temperature (273K=0°C) and q is the electron charge. The ideal current-voltage characteristic in linear and logarithmic scale is represented in Figs. A1.1 (a) and (b).

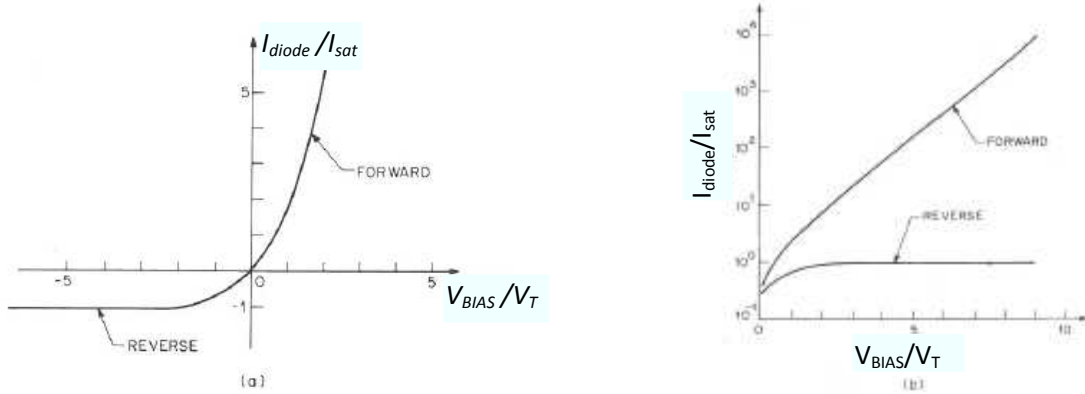


Fig. A1.1 Current-voltage characteristic of an ideal p-n junction diode **(a)** linear plot; **(b)** logarithmic plot

Since the GM-APD device described in details in this dissertation is a p-n diode working in reverse bias mode, we will discuss in details the reverse current.

The reverse bias saturation current I_{sat} is caused by the diffusion of minority electrons from the p-side to the n-side and the diffusion of minority holes from the n-side to the p-side; therefore I_{sat} depends on the density of the minority carriers as well as of the diffusion coefficient of electrons and holes as expressed by the formula:

$$I_{sat} = I_{diffusion} \equiv \frac{qAD_p p_{n0}}{L_p} + \frac{qAD_n n_{p0}}{L_n} = qAn_i^2 \left(\frac{D_p}{N_D L_p} + \frac{D_n}{N_A L_n} \right) \quad A1.2$$

where q is the electron charge, A is the junction area, n_i is the intrinsic carrier concentration, D_n and D_p are the diffusion coefficients of electrons and respectively holes, $n_{p0} \cong \frac{n_i^2}{N_A}$ and $p_{n0} \cong \frac{n_i^2}{N_D}$ are the density of minority electrons and respectively minority holes in a p-type and respectively n-type side of the junction at thermal equilibrium, N_A and N_D are the concentration of acceptor impurities and respectively the donor impurities, L_n and L_p are diffusion lengths of electrons and respectively holes.

For p-n diodes build in Si, the ideal Shockley equation can give only a qualitative agreement with reverse current-voltage characteristics of real devices. The departures from the ideal situation are mainly due to: (1) surface effects; (2) the carrier's generation in the depletion region by thermal effect; (3) the tunneling of carriers between states in the band gap. In addition, under sufficiently larger field in the reverse direction, the junction will breakdown (as a result for example, of avalanche multiplication).

The first two effects give important contributions to the reverse dark current of a p-n diode preceding the breakdown (we will call this pre-breakdown current I_{pre-BD}), therefore they will be discussed in more details in the following.

Surface effects

Surface effects occur because the entire reverse bias voltage is developed across the depletion region, and this procedure produces high electric field at the surface. The junction surface is usually contaminated with ionic impurities that get deposited during the numerous chemical treatments used in device fabrication. These ions can move under the action of the applied field. Positive ions may also invert the surface on the p-side of the junction to n-type, thus establishing a direct conducting channel over the surface. A surface leakage current flows through this channel which rises with the voltage in some unpredictable manner.

Thermal generation process

Following Shockley-Read-Hall (SRH) generation-recombination theory (Shockley & Read, 1952), in a semiconductor at thermal equilibrium, the lattice vibrations can break some bounds between neighboring atoms. As a consequence of breaking these bounds, *electron-hole pairs can be generated*. In terms of energy bands, this process can be described by the transition of a valence electron to the conduction band leaving a hole in the valence band and it is called *carrier generation*. The number of electron-hole pairs generated per cm^3 and per second is called generation rate, G_{th} . In the reverse process, called *carrier recombination*, the electron makes a transition from the conduction band to the valence band with a transition rate R_{th} .

At thermal equilibrium the recombination rate must be equal to the generation rate: $G_{th} = R_{th}$ and consequently the net transition rate, U , equals zero, i.e.: $U = R_{th} - G_{th} = 0$.

The dominant generation process in silicon is an indirect transition via localized energy states in the forbidden gap (i.e. the silicon is an indirect semiconductor). These intermediate-level states – called *generation centers* – are due to the presence of impurities and crystal defects. These defects can assume two or more charge states. In some circumstances, these centers can involve a single type of carrier. Since these generation centers can also absorb carriers, they act as recombination centers. In practice, there is no difference between generation or generation-recombination centers. Sometimes the generation-recombination centers are improperly named trapping centers. Trapping is the phenomenon by which some of the excess electron-hole pairs, originally created in equal densities, are trapped temporarily on impurities. Those electrons captured on trap level can be

subsequently emitted back to the conduction band by thermal emission. Therefore, the word “trap” is best not used when recombination or generation occurs; “recombination-generation (R-G) center” is better.

When the concentration of carriers are below those corresponding to thermal equilibrium, for instance in a depleted region (i.e. $p < n_i$, $n < n_i$; $pn < n_i^2$), the process of carrier generation will dominate over that of carrier recombination (i.e. called also Shockley-Read-Hall generation). In this case, the generation rate in the depletion region of a biased p-n junction can be expressed as following (Tyagi, 1991):

$$U = -G_{th} = -\left[\frac{n_i^2 [\exp(V_{BIAS}/V_T) - 1]}{\tau_n(p + p_1) + \tau_p(n + n_1)} \right] \quad A1.3$$

where: n_{int} is the intrinsic carrier concentration; τ_n is the electrons recombination lifetime, τ_p is the holes recombination lifetime; n and p are the electrons and respectively holes concentrations in the depletion region; n_1 and p_1 represents the equilibrium electrons and holes concentrations that would result when the Fermi level lies at the energy level of the generation-recombination center and they are defined as:

$$n_1 = n_i \exp[(E_{G-R} - E_i)/k_B T]; \quad p_1 = n_i \exp[(E_i - E_{G-R})/k_B T] \quad A1.4$$

where E_{G-R} is the energy level of the generation – recombination centers; k_B is the Boltzmann constant; T is the temperature.

When a reverse V_{BIAS} is across a p-n junction is higher than 0.1V, both electrons and holes concentrations become negligible and U can be written as:

$$U = -\frac{n_i^2}{\tau_n p_1 + \tau_p n_1} \quad A1.5$$

The concept of recombination lifetime holds when excess carriers decay as a result of recombination, and therefore during recombination an electron-hole pair ceases to exist on average after a time τ . For a p-type semiconductor, the electron recombination lifetime is expressed by:

$$\tau_n = \frac{1}{N_{G-R} v_{th,n} \sigma_n} \quad A1.6$$

where: N_{G-R} is the concentration of generation-recombination centers; $v_{th,n}$ is the thermal velocity of electrons (i.e. $v_{th,n} = \sqrt{3k_B T/m^*}$ where, m^* is the effective mass of electron; at 300K, $v_{th,n} = 2.2 \times 10^7$ cm/s); σ_n is the capture cross-section of electrons (i.e. it would be of the order of the atomic dimension, that is $\cong 10^{-15}$ cm²). For a n-type semiconductor, the holes recombination lifetime is expressed by:

$$\tau_p = \frac{1}{N_{G-R} v_{th,p} \sigma_p} \quad A1.7$$

where: N_{G-R} is the concentration of generation-recombination centers; $v_{th,p}$ is the thermal velocity of holes (i.e. at 300K, $v_{th,p} = 1.9 \times 10^7$ cm/s); σ_p is the capture cross-section of holes (i.e. $\cong 6 \times 10^{-16}$ cm²).

The recombination lifetime of minority carriers depends inversely on the density of recombination centers and the capture cross sections, but does not depends directly on the energy level of the impurity. It does depend indirectly on the energy level, because the capture cross-section tends to be highest for impurities with E_{G-R} near the center of the band-gap and lowest for E_{G-R} near the conduction or valence band. Therefore, the

generation centers are the most effective when the energy level E_{G-R} of the centers approaches the intrinsic Fermi level E_{int} , (i.e. defects with energy levels close to the middle of the energy gap).

Therefore, considering that the impurity generation-recombination level is situated at the intrinsic Fermi level E_i , so that $n_1 = p_1 = n_i$ and assuming also that $\tau_n = \tau_p = \tau_0$, then the net rate of the generation of the electrons-holes pairs in the depletion region can be expressed by the simplified expression:

$$U = -\frac{n_i}{\tau_0} \quad \text{A1. 8}$$

where τ_0 is the effective lifetime of the carriers in the depletion region.

Since the intrinsic concentration of carriers as well as the effective lifetime depends on temperature, the generation rates are temperature dependent.

The thermal generation takes place whenever there are impurities or defects in the semiconductor and it increases with increasing temperature. Since semiconductors always contain some impurities, this mechanism is always active, and it is particularly important for indirect band-gap semiconductors like silicon.

Tunneling processes

The application of a high electric field on a p-n junction stimulates the ionization of the deep-level defects and determines the emission and capture processes. The main phenomena involve are the Poole-Frenkel effect (Frenkel, 1938), the transitions via traps (trap-assisted tunneling - TAT) and direct band-to-band tunneling (Hurkx, 1992).

The effects of the tunneling phenomena are to increase the emissivity of the generation-recombination centers and they have a strong dependence on the electric field profile (usually they are active at electric fields exceeding 10^6 V/cm).

In GM-APD devices, where the maximum of the electric field (usually $4-5 \times 10^5$ V/cm) is well below a critical value, these effects are estimated to be negligible (Zappa, 2007).

Taking into account the additional effects presented previously, the total leakage current $I_{diode-pre-BD}$ of p-n diode before the breakdown voltage can be expressed as following:

$$I_{diode-pre-BD} = I_{surface} + I_{bulk} \quad \text{A1. 9}$$

where $I_{surface}$ is the surface leakage current and I_{bulk} is the bulk generation current. The I_{bulk} is given by the sum of the diffusion current $I_{diffusion}$ and the generation current $I_{generation}$ in the depletion region:

$$I_{bulk} = I_{diffusion} + I_{generation} \quad \text{A1. 10}$$

The current due to generation in the depletion region is obtained by integrating the generation rate U over the entire width of the depletion region. Thus:

$$I_{generation} = \int_0^w qA|U|dx \cong qA|U|w = -\frac{qn_iAw}{\tau_0} \quad \text{A1. 11}$$

where w is the depletion-layer width, A is the junction area, U is the generation rate, q is the electron charge, n_i is the intrinsic concentration, τ_0 is the effective lifetime of the carriers in the depletion region. The negative sign indicates that the current is in the reverse direction.

Taking into account the additional contributions to the reverse leakage current (i.e. the generation current and the surface effects), the current-voltage characteristic of a practical GM-APD diode will show the shape presented in Fig. A1.2. We can note that $I_{diffusion}$ is independent of the V_{BIAS} , but $I_{generation}$ increases with the V_{BIAS} because of the increase in the depletion region width w .

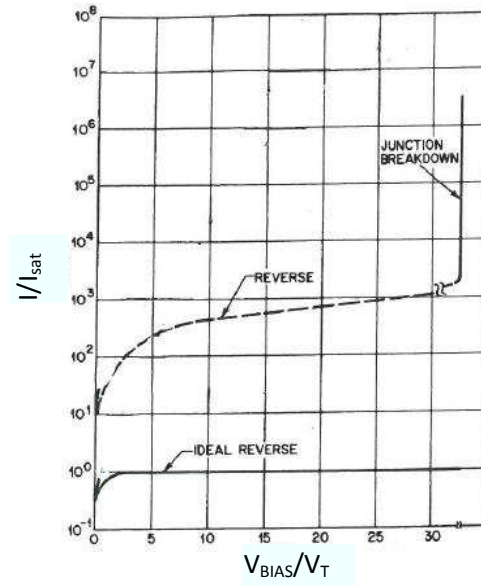


Fig. A1.2 Current-voltage characteristic for a practical Si diode; reverse leakage current takes into account generation recombination current and surface effects [Moll J.L.; (1958); The evolution of the theory of the current-voltage characteristics of p-n junctions; *Proc. IRE*, Vol. 46, 1076]

For the case of a GM-APD diode, the bulk generation leakage current is the most important component of the leakage current, because only carriers generated in the bulk can determine breakdown avalanches within the depleted region when the diode is biased over breakdown voltage V_{BD} (i.e. the working regime of a GM-APD) and therefore give contribution to the reverse dark current above breakdown $I_{post-BD}$. The bulk diffusion current gives less contribution with respect to bulk generation leakage current.

When a sufficiently high field is applied to a p-n junction diode (i.e. the applied bias voltage exceed a critical value), the junction “breakdown” and conducts a very large current. The most important mechanism of junction breakdown is the avalanche multiplication (or impact ionization): the carriers traversing the high field region gain sufficiently energy and create electron-hole pairs via collisions with bound electrons, determining charges multiplication. The critical value of the applied bias voltage over which the phenomenon of avalanche breakdown takes place is called breakdown voltage V_{BD} . The V_{BD} can be recognized in the current-voltage characteristic of a diode by the well-known *knee of the breakdown*, above which the current increases very rapidly with the bias voltage (Fig. A1.2).

The electron-hole pair generation rate G from impact ionization is related to the electron and holes ionization rates and it can be expressed by the following expression (Sze, 1981):

$$G = \alpha_n n v_n + \alpha_p p v_p$$

A1. 12

where α_n is the electrons ionization rate, n is the electron density, v_n is the electron drift velocity. Similarly, α_p is the holes ionization rate, p is the holes density, v_p is the holes drift velocity. The ionization rate α_n (α_p) is defined as the number of electron-hole pairs generated by an electron (hole) per unit distance travelled. *Both α_n and α_p are strongly dependent on the electric field.*

A physical expression for the ionization rate is given by (Thornber, 1981):

$$\alpha(\varepsilon) = \left(\frac{q\varepsilon}{E_I} \right) \exp \left\{ - \frac{\varepsilon_I}{\varepsilon \left(1 + \frac{\varepsilon}{\varepsilon_p} \right) + \varepsilon_{kT}} \right\}$$

where E_I is the high-field, effective ionization threshold energy, and ε_{kT} , ε_p , ε_I are thresholds fields for carriers to overcome the decelerating effects of thermal, optical-phonon, and ionization scattering, respectively. For Si, the value of E_I is found to be 3.6 eV for electrons and 5.0 eV for holes.

The parameters α_n and α_p are not well determined yet, and large discrepancies exist among the values extracted from the various models as we can in the Fig. A1.3 where they are represented as a function of the inverse of the electrical field (Grant, 1973). For the case of GM-APD's, the reciprocal field is of the order of $\sim 2 \times 10^{-6}$ cm/V.

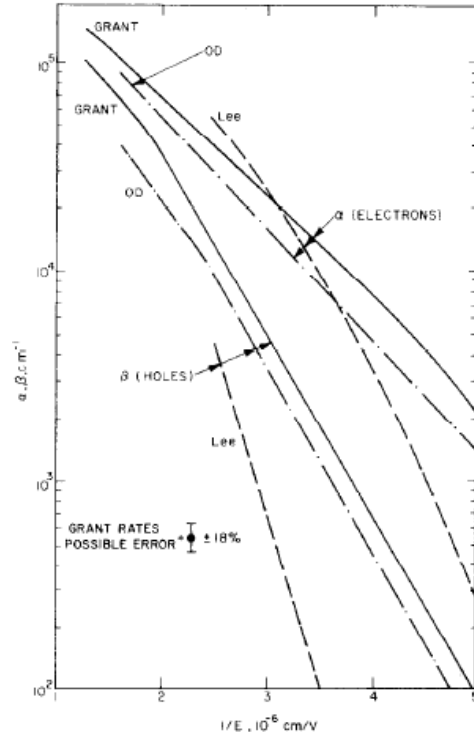


Fig. A1.3 Room temperature ionization rates in Si as a function of reciprocal field. Data are from Lee & al., 1964, van Overstraeten et al. (OD), 1970, and Grant, 1973.

When the bias voltage over the diode exceeds the breakdown voltage and the diode reaches the Geiger mode condition (i.e. the electrical field in the depletion region is high enough that electron and holes are created by impact ionization), the ionization rates of electrons and holes are so high that the electrons and holes are multiplying very fast, with a multiplication factor approaching infinity. Using quenching mechanisms, the multiplication factor is usually limited to a value of the order of 10^6 charges in one avalanche (i.e. this parameter is called gain G). The number of avalanches per second generated by carriers traversing the depleted region when the device is kept in dark conditions is called dark count rate DCR and it represents an important characteristic of a GM-APD device since it determines the intrinsic noise of the device. Therefore, the overall behavior of the GM-APD diode IV characteristic above breakdown can be explained considering the dynamics of the device and it can be expressed as:

$$I_{post-BD} \approx q \cdot G \cdot DCR \propto q \cdot \Delta V \cdot \Delta V \quad (quadratic) \quad A1. 13$$

where q is the electron charge, G is the gain (i.e. number of charges developed in one avalanche), DCR is the dark count rate, $\Delta V = V_{BIAS} - V_{BD}$ is the overvoltage. Thus, the IV characteristic of a GM-APD above breakdown features a

well defined shape (i.e. quadratic dependence with the overvoltage ΔV), as it can be seen in the experimental results in the Section 4.2.1.

The $I_{post-BD}$ is essentially given by the bulk generation current before breakdown (since surface charges cannot generate avalanches in the depletion region), multiplied by the multiplication factor G . Knowing the gain G , the value of the diode leakage current $I_{post-BD}$ of a GM-APD gives us information on DCR. Since the value of DCR is desired to be as low as possible, the number of carriers traversing the depletion region in dark conditions should be as low as possible (i.e. reducing the number of generation-recombination centers in the depletion region).

References:

- Grant, W.N. ; (1973) ; Electron and hole ionization rates in epitaxial silicon at high electric fields, *Solid State Electronics*, Vol. 16, Issue 10, 1189-1203
- Lee, C.A.; Logan, R.A.; Batdorf, R.L.; Kleimack, J.J.; Wiegmann, W.; (1964); Ionization rates of holes and electrons in silicon, *Physical review*, Vol. 134, Issue 3A, A761-A774.
- Overstraten, Van R.; Man, de H; (1970); Measurement of the ionization rates in diffused silicon p-n junctions, *Solid State Electronics*, Vol. 13, Issue 5, 583-608
- Sze, S.M.; (1981); Physics of semiconductor devices, 2nd edition, *John Willey & Sons*.
- Thornber, K.K.; (1981); Applications of scaling to problems in high field electronic transport, *Journal of Applied Physics*, Vol. 52, Issue 1, 279- 290.

Annex 2

Temperature dependence of reverse I-V characteristic; relationship with the GM-APD characteristics

As we have seen in the Annex 1, the total reverse current through a p-n junction diode before breakdown is given by:

$$I_{diode-pre-BD} = I_{surface} + I_{bulk} = I_{surface} + I_{diffusion} + I_{generation} \quad A2.1$$

The diffusion leakage current $I_{diffusion}$ is given by:

$$I_{diffusion} = qAn_i^2 \left(\frac{D_p}{N_D L_p} + \frac{D_n}{N_A L_n} \right) \quad A2.2$$

where all parameters have been described in the Annex 1. In this equation D_n , D_p , L_n and L_p are all temperature dependent, but the temperature dependence of $I_{diffusion}$ is dominated by n_i^2 .

The generation leakage current $I_{generation}$ is given by:

$$I_{generation} = -\frac{qn_i A w}{\tau_0} \quad A2.3$$

where all parameters have been described in the Annex 1. In this equation τ_0 is depending of temperature, but the temperature dependence of the $I_{generation}$ is dominated by n_i .

The surface leakage current $I_{surface}$ is generally less sensitive to temperature than $I_{generation}$.

Therefore, the temperature dependence of the both diffusion and generation leakage currents through a p-n junction before the breakdown is principally described by the temperature variation of the intrinsic carrier concentration.

The intrinsic carrier concentration is given by the expression:

$$n_i = \sqrt{N_C N_V} \cdot \exp\left(-\frac{E_g}{2k_B T}\right) \quad A2.4$$

where N_C and N_V is the effective density of states in the conduction band and respectively in the valence band, E_g is the band gap energy, k_B is the Boltzmann constant, T is the temperature. The effective densities of states depend of temperature as following:

$$N_C = 2 \left(\frac{2\pi \cdot m_e^* k_B T}{h^2} \right)^{3/2} ; \quad N_V = 2 \left(\frac{2\pi \cdot m_h^* k_B T}{h^2} \right)^{3/2} \quad A2.5$$

where m_e^* and m_h^* are the density of states effective mass of electrons and holes, h is the Plank constant, k_B is the Boltzmann constant, T is the temperature.

Therefore, the intrinsic carrier concentration can be written as:

$$n_i = K_1 T^{3/2} \exp\left(-\frac{E_g}{2k_B T}\right) \quad \text{A2. 6}$$

Where K_1 is a constant given by the expression:

$$K_1 = \sqrt{32 \left(\frac{\pi^2 k_B^2 m_e^* m_h^*}{h^4} \right)^{3/2}} \quad \text{A2. 7}$$

For Si, the band gap energy E_g is decreasing with increasing temperature as following (Sze, 1981):

$$E_g(T) = E_g(0) - \frac{\alpha T^2}{\beta + T} \quad \text{A2. 8}$$

where $\alpha = 0.473 \times 10^{-3}$ and $\beta = 636$. At $T=0\text{K}$, $E_{g,\text{Si}}(0) = 1.166\text{V}$. The temperature variation of the energy band gap of Si is represented in Fig. A2.1.

Since there are many difficulties in the independent measurements of the quantities expressing the intrinsic concentration (m_e^* , m_h^* , $E_g(T)$), the substituting of the numerical values does not lead to results in close agreement with the experimentally determined values. From measurements (Adler et al. 1964), it is found that the temperature above 50K, the intrinsic concentration of Si is given by:

$$n_i(T) = 3.88 \cdot 10^{16} \cdot T^{3/2} \cdot \exp\left(-\frac{7000}{T}\right) \text{ cm}^{-3} \quad \text{A2. 9}$$

The temperature dependence of the intrinsic carrier concentration in Si is presented in Fig. A2.2. We can observe the great variations of this parameter over many orders of magnitude in the temperature range from 100 to 300K.

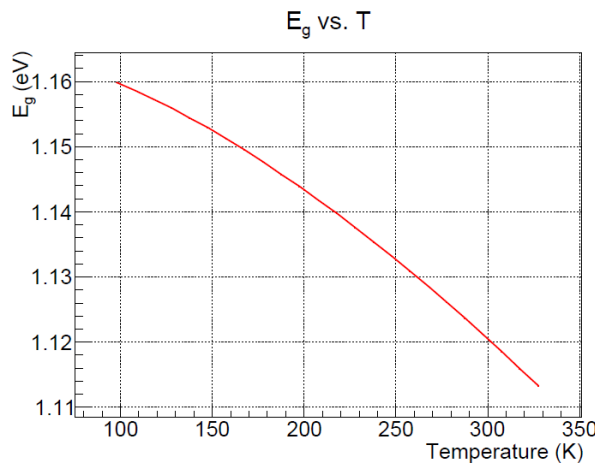


Fig. A2.1 Temperature variation of the band gap energy E_g in Si

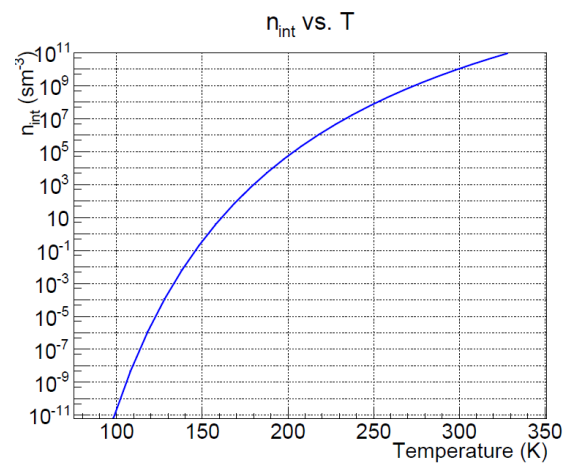


Fig. A2.2 Temperature variation of intrinsic carrier concentration n_i in Si

If we return back now to the expression of the diffusion current, substituting for n_i^2 , and assuming that all other quantities remain invariant with temperature, $I_{diffusion}$ can be expressed as:

$$I_{diffusion} \cong A_1 \cdot T^3 \cdot \exp\left(-\frac{E_{g0}}{k_B T}\right) \quad \text{A2. 10}$$

where E_{g0} is the extrapolated band gap at 0K and A_1 is the constant of proportionality. Differentiating this equation with respect to T , we obtain the following equation for the temperature coefficient of the diffusion current:

$$\frac{1}{I_{diffusion}} \frac{dI_{diffusion}}{dT} = \frac{1}{T} \left[3 + \frac{E_{g0}}{k_B T} \right] \quad \text{A2. 11}$$

If in the expression of the generation current we substitute for n_i and assuming that all other quantities remain invariant with temperature, $I_{generation}$ can be expressed as:

$$I_{generation} \cong A_2 \cdot T^{3/2} \cdot \exp\left(-\frac{E_{g0}}{k_B T}\right) \quad \text{A2. 12}$$

where E_{g0} is the extrapolated band gap at 0K and A_2 is the constant of proportionality. Differentiating this equation with respect to T , we obtain the following equation for the temperature coefficient of the generation current:

$$\frac{1}{I_{generation}} \frac{dI_{generation}}{dT} = \frac{1}{2T} \left[3 + \frac{E_{g0}}{k_B T} \right] \quad \text{A2. 13}$$

Substituting $E_{g0} = 1.166\text{eV}$ for Si, we obtain a change of about 8 percent per degree Kelvin in the reverse current at 300K.

In practice, a plot of $\log(I_{diffusion} + I_{generation})$ vs. $1/k_B T$ will yield a slope of $-E_{g0}$ for diffusion and approximately $-E_{g0}/2$ for generation dominated operation. At sufficiently high temperatures, diffusion will always dominate.

When the applied bias voltage across a p-n junction diode exceeds the critical breakdown voltage V_{BD} , the avalanche breakdown takes place (see Annex 1). Fig. A2.3 shows the reverse characteristics of a p-n junction with avalanche breakdown at two temperatures. We can observe that the breakdown voltage is increasing with the temperature, making the temperature coefficient of breakdown voltage positive.

A simple explanation of the increasing of V_{BD} with temperature is that the carriers, passing through the depletion layer under a high field, lose part of their energy to optical phonon generation after traveling each electron-phonon mean free path λ . The *mean free path λ decreases with increasing temperature* of the crystal lattice, and thus for a given distance at a fixed field, the carriers

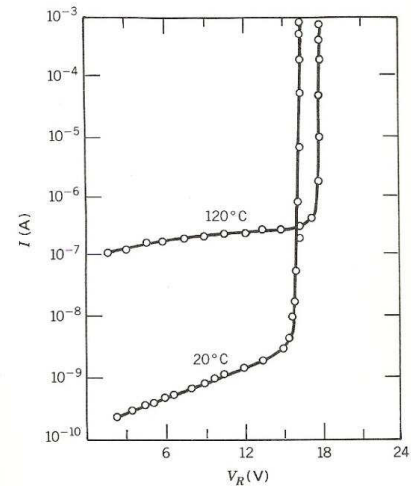


Fig. A2.3 Reverse I-V characteristic of a p-n junction with avalanche breakdown at two different temperatures

lose more energy to the crystal lattice. Under these conditions, the carriers must pass to a greater potential difference (or higher voltage) before they can acquire sufficient energy to generate an electron-hole pair. *Thus the net effect is seen to be an increase in avalanche breakdown with temperature.*

As mentioned in the Annex 1, the avalanche breakdown voltage is related to the electrons and holes ionization rates. The *dependence of the charge carrier ionization rates on the electrical field* has been expressed in terms of Baraff's theory (Baraff, 1962) using the following three parameters:

- the ionization threshold energy, E_i ; the most appropriate value of E_i is $\approx 3E_g/2=1.68V$, where E_g is the Si band gap energy.
- carrier mean free path for optical-phonon generation, λ ; the average values of λ obtained from Baraff's theory are 40 Å for holes and 65 Å for electrons;
- Raman optical phonon energy, E_r ; for Si, the $E_r=0.063$ eV.

Detailed calculations *on temperature dependence of the charge carrier ionization probability per unit distance traveled (ionization rates) on the electrical field* have been done by Crowell & Sze, 1966. They have used the average energy lost per phonon collision $\langle E_r \rangle$ in the place of E_r , and they expressed the values of λ and $\langle E_r \rangle$ as a function of temperature as following (Crowell & Sze, 1966):

$$\lambda = \lambda_0 \tanh\left(\frac{E_r}{2k_B T}\right) \quad \text{A2. 14}$$

$$\langle E_r \rangle = E_r \tanh\left(\frac{E_r}{2k_B T}\right) \quad \text{A2. 15}$$

and

$$\frac{\lambda}{\lambda_0} = \frac{\langle E_r \rangle}{E_r} \quad \text{A2. 16}$$

where λ_0 is the high-energy low temperature asymptotic value of the phonon scattering mean free path (for Si, $\lambda_0 = 76$ Å for electrons and $\lambda_0 = 55$ Å for holes). We have to note that:

$$\frac{\langle E_r \rangle}{\lambda} = \frac{E_r}{\lambda_0} = \text{const} \quad \text{A2. 17}$$

Since $1/\lambda$ is the number of collisions per unit distance traveled by a charge carrier, the energy lost per unit path length is independent of temperature.

Experimental values as well as the theoretical ones of ionization rates in Si as a function of the reciprocal field are represented in Fig. A2.4 for different temperatures (Crowell & Sze, 1966). The agreement at both 100K and 213K is satisfactory. These results can be understood qualitatively as follows. In the low field region, the ionization rate is strongly temperature dependent because the ionizing carriers travel several mean free paths without a collision. In the high field region (i.e. the case of GM-APD for which the electric field is of the order of 5×10^5 V/cm), a greater fraction of the carriers produces ionization and the energy loss per unit distance by interaction with crystal lattice (i.e. phonon generation) is more significant. This is less temperature sensitive because the energy lost per unit path length along the random walk of the carrier (i.e. the sum of ionization energy and phonon generation energy) is independent of temperature (Equation A2.17). Therefore, this further confirms *the fundamental significance of the mean free path λ and its temperature dependence (Fig. A2.5).*

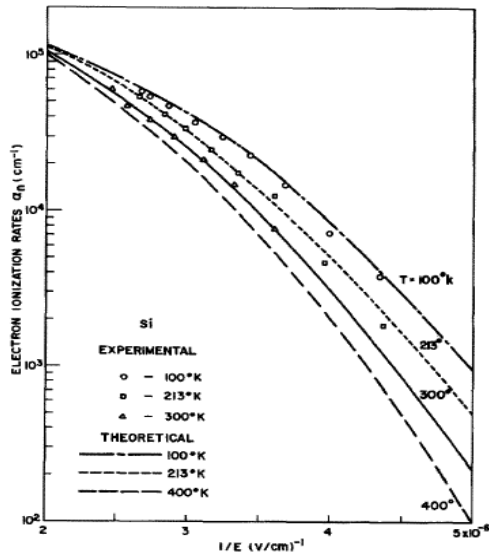


Fig. A2.4 Electron ionization rate versus reciprocal electric field in Si for different temperatures (Crowell & Sze, 1966)

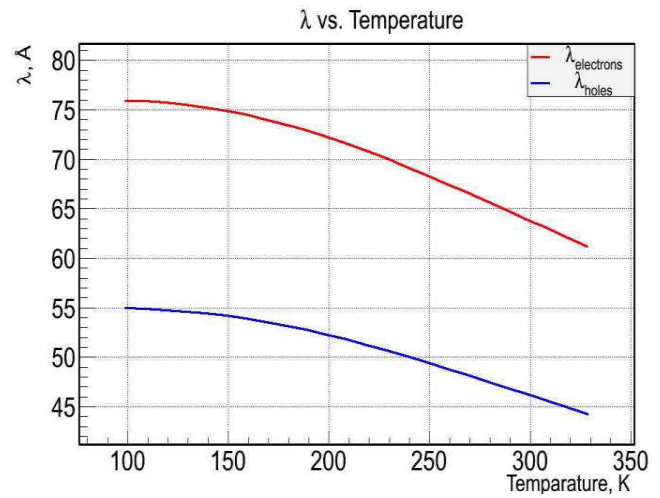


Fig. A2.5 Carrier mean free path for optical-phonon generation versus T

The predicted values of V_{BD} normalized to the room temperature value for different background doping concentrations are shown in Fig. A2.6 (Crowell & Sze, 1966). Experimental values of V_{BD} as a function of T for a p^+n abrupt junction with impurity concentration as parameter have been measured by Chang et al, (Chang et al. 1971) and they are given in Fig. A2.7. Following these plots we can observe that different background doping concentrations determine different V_{BD} values and for a given doping concentration, V_{BD} increases with increasing T .

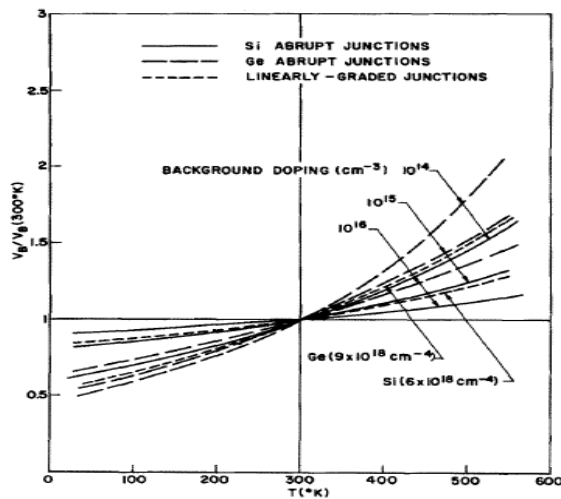


Fig. A2.6 Breakdown voltage vs temperature for Si and Ge p-n junctions. $V_{BD}(300K)$ is 2000, 330, and 60V for Si abrupt junctions with background dopings of 10^{14} , 10^{15} , and 10^{16} cm^{-3} respectively (Crowell & Sze, 1966)

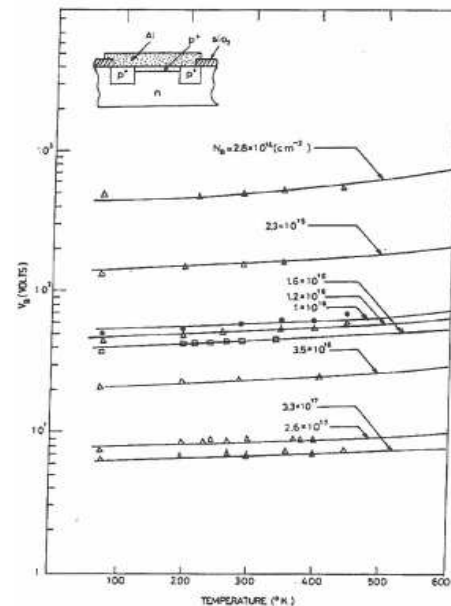


Fig. A2.7 Experimental breakdown voltage versus temperature with impurity concentration as parameter (Chang,et al. 1971)

Regarding the temperature variation of the reverse I-V characteristic of a GM-APD diode, we expect a close relationship with the temperature variation of the p-n diode characteristics as presented above:

- The pre-breakdown current I_{pre-BD} increases with increasing temperature; its temperature dependence is dominated by the temperature dependence of the $I_{generation}$ when $T \sim 20^\circ\text{C}$ and $I_{diffusion}$ for much higher temperatures (exceeding 100°C);
- The V_{BD} increases with temperature;
- The post-breakdown current $I_{post-BD}$ increases with temperature (the V_{BD} changes with temperature; therefore the $I_{post-BD}$ at different temperatures should be compared at the same overvoltage $\Delta V = V_{BIAS} - V_{BD}$); since the most important component of the post-breakdown current is the bulk generation current, the temperature dependence of the $I_{post-BD}$ will be in particular related to the temperature variation of the generation current $I_{generation}$. If we consider the dynamics of the GM-APD device over breakdown, the temperature variation of the $I_{generation}$ will determine a temperature variation of the dark count rate DCR .

References:

- Adler R.B.; Smith, A.C.; Longini, R.L.; (1964); Introduction to semiconductor physics, Vol.1, John Wiley, New York
- Baraff, G.A.; (1962); Distribution functions and ionization rates for hot electrons in semiconductors, *Physical review*, Vol. 128, Issue 6, 2507-2517.
- Chang, C.Y.; Chiu, S.S.; Hsu, L.P.; (1971); Temperature dependence of breakdown voltage in silicon abrupt pn junctions, *IEEE Transactions on Electron Devices*, Vol.18, Issue 6, 391-393.
- Crowell, C.R.; Sze, S.M.; (1966); Temperature dependence of avalanche multiplication in semiconductors, *Applied Physics Letters*, Vol. 9, Issue 6, 242-244.
- Sze, S.M.; (1981); Physics of semiconductor devices, 2nd edition, John Willey & Sons.

