Università degli Studi di Torino
Scuola di Dottorato in Scienza ed Alta Tecnologia
Indirizzo Fisica e Astrofisica
XXX ciclo

Design and Test of Sensors and Front-End Electronics for Fast Timing in High Energy Physics

---

Author:
Francesca Cenna

Tutor:
Nicolò Cartiglia

January 28, 2018

To all the wonderful people that are giving me love and understanding

Abstract

The evolution of high energy physics experiments is always imposing challenging demands on both detectors and their readout electronics. In particular, the future High Luminosity LHC (HL-LHC) environment foreseen for 2026 will require a substantial improvement in the reconstruction of particle tracks due to the pileup of 140-200 events per bunch crossing, resulting in vertices overlapping in space. In this scenario, the present detectors at CMS and ATLAS will only be capable to reconstruct vertexes with an efficiency of $\sim$85%, as a percentage of 10-15% of these vertexes will be the sum of two different events. The introduction of a fourth coordinate –namely the time coordinate– to each track measurement can drastically improve the effectiveness of vertex reconstruction, even in such critical conditions: a time resolution of about 30 ps will almost completely remove the ambiguity of assigning tracks to the correct vertex. For this purpose, both CMS and ATLAS experiments are considering the insertion of a timing layer in the Phase II upgrade, in order to assign a time stamp to minimum ionizing particles. The development of a new type of silicon detector able to provide both spatial and time measurements with good accuracy is a key component for this task, and it has been embraced as a high priority by both experiments.

In the last decades, silicon detectors allowed to cover large areas while providing a very high spatial granularity and an excellent rate capability. The employment of silicon detectors for tracking purposes is fundamental in modern particle physics experiments. However, the timing performances of these devices are limited by the signal formation process, which typically constrained their time resolution to hundreds of picoseconds.

In order to enhance the time resolution of a silicon detector two main conditions should be met: the signal should be large and its rise time short. Both can be accomplished by employing a thin sensor with a moderate charge multiplication. Silicon devices able to provide a 4-dimensional image have been called "Ultra-Fast Silicon Detectors"(UFSD). These detectors are based on the Low Gain Avalanche Detector (LGAD) technology and are currently manufactured by Centro Nacional de Mircoelectrònica, Fondazione Bruno Kessler and Hamamastu. UFSD have been chosen as timing detectors for the endcaps of the MIP Timing Detector of CMS and ATLAS and for the Precision Proton Spectrometer of the CMS-TOTEM collaboration (CT-PPS).

This thesis focuses on the development and test of Ultra-Fast Silicon Detectors and a dedicated readout integrated circuit – Time Of Flight Front-End Electronics (TOFFEE) – developed in standard 110 nm CMOS technology.

An introduction to the need of a precise time information in high energy physics is given in Chapter 1. This chapter focuses then on how a 4-dimensional tracking could improve the vertex reconstruction in presence of high pileup conditions. The timing requirements of the CMS timing layer and CT-PPS are pointed out.

The physical background and the key concepts to understand the working principle of silicon radiation detectors are presented in Chapter 2. In particular, the signal formation and the impact ionization processes are described, in order to introduce the operation of UFSD.

The main features of readout circuits for timing are reviewed in Chapter 3.

The development of two productions of Ultra Fast Silicon Detectors fabricated by Fondazione Bruno Kessler (FBK) are discussed in Chapter 4. The simulations that led to the fabrication of a first run of 275 µm thick sensors and of a timing-optimized 50 µm thick second run are described.

An integrated circuit specifically designed to read out the Ultra Fast Silicon Detector, TOFFEE, is presented in Chapter 5. The TOFFEE ASIC, based on a transimpedance amplifier and a discriminator, has been intended to produce a logic pulse in response to an UFSD signal with a time resolution of $\sim$ 50 ps. Its features have been tailored to cope with the requirements of the CMS-TOTEM Precision Proton Spectrometer (CT-PPS). The main design choices are reviewed and motivated with simulation results.

The experimental results of tests conducted on the TOFFEE ASIC are discussed in Chapter 6. Tests have been conducted on the ASIC and on the whole system composed by an UFSD read out by TOFFEE both in laboratory and with a pion beam at CERN.

Contents

List of Figures

List of Tables

Chapter 1

Timing in high energy physics

The precise measurement of time intervals is a major topic in nuclear and particle physics. The determination of very small time gaps is referred to as *timing* and concerns different time scales, from few picoseconds to microseconds. Timing techniques are used in a large variety of measurements, such as the extraction of the lifetime of excited nuclear states or elementary particles, Time Of Flight (TOF), determination of coincidences, etc. [1]. With the development of more and more powerful accelerating machines, the high energy physics environment has continuously placed challenging demands on the detectors during the years, raising the need to improve their precision. In particular, the required timing accuracy is always scaling down, reaching the current target of few tens of picoseconds.

This chapter introduces the concept of timing and its effect in the precise determination of interaction vertices at high event densities. Setting the focus on the high energy physics field, the motivations for the development of fast timing detector systems are reviewed.

1.1 Introduction to timing in physics experiments

A detector is characterized by specific purposes and in general can give information about different quantities of the particle that crosses it, such as the energy released by the particle, the time at which the particle is producing the signal, the hit position, the velocity of the particle, etc. Once that the sensor signal is acquired, the time response of the readout system should be tailored in order to get the most accurate measurement of the desired quantity. In the case of a timing measurement, this is the time of arrival of the crossing particle.

Precise time information are required in high energy physics for particle identification and tracking purposes.

1.1.1 Particle identification

Particle identification is a fundamental aspect of most high energy physics experiments. In a typical experiment, beams collide within the detectors or a single beam collides on a fixed target. As a result, many particles emerge from the interaction point producing specific signatures inside the detectors, from which the events should be reconstructed with the highest possible precision. Tracking detectors can measure the tracks of charged particles and – in presence of a magnetic field – the sign of their charge and their momentum. Neutral particles instead, are detected by calorimeters, which measure their energy and determine the nature of their interaction (electromagnetic or hadronic). Among charged particles, different hadrons have very similar interactions and therefore the most direct way to identify them is to extract their rest mass

$$m = \frac{p}{\gamma v} = \frac{p}{\gamma \beta c}, \tag{1.1}$$

where p is the particle momentum, $v = \beta c$ its velocity and $\gamma = \frac{1}{\sqrt{1-\beta^2}}$ the Lorentz factor. Since the momentum of an hadron can be measured by tracking detectors, the missing information to reconstruct the mass is the velocity. Four main techniques can be used to determine a particle's velocity:

- Time Of Flight;
- analysis of its energy loss;
- detection of its transition radiation (if the particle speed varies compared to the local speed of light);
- detection of Cherenkov radiation (for particles moving faster than the local speed of light).

Among them, Time of Flight is the technique in which timing accuracy plays a major role.

1.1.1.1 Time Of Flight in particle identification

The Time Of Flight (TOF) technique is the most direct way to determine a particle velocity and therefore it is widely used in high energy physics for particle identification. The idea of measuring the mass of a particle from the measure of the time needed to travel over a certain distance was actuated for the first time in 1911 by Hammer [2], who realized the first TOF mass spectrometer [3]. The basic concept of the Time Of Flight technique is to measure the time difference between two detector planes – typically scintillators or resistive plate chambers (RPCs), but also calorimeters in the case of neutrons –, one acting as "start"and the other as "stop"counter. Otherwise, the TOF can be defined between the particle production (synchronized with the beam-beam or beam-target collision signal given by the accelerator system) and a stop counter. The TOF can be used to distinguish between two particles with different mass and same momentum. In this case, the two particles have two different velocities and thus the difference in the time of flight allows to distinguish them. An example of a typical TOF setup is shown in Fig. 1.1 [4]: the two scintillators separated by a known distance L are hit at different times by a particle producing two signals which are amplified by a photomultiplier tube and then discriminated to produce logic signals. The logic signals are sent to a time-to-digital converter (TDC) which subtracts them and produces a time difference t. It is possible to extract the mass information from the measured time

$$m = \sqrt{\frac{c^2 t^2}{L^2} - 1} \tag{1.2}$$

and the mass resolution is

$$\left(\frac{dm}{m}\right)^2 = \left(\frac{dp}{p}\right)^2 + \gamma^4 \left(\frac{dt}{t} + \frac{dL}{L}\right)^2. \tag{1.3}$$

For a mass separation the time difference must be greater than the actual time resolution (typically $\Delta t \geq 4\sigma_t$). The difference in TOF at a distance $L = 1$ m for different kind of particles is shown in Fig. 1.2. Scintillation counters can easily provide a time

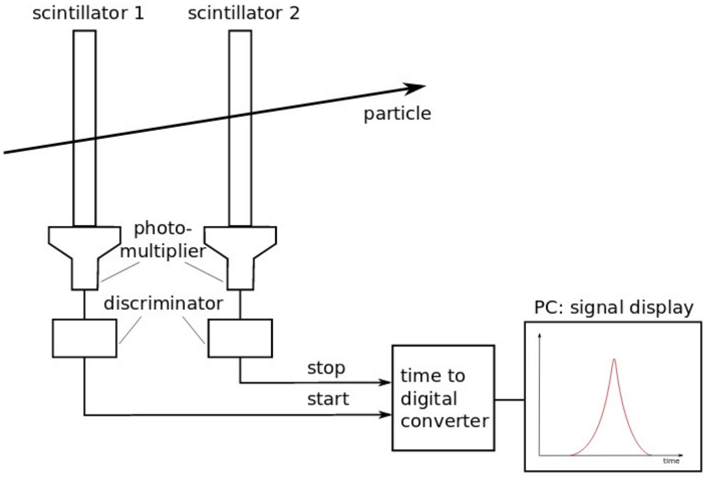


FIGURE 1.1: Typical system for TOF measurements based on scintillators and photomultiplier tubes.

resolution of $\sigma_t = 100$ ps for pion/kaon separation up to $p = 1$ GeV/c and a flight distance of 1 m [5] [6].

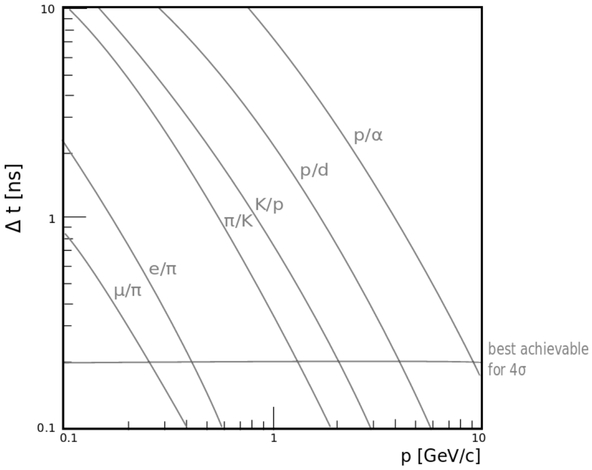


FIGURE 1.2: Mass separation capability related to Time Of Flight difference.

1.1.2 Tracking with time measurements

A time measurement can also be used to measure a spatial coordinate. This is the case in particular of the *z-by-timing* technique, used to reconstruct a spatial coordinate from a time of flight measurement. The z coordinate usually refers to the beam axis, namely positive in the "forward"direction and "negative"in the rear direction. The time difference between the signals provided by two detectors t_1 and t_2 is used to derive the position of the interaction z, as they are directly proportional, with a constant of proportionality which is nominally half the speed of light [7]:

$$z = \frac{c\Delta(t_1 - t_2)}{2}. \tag{1.4}$$

This is the case of the Precision Proton Spectrometer of the CMS-TOTEM collaboration (described in Sec. 1.5), where the z coordinate of the vertex is measured by means of the time of flight measurement of two forward protons.
This technique has already been exploited for example by the ZEUS Central Tracking Detector [8], composed by a multi-layer cylindrical drift chamber surrounding the beam: the wires of the odd-numbered layers are parallel to the chamber axis, while those of the even-numbered layers are inclined by 5° in both directions and allow an offline measurement of the z coordinate. However, in order to have a fast and online measurement of the z coordinate for trigger purposes, the z coordinate is reconstructed from a time difference.

1.1.3 Time resolution

The key concept of a timing system is its resolution, namely the smallest time interval that can be measured with accuracy. Ideally, repeated measurements of the same time t_0 form a gaussian (or pseudo-gaussian) distribution whose standard deviation σ represents the time resolution. Since the detector response is not always gaussian, it is common to extract the value of σ from the full width at half maximum (FWHM) of the measured time distribution, being $\mathrm{FWHM} = 2.35\sigma$ for normal distributions.
The performance of the timing system can be further affected by the characteristic times of the detector system [9]:

- *dead time*, i.e. the time needed by the system to record one event and be sensitive to another event;
- *recovery time*, after which the detector is able to provide a signal with full sensitivity;
- *sensitive time*, namely in a pulsed detector the time window after a trigger signal in which events can be recorded. In other words, this is the time window during which the detector has a non-zero sensitivity;
- *readout time* required by the readout system to read the event;
- *memory time*, i.e. the maximum time delay which is allowed between the crossing of a particle and the trigger signal.

The overall time resolution of a timing system composed by a sensor and its readout electronics is given by the root sum squared of the individual time resolutions of both sensor and electronics:

$$\sigma_t = \sqrt{\sigma_{t,Det}^2 + \sigma_{t,FE}^2}. \tag{1.5}$$

There are several ways to extract the time resolution of a system: among them the most common is to measure the time difference of two simultaneous signals in two separated channels and from that build the so called *coincidence curve*.

1.2 Tracking in 4 dimensions

The inclusion of timing information in the reconstruction of an event gives considerable advantages. According to the type of sensor used for this purpose, the timing information can be available at different levels of the event reconstruction: at tracking reconstruction if timing is associated to each point or at event reconstruction if timing is associated to a specific track. The first case describes a situation where a detector is providing both tracking and timing information, simplifying the reconstruction algorithm since only the hits with proper time tag are used in the pattern recognition phase. This case, however, requires a very complex readout electronics as it has to be able to measure both space and time with accuracy. The demands in terms of power consumption of the readout electronics can be very high. In Fig .1.3, a schematic view of the effect of timing information for each hit is shown: a group of seemingly random point provided by a traditional 3-dimensional tracking system is resolved into two distinct tracks and an additional random hit [10]. The second case foresees the implementation of an additional timing layer, inside or outside an existing tracking system, in order to be concurrently able to provide time information for each track with the smallest possible change to the existing hardware of the tracker [11]. In this way, the previously mentioned increase in power consumption can be minimized. The assignment of timing information to individual tracks and not to every hit has in fact a simpler implementation than the previous option. Fig. 1.4 shows two different events overlapping in space but separated in time by a small interval: the time information provided by a timing layer allows to resolve the two events that would be otherwise indistinguishable.

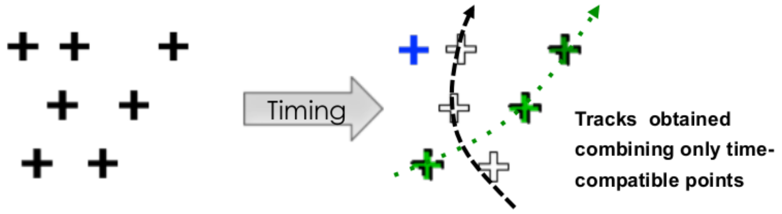


FIGURE 1.3: Effect of the addition of timing information at hit level.

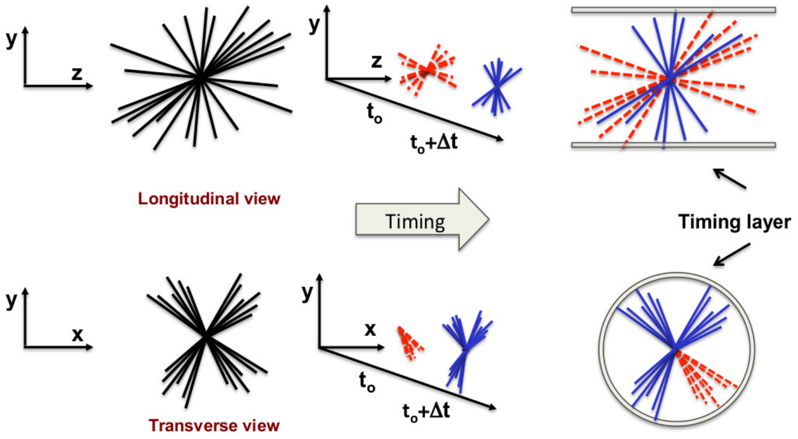


FIGURE 1.4: Event selection by means of the timing information provided by a timing layer.

1.2.1 Semiconductor detectors for 4D-tracking

Semiconductor detectors are commonly used as tracking devices due to their high spatial granularity and rate sustainability. In particular, the improvement of silicon technology has been crucial for particle physics and allowed to cover large detection areas and to handle high rates with a proper readout system. One of the very first approaches to timing with semiconductor detectors is described in [12], where thin surface barrier detectors have been proposed thanks to their fast collection time. In fact, the typical limitation of traditional silicon sensors is the signal formation process, which sets a boundary of more than 200 ps in the ability to measure the arrival time of a particle [10], which translates into more than 5 cm of spatial uncertainty if used to measure flight distances.

Currently, the NA62 experiment uses a track-timing detector (Gigatracker) [13], which employs 200 µm thick sensors with 300×300 µm^2 pixels and an expected time resolution of $\sim$ 150 ps. Recent measurements on a 100 µm thick pad sensor with an area of 2×2 mm^2 read out by a low noise amplifier achieved a time resolution of $\sim$ 105 ps [14]. Standard silicon devices can then provide good time resolutions, however it is difficult to reach less than 80-100 ps due to their small signal [15]. The option of developing a silicon sensor with enhanced signal in order to improve its timing capabilities has been explored with the Ultra-Fast Silicon Detectors (UFSD) and will be further described in this thesis.

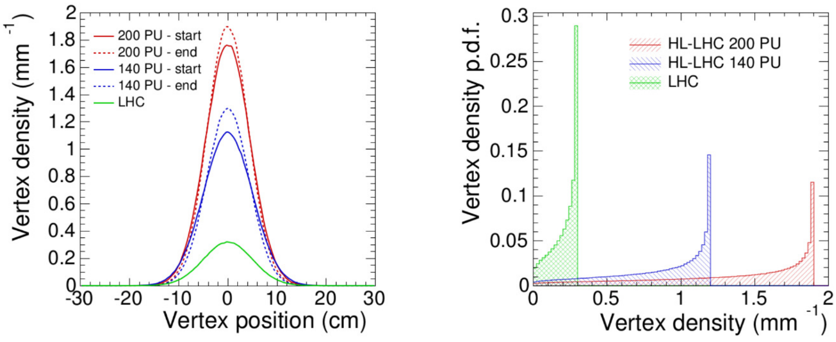
The goal is to maintain the high spacial granularity and high rate capability of silicon detectors and to provide concurrently a very precise time measurement. Since this latter is enhanced with a signal with short duration, even higher event rates can be sustained.

1.3 Effect of timing information at High-Luminosty LHC

In a typical collider experiment for high-energy physics, charged particles are detected by a tracking system that reconstructs their position along the trajectory. Traditional tracking detectors like silicon pixel detectors provide an accurate spatial reconstruction of the hits up to $\sim$ 50 overlapping events per bunch crossing, which is the current LHC scenario. In this case, a 3-dimensional tracking information is enough for a proper vertex reconstruction.

The future High Luminosity HL-LHC environment foreseen for 2025 [16] will place challenging conditions due to the high rate of concurrent collisions (pileup). The maximum luminosity that will be achieved by the LHC machine after the upgrade will be 2×10^{35} cm^{-2}s^{-1} at the beginning of each fill, but the expected instantaneous luminosity after beam focusing will lay between 5.2 and 7.2×10^{34}cm^{-2}s^{-1}. An average of 140-200 events per bunch crossing every 25 ns is expected, the root mean square (RMS) spread being approximatively 5 cm along the beam axis (Fig. 1.5.a), and a peak linear density between 1.3 and 1.9 mm^{-1} along the beam profile (Fig. 1.5.b) [17], according to CMS simulations. The transition from 50 to 200 interactions per bunch crossing will dramatically increase the fail probability of the particle-flow event reconstruction algorithms, raising the need to exploit the additional information provided by precision timing and energy deposits in the calorimeters. Indeed, about 15-20% of the amount of independent vertices will merge in absence of a proper reconstruction.

Time tagging is needed to resolve events overlapping in space but separated in time by a few tens of picoseconds by adding an extra dimension to the characterization of each vertex. If the tracking detectors provide information on the x and y coordinates, a timing detector will give a measurement of the z coordinate, i.e. along the beam axis. A visualization of the spatial distribution of the z-vertex expected at HL-LHC for a single bunch crossing is shown in Fig. 1.6. In particular, it is required



(a) (b)

FIGURE 1.5: Comparison of the spread of collision vertices along the beam direction (left) and of the vertex density distribution along the beam profile (right) at LHC (50 pileup events) and HL-LHC (140, 200 pileup events).

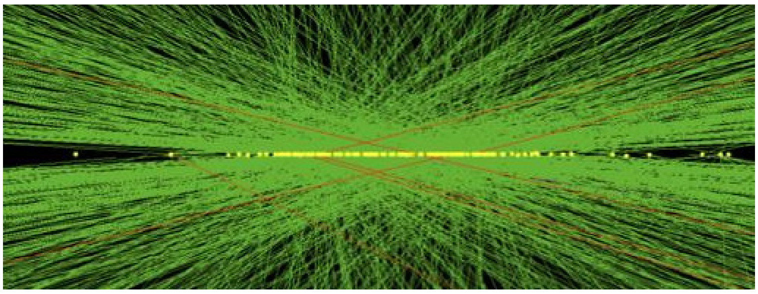


FIGURE 1.6: Z-vertex distribution at the high pileup conditions of HL-LHC: about 200 vertices are expected per bunch crossing.

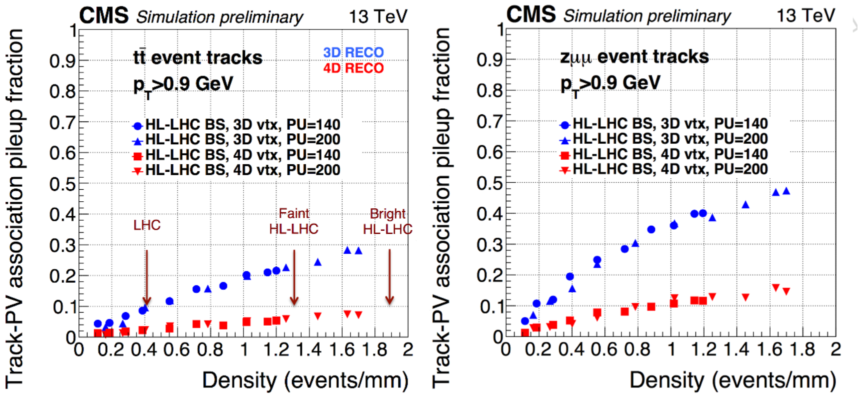
to associate tracks to the primary vertex of the hard interaction and to distinguish them from those related to pileup vertices. In fact, most interactions are not of interest for the understanding of fundamental interactions in the mass range of 0.1-1 TeV, while relevant ones occur with a probability of less than 1%. The mitigation of this pileup condition requires both offline algorithms to filter out inconsistent tracks and high granularity detectors to separate tracks. An overview of the effects of pileup at the CMS experiment in high luminosity conditions is given by the simulations results of Fig. 1.7, where the ratios between the number of tracks originating from pileup and that associated to the primary vertex of the hard interaction for $t\bar{t}$ and $Z \rightarrow \mu\mu$ events are plotted as a function of the density of events. As shown, the inefficiency due to the presence of pileup tracks would be effectively limited by precision timing.

If the distance between two vertices with coordinates z_1 and z_2 is shorter than the resolution σ_{z_0},

$$\frac{z_1 - z_2}{\sigma_{z_0}} < 1, \tag{1.6}$$

each track associated to the first vertex could be associated also to the second vertex, leading to an ambiguous vertex reconstruction.

The main impact of a precise timing information on the HL-LHC physics program will concern specific measurements like the characterization of the Higgs boson properties and the search for new phenomena such as Heavy Charged Stable particles.



(a) (b)

FIGURE 1.7: CMS simulations showing the fraction of tracks associated to the primary vertex of hard interaction originating from pileup over the total number of tracks associated to the hard primary vertex for $t\bar{t}$ events (a) and $Z \rightarrow \mu\mu$ events (b) as a function of pileup density, with 4-dimensional tracking (red) and traditional 3-dimensional tracking (blue).

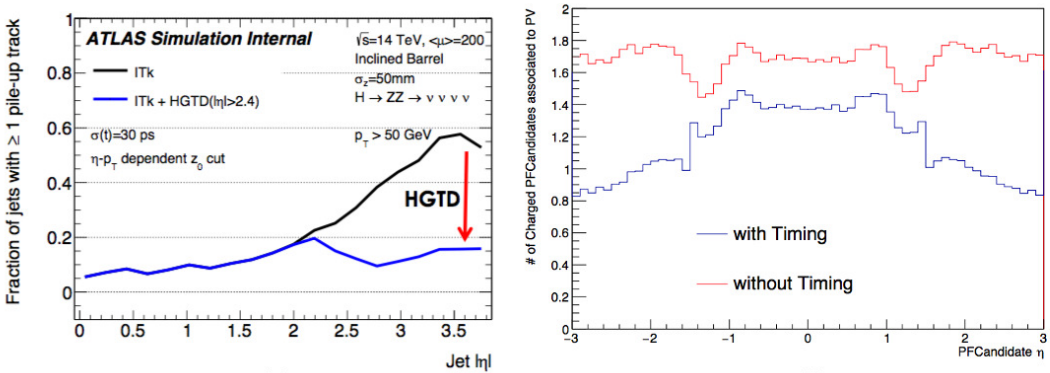
1.3.1 Timing solutions for CMS and ATLAS

Both CMS and ATLAS experiments are considering the introduction of timing layers to face the pileup issue, both enhancing track and vertex reconstruction through time measurements on minimum ionizing particles. Two different solutions have been proposed, since simulations performed by the two collaborations showed that each detector would benefit from additional information in a different pseudo-rapidity range. In particular, ATLAS is requiring a coverage in the forward region ($2.4 < \eta < 4$), while CMS in an almost complementary range (i.e. $0 < \eta < 3$). Simulation results on the wrong association of pileup tracks to vertices with and without a timing layer are shown in Fig. 1.8. ATLAS demonstrates that the jet reconstruction fails more significantly in the forward region, where the z resolution is degraded: with the addition of the High Granularity Timing Detector (HGTD) the detector will be able to obtain performances similar to that of the low η range also in the forward region. In Fig. 1.9, a simulation showing the increase of the z resolution as a function of the pseudo-rapidity justifies the choice of a timing layer in the forward region, as the resolution increases above the expected vertex density.

The CMS timing layer will be reviewed in detail in the following section.

1.4 CMS Timing upgrade

The upgrade of the CMS detector for the HL-LHC will require an improvement of the current detector for what concerns both the radiation tolerance and the capability to distinguish different events. The aim of the timing upgrade is to make the effect of HL-LHC pileup levels drop down to the LHC ones through a 30 ps reconstruction capability, thus recovering the track purity of vertices of the previous LHC scenario, i.e. up to 50 collisions per beam crossing. A precise time information should then be assigned to each particle, either charged or neutral, and to reconstructed vertices and jets. In absence of that, up to 15-20% of the vertices will merge together with the 140-200 pileup events conditions, with consequent merging of low-interest vertices, altered kinematics of the hard vertex and therefore degraded efficiency of the



(a) (b)

FIGURE 1.8: a) ATLAS simulation comparing the wrong association of pileup tracks to vertices as a function of pseudo-rapidity with and without the HGTD. b) CMS simulation comparing the average number of charged particle flow candidates associated to the primary vertex of the hard interaction as a function of the pseudo-rapidity with and without a precise timing information.

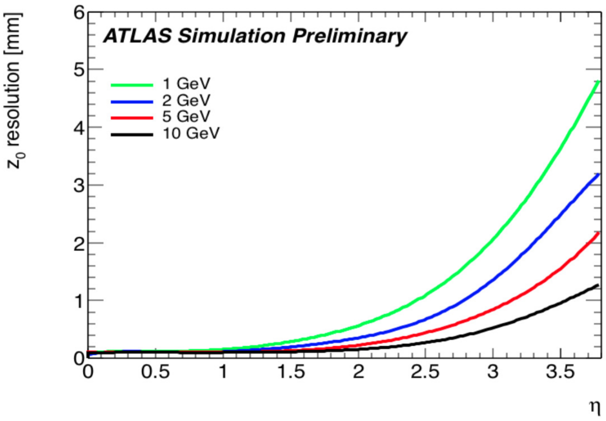


FIGURE 1.9: z_0 resolution of the Inner Tracker of ATLAS (ITk) as a function of η: for $|\eta| > 2.5$ the resolution increase above the average vertex density (1.6 vertex/mm).

algorithms used to identify the primary vertex, isolate particles, reconstruct jets and missing transverse energy. Another relevant benefit coming from a precise timing information in the vertex reconstruction is to provide additional discrimination to the High-Level Trigger, which could be useful to reduce the rejection rate of interesting events caused by a high pileup environment.

Both hardware and software improvements are thus needed: in particular, the hardware upgrade foresees increased time resolution of the calorimeters and an additional precision timing detector for minimum ionizing particles (MIPs) located between the existing tracker and calorimeters. Charged tracks will provide information to get a more precise track and vertex reconstruction and to measure the time-zero of the vertices, while calorimeters will give the timing information for neutral particles, and then the two informations will be matched. An upgrade of the current readout electronics will affect both the endcaps (CE) and the barrel of the electromagnetic calorimeter (ECAL), while the addition of a dedicated MIP Timing Detector (MTD) is still needed to reach the required timing accuracy and to extract the vertex time of minimum ionizing particles. The MIP Timing Detector has been proposed to cover both the endcap and barrel regions with detectors with high signal-to-noise ratio, exploiting different technology according to the specific requirements.

1.4.1 MIP Timing Detector

The MIP Timing Detector of CMS will involve a central barrel region based on LYSO:Ce scintillating crystals read out by silicon photomultipliers (SiPM) covering low radiation areas, and two forward regions equipped with silicon detectors with gain covering higher radiation areas. The MTD can be characterized in terms of acceptance, precision, radiation tolerance and impact on the present detectors. The timing layer is planned to cover hermetically the region up to $|\eta| = 3$ with a target resolution possibly better than 30-50 ps at an average collision rate of 200 per bunch crossing. In order to have an unambiguous time assignment avoiding double hits and a sustainable data rate, a channel occupancy of the order of percent is demanded. For what concerns radiation tolerance, the maximum expected fluence levels after 3000 fb^{-1} of integrated luminosity will be $1.55 \cdot 10^{14}$ n_{eq} cm^{-2} for the barrel and 10^{15} n_{eq} cm^{-2} for the endcaps. Finally, the additional material due to the introduction of the timing detector should have a limited impact on tracker and calorimeters and should not degrade their performance.

1.4.2 Endcaps

The Endcap Timing Layer (ETL) should cover a pseudo-rapidity range from $|\eta| = 1.6$ to $|\eta| = 2.9$ and it will be composed of a thermally isolated single layer of silicon detectors. This technology has been proposed in order to withstand the radiation levels of the forward region, since the radiation tolerance of the barrel devices cannot be granted at such fluences. The proposal of ETL foresees its installation in front of the upgraded High Granularity calorimeter (HGC), at a distance of about 3 m from the interaction point (Fig. 1.10), for a full detector area of 10.5 m^2. The separation from HGC allows to access the detector during maintenance shutdowns and to limit its influence on the calorimeter. The excess heat is removed by means of cooling pipes.

The pixel size has been chosen to be 1×3 mm^2 to limit the sensor capacitance to about 7 pF, thus allowing to achieve the target time resolution of 30-50 ps when operating the device at a gain close to 20. This cell size has also been chosen to constrain the unit cell occupancy below 5% and to limit the sensor leakage current (which is a surface and radiation related quantity, as described in Chapter 2). This choice implies a high number of sensors to cover the whole area, therefore requiring an affordable and robust technology. The most mature technology matching these requirements is that of silicon detectors with avalanche multiplication based on the Low Gain Avalanche Detector design [18][19]. The optimization of this sensor design for time purposes results in thin low-noise devices called Ultra-Fast Silicon Detectors [20], which have also been considered for the timing layer of ATLAS [21]. The effective implementation of such detectors is the subject of this work and will be further discussed, although parallel studies on alternative devices like Hyper-Fast Silicon detectors (based on Deep-Depleted Avalanche Photodiodes) are conducted [22].

The ETL detector is composed of modular units mounted on 11 aluminum ring-shaped supports, 3 in the inner part and 8 in the outer part. Each module contains 16 read-out chips bump-bonded to the sensors: a 96-channel ASIC is proposed for the inner region of the endcap where the pixel size is smaller, while 32 channels per chip are foreseen for the outer rings where the pixel size is larger. The total number of pixels per ETL sensor is 1536 in the inner rings and 512 in the outer rings.

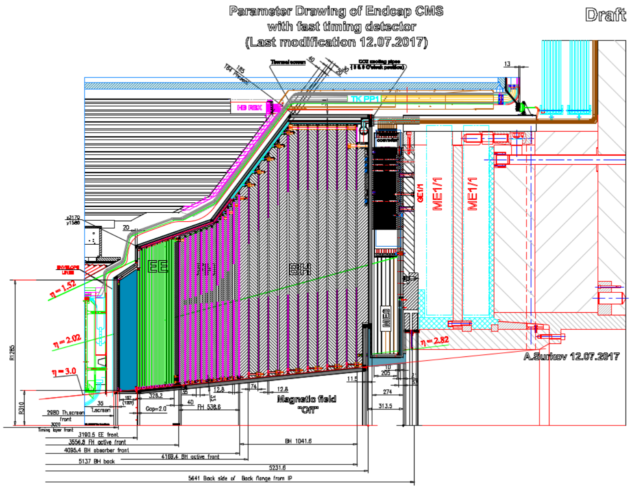


FIGURE 1.10: Placement of the Endcap Timing Layer in front of the High Granularity Calorimeter.

1.4.3 Barrel

The timing layer for the barrel region –Barrel Timing Layer (BTL)– is aimed to detect MIPs with almost full efficiency through a thin detector layer attached to the carbon fiber support of the CMS tracker covering a pseudo-rapidity between $|\eta| = 0$ and $|\eta| = 1.48$. The total area of the BTL cylinder is 40 m^2, having a length of 530 cm, an inner radius of 116.5 cm and an outer radius of 123.3 cm. The detector unit is made of a LYSO:Ce crystal with a section of 12×12 mm^2 read out by SiPMs with an area of about 4×4 mm^2. The actual crystal thickness varies with η from 3.75 to 2.4 mm in such a way that the material budget in front of the calorimeter is low and uniform and the amount of light produced in the crystals does not decrease with pseudo-rapidity. The SiPM is glued to one of the larger sides of the crystal tile with optical glue. The expected signal delivered by the detector is of about 5000 photoelectrons if a photon detection efficiency of 15% is assumed.

The detector is proven to withstand radiation up to a fluence of $2{\cdot}10^{14}$ n$_{\mathrm{eq}}$cm^{-2} [23] [24] [25], limited mainly by the radiation tolerance of SiPMs, as the transparency of the crystals is not significantly degraded above the expected fluence. SiPMs are subject to an increase of dark counts originating from charge carriers trapped into radiation-induced defects. A decrease of the cell size from 50 to 15 µm can drastically reduce the leakage current.

The readout electronics will be connected directly to the SiPM and is based on a custom chip adapted from the TOFPET2 ASIC developed for Time Of Flight Positron Emission Tomography (TOF-PET) applications [26] [27] [28] (Sec. 1.6. The aim of the readout electronics is to provide MIP timing with a precision of 20 ps with a reduced power consumption.

The BTL will share the cooling system with the tracker and the HGCAL and designed to operate at -30 ° C.

1.5 Timing in CMS-TOTEM Precision Proton Spectrometer

The CMS-TOTEM Precision Proton Spectrometer (CT-PPS) [29] is a two arm magnetic spectrometer placed in the forward regions on both sides of the CMS detector, at a distance of about 200 m from the CMS interaction point (IP5). It is aimed to study central exclusive production (CEP) in proton-proton collisions, i.e. the process $pp \rightarrow pXp$. In CEP the two protons do not dissociate, but are scattered at very small angles and interact via exchange of a photon or a gluon color singlet to produce the state X in the central region. The particular signature of the CEP events is to have two protons in the final state, scattered at very low angles with respect to the beam line [30]. An example of CEP event is shown in Fig. 1.11: a pair of leptons is produced in the central region as a result of a proton-proton collision. To distinguish these events, momentum, direction and time of arrival of the two final protons have to be measured with high precision. These two outgoing protons have lost only a small fraction of their initial momentum and are bent outside of the beam axis by LHC magnets placed between the CMS interaction point and the CT-PPS detector stations. The CT-PPS detection system will be composed by two detector stations with pixel detectors for tracking and one station with timing detectors. The detector stations are box-shaped Roman Pots (RPs) placed along the beam line and provided by the TOTEM collaboration. (Fig. 1.12) A cylindrical RP has been designed to house timing detectors and equipped with Faraday cages: this shape provides enough space to hold the detectors and simultaneously minimizes the beam coupling impedance.

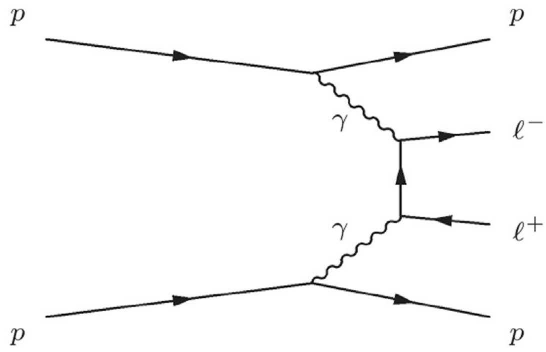


FIGURE 1.11: Feynman diagram for a central exclusive production of a pair of a di-lepton pair in a p-p collision.

The purpose of the CT-PPS is to detect forward protons with a spacial accuracy of $\sim$ 30 µm and to associate them to the correct vertex by using the *z-by-timing* technique (Eq. 1.4). On both arms, the timing detectors measure the difference between the time of arrival of the two protons of the CEP. A precise measurement of the time of arrival of the protons allows to properly associate them to the correct interaction vertex and to reduce the effect of pileup due to the high-luminosity environment. The baseline timing accuracy expected for the CT-PPS timing detector is $\sigma_t \sim 20$ ps, which leads to a precision in the z coordinate reconstruction of $\sim$ 4.2 mm and allows to retain most of the events.

Apart from the time resolution, the main requirements of the CT-PPS timing sensors concern the active area and the radiation tolerance. The overall detector area should be less than 4 cm^2, the dead region should be reduced and a slim edge is required at the beam side. Moreover, the segmentation should match the hit density

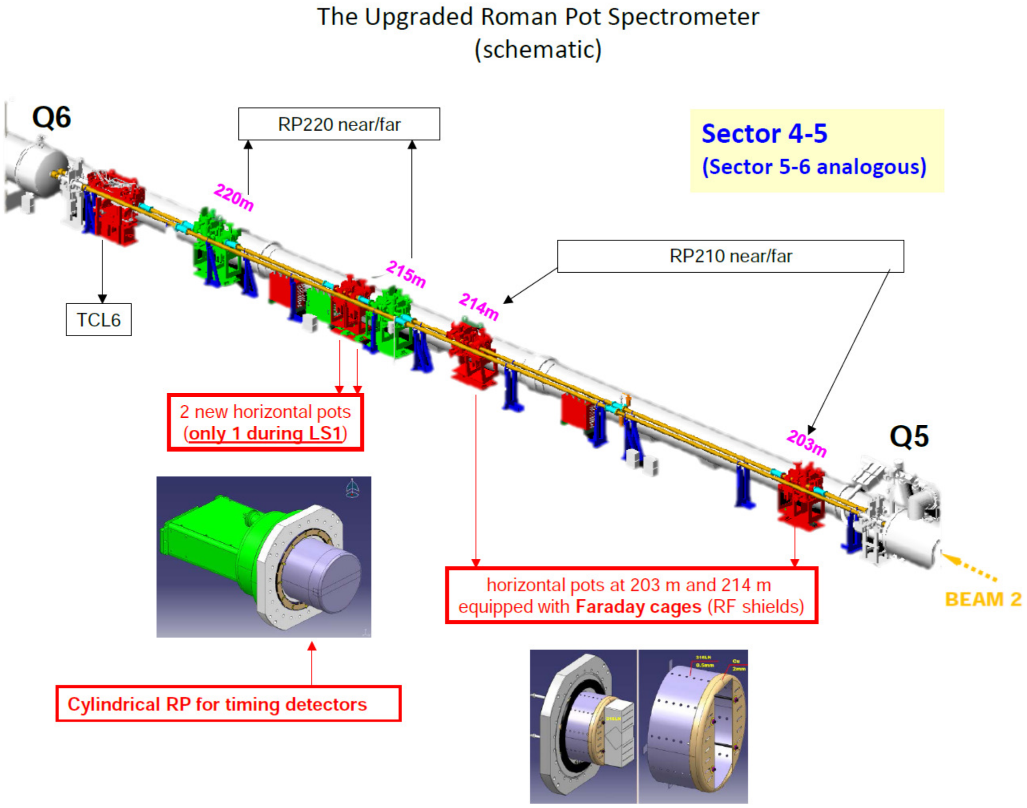


FIGURE 1.12: Overview of the beam line in the 200 m region.

distribution (fine segmentation close to the beam, coarser in the farther region). The expected dose at the RP stations has been evaluated by radiation monitors yielding ~ 100 Gray during one year of operation and a fluence of $\sim 10^{12}$ $n_{eq} \cdot$ cm^{-2} per 100 fb^{-1} of integrated luminosity.

1.5.1 Detector options

At present, the baseline for the timing system is based on diamond detectors, mounted in 2016 in the cylindrical roman pots. Other proposed solutions foresee the employing of the Ultra-Fast Silicon Detectors. For this purpose, planes of thin segmented UFSD are considered an option. A strip geometry has been specifically designed for CT-PPS and produced by Centro Nacional de Mircoelectrònica (Barcelona) and Fondazione Bruno Kessler (Trento). These detectors will be described in detail in Chapter 4.

1.5.2 Readout system

The existing CT-PPS readout system for timing detectors is composed of a preamplifier, a discriminator (NINO [31]) measuring the Time Over Threshold, a digitizer board with an High Precision Time to Digital Converter (HPTDC) and an interface to the DAQ. The expected time resolution of the whole readout chain – composed by sensor, preamplifier, discriminator and TDC – is of ~ 40 - 50 ps per detector plane. The NINO amplifier-discriminator coupled with the HPTDC is already used in the TOF detector of the ALICE experiment.

In order to properly read out silicon timing sensors, the sensor signal should be further amplified before it can be sent to the NINO input. The additional preamplifier stage should match the NINO inputs Another option is the realization of a custom integrated circuit for both amplification and discrimination stages, featuring LVDS outputs to match the HPTDC input requirements. The development of this amplifier-comparator chip specifically designed for UFSD and its performances are described in Chapter 5 and 6.

1.6 Fast timing in medical physics

A considerable contribution to the evolution of fast timing systems has been given by medical physics, in particular by Time Of Flight PET, which requires usually a time resolution below 100 ps. Modern PET scanners exploit the very high gain of SiPMs and their capability of resolving single photons. The TOF information on back-to-back photons originating from positron-electron annihilation is used to improve the background rejection of random events and to reconstruct images with enhanced signal-to-noise ratio. The SiPM signal is usually due to photons whose arrival time has a weak correlation to the time of the decay. Therefore, fast, low-noise and low-power readout systems are required. Low noise capabilities are mandatory when it is needed to resolve single photoelectrons, while the bandwidth of the circuit should be high enough so as to neglect the time walk across the dynamic range. Low power constraints are imposed by the dimension of the detectors, that are required to be as compact as possible.

The EndoTOFPET-US project [32], aimed to reach a time resolution of 200 ps in the measurement of coincident photons and a spatial resolution of $\sim$ 30 mm, led to the development of dedicated ASICs with challenging performances such as the STiC-ASIC [33] and the TOFPET-ASIC [26]. STiC is a mixed-mode, 16-channel ASIC designed in UMC 180 nm technology for TOF measurements in high energy physics and medical applications, while TOFPET is a 64-channel chip designed in standard CMOS 130 nm technology. Measurements performed on the ASICs showed a single photon time resolution of 180 ps for the STiC and 100 ps for the TOFPET.
The evolution of the TOFPET ASIC led to the TOFPET2 ASIC (fabricated in UMC 110 nm), which features a low power analog front-end and a digitization system, requiring a power consumption of less than 10 mW per channel. From tests performed on SiPMs coupled to a matrix of LYSO crystals read out by the ASIC, the system achieved a single photon time resolution of 95 ps r.m.s. biasing the SiPM at 7.5 V above the breakdown voltage [34].

Medical imaging applications like Time Of Flight PET have boosted the development of readout electronics for high resolution timing, especially for the read out of PMTs and SiPMs. The high energy physics applications benefit from this work, since the employed detectors and the required time resolution are comparable.

Chapter 2

Silicon detectors

Radiation detectors convert the energy deposited by a particle into an electrical signal which can be processed by a proper readout circuit. The electrical signal of a detector is produced by the motion of electron-hole pairs (in crystals) or electron-positive ion pairs (in gases and liquids) produced by the ionization of the incident particle. Ionization can occur in either solids (as in semiconductor detectors) or gases (as in ionization chambers). Among semiconductors, the most widely used is silicon. The foremost reason of silicon's popularity comes from its abundance, keeping the production cost of silicon-based devices relatively low. In nature, silicon exists mainly as compounds and very rarely in elemental form: monocrystalline detector grade silicon is derived from pure silica sand (SiO_2) with specific procedures during the fabrication processes of wafers.

This chapter describes the basic properties of silicon and the operating principle of a detector. The gain mechanism is also described, in order to properly understand the sensors that will be presented in the following chapters.

2.1 Introduction to silicon radiation detectors

In this section the fundamental concepts of semiconductor physics are discussed. An exhaustive description of this subject can be found in textbooks like [35], [36] and [37]. The main features of a p-n junction, generation and motion of charge carriers are also described.

2.1.1 Basic semiconductor physics

A crystalline solid has a band structure whose states are numerical solutions of Schrödinger's equation. Semiconductors have an energy region – the so called *band gap* – where no states are permitted. Energy bands above this region are called conduction bands while those below valence bands. At room temperature, the band gap value for silicon is 1.12 eV.

Intrinsic silicon is a Group IV element with four valence electrons, forming covalent bonds with four surrounding atoms. The density of free charge carriers (electrons in conduction band n and holes in valence band p) can be evaluated in similar way. The number n is given by

$$n = \int_{E_C}^{E_{top}} N(E)F(E)dE \tag{2.1}$$

where E_C and E_{top} are the energies at the bottom and at the top of the conduction band. $N(E)$ is the density of states and $F(E)$ is the Fermi-Dirac distribution which is given by

$$F(E) = \frac{1}{1 + \exp\frac{E-E_F}{kT}} \tag{2.2}$$

where E_F is the so called Fermi energy, i.e. the energy at which a possible state has a probability of being occupied equal to one half. k is Boltzmann's constant and T the absolute temperature. In the condition $E_C - E_F \gg kT$, the previous equation becomes $F(E) \approx e^{-(E-E_F)/kT}$.

The solution of Eq.2.1 can be expressed as

$$n = 2\left(\frac{2\pi m_n kT}{\sqrt{h^2}}\right)^{3/2} e^{-(E_C-E_F)/kT} = N_C e^{-(E_C-E_F)/kT}, \tag{2.3}$$

where m_n is the effective electron's mass, h Planck's constant and N_C the concentration of states in conduction band.

Similarly, the density of holes is

$$p = 2\left(\frac{2\pi m_p kT}{\sqrt{h^2}}\right)^{3/2} e^{-(E_F-E_V)/kT} = N_V e^{-(E_F-E_V)/kT}, \tag{2.4}$$

where m_p is the effective mass of holes, E_V the energy at the bottom of the valence band and N_V the related state concentration. In the intrinsic semiconductor case the concentration of electrons in conduction band and holes in valence band is equal: $n = p = n_i$, since each hole in the valence band is produced by an electron jumping to the conduction band. n_i is called the intrinsic carrier density. In this case, the Fermi level is typically placed at the middle of the band gap. Intrinsic silicon is very pure and contains a negligible amount of impurities. Its conductivity is rather low.

In device fabrication it is common to add impurities (dopants) in the crystal lattice of the semiconductor in order to control its conductivity. The doped semiconductor is also called extrinsic. Doping can be either *n-type* or *p-type* and introduces shallow levels inside the band gap. In the n-type case, a silicon atom is replaced by a Group V pentavalent or *donor* atom (arsenic, phosphorus or antimony), which has one more valence electron, creating an additional level very close to conduction band. This electron can be easily released to the conduction band. In the p-type case instead, the silicon atom is replaced with a Group III trivalent or *acceptor* atom (boron) which creates a hole near the valence band and works as a trap for electrons from valence band, as shown in Fig. 2.1. In presence of impurities, the Fermi level is no longer at the middle of the band gap: donors shift E_F closer to conduction band and acceptors closer to valence band. The introduction of impurities and the consequent creation of energy levels increase the conductivity. At room temperature, basically, all donor and acceptor states are ionized: the electron concentration is equal to the donor concentration ($n \approx N_D$) and the hole concentration is equal to the acceptor concentration ($p \approx N_A$).

In n-type silicon, electrons are called majority charge carriers while holes are minority carriers. In p-type silicon, holes are majority and electrons minority carriers. At thermal equilibrium, the concentration of positive and negative charge carriers is constant in time and follows the mass action law:

$$np = n_i^2. \tag{2.5}$$

Charge transportation in semiconductors is mainly due to two mechanisms: *drift* and *diffusion*. The first is due to the application of an external electric field, while the

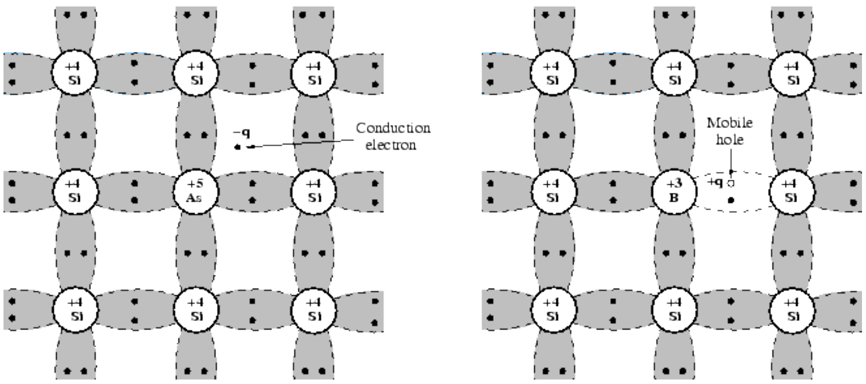


FIGURE 2.1: Bond representation of arsenic-doped n-type silicon and boron-doped p-type silicon.

second to a gradient in the carrier concentration distribution. Electrons in conduction band and holes in valence band can be considered as free particles, since they can move inside the lattice.

When no external field is applied, the free carriers move randomly and the average displacement is null. The introduction of an electric field, leads the carriers to move parallel to the field and acquire a drift velocity given by:

$$\begin{aligned} v_n &= -\mu_n E, \\ v_p &= \mu_p E, \end{aligned} \tag{2.6}$$

where v_n and v_p are the drift velocities of electrons and holes, μ_n and μ_p the respective mobilities defined as:

$$\mu_n = \frac{e \cdot \tau_n}{m_n} \text{ and } \mu_p = \frac{e \cdot \tau_p}{m_p}, \tag{2.7}$$

where $\tau_{n,p}$ is the free mean time between successive collisions and e the electron charge. The mobility depends on dopant and charge carrier concentration, temperature and electric field. For electric field values below 1 kV/cm, mobilities in silicon can be considered constant, with values of μ_n = 1350 cm^2 V^{-1} s^{-1} and μ_p = 450 cm^2 V^{-1} s^{-1} at room temperature. When the electric field increases, mobility becomes gradually more field-dependent and drift velocity tends to become constant, reaching a saturation value, as shown in Fig. 2.2. Electrons drift velocity saturates at a value of about 10^7 cm/s under an applied electric field of about 300 kV/cm. Holes saturation require much higher fields and reach a slightly lower velocity of about $9.5 \cdot 10^6$ cm/s.

Diffusion occurs when the distribution of free charge carriers is inhomogeneous. Carriers from the region with higher concentration have higher probability to flow toward the region with lower concentration than in the opposite direction. The equation describing this motion is

$$\begin{aligned} \mathbf{\Phi_n} &= -D_n \nabla n, \\ \mathbf{\Phi_p} &= -D_p \nabla p, \end{aligned} \tag{2.8}$$

where $\mathbf{\Phi_{n,p}}$ is the flux of electrons and holes, ∇n and ∇p the carrier concentration

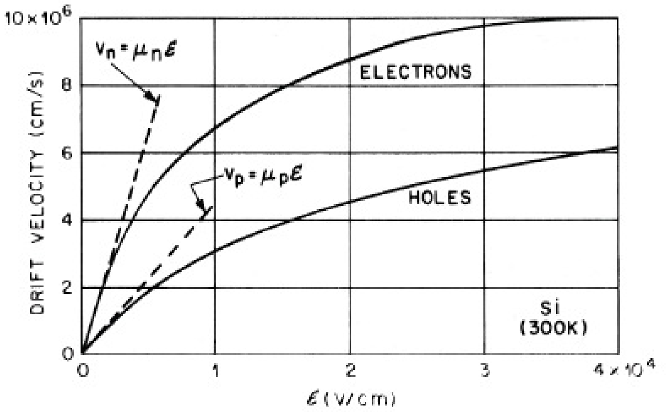


FIGURE 2.2: Dependence of the drift velocity on the electric field.

gradients. $D_{n,p}$ are the diffusion constant which are related to the mobility by Einstein's equation

$$D = \left(\frac{kT}{e}\right)\mu. \tag{2.9}$$

The total current density obtained by the combination of drift and diffusion effects is:

$$\begin{aligned} \mathbf{J_n} &= e\mu_n n\mathbf{E} + eD_n\nabla n, \\ \mathbf{J_p} &= e\mu_p p\mathbf{E} - eD_p\nabla p. \end{aligned} \tag{2.10}$$

The quantities

$$\sigma = e(\mu_n n + \mu_p p), \tag{2.11a}$$

$$\rho = \frac{1}{e(\mu_n n + \mu_p p)}, \tag{2.11b}$$

are defined as the conductivity and the resistivity of the semiconductor, respectively.

At thermal equilibrium, the carrier concentration of a semiconductor is kept constant by generation and recombination processes. In the case of a surplus of carriers, the recombination rate R becomes higher than the generation rate G, while in the case of a deficit of carriers, the system is brought back to equilibrium by generation rate overcoming recombination. The net recombination rate is defined as

$$U = G - R. \tag{2.12}$$

Recombination mechanisms can be classified according to the kind of particles involved.

- The *band-to-band recombination* is typical of radiative transitions occurring in direct bandgap semiconductors: an electron leaves its state in the conduction band and occupies the empty state in valence band associated with a hole.

- The *Shockley-Read-Hall recombination* takes place when an electron falls into a trap level and then into an empty state in the valence band. This process is the dominating one in silicon.

- The *Auger recombination* involves a third particle in the recombination process. An electron and a hole recombine in a band-to-band transition, but the resulting energy is transferred to another electron or hole. This kind of recombination is the inverse of the impact ionization mechanism described in Sec. 2.1.4.

Each of these processes have a reverse generation mechanism. Generation processes that are not associated to a recombination process involve generation due to light absorption and ionization produced by high energy beams.

2.1.2 The p-n junction

The most basic structure which can be studied to understand the properties of a silicon detector is a $p-n$ junction. It is formed by a p-doped and a n-doped regions of silicon joined together. For simplicity, it is assumed that doping is homogeneous in both sides and electrons and holes are homogeneously distributed when p and n sides are separated.

2.1.2.1 Equilibrium condition

If no external voltage is applied the system is in thermal equilibrium. When the two regions are joined holes diffuse from p to n side and electrons diffuse from n to p side. This creates an excess of positive charges in the n side and of negative charges in the p side. An electric field establishes in the region around the p-n interface: it opposes to diffusion and repels all the mobile charge carriers in that region. A *space-charge region* or *depletion region* is then formed and contains only fixed ions. In equilibrium, the potential which establishes in this zone is called built-in potential

$$V_{bi} = \frac{kT}{e} \log\left(\frac{N_A N_D}{n_i^2}\right) \tag{2.13}$$

and is equal to the difference between the Fermi levels of the n and p side $V_{bi} = (E_{Fn} - E_{Fp})/e$.

2.1.2.2 Forward bias

When an external voltage is applied, the system is no longer in equilibrium. The application of a positive potential to the p side and a negative to the n side is called *forward bias*: electrons and holes flow through the junction and the potential barrier across it decreases. The width of the depletion zone therefore decreases. With increasing forward bias, the depletion zone can become thin enough that the built-in potential cannot counteract the charge carrier motion through the junction, consequently reducing the electrical resistance. Electrons and holes flowing across the junction diffuse in the neutral region. The amount of current that may flow through the diode is indeed determined by the amount of minority carriers which diffuse in the neutral zones.

2.1.2.3 Reverse bias

By applying a negative potential to the p side and a positive one to the n side (*reverse bias*) as shown in Fig. 2.3, electrons and holes are removed from the junction and the

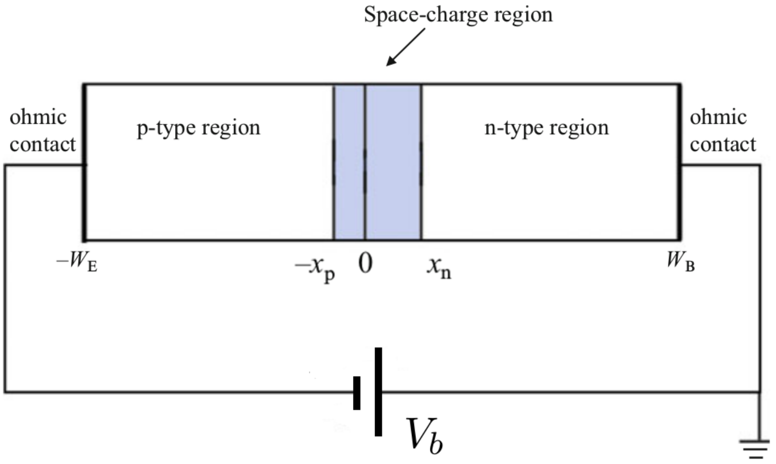


FIGURE 2.3: p-n junction at reverse bias.

width of the depletion zone increases. The current flowing through a p-n junction depends on the bias voltage V according to Shockley equation

$$I = I_0(e^{eV/kT} - 1). \tag{2.14}$$

In the case of forward bias, the current flow is dominated by the exponential term and increases rapidly with V. The width of the depletion zone in the reverse biased diode can be evaluated from Poisson's equation in the one-dimensional case

$$\frac{d^2V}{dx^2} + \frac{Ne}{\epsilon} = 0, \tag{2.15}$$

where N is the dopant concentration, ϵ the dielectric constant and e the electron charge. The width can be extracted in analog way at both n and p sides of the junction. V is the potential in the depletion region and varies perpendicularly with respect to the junction. The external reverse bias is V_b and the charge densities on the n and p sides are eN_D and eN_A, respectively. By making the assumption of an abrupt junction, calling $x_{n,p}$ the limits of the depletion zone at the n and p sides and V_j the potential at the p-n interface, the integration of 2.15 leads to

$$V = -\frac{eN_D}{\epsilon}\frac{x^2}{2} + \frac{eN_D x_n x}{\epsilon} + V_j. \tag{2.16}$$

At $x = x_n$, namely the boundary of the depletion region at the n side, the potential must be equal to the bias voltage

$$V(x_n) = V_b = \frac{eN_D^2 x_n^2}{2\epsilon} + V_j, \tag{2.17}$$

therefore the n region is contributing to the total value of the reverse potential with a quantity

$$V_b - V_j = \frac{eN_D^2 x_n^2}{2\epsilon}. \tag{2.18}$$

By inserting the p side contribution $V_j = \frac{eN_A^2 x_p^2}{2\epsilon}$ in 2.18 one has that

$$V_b = \frac{e}{2\epsilon}(eN_D x_n^2 + eN_A x_p^2). \tag{2.19}$$

The depletion width at the n and p sides are then derived from this last equation by using the charge neutrality condition $N_D x_n = N_A x_p$:

$$x_n = \sqrt{\frac{2\epsilon V_b}{eN_D(1 + N_D/N_A)}} \tag{2.20a}$$

$$x_p = \sqrt{\frac{2\epsilon V_b}{eN_A(1 + N_A/N_D)}}, \tag{2.20b}$$

and the total width is

$$w = x_n + x_p = \sqrt{\frac{2\epsilon V_b}{e}\frac{N_A + N_D}{N_A N_D}}. \tag{2.21}$$

In the case of an asymmetric junction (this is the case of a silicon detector, where the junction is between a highly doped electrode and lightly doped bulk), for example $N_A \ll N_D$ we have that

$$w \approx x_p = \sqrt{\frac{2\epsilon V_b}{eN_A}}, \tag{2.22}$$

namely the depletion region extends mostly into the less doped region. Taking into account the built-in voltage, which is always present at the junction interface, the previous equation becomes

$$w \approx x_p = \sqrt{\frac{2\epsilon(V_b + V_{bi})}{eN_A}}. \tag{2.23}$$

The reverse biased diode is the base of each silicon detector. The depletion region can be considered as a capacitor: the undepleted p and n regions are the electrodes and the depletion region is the dielectric. The reverse bias voltage controls the width of the depletion region and produces an electric field which makes the mobile charge carriers drift to the electrodes. The resulting detector capacitance is then

$$C = \epsilon\frac{A}{w} = A\sqrt{\frac{\epsilon e N}{2(V_b + V_{bi})}}, \tag{2.24}$$

which becomes independent of the bias voltage after full depletion and almost constant:

$$C = \epsilon\frac{A}{d} \tag{2.25}$$

for a sensor with bulk thickness d. The value per unit area is

$$\frac{C}{A} = \frac{\epsilon}{d} \approx 1\left[\frac{pF}{cm}\right]\cdot\frac{1}{d}, \tag{2.26}$$

being for a 100 µm thick silicon diode about 1 pF/mm^2. The dependence of the capacitance on the doping concentration N allows to extract important information on the detector doping from the capacitance-voltage characteristic:

$$\frac{1}{N} = \frac{\epsilon e A^2}{2}\cdot\frac{d(1/C^2)}{dV}. \tag{2.27}$$

2.1.3 Ionization of silicon

The ionization of a material is fundamental in particle detection since it produces free charges which will be detected by the device.
When a charged particle crosses a material, it interacts with electrons and atoms inside the medium. Atoms ionize and gradually the particle loses energy. The average energy loss of a charged particle in a certain material is estimated by the Bethe-Bloch formula:

$$-\frac{dE}{dx} = 4\pi N_A r_e^2 m_e c^2 z^2 \frac{Z}{A}\left[\frac{1}{2}\log\left(\frac{2m_e c^2\beta^2\gamma^2 T_max}{I^2}\right) - \beta^2 - \frac{\delta(\gamma)}{2}\right], \tag{2.28}$$

where E is the kinetic energy of the incident particle with charge z, T_{max} is the maximum kinetic energy with can be transmitted to a free electron in a collision, Z and A the atomic number and the atomic mass of the material, respectively. I is the mean excitation energy, N_A= 6.022 $\cdot$ 10^{23} mol^{-1} Avogadro's number, m_e the electron mass, c the speed of light, r_e the electron radius, $\beta = v/c$ is the particle velocity and $\gamma = \frac{1}{\sqrt{1-\beta^2}}$ the corresponding Lorentz factor, while δ is the high-energy corrective term for density. Fig. 2.4 shows the different stopping power exerted on different kind of particles by some materials. A particle whose energy loss is located at the minimum of this function is called Minimum Ionizing Particle (MIP). This region of the plot is fundamental since most relativistic particles have a mean energy loss close to the minimum. In silicon, minimum ionization occurs at $\beta\gamma \simeq 3$.

For thin silicon detectors (typically less than 300 µm), the probability distribution of the energy lost by a particle in a single hit follows an asymmetric spectrum called *Landau curve* [38]. Statistical fluctuations in the number of collisions are typically modeled by a Poisson distribution, while the variation in the energy transfer per scattering by a straggling function. The tail at higher energies is due to rare but very energetic events called *delta rays*: these are electrons produced by the interaction which can gain enough energy to become ionizing particles. As a result of this asymmetry, the mean energy loss is about 30% higher than the most probable. In thicker silicon layers, the mean energy released by a particle is lower and the resulting distribution is gaussian.

A MIP crossing a thin silicon device produces typically $\sim$ 75 electron-hole pairs per micron by ionization (most probable value), while the average value is 108. The energy required to produce an electron-hole pair in silicon is 3.6 eV, which is about three times the band gap value. The leftover energy causes oscillations of the lattice called *phonons*. The small energy loss of MIPs ensures that these particle cross the whole bulk of a silicon detector.

αparticles have a high and strongly velocity-dependent ionization, and therefore lose all their energy in a few microns and stop into the device. In this case, most of the energy is released at the end of the trajectory: this effect is called *Bragg peak*.

High energy physics requires the detection of charged particles, but silicon is also sensitive to electromagnetic radiation from visible (lower energy) to X-rays (higher energy). This feature makes this material useful for many other application such as imaging for either medical applications or photography. Photons can interact with silicon in three main ways, depending on their energy: photoelectric effect, Compton effect and pair production. Interaction with matter does not change their energy but attenuates their intensity: the absorption of a monochromatic beam follows Lambert-Beer's law

$$I(x) = I_0 e^{-\mu x} \tag{2.29}$$

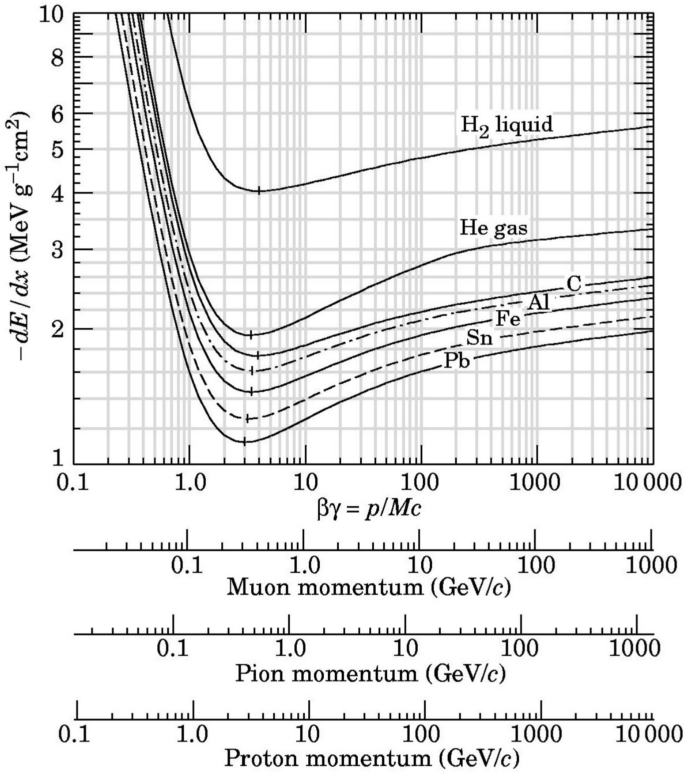


FIGURE 2.4: Energy loss for different particles and materials as a function of the particle energy.

where I_0 s the initial intensity of the photon beam, $I(x)$ is the intensity after crossing a distance x into a medium. μ is the specific absorption coefficient of that medium and is a function of the photon energy. The reciprocal of the absorption coefficient is called *absorption length* and is defined as the length at which the photon beam is attenuated by a factor $1/e$. Typically, visible and ultraviolet light produce a single electron-hole pair, while X-rays create many pairs confined in a small region.
Devices converting photons into a current (called *photocurrent*) are called *photodiodes* and are based on the *PIN* structure described in Sec. 2.3. The absorption of light produces electron-hole pairs by photoelectric effect if the light energy is greater than the band gap. The light energy is related to its wavelength according to the equation:

$$E(eV) = \frac{hc}{\lambda(nm)} = \frac{1240}{\lambda}. \tag{2.30}$$

The dependence of the absorption length on the wavelength in silicon is illustrated in Fig. 2.5, showing that silicon is sensitive to wavelengths shorter than 1000 nm. The sensitivity to light of a device is described by the *photosensitivity S*, i.e. the ratio between the photocurrent and the incident radiant power, and the *quantum efficiency QE*, which is the ratio between the number of electron-hole pairs generated and the number of incident photons. These two terms are linked each other by the relation

$$QE = \frac{S \cdot 1240}{\lambda} \cdot 100(\%) \tag{2.31}$$

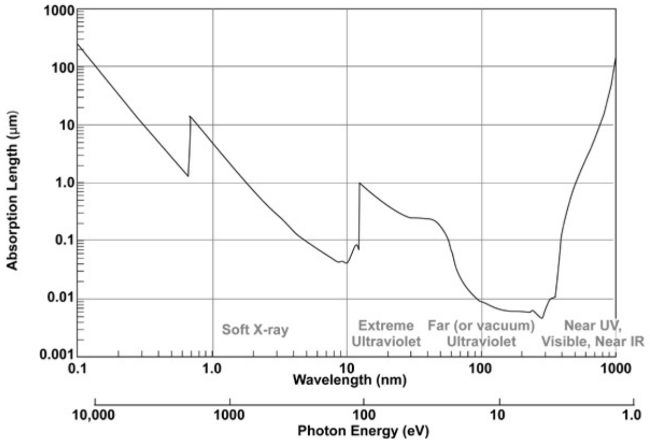


FIGURE 2.5: Wavelength dependence of the absorption length in silicon.

2.1.4 Impact ionization and avalanche multiplication

When an electron (or hole) moves into a region with high electric field inside a semiconductor, it can gain enough energy from the acceleration between collisions to produce an electron-hole pair. Part of the kinetic energy is transferred to the crystal lattice and part is used to create an additional electron-hole pair: if the energy of the carrier is higher than a threshold value E_{th}, when colliding with the semiconductor atom it will knock an electron out of the valence shell, producing the new electron-hole pair. The threshold energy for impact ionization is:

$$E_{th,e} = E_g\left(1 + \frac{m_n^*}{m_n^* + m_p^*}\right) \qquad \text{for electrons,} \tag{2.32a}$$

$$E_{th,h} = E_g\left(1 + \frac{m_p^*}{m_n^* + m_p^*}\right) \qquad \text{for holes} \tag{2.32b}$$

where $m_{n,p}^*$ is the effective mass of the electron (hole) and E_g the band gap of the semiconductor. An extended dissertation on the subject can be found in [39].

At equilibrium, the number of carriers generated must be equal to the number of carrier disappeared by recombination. In silicon, at high doping densities ($\sim 10^{10}$ cm^{-3} at 300 K) or optical injection of carriers, the decrease of charge carriers is dominated by the *Auger recombination*.

The probability of having impact ionization depends strongly on the electric field. In silicon, the characteristic field for impact ionization is 3×10^5 V/cm. Impact ionization is described by the *impact ionization coefficients* α_i (for electrons) and β_i (for holes). The impact ionization coefficients are defined as the number of electron-hole pairs generated by one electron (or one hole) along the direction of the electric field through a length of 1 cm inside the depletion layer [40] and are function of the electric field:

$$\alpha_i = \alpha_0 e^{-\frac{E_{n0}}{E}} \qquad \text{for electrons,} \tag{2.33a}$$

$$\beta_i = \beta_0 e^{-\frac{E_{p0}}{E}} \qquad \text{for holes,} \tag{2.33b}$$

where α_0 and β_0 are properly chosen parameters. Impact ionization can start an avalanche multiplication process. As previously mentioned, a high electric field is required to start the avalanche, since in low fields the kinetic energy gained between collisions by electrons and holes can only produce lattice vibrations. On average, an electron injected into a region with homogeneous electric field produces one electron-hole pair in travelling a distance $l_i = 1/\alpha_i$. Impact ionization coefficients are, in general, different for electrons and holes: α_i is sensibly higher than β_i. However, also holes can contribute to impact ionization, especially at very high field values where $\alpha_i \sim \beta_i$. In order to cause the avalanche, an extrinsic carrier needs to be injected into the avalanche region. The electron current density due to avalanche multiplication, in the case of an electron injected at the cathode of a region with length L and high electric field satisfies

$$\frac{dj_n}{dx} = (\alpha_i - \beta_i)j_n + \beta_i j \tag{2.34}$$

where j_n is the electron current and $j_n(0) = j$ is the boundary condition accounting for electron injection. A *multiplication factor M* can be defined for both electrons and holes in such a way that

$$M_n = \frac{j_n(L)}{j_n(0)} = e^{\alpha_i L} = \frac{1}{1 - \int_0^L \alpha_i \exp[-\int_0^x (\alpha_i - \beta_i)dx']dx} \quad \text{for electrons,} \tag{2.35a}$$

$$M_p = \frac{j_p(0)}{j_p(L)} = e^{\beta_i L} = \frac{1}{1 - \int_0^L \beta_i \exp[-\int_0^x (\alpha_i - \beta_i)dx']dx} \quad \text{for holes.} \tag{2.35b}$$

The condition at which this coefficient becomes infinite is called *avalanche breakdown*. In particular, the condition that leads to avalanche breakdown with the highest probability is $\alpha_i = \beta_i$, while with $\alpha_i \gg \beta_i$ or $\alpha_i \ll \beta_i$ the maximum multiplication can be reached.

Some devices, like avalanche photodiodes (Sec. 2.3.2) make use of the avalanche multiplication phenomenon to obtain an amplification of the signal: in this particular case the signal is produced by a single photon and would be too small to be seen by a readout circuit. Low Gain Avalanche Diodes (LGAD) are the subject of this thesis and will be described in detail in Chapter 4.

2.2 Operating principle of silicon detectors

This section will focus on how a silicon detector operates. The signal formation and charge collection mechanisms are explained. A detailed dissertation on this subject can be found for example in [36], [41].

2.2.1 Depletion voltage and electric field

As introduced in section 2.1.2.3, a silicon detector can be considered as a capacitor formed by a reverse biased p-n junction. Its sensitive area is given by the depletion region w which increases with the reverse bias voltage. At full depletion, the bulk is completely emptied of mobile carriers: the depletion width is equal to the detector thickness $w = d$. This condition is reached when the externally applied voltage is equal to

$$V_{dep} = \frac{eN_A w^2}{2\epsilon} - V_{bi}. \tag{2.36}$$

The capacitance decreases with the square root of the bias voltage until full depletion is reached, and then it is constant:

$$C(V) = \begin{cases} A\sqrt{\frac{\epsilon e N}{2(V_b+V_{bi})}}, & V_b \leq V_{dep}. \\ \frac{\epsilon A}{d}, & V_b > V_{dep}. \end{cases} \tag{2.37}$$

Ionizing particles crossing the detector produce electron-hole pairs inside the detector bulk. The drift motion of these electrons and holes produce the signal and depends on the electric field in the bulk (Eq. 2.6). For $V_b \leq V_{dep}$, the electric field is maximum at the junction and is zero at the opposite contact:

$$E(x) = \frac{2(V_b + V_{bi})}{w}\left(\frac{x}{w} - 1\right). \tag{2.38}$$

If the bias voltage is lower than the depletion voltage, the bulk is only partially depleted: electrons and holes produced in the undepleted region recombine immediately and cannot contribute to the signal. If the bias voltage exceeds the depletion voltage, the bulk is fully depleted and the overbias $V_b - V_{dep} + V_{bi}$ adds a uniform offset to the electric field given by $(V_b - V_{dep} + V_{bi})/d$. Both cases are shown in Fig. 2.6.

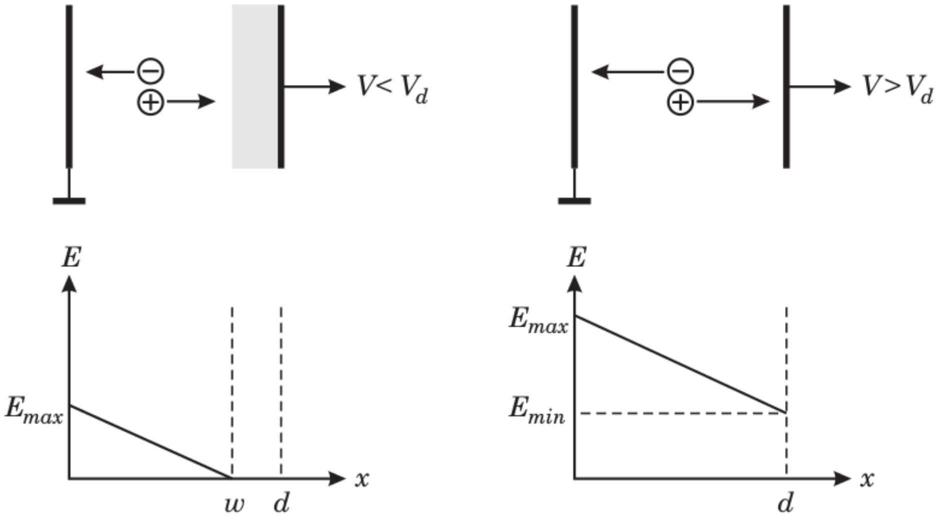


FIGURE 2.6: Electric field for a reverse-biased diode when it is partially depleted (left) and with overbias (right) [36].

2.2.2 Induced charge and signal formation

A charge moving in the sensitive volume of the sensor produces a signal current. This current is induced by the moving charge on the electrodes, because the charge induced on each electrode varies in time. The charge induced on an electrode can be calculated integrating the electric field on a Gaussian surface surrounding the electrode and has opposite sign with respect to the "real "charge. Since a sensor can be thought as two parallel electrodes, a charge q situated between the two electrodes induces a charge on each electrode that is a fraction of $-q$, in such a way that the sum of the charges induced on the two electrodes is $-q$. The portion of charge induced on each electrode depends on the position of the charge between the two parallel plates: the higher is the number of field lines enclosed by the integration surface, the higher is the charge induced, as shown in Fig. 2.7. The current induced on an electrode by a moving charge has been quantitatively described by the studies performed by Shockley [42] and Ramo [43] and summarized in the *Shockley-Ramo's theorem*:

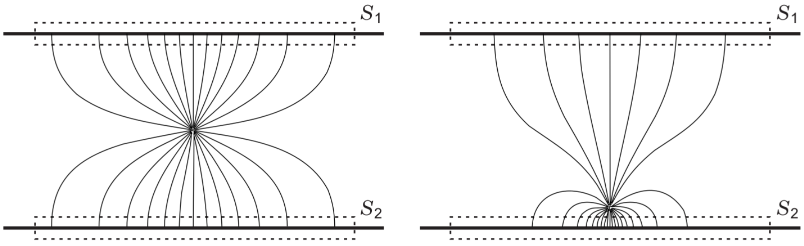


FIGURE 2.7: A charge q placed between two parallel electrodes induces a charge on the electrodes which changes with position: if the charge is exactly between the two electrodes (left), it induces $-q/2$ on each; if it is placed closer to the bottom one (right), the charge induced on this electrode is higher than the charge induced on the upper one, since the lower electrode has a higher number of field lines terminating on it [36].

$$i(t) = -q\vec{v}(t) \cdot \vec{E}_W, \tag{2.39}$$

where $\vec{v}$ is the charge velocity and $\vec{E}_W$ is the *weighting field*. The weighting field describes the coupling between the charge and the electrode, is only geometry dependent and has the dimension of $[\mathrm{m}]^{-1}$. It is determined by applying a unit potential to the considered electrode: in the basic detector case (i.e. with parallel plate geometry), where only two electrodes are present, the weighting field is established by applying a unit potential to the reading electrode and grounding the other. This can also be applied to detectors with a higher number of electrodes such as strip and pixel detectors, with the difference that more complex geometries have a more complex weighting field and different current signals.
In any case, the continuity equation has to be satisfied:

$$\sum_k i_k = 0, \tag{2.40}$$

that is the sum of all currents from the electrodes must be zero. With two electrodes, the current on each is equal with opposite sign ($i_1 = -i_2$). With more electrodes, the current from one electrode has to be balanced by the sum of the currents from the others, therefore signal currents can have different shapes.

2.2.3 Charge carriers motion and charge collection

In the simplest case of a detector with parallel plate geometry, if the bias voltage V_b is much higher than the depletion voltage, the electric field between the two electrodes can be considered uniform:

$$E = \frac{V_b}{d} \tag{2.41}$$

where d is the electrode distance. A charge carrier (produced for example by ionizing radiation or light) inside the detector moves with drift velocity

$$v = \mu E = \mu \frac{V_b}{d} \tag{2.42}$$

and couples to the collection electrode by the weighting field

$$E_W = \frac{1}{d}. \tag{2.43}$$

The current induced on this electrode is therefore

$$i = qvE_W = q\mu \frac{V_b}{d^2} \tag{2.44}$$

and the time needed by a charge produced at the opposite electrode to cross the whole detector bulk, i.e. the *collection time* is

$$t_c = \frac{d}{v} = \frac{d^2}{\mu V_b}, \tag{2.45}$$

while the induced charge $Q = it_c$ is equal to q.

By applying this reasoning to an electron-hole pair produced, for example, at a certain distance x from the electrode with positive bias, electrons have to drift over a distance x to be collected, while holes have to cross $(d - x)$ before reaching the negative electrode. Electrons and holes drift in opposite directions and with different velocities: the current induced on the two electrodes has therefore the same sign and the induced charges are respectively

$$Q_e = e\mu_e \frac{V_b}{d^2} \frac{xd}{\mu_e V_b} = e\frac{x}{d} \qquad \text{for electrons} \tag{2.46a}$$

$$Q_h = e\mu_h \frac{V_b}{d^2} \frac{(x-d)d}{\mu_h V_b} = e\left(1 - \frac{x}{d}\right) \qquad \text{for holes.} \tag{2.46b}$$

If the electron-hole pair is produced exactly in the middle of the sensor ($x = d/2$), electrons will be collected in a time which is about 3 times shorter than for holes due to the mobility values ($\mu_h \approx \mu_e/3$). When all electrons have been collected, they have induced the charge $e/2$. Holes still continue to induce charge until they are all collected and the total induced charge is e.

2.2.4 Leakage current

Besides the current induced by a charge moving inside the bulk, a certain current flows inside the reverse biased sensor even in the absence of radiation. This current is known as *leakage current* or *dark current*. The typical shape of a current-voltage characteristic of a sensor is shown inf Fig. 2.8: before reaching full depletion, the current increases with the square root of the bias voltage (volume generated current). And then, a surface contribution adds when the space charge region reaches a defined structure. After full depletion, the current increases very slowly until electrical breakdown. Most detectors operate in this plateau region and sensibly below the breakdown voltage. Nevertheless, the operation voltage of a sensor is purpose-specific, for example it could be useful to operate a detector slightly before breakdown if avalanche is needed, or even after breakdown like in avalanche photomultipliers.

The two breakdown mechanisms observed in a p-n junction are *Zener* and *avalanche breakdown*. The first is due to quantum mechanical tunneling of charge carriers through the junction and occurs when both sides of the junction are heavily doped. The second is due to charge carriers multiplication in the high electric field region at the junction and occurs in lightly doped junctions.

In principle, breakdown is limited to the regions where the electric field is higher (local breakdown) and the current increase is relatively controlled. At this point, if the voltage is further increased, hard breakdown occurs and this is an irreversible condition that destroys the device.

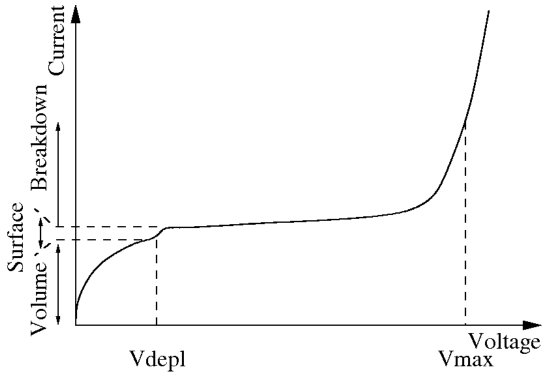


FIGURE 2.8: Typical sensor IV-curve [41]

2.3 PIN diode and avalanche diodes

This section describes the main features of a PIN diode, which is the structure at the base of silicon radiation detectors. Then, avalanche diodes are described, whose operating principle is fundamental to properly understand the sensors which are the subject of this thesis.

2.3.1 PIN diode electrical model

As described in Sec. 2.1.2, the simplest sensor unit is formed by two thin layers of highly doped semiconductor (p- and n-type) separated by a lightly doped, nearly intrinsic bulk (*i*-layer, often called π-layer if p-type) operated under reverse bias. According to Eq. 2.36, the low doping concentration of the bulk allows to reach full depletion at low reverse bias voltages, so that the overall detector capacitance (junction capacitance) is determined by the bulk thickness as described by Eq. 2.25.

A silicon sensor can be represented as a current source in parallel to a capacitance C_{det} determined by the junction capacitance. The current signal produced by the motion of charge carriers generated by an impinging particle is represented by the current source I_{sig}. Furthermore, the sensor is characterized by a leakage current I_{leak} which follows the diode law 2.14 and parasitic resistances modeled by a parallel shunt resistance R_{sh} and a series resistance R_s. The series resistance accounts for the contact resistance between silicon and the metal contacts, while the shunt resistance depends on manufacturing defects in the device and on its design.

The overall sensor model is sketched in Fig. 2.9. The signal adds to the normal dark current in the diode so that the total output current is given by

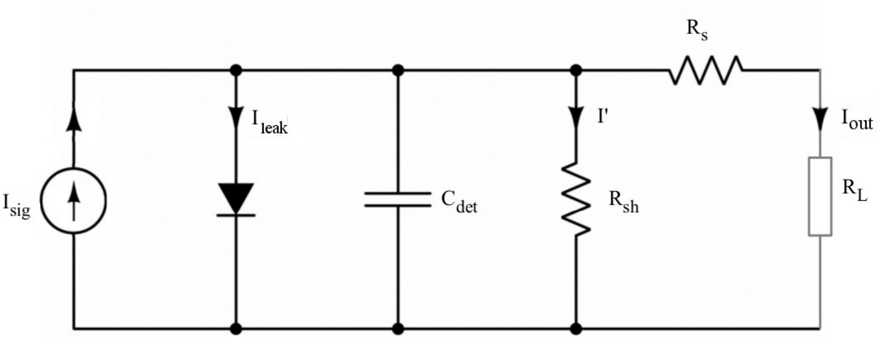


FIGURE 2.9: Silicon sensor equivalent circuit connected to a generic load resistor

$$\begin{aligned} I_{out} =& I_{sig} - I_{leak} - I' = \\ & I_{sig} - I_0\big(e^{eV/kT} - 1\big) - I', \end{aligned} \tag{2.47}$$

where V_D is the voltage across the diode and I' is the current flowing through the shunt resistance. The signal affects in the sensor IV-curve by shifting it downwards as shown in Fig. 2.10.

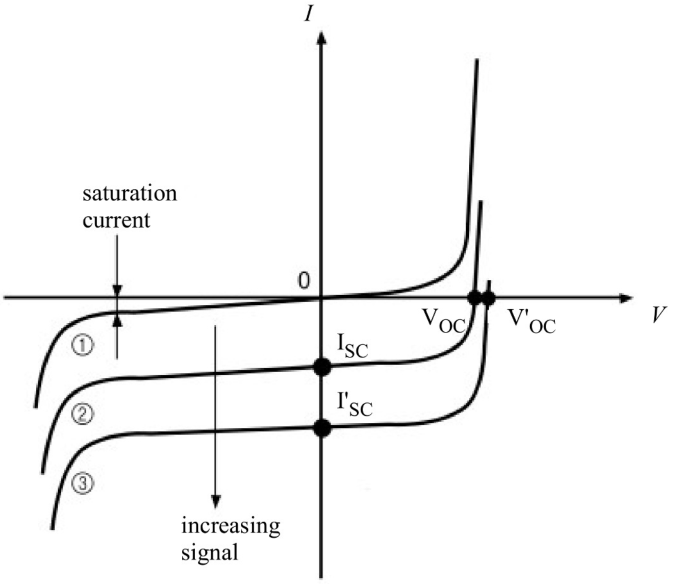


FIGURE 2.10: Sensor IV-curve is shifted downwards with increasing signal level. V_{OC} slightly varies when the signal changes.

Other relevant quantities in the analysis of the electrical model of a PIN diode are the open circuit voltage V_{OC} and the short circuit current I_{SC}. The open circuit voltage is defined as the output voltage at null output current (I_{out} = 0) [44]

$$V_{OC} = \frac{kT}{q} \ln \left(\frac{I_{sig} - I'}{I_0} + 1 \right), \tag{2.48}$$

changes logarithmically with the signal but is strongly temperature dependent.
The short circuit current is the output current in absence of load resistance and null output voltage (i.e. when the output is shorted)

$$I_{SC} = I_{sig} - I_0 \left[\exp \left(\frac{eI_{SC}R_s}{kT} \right) - 1 \right] - \frac{I_{SC}R_s}{R_{sh}} \tag{2.49}$$

The possibility to neglect the parasitic resistors without making too rough an estimate of the real behavior depends on the sensor geometry and the operating conditions, that is on the bias voltage and the leakage current at a given bias voltage. In a non-irradiated device, the leakage current at normal operation conditions (IV-curve plateau) is usually in the range from some tens of pA to nA. R_s is typically of several Ohms, while R_{sh} between 10^7 and 10^{11} Ohms. It can be seen that, in absence of radiation, the IV-curve has a linear behavior in a small voltage range. The shunt resistance can be seen as the slope of this line.

2.3.1.1 Noise characteristics

With all the previously described elements, the sensor is obviously producing a certain amount of noise. When connected to a readout circuit, the sensor noise appears at the circuit input. The noise analysis should take into account the following components, assuming that the current noise is represented by a current generator connected in parallel and the voltage noise by a voltage source in series:

- fluctuations in the leakage current, due to electron emission statistics, i.e. shot noise represented by a current source in parallel with the detector and evaluated as $< i_{nD}^2 >= 2eI_0B\big(e^{eV/kT} + 1\big)$;
- thermal noise coming from series resistance, due to velocity variations of the charge carriers $< i_{ns}^2 >= 4kTB/R_s$;
- thermal noise from shunt resistance $< i_{nsh}^2 >= 4kTB/R_{sh}$.

The frequency spectrum of thermal and shot noise is white. In addition, there is a frequency-dependent voltage noise component due to trapping and detrapping mechanisms inside resistors, dielectrics and semiconductors given by

$$< e_{nf}^2 >= \frac{A_f}{f}, \tag{2.50}$$

where A_f is a device specific coefficient located between 10^{10} – 10^{12} V^2 [36].

2.3.2 Avalanche diodes

A PIN diode featuring the gain mechanism described in Sec. 2.1.4 is called avalanche diode. Such a device is widely used to detect photons and therefore called "avalanche photodiode"(APD): its schematic cross section is depicted in Fig. 2.11.
The avalanche multiplication is accomplished by a p-doped layer generating a high

electric field which multiplies charge carries by impact ionization, with a gain ranging between 10^2 and 10^6. The APD presents an absorption region A and a multiplication region M. Photons are absorbed in the A region, and the produced electron-hole pairs start to drift toward the p^+ and n^+ electrodes. In the p multiplication region, carriers are multiplied. The effectiveness of the electron-hole pair production of a photodiode can be evaluated by the quantum efficiency described in Sec. 2.1.3. The leakage current of an APD is higher than the one of a PIN diode due to the fact that the bulk leakage current of the APD is multiplied by its gain (and thus its noise is higher). It is therefore useful to distinguish the surface and the bulk components of the device leakage current, $I_{leak,s}$ and $I_{leak,b}$, respectively: component

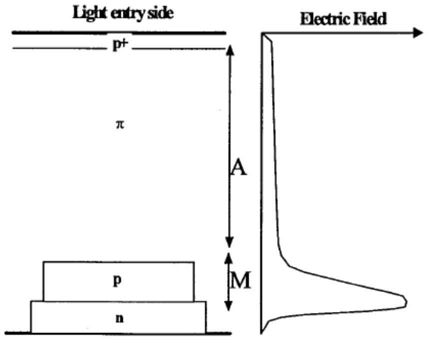


FIGURE 2.11: Avalanche photodiode cross section. The avalanche is produced by impact ionization in the high electric field region at the junction [45].

$$I_{leak} = I_{leak,s} + I_{leak,b} \cdot M, \tag{2.51}$$

where M is the gain.

The total output current of an APD is given by the sum of the multiplied photocurrent and the multiplied bulk leakage current.

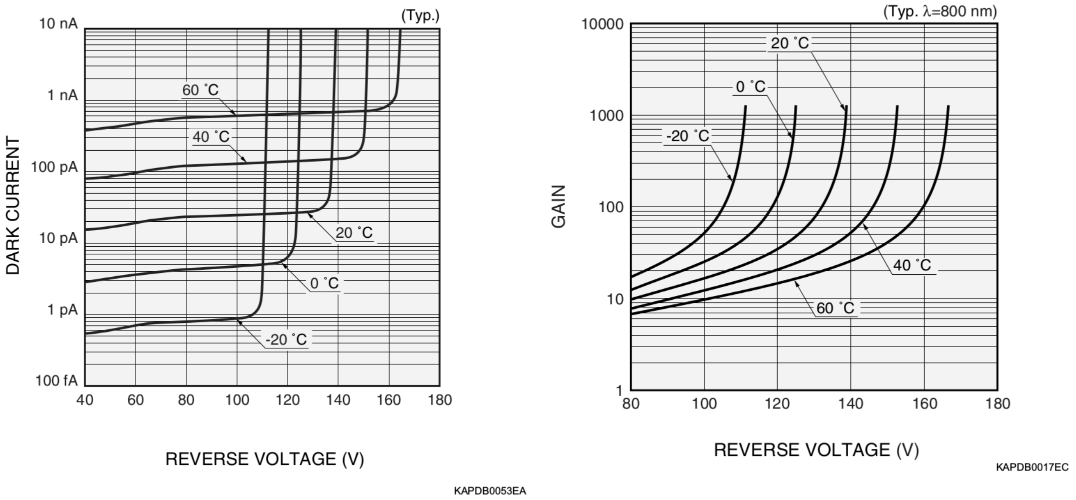


FIGURE 2.12: Temperature dependence of the current-voltage (left) and gain-voltage (right) characteristics [46].

2.3.2.1 Gain vs. bias voltage

The gain of an avalanche diode depends on the electric field in the multiplication layer and increases with increasing bias voltage (Fig. 2.12) up to a maximum value which depends on the photocurrent. Indeed, at a certain voltage value, the current flow through the series and load resistances increases, leading to a decrease in the voltage across the multiplication layer. The output current is then no longer proportional to the amount of light received. In any case, a device operated at a relatively high bias voltage could be subject to micro-discharges attributed to local breakdown due to inhomogeneous and irregular electric field. This situation should be prevented as it leads to fake signals.

Furthermore, the gain vs. voltage characteristics is also temperature dependent, namely at a given voltage value the gain decreases with increasing temperature. At higher temperatures, stronger lattice vibrations make the carriers collide with the lattice before they have reached enough energy to start the avalanche process.

2.3.2.2 Noise and excess noise factor

The avalanche mechanism of APDs generates excess noise due to statistical fluctuations, described by the *excess noise factor F*. Even if the gain is constant, the ionization process of the individual carrier is not uniform: the multiplication noise (excess noise) is then added during the multiplication process [46].

As previously mentioned, the shot noise of an avalanche diode is higher than the one of a PIN and is given by

$$< i_n^2 >= 2e(I_{sig} + I_{leak,b})BM^2F + 2eI_{leak,s}B, \tag{2.52}$$

where I_{sig} is the photocurrent at unitary gain and B is the system bandwidth.

The excess noise factor can be calculated using the model developed by McIntyre [47] as a function of the multiplication coefficient M and the ratio between the electron and hole ionization rates (defined in Eq. 2.33) denoted as k

$$F = k \cdot M + (1-k)\big(2 - \frac{1}{M}\big), \tag{2.53}$$

which can be effectively approximated for $k < 0.1$ and $M > 10$ with

$$F = 2 + k \cdot M. \tag{2.54}$$

The k factor is strongly electric field dependent and therefore is also a function of the doping profile.

Another empirical way to calculate F is by using the formula

$$F = M^x, \tag{2.55}$$

where x is derived from a log-normal fit of measured values of F at different gain values. Since x is dependent on the gain, the use of this approximation could lead to misleading results.

The same operating principle of APDs applies to Low Gain Avalanche Detectors and then to the Ultra-Fast Silicon Detectors, which will be described in Chapter 4.

In these two cases, the gain is kept moderate (around 15-20) in order to maintain the multiplication noise as low as possible.

2.4 Radiation damage

Radiation can affect semiconductors in two different ways: ionization damage and displacement damage. The first is due to charge carriers produced by ionization in the oxide layers (mainly SiO_2): since the oxide is an insulator, charges cannot be removed and form local concentrations. The oxide is used on the surface of most silicon devices as gate insulator in MOS transistors and as insulating layer on both wafer surfaces in detectors. The second is due to non ionizing energy loss by incident particles which displace silicon atoms from their location in the lattice, resulting in point defects and clusters.

Both detectors and electronics are then influenced by radiation damage: the following paragraphs will explain the two mechanisms and the relative consequences.

2.4.1 Ionization damage

Electrons and holes are produced by ionization in the oxide, too. Electrons move easily to the most positive electrode, while holes have a slow and complex motion causing an increase of fixed positive charges trapped in the oxide. In particular, charges can be trapped in the oxide-silicon interface and in the oxide.

In the first case, traps deriving from the lattice mismatch at the oxide-silicon interface or impurities can capture holes. The result on the device behavior is a possible increase of the surface recombination and then of the surface current.

The second configuration is due to the presence of charges in the oxide, either fixed, trapped or mobile. Fixed oxide charges are produced by positive charges near the silicon-oxide interface, attracting negative charges on the opposite side and affects the voltage handling capabilities. Trapped charges of both signs arise from ionizing radiation and can be annealed at low temperature, while mobile oxide charges are mainly due to ionic impurities like Na^+, Li^+ or H^+.

The presence of positive charges in the oxide which attract negative charges has then the following effects: reduction of insulation between n^+ regions, increase of the parasitic capacitance between confining regions and consequent increase of the noise, alteration of the electric field profile at the interface compromising voltage handling of the sensor.

In addition, ionizing radiation can increase the interface state concentration in the oxide, leading to an increase of surface recombination velocity and current and thus of the surface related leakage current.

2.4.2 Displacement damage

The displacement damage depends on the non ionizing energy loss and the entity of the damage is not proportional to the total absorbed energy. It varies with the particle type and energy, depending also on the momentum transfer.

This damage is produced by hadrons (like neutrons, protons, pions) or high energy leptons, which displace an atom of the semiconductor (*primary knock-on atom*, PKA) from its lattice position.

In silicon, the minimum energy required to dislocate an atom is $E_{min} \sim 15$ eV, while the displacement energy (i.e. the minimum energy to remove an atom from the lattice) is $E_d \sim 25$ eV [48].

If the energy imparted by the impinging particle is lower than E_d, it can produce only lattice oscillations and no damage. Conversely, if the energy is higher than E_d, the PKA is removed from its former site becoming an *interstitial* and leaving an empty space called *vacancy*. The vacancy-interstitial system is called "Frenkel pair ". Interstitials and vacancies can easily move through the silicon lattice: part of them recombine at room temperature and the resulting damage is null, but the remaining interact with each other and with the impurities already present in silicon and form permanent point defects.

Moreover, if the particle energy is much higher than the displacement energy (typically $\sim$ 5 keV), the recoil atom loses energy via ionization and displacement along its path. Ionization does not alter the silicon lattice, while non ionizing interaction prevail and produce defect clusters at the end of the recoil range, as shown in Fig. 2.13. Clusters and point defects are responsible of the damage of the silicon bulk, altering its original electrical properties.

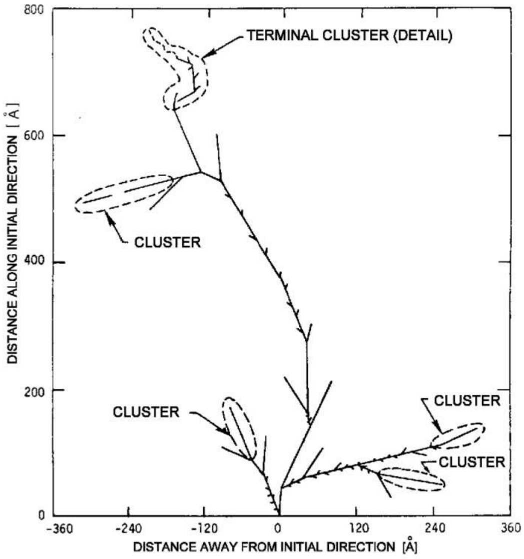


FIGURE 2.13: Monte Carlo simulation of a recoil atom track with an initial energy of 50 keV [49]

In a nonrelativistic approach, the maximum energy E_{Rmax} that can be transferred by a particle with mass m_p and kinetic energy E_p to the recoil atom is:

$$E_{Rmax} = 4E_p \frac{m_p m_{Si}}{(m_p + m_{Si})^2}. \tag{2.56}$$

With this equation, one can deduce the kinetic energy required to produce a Frenkel pair for different kind of particles, being $\approx$ 185 eV for neutrons, $\approx$ 186 eV for protons and $\approx$ 255 keV for electrons.

2.4.2.1 The NIEL scaling hypothesis

The displacement damage induced by different particles and different energies can be related with the *Non Ionizing Energy Loss (NIEL) scaling hypothesis* [50]. It is based on the assumption that any change in the material due to displacement damage scales linearly with the energy transferred in the displacing collisions: the concentration of the defects is independent of the properties of the incident particle and the

amount of the defects is proportional to the NIEL.
The *displacement damage cross section*

$$D(E) = \sum_{\nu} \sigma_{\nu} \int_{0}^{E_{Rmax}} f_{\nu}(E, E_R) P(E_R) dE_R \tag{2.57}$$

allows to calculate the NIEL for a particle with energy E: each interaction producing displacement damage release a PKA with energy E_R. The fraction of recoil energy transmuted into displacement damage can be analytically derived by using the *Lindhard partition function* $P(E_R)$. The number of possible interactions between the particle and the silicon atoms that are being displaced is referred as ν. For each interaction, σ_ν is the corresponding crosss section and $f_\nu(E, E_R)$ is the probability that the particle with energy E has to produce a PKA with recoil energy E_R.

The *hardness factor* κ

$$\kappa = \frac{\int D(E)\phi(E)dE}{D(E_n = 1MeV) \cdot \int \phi(E)dE} \tag{2.58}$$

allows to compare the damage produced by any kind of particle with spectrum $\phi(E)$ to the damage produced by 1 MeV neutrons with the same fluence. The resulting 1 MeV neutron equivalent fluence Φ_{eq} is defined as

$$\Phi_{eq} = \kappa\Phi = \kappa \int \phi(E)dE. \tag{2.59}$$

Fig. 2.14 shows the 1 MeV neutron equivalent displacement damage as a function of the energy for protons, neutrons, pions and electrons.

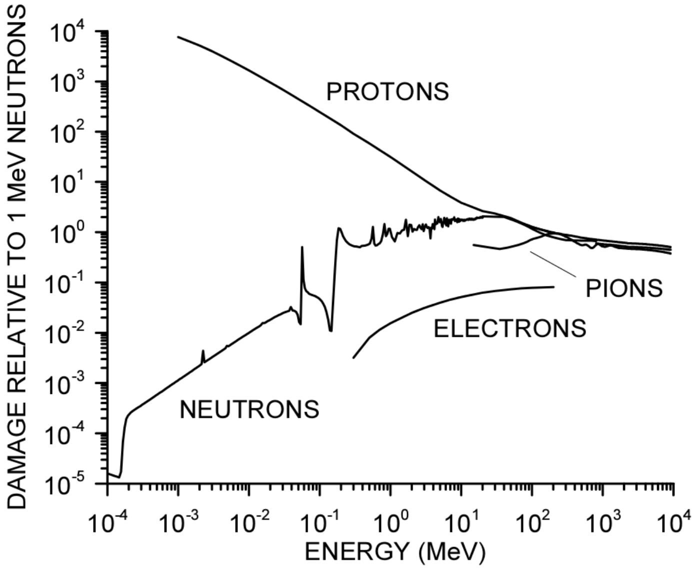


FIGURE 2.14: 1 MeV neutron equivalent displacement damage for protons ([51], [52], [53]), neutrons (gri96, [54], [53]), pions ([53]) and electrons ([52])

2.4.2.2 Effects of bulk damage

Defects in silicon lattice create energy levels in the band gap between valence and conduction band that emit and capture electrons and holes (like acceptors and donors). At thermal equilibrium, the charge state of defects is determined by the Fermi level: if the Fermi level is above the defect level, acceptors have negative charge and donors are neutral; conversely, if the Fermi level is below the defect level, acceptors are neutral and donors have positive charge. The ionization energy needed to emit an electron into the conduction band is equal to the distance between the bottom of conduction band E_C and the defect level $\Delta E_t = E_C - E_t$.
The main effects of bulk damage are shortly described in the following.

- **Increase of the leakage current:** the presence of energy levels in the band gap increase the generation recombination rate, and thus the leakage current of the device.
- **Change of effective bulk doping:** type inversion increase of depletion voltage
- **Reduction of charge collection efficiency:** energy levels in the forbidden gap can also act as traps for charge carriers produced in the bulk by ionization, affecting the charge collection efficiency. Charge carriers are locked into the trap levels and then released after a certain time τ_{eff} which is inversely proportional to the irradiation fluence Φ_{eq}:

  $$\frac{1}{\tau_{eff,(n,p)}} = \beta_{n,p}(t,T)\Phi_{eq}. \tag{2.60}$$

  This quantity is referred to as trapping probability and depends on time, temperature and irradiating particle through the *trapping constant* $\beta_{n,p}(t,T)$.

Chapter 3

Front-end electronics for timing

3.1 Readout systems for silicon radiation detectors

A radiation detector converts part of the energy deposited inside its sensitive volume by radiation into an electrical signal, which can be then fed to dedicated readout electronics. The readout of a silicon detector can be implemented in several ways. Readout circuits differ from each other according to the specific purpose to which they are dedicated and can be either discrete or integrated. Discrete circuits are still widely used, although integrated circuits (ICs) are often the best choice due their small area and lower power consumption. At present, the majority of front-end circuits for radiation sensors is fabricated in CMOS technology [55], which allows to stack in a few square millimeters several identical channels performing the same operation simultaneously, such as the readout of a strip or pixel detector. The scaling of the MOS transistor technology allows in principle to increase the speed performance of the individual transistors and the density of devices on a single IC, reducing the power consumption and the overall manufacturing cost. An integrated circuit specifically optimized for a particular purpose is called Application Specific Integrated Circuit (ASIC) and typically performs several operations such as amplification, filtering, analog-digital conversion and data transmission, while the actual readout and data acquisition is often performed by commercial Field Programmable Gate Arrays (FPGA), which can be programmed to execute algorithms of different complexity.

3.1.1 Sensor electrical model

The first step to study the system composed of a detector and its readout system is to approximate the sensor with a realistic equivalent model in order to treat it as an electronic circuit, as already discussed in Sec. 2.3. The sensor is as usual idealized with a parallel plate capacitor: one plate is connected to the high voltage supply and the other to the input of an inverting voltage amplifier. In Fig. 3.1 a negative high voltage is connected to the p electrode of the sensor while the n electrode is connected to the amplifier. This system is then collecting electrons, resulting in a negative charge moving towards the amplifier input and therefore the input node voltage moves downwards. The signal at the amplifier input has then a negative polarity, while the output of the inverting amplifier is a positive voltage signal. The opposite case, i.e. when the sensor is collecting holes, leads to an input signal with positive polarity and a negative amplifier output. In the first case, the sensor signal is faster due to the fact that electrons drift velocity is higher, and is therefore the preferable situation in many applications.

The small signal equivalent circuit of a sensor collecting electrons is shown in

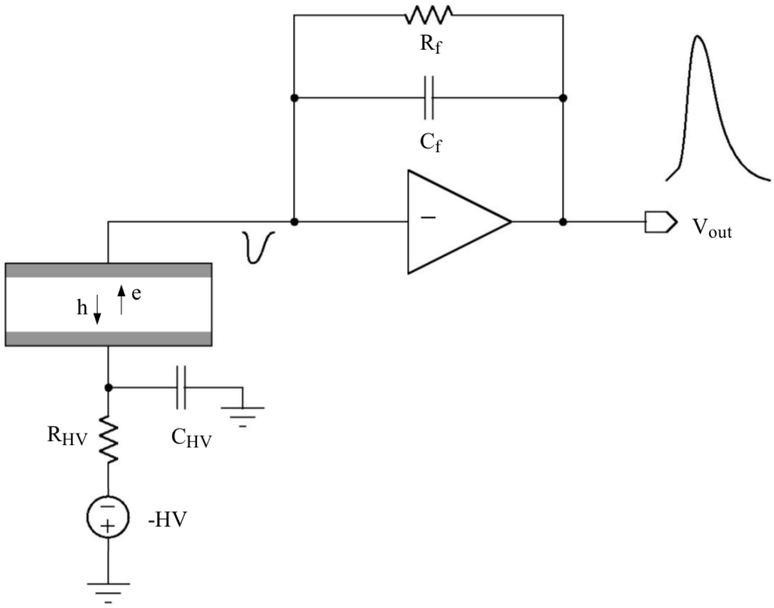


FIGURE 3.1: Schematic view of a sensor connected to an inverting voltage amplifier.

Fig. 3.2. The time-dependent current source I_{in} represents the sensor signal connected in parallel to a capacitor C_{det}, whose value comprises all the contributions to the sensor capacitance including neighbors and parasitics. Since the signal is produced by electrons, it should be represented by a flow of positive charges moving away from the amplifier input. The leakage current is modeled by a DC current source I_{leak}, which could be neglected if the signal is substantially higher. The sensor output impedance is represented by the resistor R_{det} but is typically negligible. The connections between the sensor and the readout circuit (usually wire or bump bonding or specific connectors) are characterized by a parasitic inductance L_c and a resistance R_c. These two latter components can be omitted in a preliminary analysis of the circuit behavior but their entity should be estimated in order to have a realistic overview of the whole system.

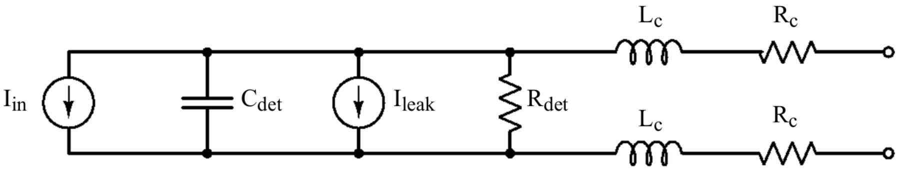


FIGURE 3.2: Equivalent circuit for sensor with negative polarity signal.

3.1.2 Front-end amplifier

The first element of a readout system is the amplifier, often implemented through the cascade of a *preamplifier*, connected directly to the sensor, and a *shaper*. The typical block scheme of a front-end amplifier is shown in fig. 3.3: the first block is the preamplifier, followed by a CR-RC shaper.

3.1.2.1 Charge Sensitive Amplifier

In most front-end circuits for radiation detection, the preamplifier is implemented as a "Charge Sensitive Amplifier"(CSA), which is relevant for the purpose of this thesis, although other topologies such as current sensitive amplifiers can be used.
In such a configuration, the CSA has the role to convert part of the sensor charge Q_{in} into a voltage step, integrating over the the feedback capacitance C_f. At low frequencies, the CSA has a voltage gain $-A_0$ and the output voltage is $V_{out} = -A_0 V_{in}$. Since A_0 is a finite value, only a fraction of the input charge is integrated by the feedback capacitance. The feedback resistance R_f, instead, is assumed to be noiseless and high enough to neglect its contribution to the signal processing, having the only purpose to set the proper DC path to bias the input transistor of the amplifier. Another useful simplification is to treat the sensor signal as a Dirac delta $I_{in}(t) = Q_{in}\delta(t)$, which reads in the Laplace domain $I_{in}(s) = Q_{in}$ (where $s = \jmath\omega$). At the input node, the Kirchoff current law in the Laplace domain leads to [55], [56]

$$I_{in}(s) + I_{leak} + I_{CSA} = 0, \tag{3.1}$$

which then becomes, neglecting the leakage current:

$$I_{in}(s) + sC_{det}V_{in} + sC_f[V_{in}(s) - V_{out}(s)] = 0 \tag{3.2}$$

and the resulting CSA transimpedance

$$T(s) = \frac{V_{out}(s)}{I_{in}(s)} = \frac{1}{s}\left[\frac{A_0}{(1+A_0)C_f + C_{det}}\right]. \tag{3.3}$$

The voltage step produced by the CSA has an amplitude equal to

$$\Delta V_{out,CSA} = -\frac{Q_{in}A_0}{(1+A_0)C_f + C_{det}}. \tag{3.4}$$

The total capacitance resulting at the input of the preamplifier is then

$$C_{in} = (1+A_0)C_f + C_{det} \tag{3.5}$$

which is the parallel of two components: the feedback capacitance multiplied by the gain due to Miller effect and the detector capacitance. The signal charge Q_{in} is then distributed between the sensor capacitance C_{det} and the input capacitance C_{in}:

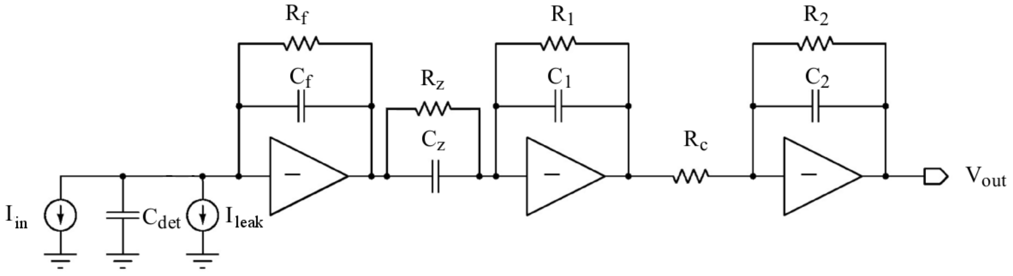


FIGURE 3.3: Block scheme of a front-end amplifier.

the amount of charge which is actually integrated by the feedback capacitance is $Q_f = [(1+A_0)C_f]V_{in}$, while the remaining is lost inside the sensor. It is thus required an input capacitance higher than the sensor capacitance. The quantity Q_f/Q_{in} defines the charge collection efficiency of the preamplifier, i.e. the fraction of the input charge that is really contributing to the signal.

Another important parameter of the charge sensitive amplifier is the input impedance:

$$Z_{in} = \frac{Z_f}{1+A_0} \approx \frac{Z_f}{A_0} \quad (|A_0| \gg 1), \tag{3.6}$$

where Z_f is a generic feedback impedance. In practice, the value of the feedback resistance is finite and not negligible. A finite value is chosen to provide proper discharge of the feedback capacitance in order not to saturate the amplifier with the accumulated charge. The feedback impedance of the preamplifier is then the parallel between the feedback capacitor and the feedback resistor:

$$Z_f = \frac{R_f}{1+sR_fC_f}, \tag{3.7}$$

and therefore the CSA transimpedance becomes

$$T(s) = \frac{A_0R_f}{(1+A_0)\left[1+s\left(R_fC_f+\frac{R_fC_{det}}{1+A_0}\right)\right]}. \tag{3.8}$$

Since $A_0 \gg 1$, the two terms A_0 and $1+A_0$ cancel out. In the realistic case of an amplifier with finite bandwidth determined by the unity gain frequency $f_{ug} = \omega_0A_0/2\pi$, the transimpedance has the form

$$T(s) = \frac{R_f}{1+s\left(\frac{1}{\omega_0A_0}+R_fC_f+\frac{C_{det}R_f}{A_0}\right)+s^2\left[\frac{1}{A_0\omega_0}(C_{det}R_f+R_fC_f)\right]}. \tag{3.9}$$

3.1.2.2 Shaper

The second stage of the front-end amplifier of Fig. 3.3 is a high-pass filter (differentiator). The voltage at the output of the second stage is

$$V_{outD}(s) = \frac{Q_{in}C_z}{C_f}\frac{R_1}{1+sR_1C_1} \tag{3.10}$$

in the Laplace domain. In the ideal case of a preamplifier producing a voltage step the CR filter derives the step and provides its return to the baseline.

In order to soften this sharp signal, an integration is performed by the low-pass filter of the third stage, which shapes further the output to

$$V_{outI}(s) = \frac{V_{outD}}{R_c}\frac{R_2}{1+sR_2C_2}. \tag{3.11}$$

The low-pass filter cuts the high-frequency components of the waveform smoothing the signal and optimizes the signal-to-noise ratio.

The resulting output of front-end is given by

$$V_{out}(s) = \frac{Q_{in} C_z}{C_f} \frac{R_1}{1 + sR_1 C_1} \frac{1}{R_c} \frac{R_2}{1 + sR_2 C_2}. \tag{3.12}$$

If the second and third stages have equal time constants $\tau = R_1 C_1 = R_2 C_2$ the output voltage in the Laplace domain is

$$V_{out}(s) = Q_{in} \frac{C_z}{C_f} R_1 \frac{R_2}{R_c} \frac{1}{1 + s\tau^2} \tag{3.13}$$

and in the time domain

$$V_{out}(t) = Q_{in} \frac{C_z}{C_f} R_1 \frac{R_2}{R_c} \frac{1}{\tau} \left(\frac{t}{\tau} \right) e^{-\frac{t}{\tau}}. \tag{3.14}$$

If the preamplifier has a finite feedback resistance, the voltage at its output is no longer a step but gives a contribution to the signal shape. The CSA transimpedance function has then a pole in $-\frac{1}{R_f C_f}$ instead than in the origin, and this is no longer canceled by zero of the differentiator. An additional resistor in parallel to C_z can be added to introduce a zero matching the pole of the CSA in such way that the $R_f C_f = R_z C_z$. The response of the full chain to a delta-like current in the time domain is

$$V_{out}(t) = \frac{Q_{in}}{C_1} \frac{R_f R_2}{R_z R_c} \left(\frac{t}{\tau} \right) e^{-\frac{t}{\tau}}. \tag{3.15}$$

The effective filter implementation should match the specific purpose of the detector, namely energy or time measurement, as described in the following. Trade-offs to optimize both are also a third option. In any case, any additional stage introduces poles in the transfer function, which should be adequately compensated in order to maintain the system stability.

3.1.3 Noise in a front-end amplifier

The noise affecting a front-end amplifier originates from both inside and outside the amplifier. The devices composing the amplifier are noise sources, as well as the sensor (leakage current and bias resistors), but the main contributors are the input transistor of the amplifier (here supposed to be a MOS) and the sensor leakage current. Noise sources can be represented by either current (connected in parallel) or voltage generators (connected in series). The amplifier is characterized mainly by a series noise with both white and frequency-dependent components. The first is due to thermal noise, i.e. the noise produced by statistical fluctuations of the charge carriers velocity, which has a white spectrum, while the second by the flicker noise, also called $1/f$ noise as it decreases with frequency.

Given that the input transistor is the main source of series noise in the amplifier, different components can be distinguished: the main is the noise originating from the resistive channel of a MOS, followed by the one coming from gate and bulk resistances. The spectral density of the thermal noise produced by the input transistor is

$$v_{nw}^2 = 4 k_B T \alpha_w \gamma_1 \frac{1}{g_{m1}}, \tag{3.16}$$

where α_w is the excess noise factor, γ_1 is the coefficient measuring the level of inversion of the MOS channel (approximated by $1/2$ in weak inversion and by $2/3$ in strong inversion) and g_{m1} the device transconductance.

The spectral density of flicker noise is instead:

$$v^2_{n1/f} = \frac{K_f}{f}, \tag{3.17}$$

where K_f is a constant specific of the device that depends on the gate capacitance, channel width and length. The sensor leakage current introduces a shot noise with spectral density

$$i^2_n = 2eI_{leak}, \tag{3.18}$$

which can be represented as a current source in parallel to the amplifier input. The feedback resistor produces white noise whose expression referred to the amplifier input is $i^2_{Rf} = 4k_BT/R_f$, but its value can be properly chosen to contribute less than the sensor leakage current, i.e. larger than the minimum value of $\frac{2k_BT}{eI_{leak}}$.

The amplifier with the input-referred noise sources is shown in Fig. 3.4. The relevant capacitances connected between the amplifier input and ground are the detector capacitance, the gate-source capacitance and the drain-source capacitance of the input transistor, gathered in the capacitance

$$C_T = C_{det} + C_{gs1} + C_{ds1}. \tag{3.19}$$

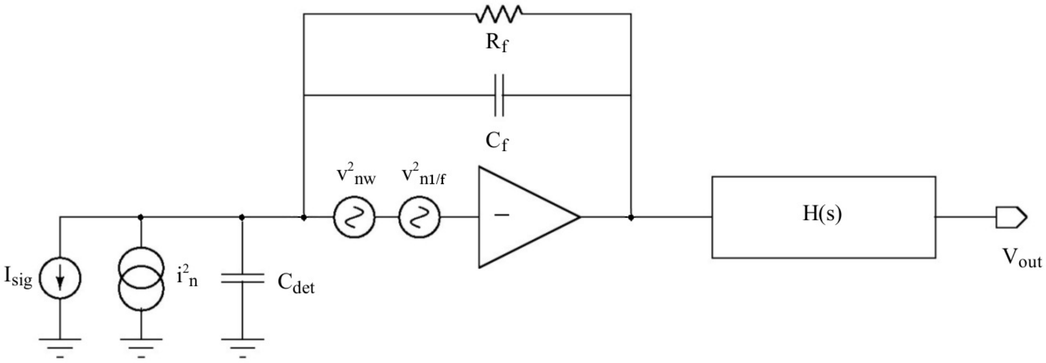


FIGURE 3.4: Preamplifier with series and parallel noise generators followed by a filter with transfer function $H(s)$.

The front-end noise is commonly expressed in terms of equivalent noise charge (ENC), i.e. the charge (in electrons) for which the noise and the signal are equal. Since an electron produces an output voltage with peak at e/C_f, the noise spectral density has then to be divided by $(e/C_f)^2$. The total ENC is given by summing in quadrature the three contributions (current, white and flicker noise):

$$ENC = \sqrt{ENC^2_i + ENC^2_w + ENC^2_f}. \tag{3.20}$$

3.2 Time pick-off circuits

The aim of a time pick-off circuit is to give an optimal determination of the time of occurrence. The output of an ideal circuit is independent on the shape and amplitude of the signal [57]. A time measurement performed by such a time pick-off circuit is affected by an error defining its capability to resolve two different events, i.e. the resolution of the system. This error arises from different sources: the most important are *jitter*, *time walk* and *drift*. A time pick-off circuit produces a logic output related in time to the beginning of the signal. The logic output is created by the discriminator (or comparator) which produces a logic "high "when the signal crosses a given threshold value V_{th} and a "low "otherwise. Since the instantaneous signal level is modulated by the noise, the time of threshold crossing is subject to gaussian fluctuations with a variance called *jitter* given by:

$$\sigma_t = \frac{\sigma_n}{\frac{dV}{dt}|_{V_{th}}} \approx \frac{t_r}{S/N}, \tag{3.21}$$

where t_r is the signal rise time, defined as the interval between the times at which the signal reaches the 10% and 90% of its total amplitude V_0.

At the output of the amplifier of Fig. 3.4, the voltage signal in the time domain is given by the Laplace transform of Eq. 3.9:

$$V(t) = V_0(e^{-t/\tau_f} - e^{-t/\tau_r}), \tag{3.22}$$

where τ_r is determined by the unity gain frequency of the amplifier f_{ug}:

$$\tau_r = \frac{1}{2\pi f_{ug}}, \tag{3.23}$$

while $\tau_f = C_f R_f + C_{det} R_f / A_0$. If $\tau_f \gg \tau_r$, for low values of t the output voltage becomes

$$V(t) = V_0(1 - e^{-t/\tau_r}). \tag{3.24}$$

The rise time can be expressed as

$$t_r = \tau_r \ln 9 = \frac{0.35}{f_{ug}}. \tag{3.25}$$

If a realistic detector signal with collection time t_c is sent into an amplifier with rise time t_r, the overall rise time is

$$t_{r,tot} = \sqrt{t_c^2 + t_r^2}. \tag{3.26}$$

The electronic noise at the amplifier output is $v_n \propto \sqrt{f_{ug}} \propto \sqrt{\frac{1}{t_r}}$ while the slope is

$$\frac{dV}{dt} \propto \frac{1}{t_r} \propto f_{ug} \tag{3.27}$$

and can be approximated by the ratio between the amplitude V_0 and the total rise time, $V_0/t_{r,tot}$. Moreover, it is proportional to the bandwidth f_{ug}, while the noise is proportional to the square root of the bandwidth. Therefore, an increase in the bandwidth of the amplifier leads to a more pronounced increase in the signal slope rather than in the noise. The lower limit to the overall rise time is in any case set by the detector rise time: if the amplifier is faster than the input signal, it just increases the noise without improving the overall performance of the system.
At the amplifier output, the electronic noise is $v_{no}^2 = \int e_{ni}^2 |A(\jmath\omega)|^2 df = e_{ni}^2 \Delta f_n$, where the noise bandwidth Δf_n is inversely proportional to the amplifier rise time

$$\Delta f_n \propto \frac{1}{t_r}. \tag{3.28}$$

It has to be noted that the noise at the input is amplified by the amplifier gain $A(\jmath\omega)$. The jitter is given by

$$\sigma_t = \frac{V_{no}}{dV/dt} \approx \frac{V_{no}}{V_0/t_{r,tot}} \propto \frac{1}{V_0}\frac{1}{\sqrt{t_r}}\sqrt{t_c^2+t_r^2} = \frac{\sqrt{t_c}}{V_0}\sqrt{\frac{t_c}{t_r}+\frac{t_r}{t_c}}, \tag{3.29}$$

which takes on a minimum at $t_r = t_c$, i.e. when the amplifier rise time is equal to the detector rise time.

Moreover, amplitude variations in the input signal result in a displacement of the output pulse, since greater signals cross the threshold earlier than small ones. This is referred to as *time walk*. For a given rise time, the time walk can be mitigated by measuring the pulse height and correcting the time measurement with software algorithms (*time walk compensation*).

3.2.1 Leading edge timing

A time measurement can be performed with a leading edge discriminator: the input signal produces the output pulse when crossing a fixed threshold [58]. A discriminator or comparator is defined as a circuit whose binary output is based on the comparison of two analog inputs, as shown in Fig. 3.5. The output of the discriminator assumes the "high "value v_{OH} when the difference between the noninverting and inverting inputs $v_P - v_N$ is positive, and the "low "value v_{OL} when the difference is negative. A real discriminator is also described by its gain:

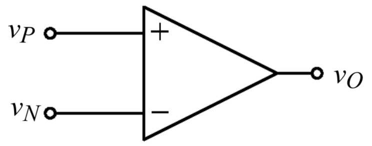


FIGURE 3.5: Circuit symbol for a discriminator.

$$A_v = \frac{V_{OH} - V_{OL}}{V_{IH} - V_{IL}}, \tag{3.30}$$

where v_{IH} and v_{IL} is the input voltage difference which is needed to saturate the output at the upper and lower limit, respectively and defines the minimum change at the input needed to perform the transition between the two possible output states (resolution of the discriminator). Moreover, the discriminator is characterized by a delay to produce the output after the input has been received, called *propagation delay time* ΔT (Fig. 3.6). This delay varies with the rise time and the amplitude of the input (the larger the input, the smaller the delay) and is a source of time walk error. Another source of time walk is the so called *charge sensitivity*: the discriminator needs an additional amount of charge to be triggered after the signal crosses the threshold. This situation can be summarized with Fig. 3.7: the signals A and B have same time of occurrence and same rise time but different amplitudes, resulting in a time walk of the output; signals B and C have same amplitude but different rise times, causing a displacement in the output pulse increasing with the length of the rise time. The

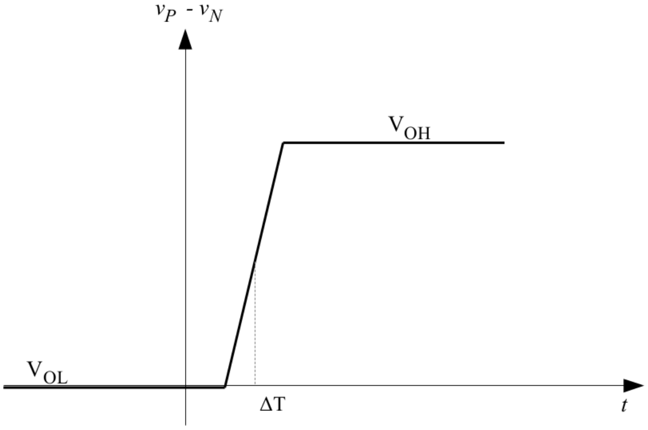


FIGURE 3.6: Propagation delay time of a leading edge discriminator, defined as the time required by the output to reach 50% of the signal since the threshold has been crossed.

shaded area represents the additional charge needed by the discriminator to switch. The three signals cross the threshold at the times t_1, t_2 and t_3 respectively, while the times at which the discriminator is actually switching are t_{10}, t_{20} and t_{30} and the delay from the time of threshold crossing is variable. The transfer function of the discriminator $A_v(s)$ can be modeled with a first order low pass behavior

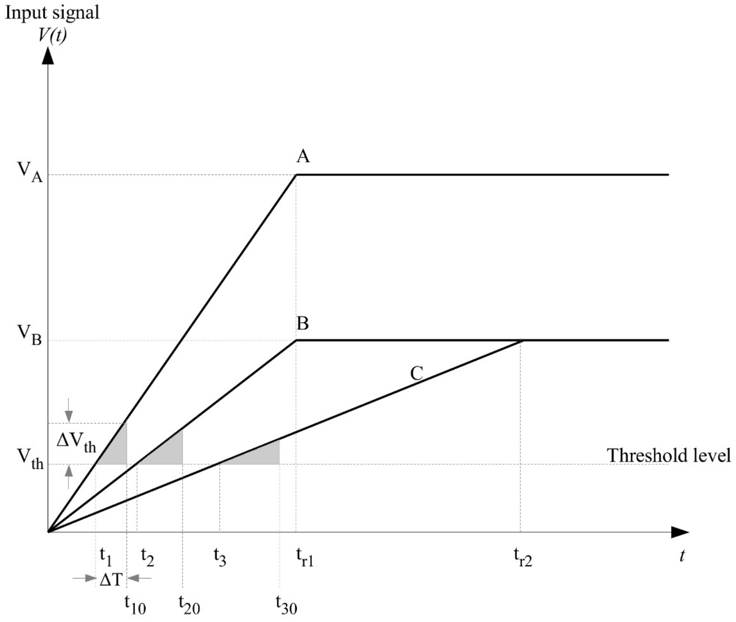


FIGURE 3.7: Variations in amplitude and rise time and charge sensitivity producing time walk in a leading edge discriminator.

$$A_v(s) = \frac{A_v(0)}{1 + s\tau_P}, \tag{3.31}$$

where $A_v(0)$ is the dc gain of the discriminator and τ_P is the time constant of the pole relative to the -3 dB frequency. The resolution of the discriminator, i.e. the minimum input voltage that the discriminator can amplify is

$$V_{in}(min) = \frac{V_{OH} - V_{OL}}{A_v(0)}. \tag{3.32}$$

By applying to the discriminator an input V_{in}, which can be expressed in terms of $V_{in}(min)$ as $V_{in}(s) = kV_{in}(min)/s$, it is possible to derive the propagation delay ΔT needed by the output to reach 50% of the maximum amplitude

$$\frac{V_{OH} - V_{OL}}{2} = \mathcal{L}^{-1}[A_v(s)V_{in}(s)]|_{\Delta T} \tag{3.33}$$

$$\frac{V_{in}(min)A_v(0)}{2} = kV_{in}(min)A_v(0)\big(1 - e^{-\frac{\Delta T}{\tau_P}}\big) \tag{3.34}$$

$$\Delta T = \tau_P \ln\left(\frac{2k}{2k-1}\right). \tag{3.35}$$

The effect of noise is to produce additional uncertainty on the threshold crossing time t_c. By assuming its distribution as gaussian with zero mean and standard deviation σ_v, the error in threshold crossing time can be estimated with:

$$\sigma_{t_c} = \frac{\sigma_v}{\frac{dV(t)}{dt}|_{t=t_c}}, \tag{3.36}$$

which becomes, in terms of a linear input signal with peaking time t_r, $V(t) = At/t_r$:

$$\sigma_{t_c} = \frac{\sigma_v t_r}{A}. \tag{3.37}$$

3.2.1.1 Time over threshold correction

The Time Over Threshold (ToT) technique is a common and affordable method to extract the signal amplitude information from the time spent by the signal above the threshold. In general, the signal amplitude is related to the time spent by this signal over a fixed threshold. A leading edge discriminator records the rising and falling edge of an input signal: the ToT is defined as the difference between the threshold crossing times of falling and rising edge. By considering a simple triangular signal with amplitude A, rise time t_r and fall time t_f (3.8), the threshold crossing times for the rising edge – often referred to as Time of Arrival (ToA) – is

$$T_1 = \frac{V_{th} t_r}{A} \tag{3.38}$$

while for the falling edge is

$$T_2 = t_r + t_f - \frac{V_{th} t_f}{A}. \tag{3.39}$$

The time over threshold is

$$ToT = T_2 - T_1 = (t_r + t_f)\left(1 - \frac{V_{th}}{A}\right) \tag{3.40}$$

and therefore the relationship between amplitude and ToT is:

$$\frac{1}{A} = \frac{1}{V_{th}}\left(1 - \frac{ToT}{t_r + t_f}\right). \tag{3.41}$$

If the time of arrival measured by the discriminator is $T_{1,disc}$, the “corrected ”time of arrival is

$$T_0 = T_{1,disc} - \frac{V_{th}t_r}{A} = T_{1,disc} + t_r\left(1 - \frac{ToT}{t_r + t_f}\right) \tag{3.42}$$

and the resulting time walk is

$$t_w = t_r\left(\frac{ToT}{t_r + t_f}\right). \tag{3.43}$$

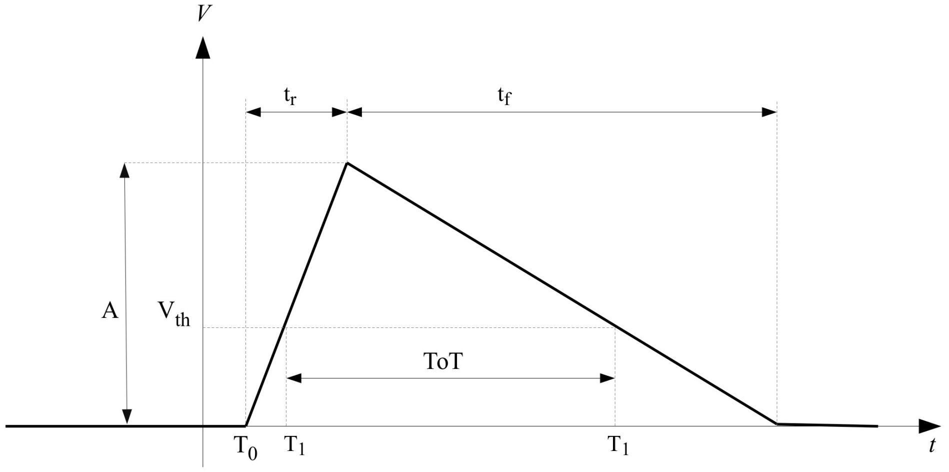


FIGURE 3.8: Time Over Threshold for a triangular signal with rise time t_f and fall time t_f.

It has to be noted that generally, the relationship between ToT and time walk is not linear due to the fact that the signal shape is not perfectly triangular and to non-linearity effects like charge sensitivity, requiring the calibration technique to be more complex.

Keeping into account the jitter on the rising and falling edges due to noise (being $\sigma_{T_1} = \sigma_v t_r/A$ and $\sigma_{T_2} = \sigma_v t_f/A$, respectively), the jitter on the ToT is then $\sigma_{ToT} = \sqrt{\sigma_{T_1}^2 + \sigma_{T_2}^2}$ and the resulting error on the corrected time of arrival is

$$\sigma_{T_0} = \sqrt{\sigma_{T_1}^2 + \left(\frac{t_r}{t_r + t_f}\right)^2 \sigma_{ToT}^2} = \frac{\sigma_v t_r}{A}\sqrt{1 + \frac{t_r^2 + t_f^2}{(t_r + t_f)^2}}. \tag{3.44}$$

The uncertainty on the corrected time is thus dependent on the jitter on the leading and trailing edge. Due to the typical asymmetric shape of the signals, the jitter on the trailing edge is generally higher than the one on the leading edge and is therefore the main contribution to the ToT jitter The advantage of the ToT technique is its simple implementation, but its reliability depends on the relationship between ToT and time walk: if this relationship changes (for example due to detector aging, or to signal shape variations due to changes in the rise and fall times), it requires to be periodically calibrated.

3.2.2 Zero-crossing timing

A technique to reduce time walk due to amplitude variations is the zero-crossing discrimination, which tracks the time of zero-crossing of a bipolar signal. If the signals have the same peaking time, they also have the same zero-crossing time. This method can be summarized as follows: a shaper produces an unipolar signal whose time constants can be considered equal for simplicity

$$V_u(T) = V_0 T e^{-T}, \tag{3.45}$$

where T is the time normalized to the time constant $T \equiv t/\tau$. This signal is then derived to produce a bipolar one

$$V_b(T) \equiv \frac{dV_u(T)}{dt} = V_0 e^{-T}(1-T), \tag{3.46}$$

which crosses zero when the signal reaches the peak amplitude, i.e. at $t = \tau$. Then, it is fed into both a zero-crossing discriminator and a threshold discriminator, as shown in Fig. 3.9. The threshold discriminator is set high enough that triggers on noise are suppressed and produces an output pulse which is then stretched. To sense the zero-crossing time, the threshold of the zero-crossing discriminator must be set to zero. This means that the discriminator is always triggering on the noise. and then the output of the zero-crossing circuit is given by the logic AND of the zero-crossing discriminator output and the delayed output. The time resolution is set by the zero-crossing discriminator, since the threshold discriminator switches before the zero-crossing one and just distinguishes valid signals from noise. However, this method does not allow to compensate for rise time variations, as it is supposed to be the same for the different signals.

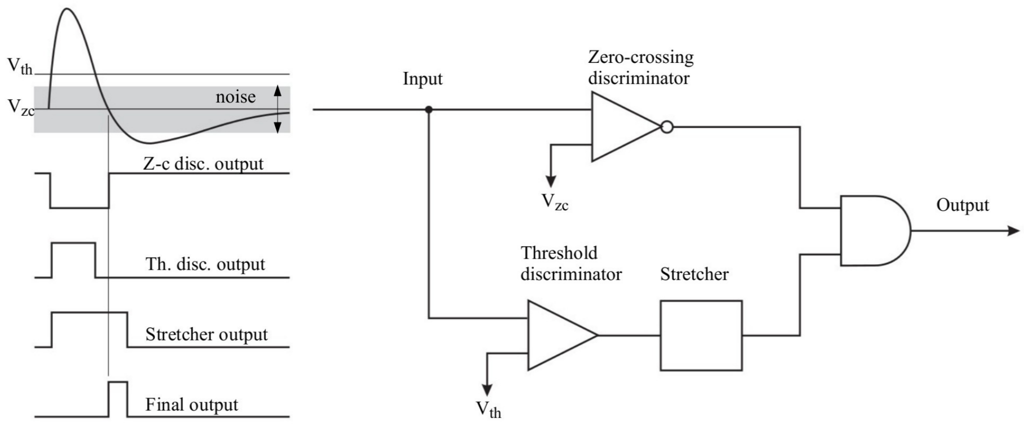


FIGURE 3.9: Zero-crossing circuit scheme [36].

3.2.3 Constant fraction timing

The technique of constant fraction discrimination allows to compensate for both amplitude and rise time variations described in Sec. 3.2.1. The constant fraction discriminator (CFD) uses indeed a constant fraction f of the input pulse to determine the timing of the output pulse relative to the input signal [59]. The input signal is split: one half is attenuated (resulting in a certain fraction of the original amplitude) and inverted, while the other half is delayed of a time t_d greater than the rise time. By considering a linear signal with initial amplitude A, the leading edge is always $V(t) = At/t_r$, the attenuated and inverted signal thus being

$$V_{att} = -fV(t) \tag{3.47}$$

and the delayed one

$$V_{del} = V(t - t_d). \tag{3.48}$$

The two signals are then compared, resulting in a bipolar output (Fig. 3.10)

$$V_{CFD} = V(t - t_d) - fV(t) \tag{3.49}$$

whose zero crossing is detected and then the relevant signals are selected by a leading edge discriminator. For a linear signal, the zero crossing time is derived as follows:

$$V(t_{zc} - t_d) = fV(t_{zc}), \tag{3.50}$$

$$\frac{A(t_{zc} - t_d)}{t_r} = \frac{fAt_{zc}}{t_r}, \tag{3.51}$$

$$t_{zc} = \frac{t_d}{1 - f}. \tag{3.52}$$

Since this depends on the delay and fraction only, it removes the dependence on rise time and amplitude. However, real signals are not perfectly linear, limiting the effectiveness in removing the rise time dependence. This effect could be mitigated by employing small values of f and t_d, provided that the negative signal amplitude is high enough to trigger the discriminator.

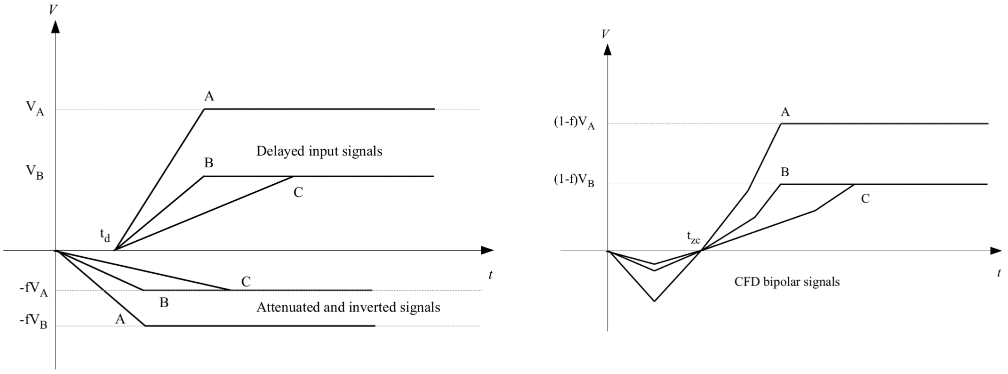


FIGURE 3.10: The combination of delayed and attenuated inverted signals establishing the zero crossing time of a CFD.

The jitter on the zero-crossing time can be evaluated as

$$\sigma_{t_{zc}} = \frac{\sigma_v t_r \sqrt{1+f^2}}{A(1-f)}, \tag{3.53}$$

a higher value than the one for the leading edge discriminator, the latter less affected by charge sensitivity due to a more pronounced signal slope. Despite its complexity, CFD is still a convenient choice when amplitude and rise time variations need to be minimized.

3.3 Time to Digital conversion

A Time to Digital Converter (TDC) is a device producing a digital code related to a time interval with excellent time resolution. The TDC recognizes an event and provides a digital representation of the time of occurrence.

The simplest method to perform the time-to-digital conversion is to combine a clock generator with a counter [9]. Start and stop signals determine the interval in which the clock pulses should be counted. The speed of the counter is the actual limitation to this technique.

One possible implementation of a TDC is a two-step approach making use of a time-to-amplitude converter (TAC) followed by an analog-to-digital converter (ADC). The time interval is converted into a voltage by charging a capacitor through a switchable current source, which is turned on and off by the start and stop signals, respectively. An example is shown in Fig. 3.11: the memory capacitor C is charged by the current I_T for the time $T_{stop} - T_{start}$, producing the voltage $V = Q/C = I_T(T_{stop} - T_{start})/C$ which is further digitized by the following ADC. For an N-bit ADC with a resolution V_{LSB}, the maximum time interval that can be measured with this technique (also called "dynamic range"DR) is:

$$DR = 2^N \cdot T_{LSB}, \tag{3.54}$$

while the minimum time interval that can be measured is T_{LSB}. The performance is mainly limited by analogue constraints of the two building blocks.

Digital TDCs overcome the dynamic range limitation of the analogue version and are therefore widely used. In this case, there is no analog conversion step [60]. In the fully digital approach, the measurement resolution is increased beyond the maximum clock frequency by dividing each clock cycle in a number of smaller intervals. Multiple phases of the reference clock are used to divide the clock interval. The measured time interval quantized with a TDC is

$$\begin{aligned} \Delta T &= NT_{CP} - (T_{CP} - \Delta T_{start}) + (T_{CP} - \Delta T_{stop}) \\ &= NT_{CP} + N_1\frac{T_{CP}}{k} - N_2\frac{T_{CP}}{k} - \epsilon_T, \end{aligned} \tag{3.55}$$

where T_{CP} is the reference clock period, $\Delta T_{start} = N_1 T_{CP}/k - \epsilon_1$ and $\Delta T_{stop} = N_2 T_{CP}/k - \epsilon_2$ are the time intervals between the start and the stop signals and the rising edge of the following clock signal. k describes how many partitions of the initial clock cycle are performed, N is the count of the reference clock cycle, N_1 and N_2 the position of the start and stop signals with respect to the reference clock cycle. The quantity $\epsilon_T = \epsilon_2 - \epsilon_1$ is the quantization error and is in the range between $-T_{CP}/k$ and T_{CP}/k, being the TDC resolution $T_{LSB} = T_{CP}/k$.

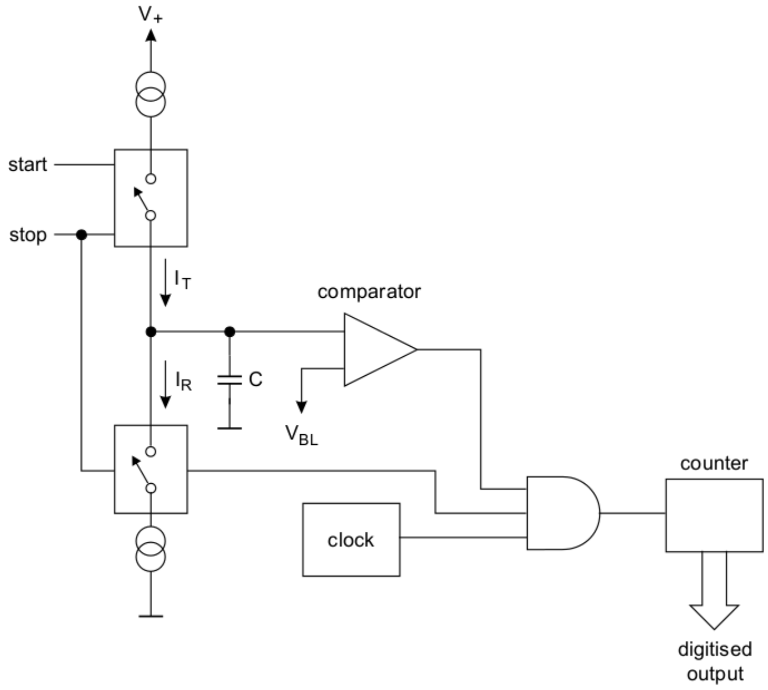


FIGURE 3.11: TDC composed by time-to-amplitude converter (TAC) and analog-to-digital converter (ADC) [9].

A very common implementation of a fully digital TDC uses digital delay lines to measure a time interval: the original clock reference is used as start signal and is delayed in a chain of digital delay blocks. When the stop signal arrives, the delayed versions of the start signal are sampled in parallel: the delay blocks that have been reached by the start signal produce a logic high value, while the ones that still have not been reached produce a low value. The time interval is then measured from the high-low transition as the number of high outputs:

$$\Delta T = NT_{LSB} + \epsilon. \tag{3.56}$$

In this case the resolution T_{LSB} is determined by the delay of the single elements of the delay line and the related quantization error is ϵ.

Another implementation of the digital TDCs makes use of Delay Locked Loops (DLL). This method allows to increase the dynamic range without employing an excessive number of delay blocks. The reference clock provides a coarse time measurement and is still sent to a delay line, whose output increments a counter. Simultaneously, this signal is compared with the reference clock phase. The DLL forces the total delay to be equal to the reference clock. The delays are incremental and controlled by the DLL.

3.4 Filter optimization

The purpose of the shaper is to manipulate the signal maximizing a given quantity, namely the signal-to-noise ratio (SNR) for energy measurements and the jitter (and therefore the slope-to-noise ratio, SLNR) for time measurements. In the following both filters optimized for energy and time measurements are analyzed, as the two analyses are connected [57] [55].

3.4.1 Optimization for energy measurements

The preamplifier of Fig. 3.4 can be considered. The noise generators to be considered are only the series white noise v_{nw}^2 and the parallel noise i_n^2, while the flicker noise is neglected. The detector delivers the charge Q_{in}, while the amplifier provides an impulse response

$$V_{out}(t) = -Q_{in}S(t), \tag{3.57}$$

or in the Laplace domain

$$V_{out}(s) \simeq -\frac{Q_{in}R_f}{1 + sR_f\left(\frac{C_T}{A_0} + C_f\right)}, \tag{3.58}$$

where C_T contains the contributions of all capacitances between the amplifier input and ground as in Eq. 3.19 [61]. The preamplifier is then followed by a filter with transfer function $H(s)$ [62]. In the time domain, the voltage signal at the output of the filter is given by

$$V_{F,E}(t) = Q_{in}S(t) * H(t) = Q_{in}\int_{-\infty}^{+\infty} H(\jmath\omega)S(\omega)e^{\jmath\omega t}df, \tag{3.59}$$

while the noise is

$$\sigma_n^2 = \int_{-\infty}^{+\infty} N(\omega)|H(\jmath\omega)|^2 df, \tag{3.60}$$

where $N(\omega)$ is the noise spectral density introduced by the preamplifier. The signal-to-noise ratio at the filter output observed at the time t is

$$SNR^2 = Q_{in}^2 \frac{\int_{-\infty}^{+\infty} H(\jmath\omega)S(\omega)e^{\jmath\omega t}df}{\int_{-\infty}^{+\infty} N(\omega)|H(\jmath\omega)|^2 df} \tag{3.61}$$

and an upper limit at

$$SNR_{max}^2 = Q_{in}^2 \frac{\int_{-\infty}^{+\infty} |S(\omega)|^2}{\int_{-\infty}^{+\infty} N(\omega)} df, \tag{3.62}$$

which is independent on the filter transfer function $H(s)$. The maximum SNR^2 is then obtained for

$$H(\jmath\omega) = k\frac{S^*(\omega)}{N(\omega)}e^{\jmath\omega t_m}, \tag{3.63}$$

where S^* is the complex-conjugate of S.
The noise spectral density $N(\omega)$ can be evaluated taking into account that the noise at the preamplifier output comes from both white (series noise) and frequency-dependent (parallel noise) components:

$$v_{out,s}^2 = v_{nw}^2 \left|\frac{C_T}{C_f}\right|^2, \tag{3.64a}$$

$$v_{out,p}^2 = \frac{i_n^2}{\omega^2 C_f^2}, \tag{3.64b}$$

becoming therefore

$$N(\omega) = v_{out,s}^2 + v_{out,p}^2 = \left(\frac{C_T}{C_f}\right)^2 \left(v_{nw}^2 + \frac{i_n^2}{\omega^2 C_T^2}\right) = A^2\left(v_{nw}^2 + \frac{i_n^2}{\omega^2 C_T^2}\right), \tag{3.65}$$

where $A = C_T/C_f$. However, the mathematical procedure to derive the filter that optimizes the signal-to-noise ratio requires the noise spectral density to be white. Thus, a whitening filter with transfer function $H_w(\jmath\omega)$ is added to make the noise at its output white again. The noise spectral density at the output of the whitening filter becomes therefore a constant

$$W^2 = N(\omega)|H_w(\jmath\omega)|^2, \tag{3.66}$$

where

$$|H_w(\jmath\omega)|^2 = \frac{W^2}{N(\omega)} = \frac{W^2\omega^2 C_T^2}{A^2(i_n^2 + v_{nw}^2\omega^2 C_T^2)} \tag{3.67}$$

and therefore the transfer function of the whitening filter is

$$H_w(\jmath\omega) = \frac{W}{Av_{nw}}\frac{\jmath\omega C_T}{1+\jmath\omega C_T\frac{v_{nw}}{i_n}} = H_0\frac{\jmath\omega\tau_c}{1+\jmath\omega\tau_c}, \tag{3.68}$$

namely the filter is composed by an ideal amplifier with gain $H_0 = \frac{W}{Av_{nw}}$ and a high-pass filter with time constant $\tau_c = C_T v_{nw}/i_n$ called *noise corner time constant*. Indeed, at the frequency $\omega_c = 1/\tau_c$ the series noise and the parallel noise at the preamplifier output are equal.

The voltage signal at the output of the whitening filter is then in the Laplace domain

$$V_w(s) = V_{out}(s)H_w(s) = \frac{Q_{in}}{sC_f}H_0\frac{s\tau_c}{1+s\tau_c} \tag{3.69}$$

and in the time domain

$$V_w(t) = \frac{Q_{in}}{C_f}H_0 e^{-\frac{t}{\tau_c}}u(t) = \frac{Q_{in}}{C_T}\frac{W}{Av_{nw}}e^{-\frac{t}{\tau_c}}u(t), \tag{3.70}$$

where $u(t)$ is the unit step function. This impulse response has a maximum at $t = 0$, after which it decreases exponentially. Given a measurement time t_m, at a time $t - t_m$ one has that

$$V_w(t-t_m) = \frac{Q_{in}}{C_T}\frac{W}{v_{nw}}e^{-\frac{t-t_m}{\tau_c}}u(t-t_m). \tag{3.71}$$

It is then possible to derive the filter to be applied in order to maximize the SNR, as the noise has no longer frequency-dependent components. The impulse response of this filter can be obtained from 3.63 supposing a white noise spectral density W:

$$H(t) = \frac{k}{W^2}s(t_m - t) = Ke^{\frac{t-t_m}{\tau_c}}u(t_m - t)., \tag{3.72}$$

where K is an arbitrary constant. The total output of the system composed by preamplifier, whitening filter and shaper is obtained by the convolution of $V_w(t)$ and $H(t)$:

$$\begin{aligned} V_{op,E}(t) &= \int_0^{+\infty} V_w(x)H(t-x)dx \\ &= \frac{KQ_{in}W}{C_T v_{nw}}\int_0^{+\infty} e^{-\frac{x}{\tau_c}}u(x)e^{\frac{t-t_m-x}{\tau_c}}u(x-t+t_m)dx \\ &= \frac{K\tau_c Q_{in}W}{2C_T v_{nw}}\left[e^{-\frac{t-t_m}{\tau_c}}u(t-t_m) + e^{\frac{t-t_m}{\tau_c}}u(t_m-t)\right]. \end{aligned} \tag{3.73}$$

whose maximum, located at the time t_m, is

$$V_{op,E,max} = \frac{KQ_{in}W\tau_c}{2C_T v_{nw}} \tag{3.74}$$

The impulse response of the whole chain has the shape of Fig. 3.12.a: an infinite cusp peaking at the measurement time.
In order to obtain the SNR at the peak value the maximum of the impulse response has to be divided by the total output noise, the latter being

$$\sigma_n^2 =< v_{out}^2 >= W^2 \int_0^{+\infty} |H(\jmath 2\pi f)|^2 df = K^2 W^2 \frac{\tau_c}{2} \tag{3.75}$$

with an RMS value of $\sigma_n = KW\sqrt{\tau_c}/2$. The resulting signal-to-noise ratio is then

$$SNR_{peak} = \frac{Q_{in}}{C_T}\frac{\sqrt{\tau_c}}{v_{nw}} = \frac{Q_{in}}{\sqrt{2C_T v_{nw} i_n}} \tag{3.76}$$

and the equivalent noise charge, defined as the input charge giving a SNR of 1 is [55]

$$ENC = \sqrt{C_T v_{nw} i_n} = \frac{C_T v_{nw}}{\sqrt{\tau_c}}. \tag{3.77}$$

The previous discussion however, has the following limitations: first, it assumes that the whole information contained in the input signal is used, which requires an infinite delay between the time of occurrence and the measurement time. Second, since the duration of the filter response is infinite, it is not possible to perform successive measurements. Moreover, no finite duration of the input signal is considered and the noise spectrum is supposed to be perfectly white. For what concerns the duration of the filter, it can be limited by introducing a time boundary:

$$H_b(t) = H(t)u(t - t_m + t_b), \tag{3.78}$$

namely the shaper output assumes a finite duration and only a part of the input signal is used. In practice, this is pursued to avoid the overlapping of the signals coming from subsequent events (the so-called “pileup”. The response of this filter is

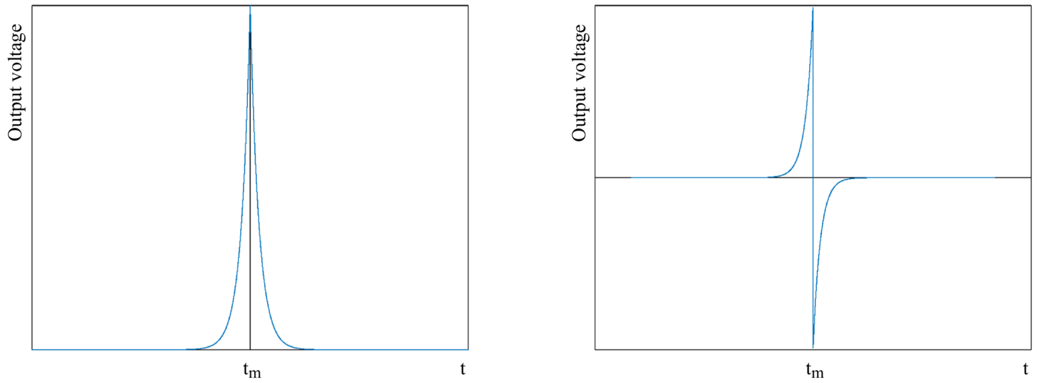
$$\begin{aligned} V_{b,E} &= \int_0^{t_1} V_w(x)H_b(t-x)dx \\ &= \frac{KQ_{in}\tau_c}{2C_T v_{nw}}\left[\left(e^{\frac{t-t_m}{\tau_c}} - e^{-\frac{t-t_m+2t_b}{\tau_c}}\right)u(t_m - t)u(t - t_m + t_b)\right. \\ &+ \left.\left(e^{-\frac{t-t_m}{\tau_c}} - e^{\frac{t-t_m-2t_b}{\tau_c}}\right)u(t - t_m)u(t_m - t_b - t)\right] \end{aligned} \tag{3.79}$$

and the related SNR is

$$SNR_b = SNR_{op}\sqrt{1 - e^{\frac{2t_b}{\tau_c}}}. \tag{3.80}$$

3.4.2 Optimization for timing

For timing applications, the minimization of the jitter implies the maximization of the slew rate with respect to the noise [63]. The filter optimized for time measurements should then shape the signal in such a way that the slope-to-noise ratio is



(a) (b)

FIGURE 3.12: Impulse response of the signal processing chain optimized for energy (left) and timing (right). The function of the timing-optimized filter is the time-derivative of the energy-optimized one.

maximized at the instant of the measurement. The filter that optimizes a time measurement can be derived from the analysis of the filter optimizing energy measurements, taking into account that the slope of a signal is the time derivative of the signal itself, i.e. by using $Q_{in}dS(t)/dt = Q_{in}\jmath\omega S(\omega)$ instead of $Q_{in}S(t)$:

$$SLNR^2_{max} = Q^2_{in}\int_{-\infty}^{+\infty}\frac{|\jmath\omega S(\omega)|^2}{N(\omega)}df, \tag{3.81}$$

which becomes in the case of white noise

$$\begin{aligned} SLNR^2_{max} &= \frac{Q^2_{in}}{W^2}\int_{-\infty}^{+\infty}\frac{dS(t)^2}{dt}dt \\ &= \frac{Q^2_{in}}{W^2}\int_{-\infty}^{+\infty}|\jmath\omega S(\omega)|^2 df, \end{aligned} \tag{3.82}$$

and the response of the filter, instead of $H(t)$, is its time derivative

$$\begin{aligned} Z(t) &= K\frac{dS(t_m - t)}{dt} = \frac{dH(t)}{dt} \\ &= K\left[\frac{e^{\frac{t-t_m}{\tau_c}}}{\tau_c}u(t_m - t) - e^{\frac{t-t_m}{\tau_c}}\delta(t - t_m)\right] \end{aligned} \tag{3.83}$$

The resulting filter has the impulse response given by the convolution of the whitening filter (3.70) and 3.83:

$$\begin{aligned} V_{op,t}(t) &= V_w(t) * Z(t) \\ &= \frac{KQ_{in}W}{2C_T v_{nw}}\left[e^{\frac{t}{\tau_c}}u(t_m - t) - e^{-\frac{t}{\tau_c}}u(t - t_m)\right] \end{aligned} \tag{3.84}$$

and the shape shown in Fig. 3.12.b, being the time derivative of Fig. 3.12.a. It can be demonstrated that the SLNR at the output of this filter is infinite:

$$
\begin{aligned}
SLNR &= \frac{Q_{in}^2}{W^2}\int_{-\infty}^{+\infty} |\jmath\omega S(\omega)|^2 df \\
&= \frac{Q_{in}^2}{2\pi W^2 \tau_c}\int_{-\infty}^{+\infty} \frac{\omega^2\tau_c^2}{1+\omega^2\tau_c^2} d(\omega\tau_c) \\
&= \frac{Q_{in}^2}{2\pi W^2 \tau_c}(\omega\tau_c - \arctan\omega\tau_c)|_{-\infty}^{+\infty} \\
&= \infty.
\end{aligned}
\tag{3.85}
$$

The ideal filter for timing suffers from the same limitations of the ideal filter for energy, its response being not limited in time. The addition of a time boundary degrades the achievable SLNR but leads to feasible solutions and limits the pileup problem and is obtained by replacing $Z(t)$ with

$$
Z_b(t) = Z(t)u(t - t_m + t_b). \tag{3.86}
$$

Chapter 4

Ultra Fast Silicon Detectors

Ultra Fast Silicon Detectors (UFSD) [20] have been proposed with the aim of enhancing concurrent space and time resolution. In order to increase the time resolution it is crucial to maximize the *slope-to-noise ratio*, which can be achieved by a sensor providing a large signal with a short duration in time. The collection time can be reduced by employing a thin bulk: this, however, leads to a decrease of the signal amplitude since the sensitive volume in which electron-hole pairs can be created is rather small. An internal charge multiplication is then used to overcome this reduction of collected charge and to tailor the signal shape in order to increase its slope.

UFSD are thin pixelated n-on-p silicon detectors based on the "Low Gain Avalanche Detector"(LGAD) design developed by CNM Barcelona [18]. UFSD are currently produced by CNM, FBK [64],[65] and HPK. The goal of UFSD sensors is to provide simultaneously high granularity for spatial measurements (i.e. segmenting the read-out electrode) and capability for high rate data collection and accurate time measurements (requiring signals with short duration).

In the last years, several geometries, gain values and bulk thicknesses have been explored within the design of LGAD. Since not all combinations of the previous characteristics are suitable for timing applications, the wording UFSD refers to those LGAD sensors optimized for precise time measurements.

First applications of UFSD are envisioned in LHC upgrades, in cases where the excellent time resolution coupled with good spatial resolution helps to drastically reduce pileup effects due to the large number of individual interaction vertices. As introduced in Chapter 2, UFSD have been proposed for the endcaps of the MIP Timing Detector of CMS. In addition, CMS-TOTEM considered UFSD as timing detectors for the high momentum - high rapidity Precision Proton Spectrometer (CT-PPS), placed in roman-pots at about 200 m from the primary interaction region, while ATLAS is proposing these sensors as one of the technical options for the High Granularity Timing Detector (HGTD) located in front of the forward calorimeter (FCAL). In all the previous cases, the radiation requirements will be high (few times 10^{15} neq/cm^2) and the required time resolution will be around 30 ps, achievable by stacking up in a series up to four sensors. The expected segmentation would be a few mm^2. The LGAD technology couples the best characteristics of standard silicon sensors with the main feature of APDs [66].

The time resolution of UFSD produced by CNM has been proofed in beam tests with 180 GeV pions and led to ~ 16 ps when read out with a discrete component board developed at the University of California Santa Cruz (UCSC) [67].

4.1 Operating principle

In silicon, as mentioned in Sec. 2.1.3, the electron drift velocity saturates to a value of $\sim 10^7$ cm/s. This means that, in a sensor of about 300 μm of thickness collecting electrons, the charge collection time is ~ 3 ns. Thus, a faster sensor can be obtained by reducing the sensor thickness. Since a thinner sensor leads to a smaller collected charge, this will not be enough to reach an adequate signal-to-noise ratio and then a good time resolution. This means that, in order to build a detector which can be considered "fast", it has to be thinner than average sensors but also able to work with a smaller collected charge. For this reason, an internal charge multiplication mechanism is used to enlarge the signal and enhance time measurements. The charge multiplication is based on the impact ionization mechanism described in Sec. 2.1.4 and occurs when charge carriers move into an electric field of the order of $E \sim$ 300 kV/cm. This effect is already used in silicon sensor technology in devices like Avalanche Photo-Detectors (APDs) with a gain of about 10^2 and in Silicon Photo-Multipliers (SiPM) with a gain around 10^6. LGAD-based sensors like UFSD feature a gain of 10-20.

The choice of an appropriate electric field is a crucial point in the development of UFSD: the field should be high enough to determine charge multiplication and sufficiently low to prevent breakdown. The high electric field zone is generated with a p^+ implant (boron or gallium) under the n electrode, which creates a large increase in the doping concentration close to the junction. The comparison between the structure of a traditional n-in-p silicon detector and a LGAD is shown in Fig. 4.1. The electric field in a 300 µm thick LGAD at three different bias voltages is shown in Fig. 4.2 and plotted alongside the electric field of a PiN diode at a bias voltage of 600 V. The electric field of an LGAD has two distinct zones: a low-field zone in the drift volume ($E \sim 30$ kV/cm) and a narrow high-field zone in the multiplication

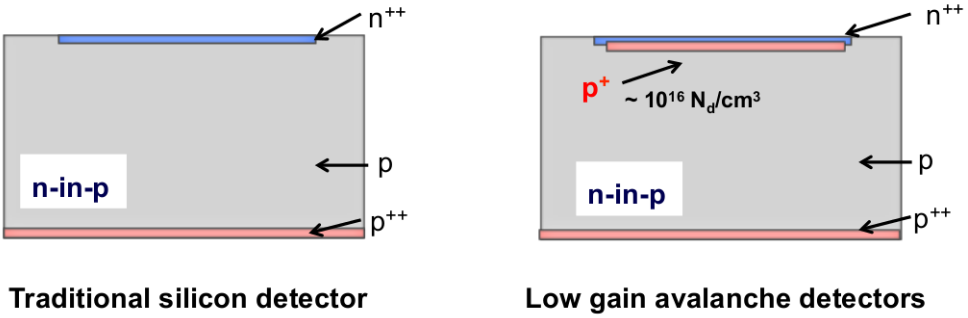


FIGURE 4.1: Left: n-in-p silicon sensor. Right: the LGAD design, obtained from a n-in-p sensor with the introduction of a thin p^+ layer below the junction.

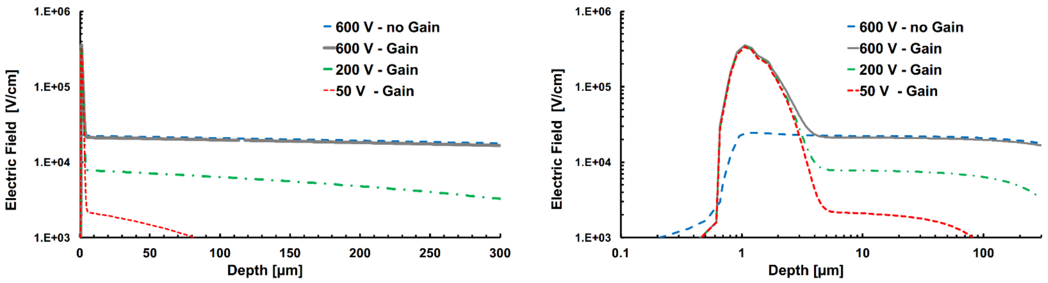
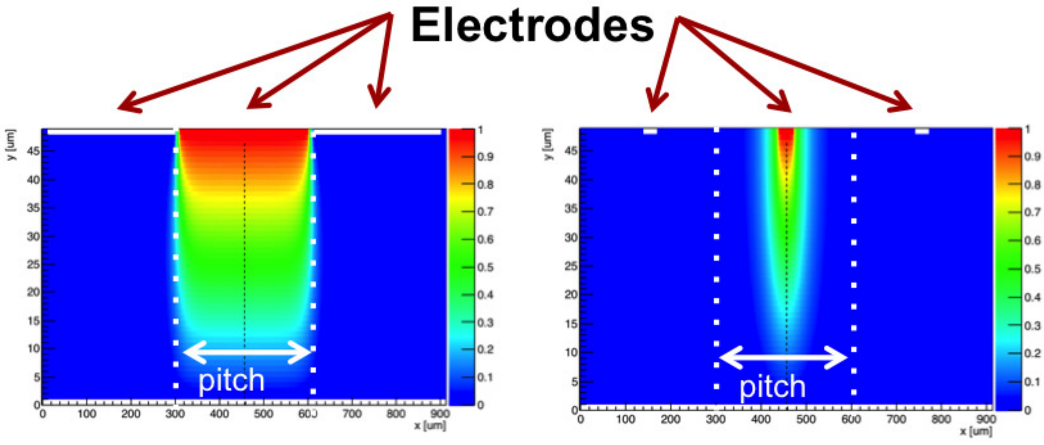


FIGURE 4.2: Electric field of a 300 µm thick LGAD at bias voltages of 50, 200 and 600 V compared to the electric field of a sensor without gain (PiN) in linear scale (left) and logarithmic scale (right).

region ($E \sim 300$ kV/cm), enclosed in about 2 μm of space. The p-type bulk is chosen since in a n-in-p device the multiplication is initiated by electrons drifting toward the n electrode. Compared with a p-in-n device where the multiplication is started by holes drifting to the p electrode, the first design allows a better control on the multiplication process: indeed, electrons have a higher ionization coefficient than holes and thus the multiplication can be initiated at lower field strengths. Moreover, in the n-in-p design the electric field can be tuned in such way that the multiplication is only driven by electrons, avoiding to operate the sensor in avalanche mode. In this way, the gain is less sensitive to the electric field value inside the device and the multiplication noise (excess noise factor) is reduced [10].



Wide Strip

Thin Strip

FIGURE 4.3: Weighting field for two strip configurations: wide strips (left) and thin strips (right) simulated with *Weightfield 2*.

4.1.1 Signal shape

The program *Weightfield 2* [68] allows to simulate the output signal of UFSD as a function of several parameters, such as gain value, sensor thickness, electrode segmentation and electric field. The simulated current produced by a MIP in a UFSD sensor with 50 μm of thickness and a gain of 10 is shown in Fig. 4.4: the total current (green) is given by the contribution of the initial electrons (red) and holes (blue) drifting toward the n and p electrodes respectively, gain electrons (purple) and gain holes (light blue) produced by initial charge carriers crossing the gain layer. The multiplication starts when an electron enters the gain layer at the n-p junction. The rise time of the UFSD signal is determined by the drift time of electrons, increasing the current when drifting to the n electrode. The gain electrons are quickly collected as they are produced near the cathode of the diode, while gain holes drifting to the anode have to cross the whole bulk and thus generate a large current. The signal of a 50 μm UFSD device has a total length of ~ 1.2 ns: the total collection time of a sensor of the same thickness without charge multiplication would be around 600 ps (500 ps for electron collection and an additional 100 ps due to the slower hole drift), which is about doubled due to the gain hole component.

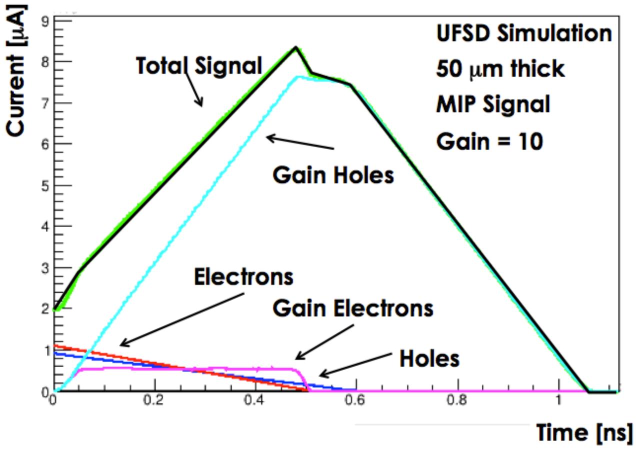


FIGURE 4.4: Current signal of a 50 μm thick UFSD simulated with *Weightfield 2*.

Variations in the signal shape are due to non-uniform charge deposition produced by Landau fluctuations and to signal distortions coming from non uniform weighting field and non-saturated drift velocity.

4.1.1.1 Landau fluctuations

As said in Chapter 2, the charge distribution created by an ionizing particle into a sensor varies at each event, being subject to statistical fluctuations, and follows mainly a Landau distribution. Such variations lead to both a change in the signal magnitude (due to the amount of charge released and at the basis of the time walk effect) and an irregular signal shape. This effect is common to both sensors with and without gain.

Two simulated examples of non uniform energy deposition and subsequent changes in the signal shape for a sensor with gain= 1 are shown in Fig. 4.5: *Weightfield 2* allows to estimate the effect of the “Landau noise”on the current signal. In addition, as the thickness of the silicon layer decreases, the energy loss distribution is subject to a decrease in the most probable value and a increase of distribution width [69]. In particular, in a 50 μm thick sensor, the number of electron-hole pairs produced per micron is about 57 e-h pairs/μm.

4.1.1.2 Signal distortion

From the Shockley-Ramo’s theorem (Eq. 2.39) the key points in the design of a timing sensor emerge. For what concerns the drift velocity, it has to be kept as constant as possible within the sensor bulk, since non-uniformities in drift velocity cause variations in the signal shape as a function of the hit position [15]. Therefore, it is useful to have an electric field inside the sensor which is high enough to move charge carriers with saturated drift velocity. At room temperature, the drift velocity saturates for an electric field of $\sim$ 30 kV/cm. When cooling the sensor, the field magnitude that allows saturated drift velocity is lower and the value of the velocity is higher.

Moreover, the weighting field varies as a function of the electrode width. In Fig. 4.3, two situations are depicted: in the first, the sensor has wide strips and a

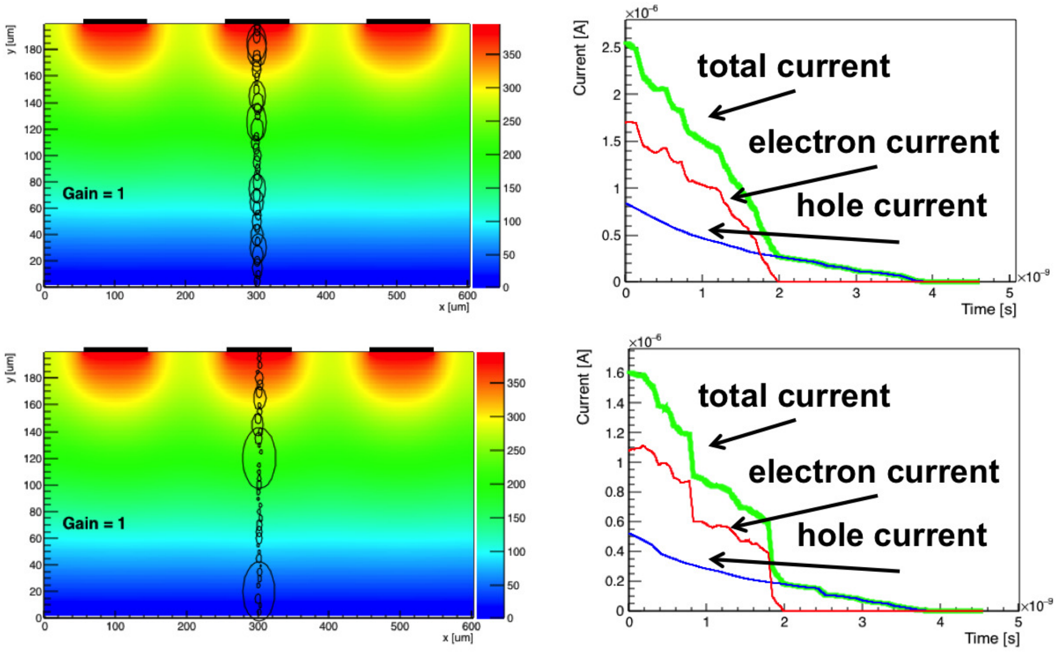
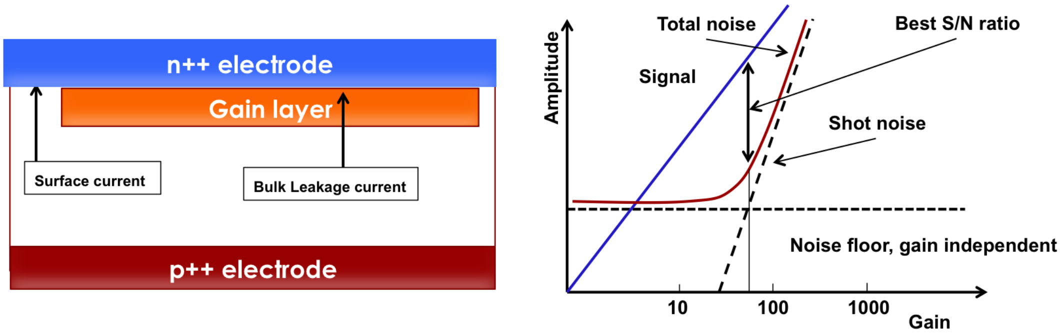


FIGURE 4.5: Simulation of different energy depositions in a 200 µm thick sensor without gain (left) and their effect on the current signal (right) [11].

pitch (i.e. the distance between each strip) very close to the strip width and therefore very small dead space; in the second, the strip width is much smaller compared with their distance. The resulting weighting field between strips, which is maximum under the strip, is very uniform in the first case, while varies rapidly in the thin strip case. This means that with thin strips, due to the weighting field shape, particles hitting near the center of the strip produce a much steeper and earlier signal [70].

4.1.2 Noise

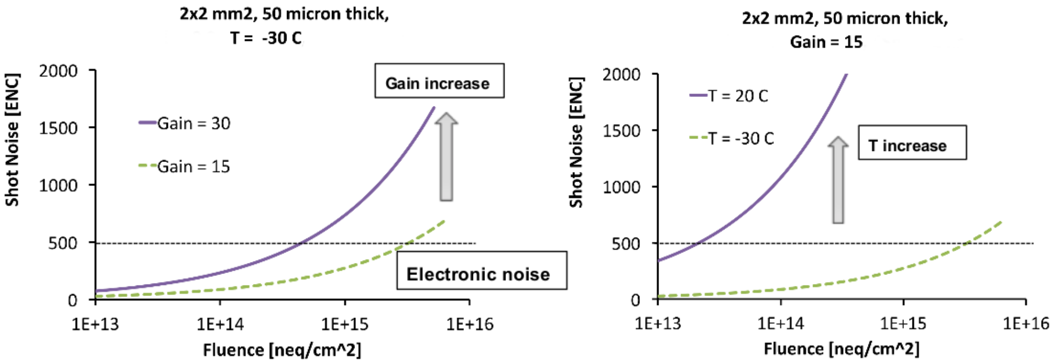
As explained in Sec. 2.3.1.1, the noise of a silicon sensor is due to two main components: shot noise and the frequency-dependent voltage noise due to trapping and detrapping. In devices with charge multiplication, the shot noise is further increased by the gain, becoming easily the main noise source. As described in Sec. 2.3.2.2, the sensor leakage current is the sum of two currents: (i) the surface current, which is not multiplied by the gain since it is does not cross the bulk and (ii) the bulk current, multiplied by the gain (Fig. 4.6.a). The relative importance of the two noise components and their comparison with the signal level are shown in Fig. 4.6.b: at low gain, the noise is fairly constant, then shot noise increases linearly with gain. The signal amplitude has a linear increase with the gain, too. However, this increase is slower than the one of shot noise, which becomes eventually the dominant contribution. For a gain ≤ 20, the signal is substantially higher than the shot noise and this latter has about the same magnitude of the electronic noise. In unirradiated sensors, shot noise is typically lower than the electronic noise, but it may prevail in irradiated devices. In Fig. 4.7 the shot noise for a sensor with an area of 4 mm^2 and a thickness of 50 µm is evaluated as a function of the fluence Φ, assuming an integration time of 2 ns and an electronic noise level of $\sim 500e$. Shot noise prevails on electronic noise with increasing gain and temperature, showing that this effect can be controlled by



(a) (b)

FIGURE 4.6: a) Shot noise mechanism in sensors with gain. b) Signal and noise amplitude dependence on the gain: [46].

keeping the gain and the temperature low. Moreover, since the leakage current increases with temperature (a factor 2 every 7 degrees) and with the electrode volume V according to $I_{leak} = \alpha\Phi V$ (where $\alpha = 3 \cdot 10^{-17}$ cm^{-1}), the shot noise can also be limited by keeping the volume under each electrode small [15].



(a) (b)

FIGURE 4.7: a) Increase of shot noise as a function of fluence for a gain of 15 and 30. b) Increase of shot noise with fluence for two different temperatures [15].

4.1.3 Time resolution

Taking into account the previously described effects and those described in Chapter 3 (namely, jitter and time walk), the time resolution of a system based on a UFSD sensor and a readout chain composed of an amplifier, a discriminator and a Time-to-Digital Converter can be summarized in the formula

$$\sigma_t^2 = \sigma_{TimeWalk}^2 + \sigma_{LandauNoise}^2 + \sigma_{Distortion}^2 + \sigma_{jitter}^2 + \sigma_{TDC}^2. \tag{4.1}$$

The variable energy deposition of a particle determining amplitude variations enters the terms $\sigma_{TimeWalk}$ and $\sigma_{LandauNoise}$. The term $\sigma_{Distortion}$ accounts for signal variabilities introduced by non-uniformities in the weighting field and in carriers drift velocity. The last two terms describe the uncertainties introduced by electronic noise and signal digitization. Jitter and time walk can be minimized by detectors with high slew rate, low intrinsic noise and a dedicated low noise readout. The signal distortion can be mitigated by keeping the carriers drift velocity properly saturated and modeling the device in such way that the weighting field uniformity is maximized (i.e. by employing a geometry where the active area is substantially larger than the sensor thickness). The time uncertainty due to TDC binning is usually negligible, as common TDCs provide an overall resolution of 25 ps.

4.2 Device Modeling

The production of a sensor requires dedicated simulation phases to predict its behavior. The physical behavior of a device can be represented with an abstract mathematical model that approximates its operation. Such a model can be either analytical or a system of coupled differential equations that need to be solved numerically. The analytical modeling represents a physical property in terms of approximated closed-form expressions, while numerical modeling implies the numerical solution of the differential equations describing the physics of the device on a given geometrical domain. Both are complementary techniques that are often used together in semiconductor design and should be compared or calibrated with experimental data. The numerical modeling can be accomplished with a Technology Computer Aided Design (TCAD), which is a branch of Electronic Design Automation (EDA) modeling semiconductor fabrication and device operation.

The physical simulation of a semiconductor device is based on two main components: the charge transportation, namely the motion of charges due to electric field and diffusion, and the electric field set by a charge distribution. The most common model in device simulation is the Drift-Diffusion Model (DDM) [71]. It is a system of coupled partial differential equations consisting in Poisson's equation and two continuity equations describing the charge conservation for electrons and holes:

$$\begin{cases} -\nabla \cdot (\epsilon \nabla \varphi) - \rho = 0, \\ e\frac{\partial n}{\partial t} - \nabla \cdot \mathbf{J_n} = e(G-R), \\ e\frac{\partial p}{\partial t} + \nabla \cdot \mathbf{J_p} = e(G-R) \end{cases} \tag{4.2}$$

where ϵ is the electric permittivity of the material, φ the electric potential inside the device, n, p are the electron and hole concentrations per volume unit, ρ the charge density per volume unit, $\mathbf{J_n}$ and $\mathbf{J_p}$ the current density vectors per surface unit, $(G-R)$ is the net generation-recombination rate. The previous equations are solved at each node of the discretized domain.

4.3 FBK UFSD1 design

The design of the first production of FBK UFSD sensors started in May 2014. Simulations have been performed with the commercial TCAD Synopsys Sentaurus. Different simulation domains have been used according to the type of analysis. 2D simulations allow to predict electrical characteristics of the device like leakage current and breakdown voltage, while dynamic simulations to evaluate gain and charge collection should be accomplished by implementing a cylindrical geometry in 2D simulations [64]: this allows to keep the lateral spread of charge clouds into account and thus to have a realistic estimate of the charge collection, at the price of a more reasonable simulation time with respect to 3D geometry.

All the following simulations use the standard avalanche model implemented in Sentaurus (Van Overstraeten-De Man [72]), unless otherwise specified.

4.3.1 Preliminary simulations

Simulations on 275 µm thick sensors started with an exploratory phase of preliminary simulations aimed at defining the correct doping concentration of the multiplication layer. This phase focused on the study of the achievable charge multiplication and the breakdown voltage of a simplified planar structure: the device is a 20 µm wide diode with the addition of a p^+ implant below the n electrode and has no edge termination.

The most common method to introduce dopants into the near-surface region of silicon wafers is by ion implantation. A first-order model for an implant doping profile is given by a gaussian

$$N(x) = N_p e^{-\frac{(x-R_p)^2}{2\Delta R_p^2}}, \tag{4.3}$$

where N_p is the peak concentration of the doping profile [atoms/cm^3], R_p the position of the peak (i.e. the implantation range) and ΔR_p the straggling of the gaussian, as shown in Fig. 4.8.

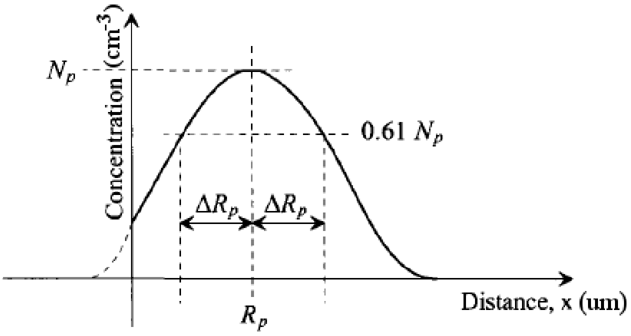


FIGURE 4.8: An ideal gaussian implant profile.

The doping of the bulk is p-type (boron) and set to the constant value of $2.5 \cdot 10^{12}$ cm^{-3}, while all doping profiles of electrodes and gain layer are gaussian distributions. The p electrode and the multiplication layer are doped with boron, while the n electrode is doped with phosphorus. The resulting doping of the $n-p$ junction is shown in Fig. 4.9.

The current vs voltage characteristic has been simulated for different values of the peak concentration of the gain layer. The device is reverse biased by applying a negative voltage to the p electrode, while the current at the n electrode is read. For the same concentrations, the current signal for a MIP hitting the sensor has been simulated and the gain value extracted. Optical simulations with a light beam with proper wavelength (for a MIP the value $\lambda = 1060$ nm has been used) impinging on the central coordinate of the sensor are suited for this purpose. The gain is evaluated as the ratio between the charge collected by a device with charge multiplication and the charge collected by a PIN diode:

$$\text{Gain} = \frac{Q_{UFSD}}{Q_{PIN}}, \tag{4.4}$$

where for a 275 µm thick device $Q_{PIN} \sim 80e \cdot 275 \simeq 22000$. The charge collected by the sensor is calculated by the simulator as the integral over time of the current signal. The IV characteristics of the device for different peak concentrations of the gain layer are shown in Fig. 4.10. The simulations demonstrate that the device was properly multiplying charge, therefore more realistic doping profiles have been introduced.

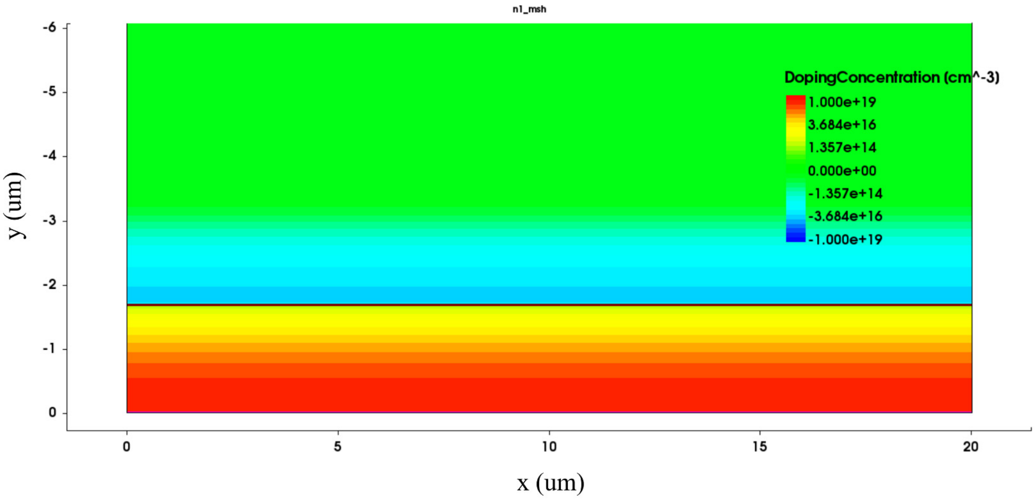


FIGURE 4.9: Doping concentration of the $n - p$ junction side showing the n^{++} electrode (red) and the p^{+} gain layer (blue) implemented with a gaussian profile.

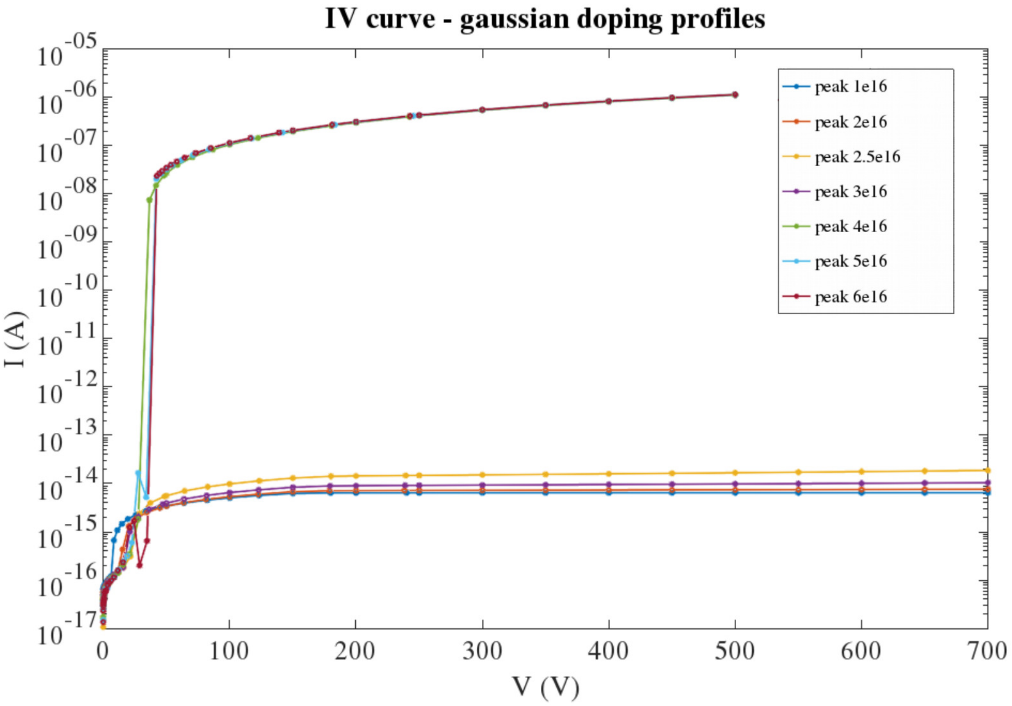


FIGURE 4.10: IV curve for different peak concentrations of the gain layer with a gaussian-shaped doping profile.

4.3.2 IV curves and charge collection with FBK doping profiles

The following step has been to introduce "real"doping profiles received from the foundry, which can be loaded into the simulator. The multiplication layer is achieved with a boron implant, while the n electrode is obtained with arsenic. The gain layer is scaled by a multiplication factor obtaining different splits, whose IV-characteristics are shown in Fig. 4.11. Each split is identified by this multiplication number, which is referred to as "gain scale". From the IV-curve it is possible to notice that lower gain scales (i.e. up to 0.8) have a higher breakdown voltage: this hints at the fact that the multiplication that can achieved with these scales is too low, and thus they have been excluded from further analyses.

Again, optical simulations have been used to predict the signal shape and the gain:

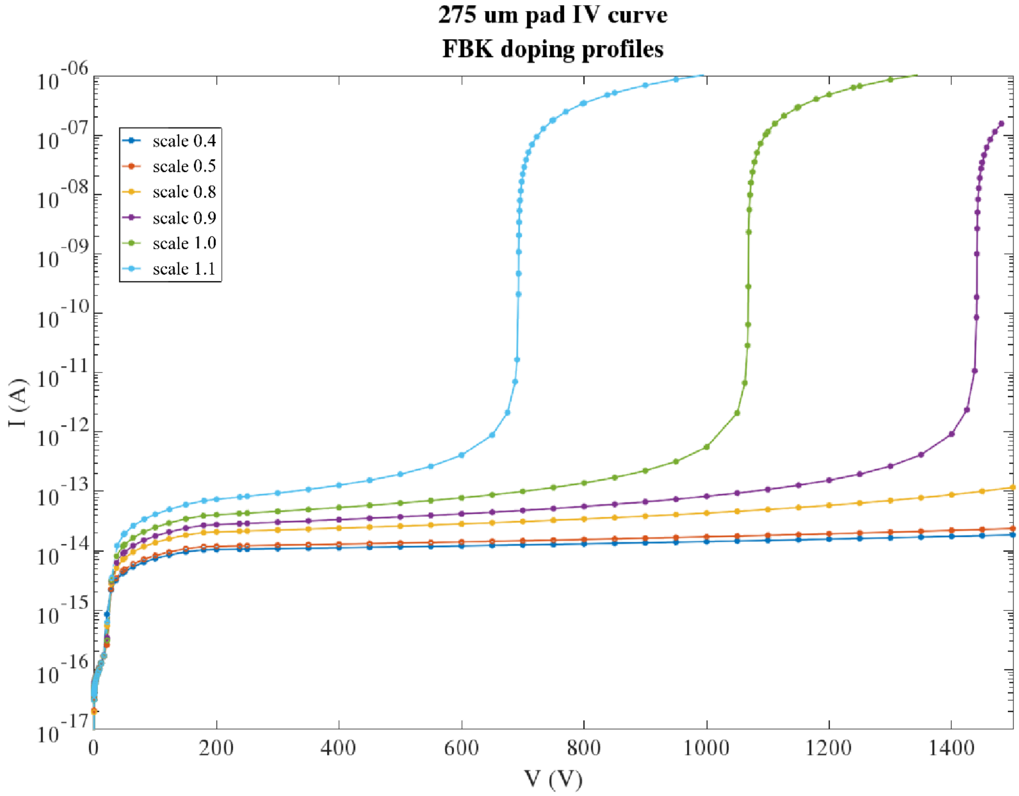
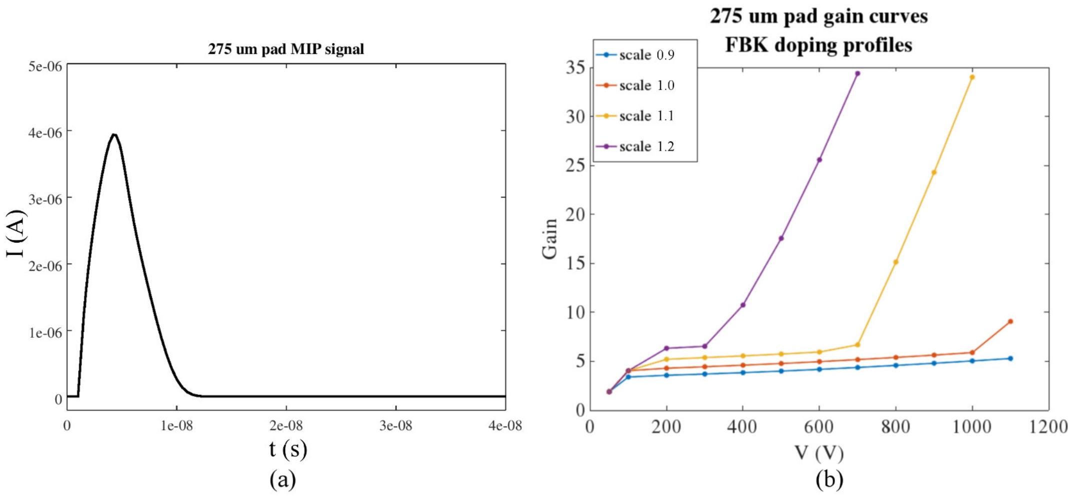


FIGURE 4.11: Leakage current as a function of reverse bias voltage simulated with FBK doping profiles at different gain layer scales.

in Fig. 4.12.a the current signal has a total length of $\sim$ 10 ns, which is mostly determined by the drift time of the holes and in agreement with the expectations. The time at which the signal reaches the peak, instead, is set by the drift of electrons. Fig. 4.12.b shows the gain curve of the device for different scales. The curve corresponding to a scale of 0.9 indicates that the achievable gain is still too low, while the one corresponding to a scale of 1.2 is too high: the device reaches breakdown before charge carriers reach the saturation of drift velocity. This study is particularly useful as it allows to determine the optimal value for the gain layer dose, as the difference between gain scales is not linear.



(a) (b)

FIGURE 4.12: a) Current signal as a function of time for a 275 µm thick device hit by a MIP for a gain scale of 1.0. b) Gain as a function of the reverse bias voltage for different gain scales.

4.3.3 Electrode segmentation strategy

The segmentation of the collecting electrode clearly improves the tracking resolution. However, the segmentation of the n electrode implies the segmentation of the multiplication layer. This could lead to a non-uniformity in the electric and weighting fields, possibly degrading the timing performance.
Highly segmented detectors like pixels require thus a p-type substrate, a large and uniform multiplication junction on one side and ohmic pixels on the other side [64]. The multiplication region can also be patterned into macro-pixels with an area of about 1 mm^2 without a drastic reduction the geometrical efficiency (estimated to be $\sim 5\%$.
The double sided device, whose schematic cross section is shown in Fig. 4.13, has been designed with a pixel side (on the top) that can be realized with a simple boron implant, and a rather complex "gain"side. This latter (bottom side of the figure) is obtained with four different layers realized combining ion implantation and thermal diffusion. A deep n$^+$ termination with a metal plate (JTE) to prevent edge breakdown is also used. In addition, the isolation of the gain macro-pixel is achieved with a p-stop, which is preferable to p-spray as it does not overlap with the doping profile of the gain layer. Early peripheral breakdown is prevented by adding a multiple guardring at the gain side.

4.3.4 Dead area studies

The study of the inactive area between pixels is of fundamental importance in the fabrication of sensors. Assuming a pixel with a surface of 200×200 µm^2 surrounded by a 50 µm wide guardring, the active area is given by the pixel size while the total area is given by the pixel size plus twice the guardring width. The fill factor i.e. the ratio between the active area and the total area is thus

$$F.F. = \frac{200 \times 200}{300 \times 300} = \frac{4}{9}, \tag{4.5}$$

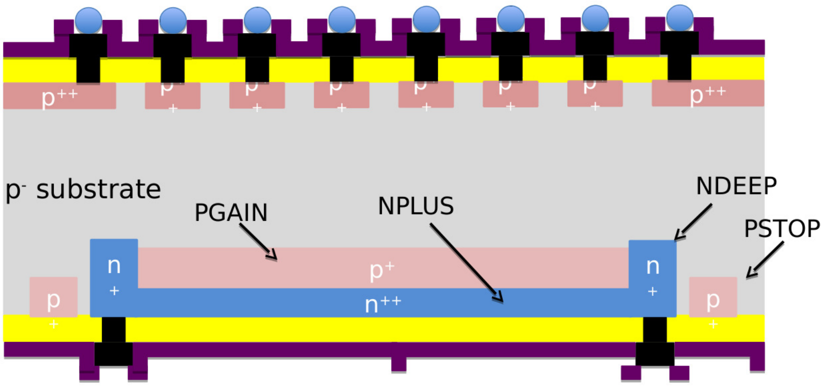


FIGURE 4.13: Schematic cross section of double-sided device.

leading to the fact that the inactive area can be a considerably high percentage of the total.

The simulation of two adjacent pixels of 100 μm width separated by two p-stop implants allows to investigate the signal of the two pixels and their spacing. The device is hit by a MIP along the y direction. The hit coordinate is varied to study the signal of one of the two pixels when the particle hits different areas of the device. The edge of each electrode is terminated by a n-type implant (*ndeep*) with different widths. When read out, the hit pixel should produce a unipolar signal while the signal of the neighbor pixel should be bipolar. The p-stop size and their distance are fixed to 5 μm and 6 μm, respectively. The *ndeep* can be considered as part of the dead area, since it acts as edge termination for the gain layer and then a particle hitting this implant does not undergo charge multiplication. Therefore, the width of the *ndeep* implant is varied in order to study the effect of reducing the dead space between pixels: the considered widths are 16, 10 and 5 μm.

The simulated structure is shown in Fig. 4.14 in the case of a *ndeep* width of 16 μm. By

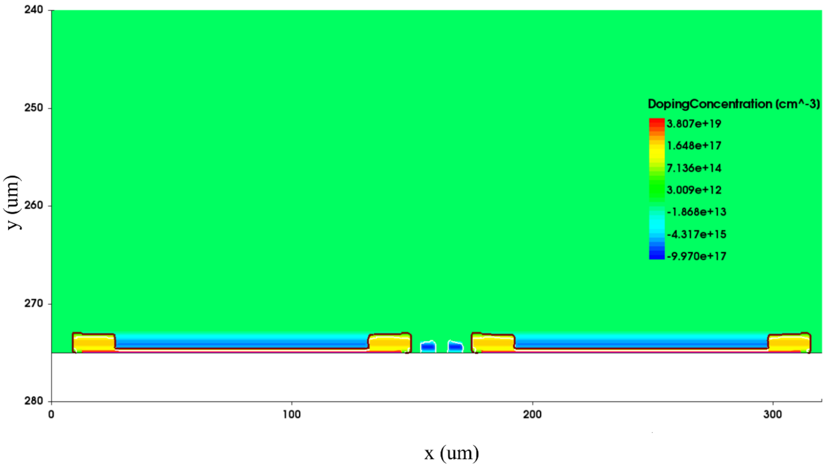


FIGURE 4.14: Simulated structure for the study of the signal of two adjacent pixels. In the picture, the electrodes have a width of 100 μm, the *ndeep* size is 16 μm and the p-stop width is 5 μm.

evaluating the gain as a function of the x position as shown in Fig. 4.15, it is possible to confirm that the *ndeep* is actually dead area, as charge collected when the particle hits this implant corresponds to a gain of less than 1. The current signals of the two adjacent pixels are compared in Fig. 4.16: the hit pixel is producing an unipolar signal (solid line), while the neighbor one produces a bipolar signal (dashed line) as expected. Moreover, the *ndeep* width can be decreased without affecting the signal, as the two pixels are effectively separated.

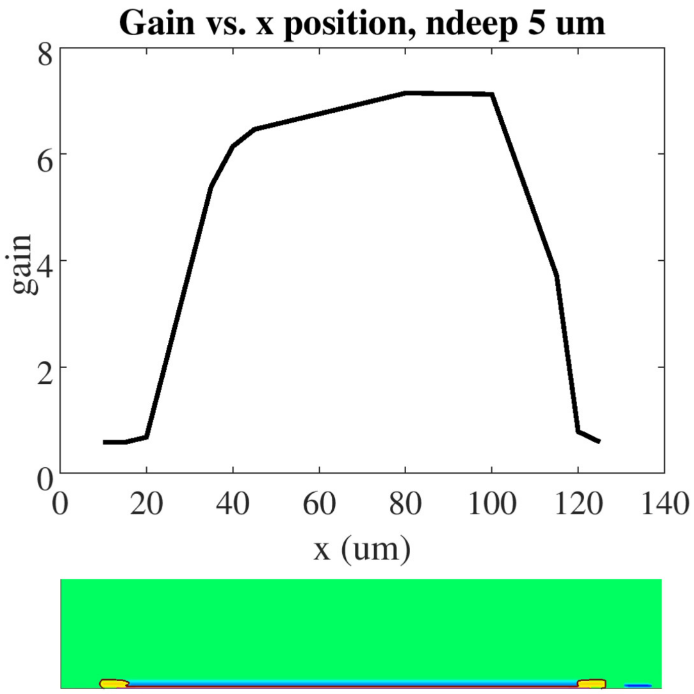


FIGURE 4.15: Gain as a function of the x coordinate for a width of the *ndeep* implant of 5 µm.

4.3.5 AC-coupled devices

In order to allow the electrodes segmentation without the need of segmenting the gain layer, the option to design a device with AC coupling has also been explored since it allows to have no dead area due to the presence of guardrings. In this way, it is possible to pixelate the n side while keeping the n-p junction side unaltered. For this purpose, three AC pads with equal width and spacing have been simulated. The AC pads are made of one conductive plate (the aluminum contact covering the boron-doped p electrode), an insulator layer acting as dielectric composed by the stack of a 120 nm layer of silicon dioxide (SiO_2) and a 120 nm layer of silicon nitride (Si_3N_4) and a second aluminum at the top, forming thus a parallel plate capacitor.

Simulations have been performed to evaluate the signal of the three pads when only the one in the middle is hit by a particle. AC pads produce a bipolar signal, due to the fact that there is no net charge collection as only image charges are involved. Although bipolar, the hit pad should produce a bigger signal while the neighbors should produce a smaller signal, Fig. 4.17 and 4.18 show the signal produced by AC pads when hit by the same amount of light for a spacing between pads of 20 µm and 30 µm, respectively. The situation is symmetric, then the signal of the leftmost pad is equal to the rightmost one. The signal is generated by a photon beam with 1060 nm wavelength and an intensity of 50 W/cm^2.

For 300 µm wide pads there is no substantial difference between the signal with 20 and 30 µm spacing, and the hit pad signal magnitude is substantially higher than

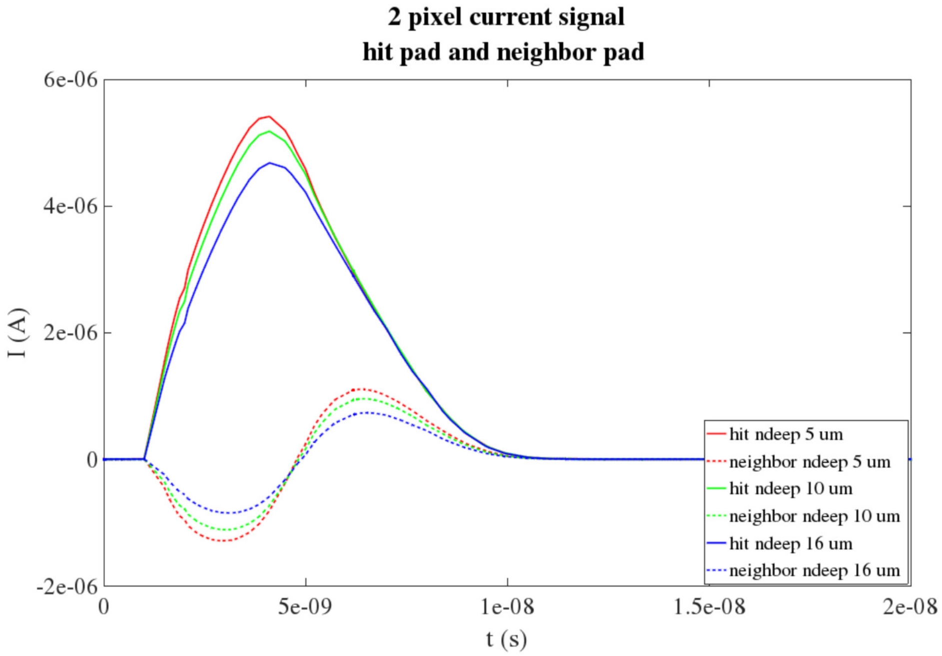


FIGURE 4.16: Current signal for hit pixel and neighbor pixel at different *ndeep* sizes.

the neighbor pad one. In the case of a 80 µm wide pad, a spacing of 20 µm optimizes the relative magnitude of the hit pad signal with respect to the neighbor pad, but the signal provided is much lower compared with the 300 µm pad. Moreover, the difference in magnitude of hit and neighbor pad signal is more pronounced for wider pads. In conclusion, the combination of 300 µm wide pads and 20 µm spacing optimizes the signal and minimizes the dead space.

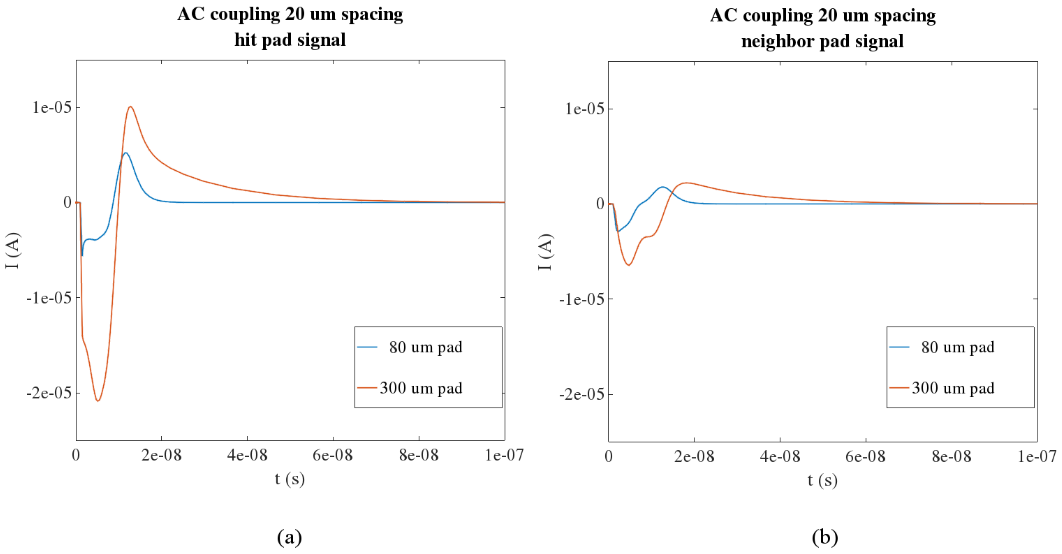


FIGURE 4.17: Comparison of the current signal of AC pads with 80 µm and 300 µm wide pads and 20 µm spacing produced by the hit pad (left) and the neighbor pad (right).

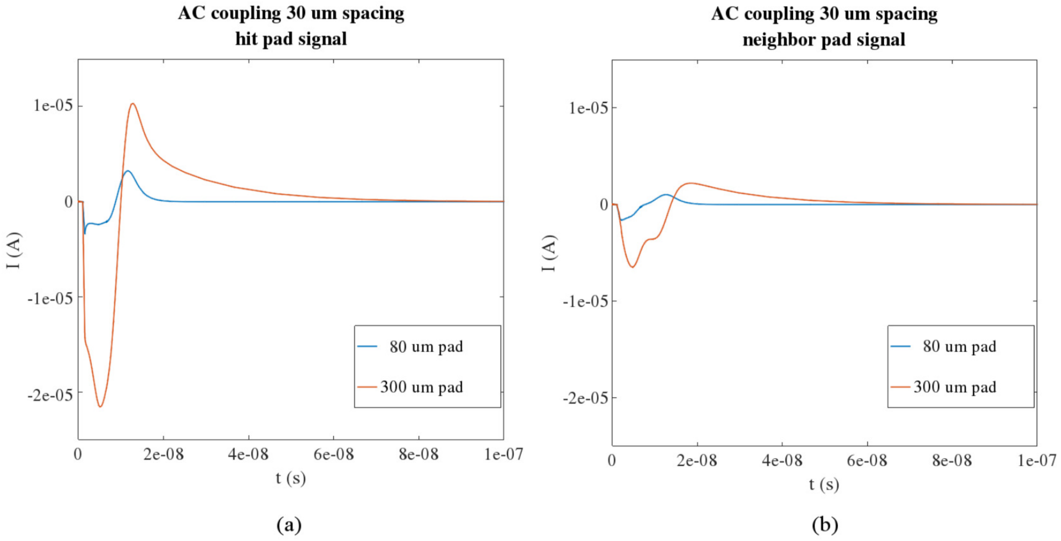


FIGURE 4.18: Comparison of the current signal of AC pads with 80 µm and 300 µm width and 30 µm spacing produced by the hit pad (left) and the neighbor pad (right).

4.3.6 Wafer layout

The final wafer layout is shown in Fig. 4.19. The featured device structures are:

- Single diodes with area ranging from 2.5 to 100 mm^2,
- 6 × 6 diode array with diode area comprised between 2.5 and 25 mm^2,
- 10 × 10 diode array with diode area between 2.5 and 25 mm^2,
- DC-coupled strips,
- AC-coupled strips with and without p side patterning,
- AC pads with AC pad size between 100 and 900 µm.

Five different splits of the gain layer dose and two p electrode doses have been produced. It is necessary to have several dose splits due to uncertainties in the fabrication process. About 1% deviation between simulations and effective dose of the gain layer after implantation and activation of the dopant is expected. Moreover, two wafers for each dose split allow an additional margin of error The combinations realized in the UFSD1 production are listed in Table 4.1. For testing purposes, diodes without gain are implemented on each wafer.

4.3.7 Outcome of UFSD1 production

The effective gain of different dose splits produced for UFSD1 have been investigated with laboratory measurements and compared with simulation predictions. For a given dose split, an infrared laser with fixed intensity is used to produce the signal and the gain has been measured at different bias voltages. The gain is evaluated as the ratio between the area of the UFSD signal acquired on the oscilloscope and the area of the signal produced by the diode without gain. Both results of measurements and simulations are plotted in Fig. 4.20. The data indicate a good agreement between measurements and simulations for Wafer 3 and Wafer

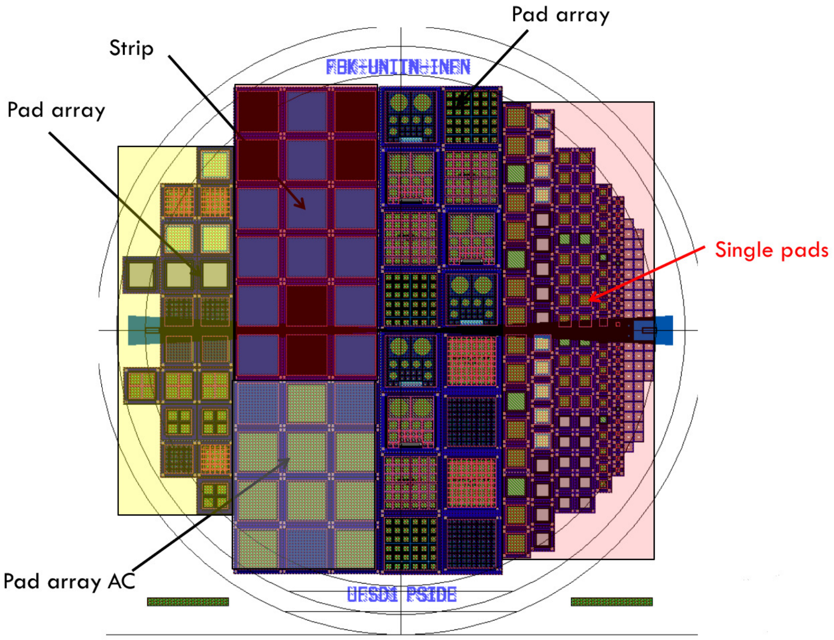


FIGURE 4.19: Final wafer layout for the FBK-UFSD1 production.

TABLE 4.1: Dose splits of gain layer and p electrode in UFSD1 wafers.

Wafer n.	Gain dose	p^+ dose
1, 2	1.0	4
3, 4	1.05	4
5	1.05	5
6, 7	1.1	4
8	1.1	5
9, 10	1.15	4
11	1.15	5
12, 13	1.2	4

10, demonstrating that the effective implanted dose is very close to the simulated one. Measurements performed on Wafers 1 and 7 indicate that the effective doping is lower than the predicted one. This effect can be ascribed to process-related variations that can occur during implantation and diffusion of the dopant, proving the fact that more wafer for a given dose splits are needed.

In conclusion, the UFSD1 production allowed to learn how to produce sensors with charge multiplication with the proper tuning of the gain layer doping, although they cannot be considered yet as fast timing devices due to their thickness and their consequently long collection time. The optimization of such device for timing purposes will be realized in the following production by employing a 50 µm thick wafer.

4.4 FBK UFSD2 design

In order to decrease the sensor thickness to 50 µm, the baseline is to simulate a thinner device with the same doping profiles used for the 275 µm sensors. However, given the same doping concentration, the breakdown voltage of a thinner device

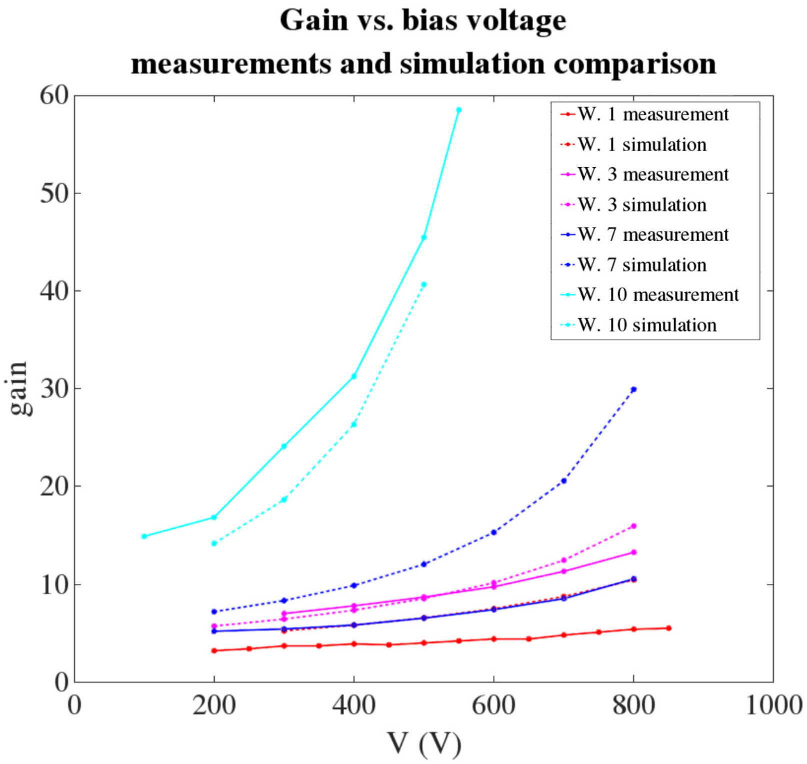


FIGURE 4.20: Measured gain as a function of the bias voltage for different UFSD1 wafers compared with simulation results.

will be lower, due to a higher electric field. It is required then to reevaluate the IV-characteristics. The goal is to have a breakdown voltage between 200 and 300 V and a gain between 10 and 20.

This second UFSD production also features two different dopants for the gain layer – boron and gallium – in order to compare their radiation hardness. In this work, simulations with boron-doped gain layer are presented, although several studies have been performed on a gallium-doped gain layer since the behavior of this latter dopant is considerably different. It has to be noted that during the process steps, the dopant undergoes a “diffusion”(also called *drive-in*) after being implanted. Diffusion is carried out at temperatures higher than 800°C and is used to activate the dopant. After diffusion, the doping profile experiences substantial changes, assuming a wider shape and a lower value at the peak. The electrical properties of the device are thus considerably different before and after diffusion and have been studied with dedicated simulations. Furthermore, a too low diffusion temperature could lead to a partial activation of the dopant.

4.4.1 Breakdown voltage and gain curves

Around the gain scale values chosen for the UFSD1 production, several gain scales have been investigated to understand the best matching between gain scale and breakdown voltage. Fig. 4.21 demonstrates that the breakdown can be reached even below 200 V of reverse bias voltage for the 1.05 scale, so far used for thicker devices. Lower gain scales have been considered in order to grant device operation above 200 V. The breakdown voltage as a function of the gain layer doping is shown in Fig. 4.22: it can be seen that a gain scale around 0.95 ensures the desired breakdown voltage range, i.e. between 200 and 300V. The actual behavior of charge multiplication as a function of the bias voltage has to be evaluated by means of the gain curve. The preliminary gain $vs.$ bias voltage curves for a diffused boron-doped gain layer are shown in Fig. 4.23.

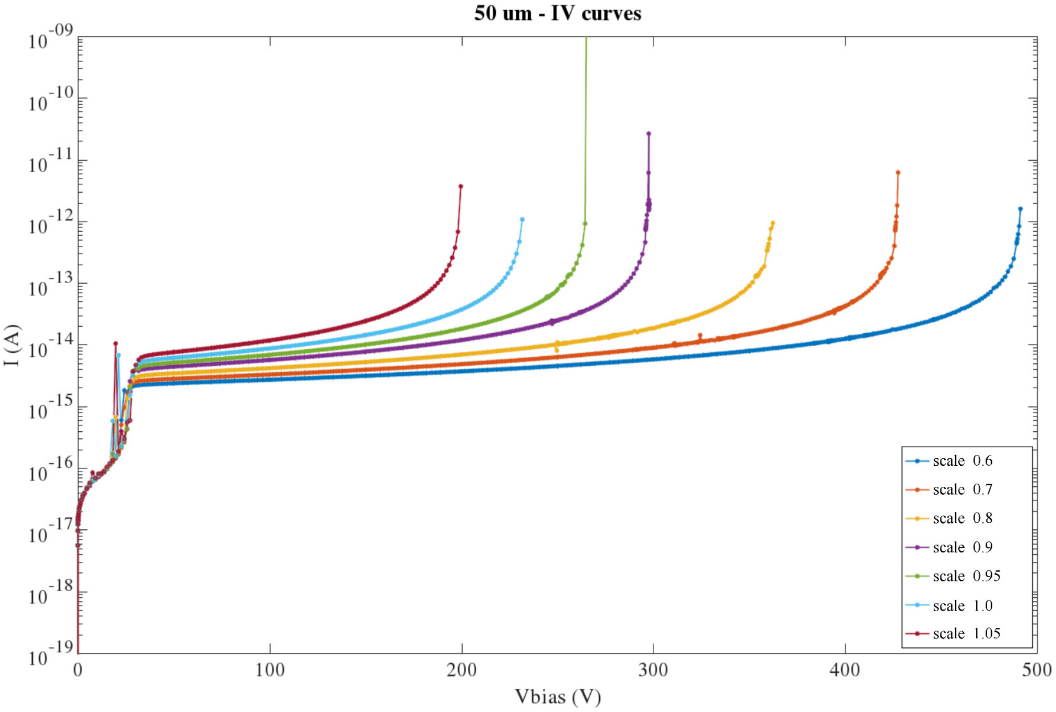


FIGURE 4.21: Simulated IV curves for different doping concentrations of the gain layer in a 50 μm device.

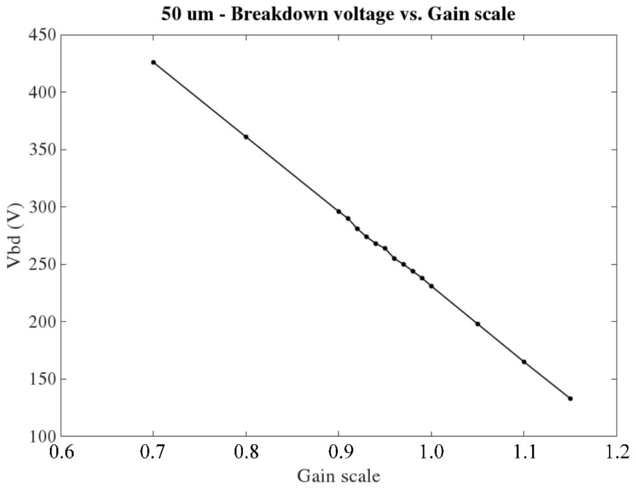


FIGURE 4.22: Breakdown voltage as a function of the gain layer doping for a 50 μm device.

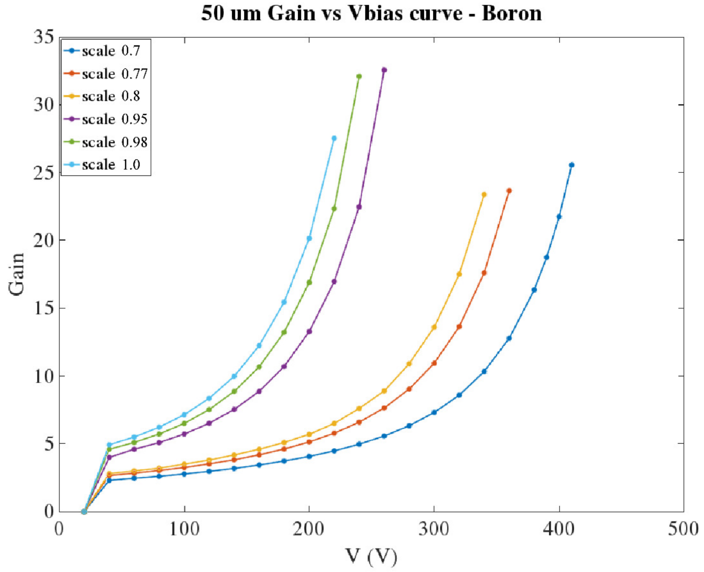


FIGURE 4.23: Gain curve for different boron gain scales.

4.4.2 Dead area studies

Similarly to Sec. 4.3.4, studies on the feasibility of a reduction of the dead area between electrodes have been performed. For this purpose, simulations aimed at reducing the inactive space between adjacent electrodes have been performed. The edge of single pixel is simulated by inserting an appropriate termination to the gain layer, implemented with an n-type implant (*ndeep*). The electrode is surrounded by a 5 µm wide p-type implant ("p-stop") and a n-type guardring. The distance between the guardring and the p-stop is fixed, as well as the distance between the p-stop and the n electrode. The guardring and the ndeep width are progressively reduced and the gain is evaluated along the x coordinate. The gain as a function of the x coordinate is plotted in Fig. 4.24 for different guardring and ndeep sizes. Simulations show that from an electrical point of view, the guardring and ndeep sizes can be decreased without anticipated edge breakdown. The ultimate limit to each segment size is set by the mask alignment, requiring a 5 µm tolerance.

4.4.3 Diffused and non diffused gain layer dopant

The effect of drive-in on the gain layer has been studied in gain $vs.$ bias voltage simulations. The Okuto avalanche model [73] has been implemented in these simulations, since it leads to more reliable results at low temperatures. Fig. 4.25 shows the gain as a function of bias voltage for a boron profile without diffusion, while Fig. 4.26 with diffusion. In the non diffused case, the gain curve is steeper and the breakdown occurs earlier, as a result of a narrower and sharper doping profile.

From process simulations, gallium revealed to have a wider and less pronounced doping profile immediately after implantation with respect to boron, leading to less steep gain curves and a lower sensitivity to bias voltage changes, reaching breakdown at higher voltages than boron. For the sake of completeness, a comparison between the gain curves of as-implanted boron and gallium is illustrated in Fig. 4.27, showing the clear difference between the behavior of two dopants before diffusion.

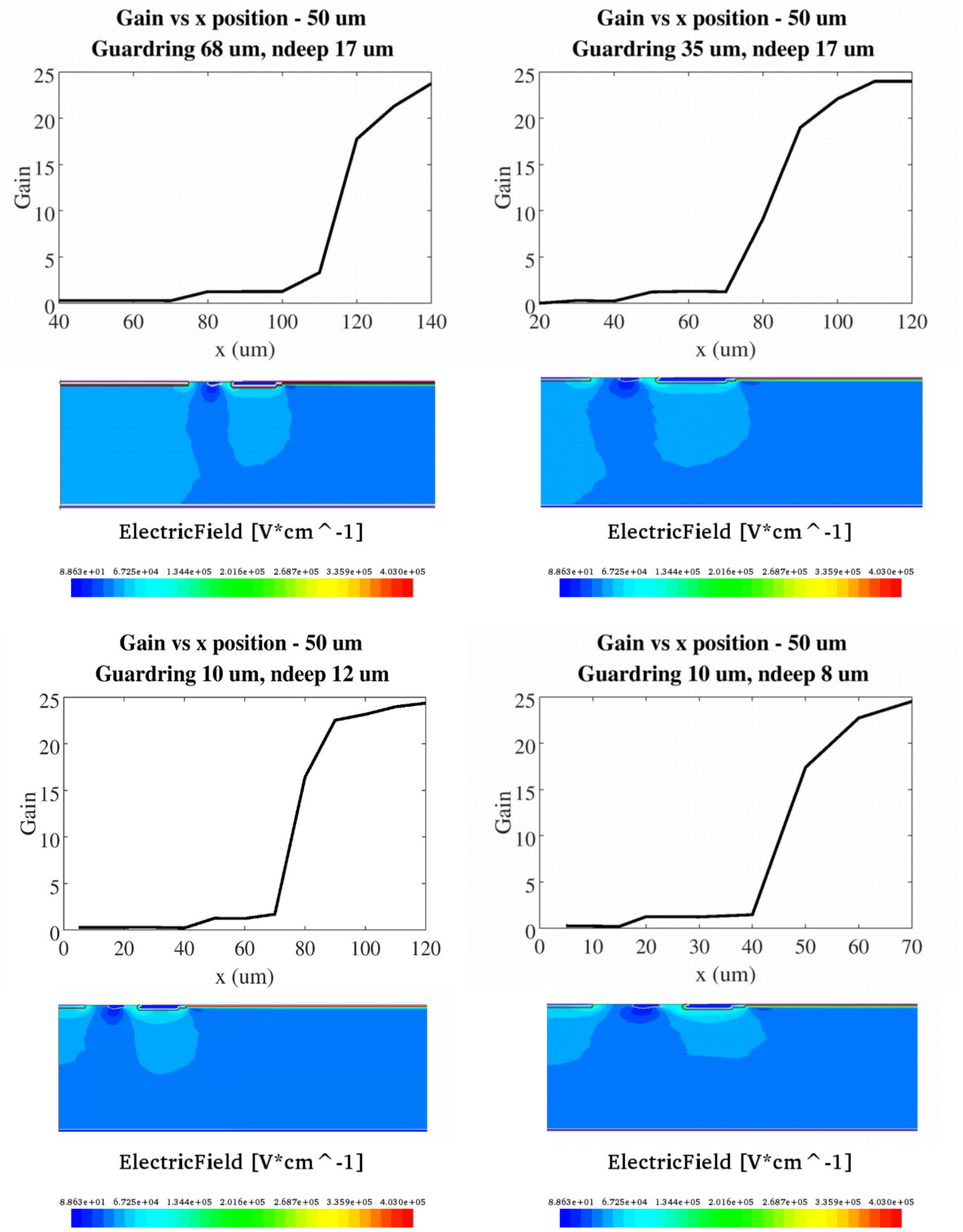


FIGURE 4.24: Gain as a function of the x coordinate for different guardring and electrode termination widths.

These results suggest that as-implanted boron can be implemented only at lower doses, while higher dose splits require diffusion. In practice, all dopants need to be activated with thermal diffusion: the “as-implanted”case refers to a lower-bound diffusion temperature, while the “diffused”corresponds to the higher diffusion temperature.

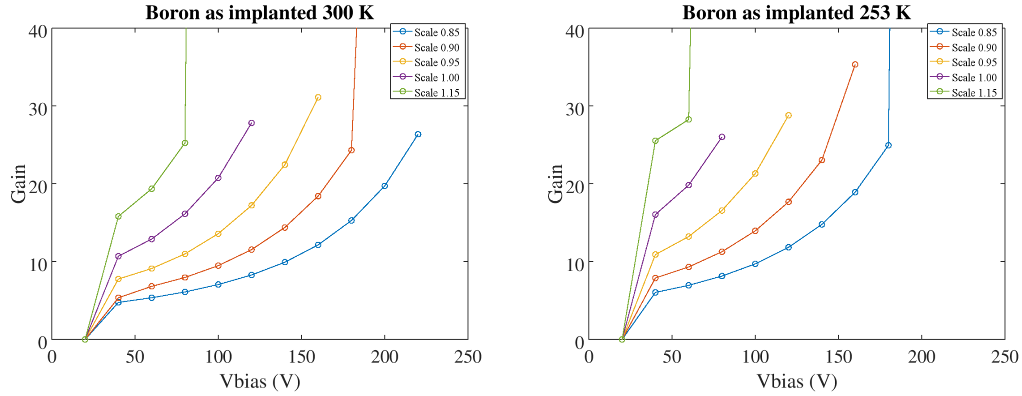


FIGURE 4.25: Gain curve for “as implanted” boron doped gain layer at different doping scales at 300 K (left) and 253 K (right).

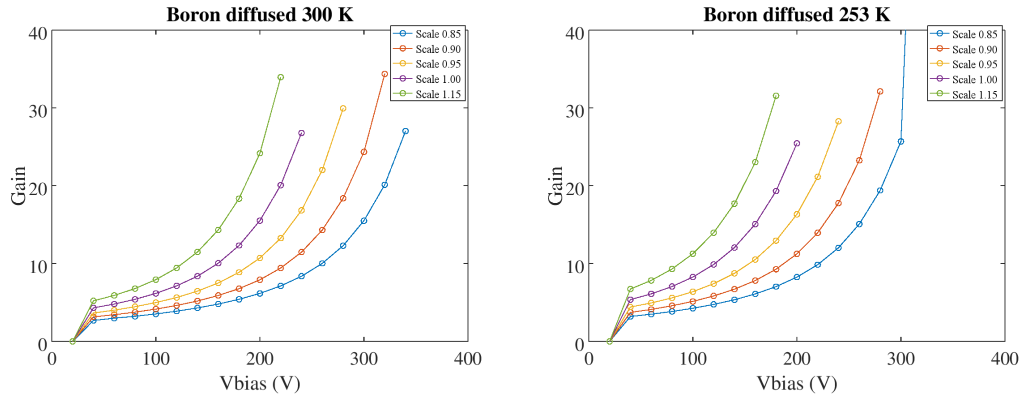
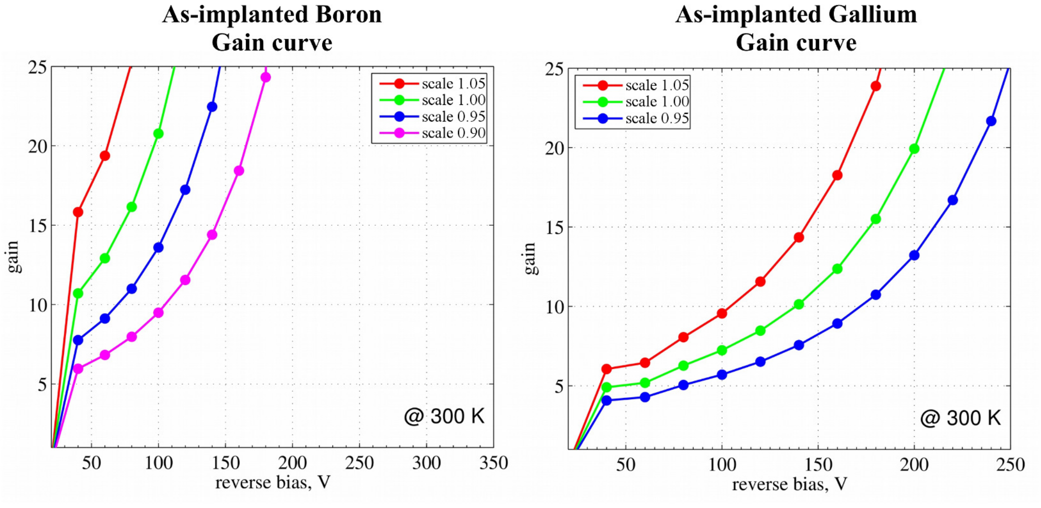


FIGURE 4.26: Gain curve for “diffused” Boron doped gain layer at different doping scales at 300 K (left) and 253 K (right).



(a) (b)

FIGURE 4.27: Gain curve for as-implanted boron (left) and as-implanted gallium (right) at an operating temperature of 300 K.

TABLE 4.2: Dose and dopant type splits for UFSD2 production.

Wafer n.	Dopant	Gain dose	Carbon	Diffusion
1	Boron	0.85	-	Low
2	Boron	0.9	-	Low
3	Boron	0.9	-	High
4	Boron (carbonated)	0.9	Low	High
5	Boron (carbonated)	0.9	High	High
6	Boron (carbonated)	0.95	Low	High
7	Boron (carbonated)	0.95	High	High
8	Boron	0.95	-	High
9	Boron	0.95	-	High
10	Boron	1.0	-	High
11, 12	Gallium	0.9	-	Low
13, 14	Gallium	1.0	-	Low
15	Gallium (carbonated)	1.0	Low	Low
16	Gallium (carbonated)	1.1	High	Low
18, 19	Gallium	1.1	High	Low

4.4.4 Production summary

The production has been carried out with four different gain layer configurations: boron, gallium, carbonated boron and carbonated gallium. Each configuration has different dose splits. The option of a carbonated gain layer dopant has been pursued to investigate its radiation hardness. Two carbon concentrations, "high"and "low" (with a factor 10 difference) have been realized.
Moreover, two diffusion temperatures named "high"and "low"have been considered. Given the different behavior of gain as a function of bias voltage for boron and gallium, the wafers with a gallium-doped multiplication layer have been processed with low diffusion temperature, since a high diffusion temperature further decreases the steepness of the gain curve. In the case of boron gain layer, low temperature has been applied only to the less doped wafers, while higher dose splits required high diffusion temperature.
The gain layer splits that have been implemented on the produced wafers are listed in Table 4.2.

4.4.5 Measurements on UFSD2

The UFSD2 FBK production was completed in June 2016 and has been made available for testing. The doping profiles have been studied by means of current $vs.$ voltage characteristics: measures performed on different wafers led to the results shown in Fig. 4.29. Measurements refer to all four main combinations of boron/gallium with and without carbon addition: Wafer 1 (boron, low doping), Wafer 6 (carbonated boron, low C concentration), Wafer 8 (boron, medium-high doping), Wafer 14 (gallium, medium doping), Wafer 15 (carbonated gallium, low C concentration) and Wafer 18 (gallium, high doping). Fig. 4.30.a shows the gain curves simulated with the doping profiles extracted from Monte Carlo simulations (solid line) are compared with measurements (squares), while Fig. 4.30.b shows the doping profiles extracted from measurements are used as input parameters for the simulator and the

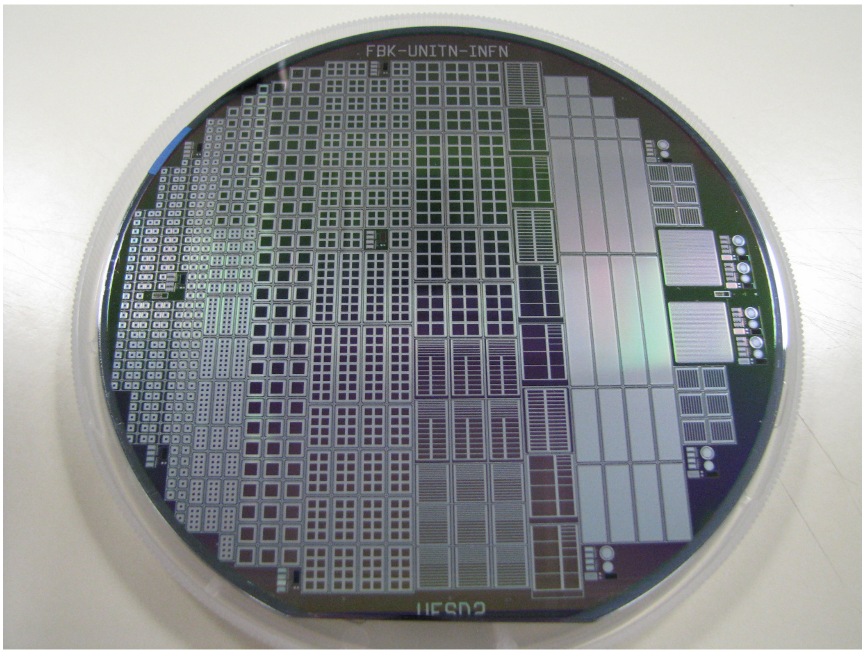


FIGURE 4.28: UFSD2 wafer.

corresponding gain (solid line) is plotted alongside the gain evaluated by CV measurements (squares). From measurements and simulations there emerged that the effective gain achieved by boron is lower than the initially predicted one, while the actual gallium implant is deeper than the predicted one and thus exhibits a higher gain than expected.

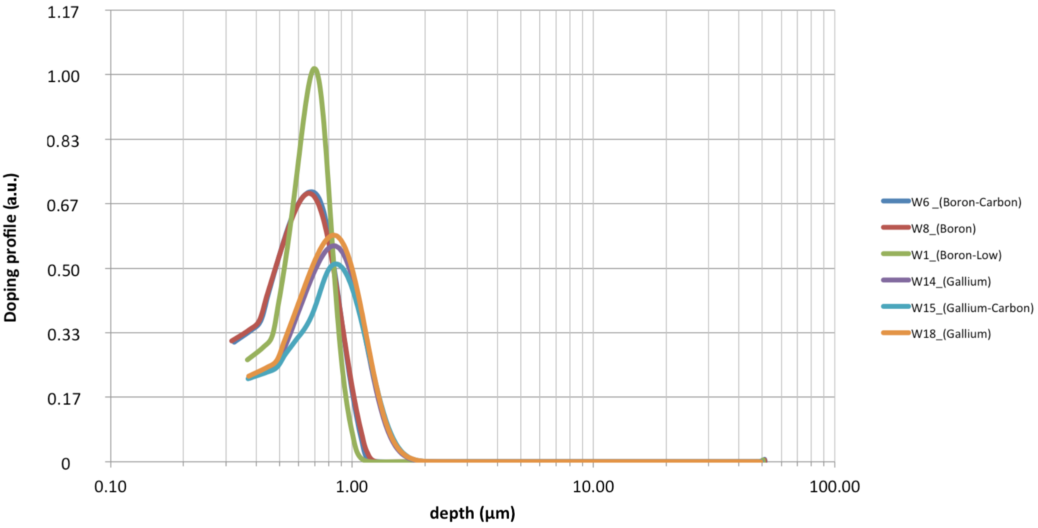
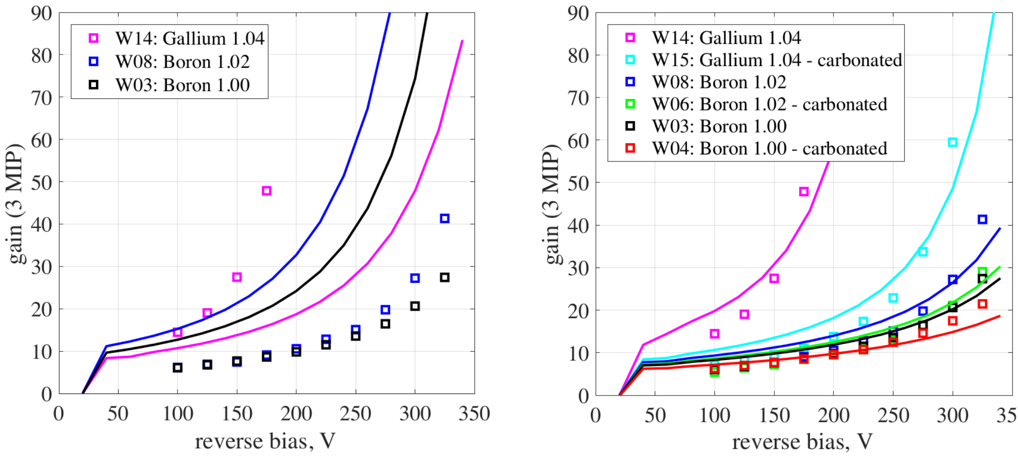


FIGURE 4.29: Doping profile extracted by C-V measurements for different wafers from the UFSD2 production.

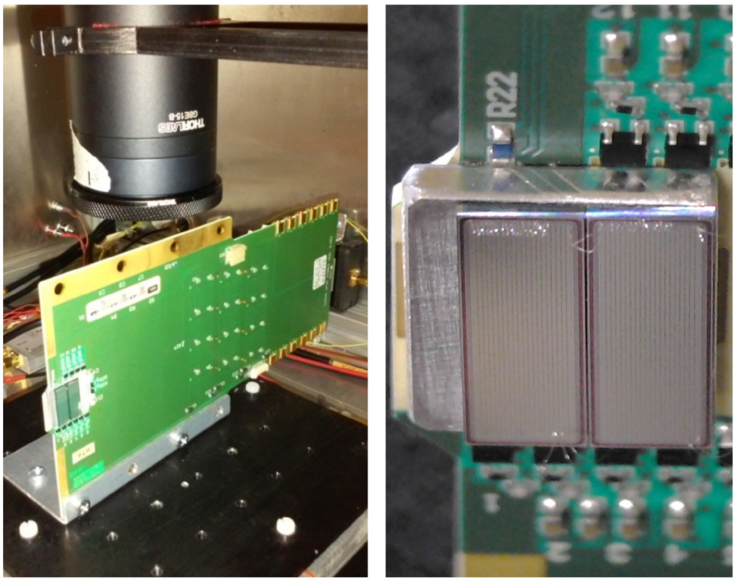
The dead area between adjacent electrodes has been characterized for strip sensors from UFSD2 production [75]. The sensor strips have a width of 150 µm and a pitch of 220 µm. Tests have been performed by means of the "Edge-Transient Charge Technique"(Edge-TCT), consisting in the injection of a focused laser beam on the edge of the sensor. The laser has a wavelength of 1060 nm and a spot diameter of 12 µm. A clear distinction of the two neighboring strips has been obtained, as shown in



(a) (b)

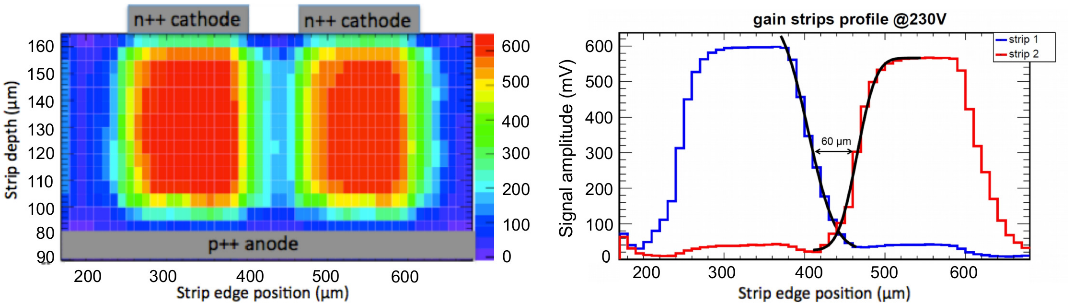
FIGURE 4.30: a) Gain *vs.* bias voltage curves simulated using the doping profiles obtained with Monte Carlo simulations (solid line) compared with measurements (squares). b) Gain as a function of bias voltage simulated using the doping profile extracted from CV measurements (solid line) compared with measurements (squares) [74].

Fig. 4.32.a. The projection of the signal amplitude between the two strips (Fig. 4.32.b) demonstrates that the dead area at half height of the amplitude distribution is ~ 60 µm and compatible with the sensor layout.



(a) (b)

FIGURE 4.31: a) Edge-TCT setup for strip characterization. b) Strip sensor under test.



(a) (b)

FIGURE 4.32: a) Strip edge scan acquired with TCT setup: the strip edge position is plotted along the horizontal axis, the strip depth along the vertical axis, while the color scale represents the signal magnitude. b) Signal amplitude as a function of the edge position between two neighboring strips [75].

Chapter 5

TOFFEE

TOFFEE (Time Of Flight Front-End Electronics) is an 8 channels front end ASIC for UFSD read out, designed in standard 110 nm CMOS technology [76]. Each channel is independent and the signal processing chain, depicted in Fig. 5.1, is composed by:

- transimpedance amplifier
- single threshold discriminator
- stretcher
- LVDS driver.

The time to digital conversion is performed by the HPTDC, a general purpose data driven multi-hit time-to-digital converter described in Sec. 5.6. The dead time between consecutive hits is ~ 5 ns, so that a minimum pulse width is required in order to enable the measurement of the leading and trailing edges of the pulse. A stretcher is used to widen the discriminator output in order to produce a pulse always wider than 5 ns, as required by HPTDC. TOT technique is used for time walk correction. The total power budget per channel is less than 30 mW, constrained by thermal dissipation in the RPs used for the CT-PPS detector. The output format matches HPTDC inputs through LVDS links (32 pairs per detector module), which consume 10 mW of the power budget per channel. The CT-PPS UFSD detector module produced by CNM, shown in Fig. 5.2, is made of two arrays of 16 pixels (or "fat strips") each. The sensor has an overall area of 12×6 mm^2, a thickness of 50 µm and a nominal gain of ~ 15 at a bias voltage of 200 V. The design of this device is asymmetric: thicker pixels (shown in the left side of the picture) will be located away from the beam spot, while thinner pixels (right side) will provide a higher segmentation in the region closer to

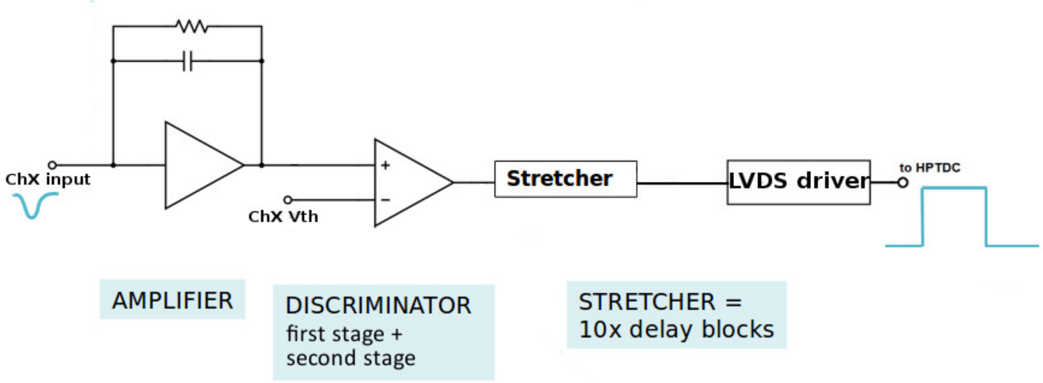


FIGURE 5.1: TOFFEE signal processing chain

the beam. The pixel length on the upper side of the figure is 3.1 mm and 2.9 mm on the lower side. The 16 thicker pixels have an area of ~ 3 mm^2 (1 mm of width) for a capacitance of ~ 6 pF. The 16 thinner pixels have an area of ~ 2 mm^2 (0.5 mm of width) and a capacitance of ~ 4 pF. Moreover, the bottom-left pixel is 0.95 mm wide and the top-right one is 0.45 mm wide. Then sensor features an inactive gap between each pixel of 50 µm and a slim edge of ~ 200 µm on the beam side.

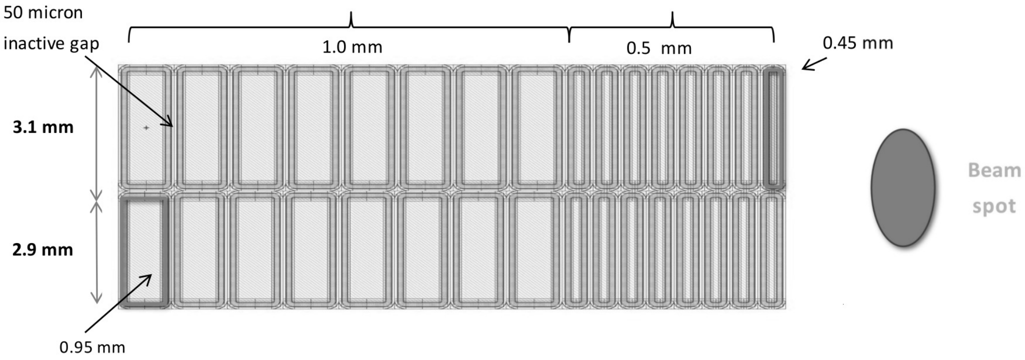


FIGURE 5.2: CT-PPS UFSD sensor

Both sensor and readout electronics will be placed on the same board: each pixel will be wire-bonded to one channel of the readout chip and therefore it is of crucial relevance to minimize the bond length. To achieve this, a 8-channel ASIC has been purposed. The whole system of 32 pixels will be thus read out by 4 ASICs, as shown in Fig. 5.3.

The power consumption of the whole 32-channel detector plane is ~ 1.3 W. The stack of four detector planes into the roman pot consumes ~ 5.2 W, which is sensibly below the 50 W power constraint of the roman pots.

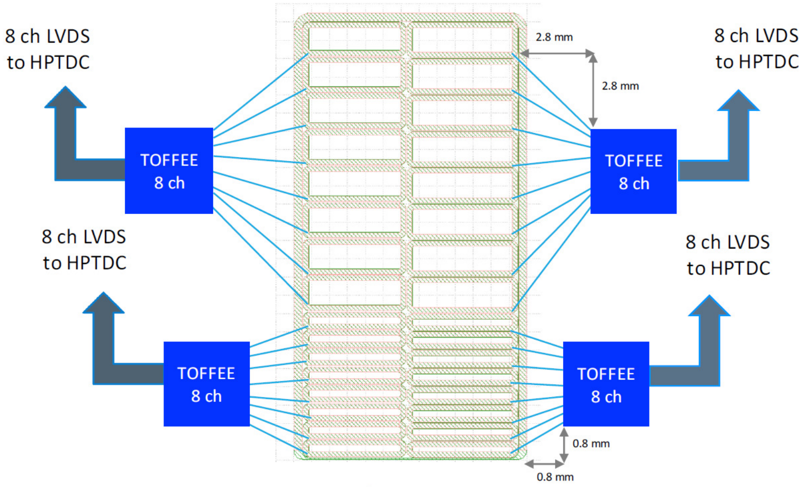


FIGURE 5.3: CT-PPS UFSD sensor with 32 strips read out by four TOFFEE ASICs

5.0.1 Input signal

The sensor signal and the input charge distribution have been simulated with *Weightfield 2*. The particular signal shape and components are described in Sec. 4.1.1 and shown in Fig. 4.4. The typical UFSD signal for a 50 µm device has a total length of ~ 1.2 ns. Fig. 5.4 shows the sensor signals for different charge values in the relevant range of the Landau distribution: the black line refers to the charge produced inside the sensor by a MIP, corresponding to a total charge of 8 fC delivered by the sensor. The average UFSD signal can be approximated by a trapezoidal signal with a rise time and a fall time of 450 ps, a minor base b of 300 ps and an area equal to the input charge Q_{in}. The major base B is then 900 ps and the height is given by

$$h = \frac{2Q_{in}}{B+b} \simeq 1.33 \cdot 10^9 [s^{-1}] \cdot Q_{in}, \tag{5.1}$$

leading to a peak current of $\sim$ 10 µA for a 8 fC signal.
Fig. 5.5 shows the charge distribution for a sensor with gain 15 fitted with a Landau distribution. The shape of the Landau distribution is strongly asymmetric and its long tail at higher energies shows that a very wide range of charge values are allowed but have a low probability.

5.1 Amplifier

In order to determine the specifications for the amplifier in terms of gain and bandwidth, the circuit in Fig. 5.6.a is considered. The input impedance Z_{in} of this amplifier can be evaluated to obtain the impulse response of the circuit with the addition of an input capacitance (Fig. 5.6.b).

The transfer function of the amplifier is modeled with a first order low-pass filter with a pole in ω_0:

$$A(\jmath\omega) = -\frac{A_0}{1+\frac{\jmath\omega}{\omega_0}}, \tag{5.2}$$

where $-A_0$ is the DC gain. The impulse response is then expressed as

$$V_{out,0} = A(\jmath\omega)V_{in,0} = A(\jmath\omega)\left(I_0 \frac{R_f}{1+\jmath\omega R_f C_f} + A(\jmath\omega)V_{in,0}\right). \tag{5.3}$$

By solving this equation with respect to V_{in}, the input impedance is obtained as

$$Z_{in}(\jmath\omega) = \frac{V_{in,0}}{I_0} = \frac{R_f}{1+A_0}\frac{1}{1+\jmath\omega R_f C_f}\frac{1+\jmath\omega/\omega_0}{1+\frac{\jmath\omega/\omega_0}{1+A_0}}. \tag{5.4}$$

In the hypothesis of $A_0 \gg 1$, the term $1 + A_0$ can be approximated with A_0. With an input capacitance C_{det}, the input impedance becomes

$$\frac{V_{in}}{I_{det}} = \frac{1}{\jmath\omega C_{det}} /\!/ Z_{in} = \frac{Z_{in}}{1+\jmath\omega C_{det} Z_{in}} \tag{5.5}$$

and therefore, considering I_{det} a δ-like signal, the impulse response is

$$V_{out}(\jmath\omega) = -I_{det}\frac{R_f}{1+\jmath\omega\left(\frac{1}{\omega_0 A_0}+\frac{C_{det}R_f}{A_0}+R_f C_f\right) - \omega^2\left[\frac{1}{A_0\omega_0}(C_{det}R_f + R_f C_f)\right]}. \tag{5.6}$$

The detector can be modeled as a current source in parallel to a capacitance C_{det}, which is $\sim$ 6 pF for larger strips and $\sim$ 4 pF for thinner strips. Recalling that the

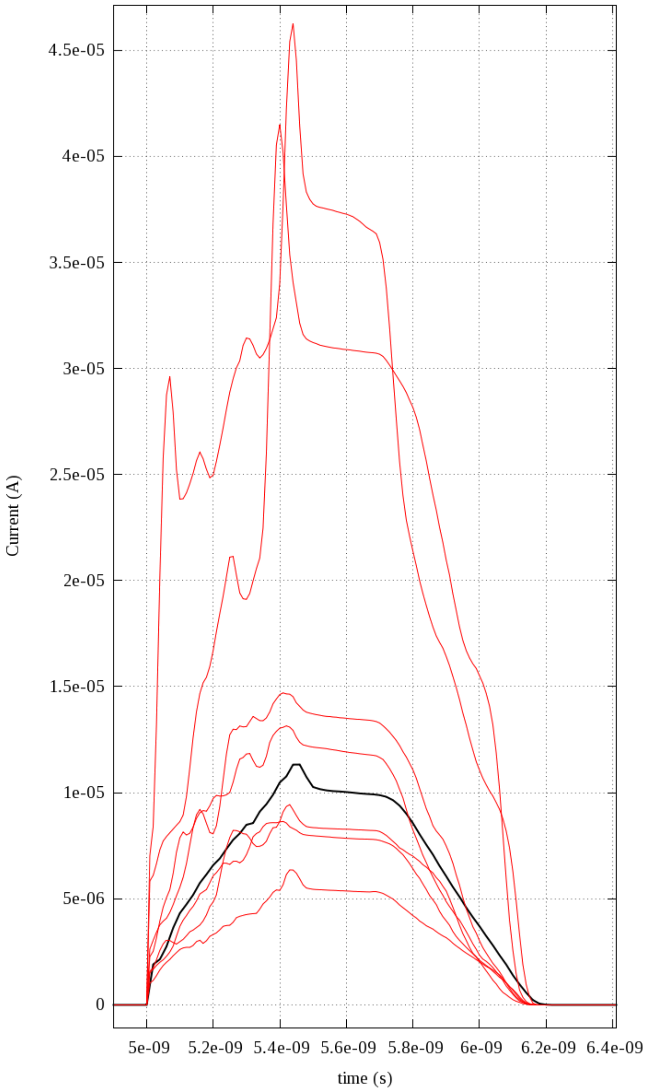


FIGURE 5.4: Sensor signal for different charge values extracted from Landau distribution. The sensor has a nominal gain of 15. The black line corresponds to the signal produced by a minimum ionizing particle

collection time and the peak current for a MIP signal are 1.2 ns and 10 µA, respectively, a transimpedance of 20 kΩ is a reasonable value for the amplifier. This transimpedance will lead to a voltage amplitude of about 200 mV at the amplifier output for the average signal, and up to 800 mV for the maximum signal. Due to the ballistic deficit, the effective voltage amplitude will be lower. In the final design the feedback resistance, and hence the transimpedance value at middle frequency, have been set to 24 kΩ.

In timing applications, it is crucial for the detector to deliver the overall charge of the signal to the amplifier in the shortest time possible. In this way, the signal amplitude at the output of the amplifier will be maximum. Considering the realistic case in which

$$I_{det}(\jmath\omega) = \frac{I_{det}}{1 + \jmath\omega t_c} \tag{5.7}$$

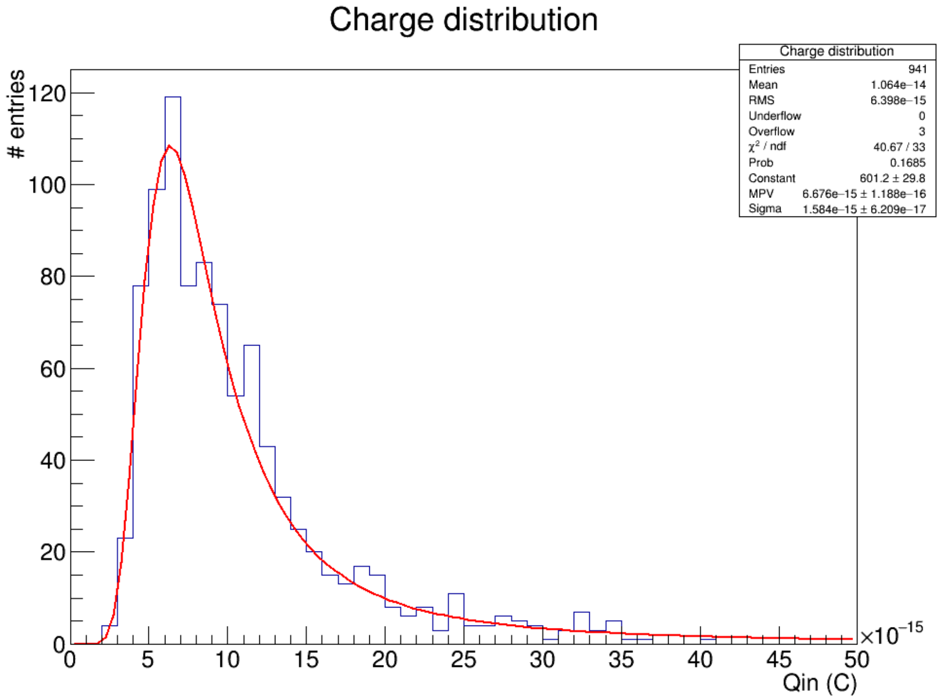


FIGURE 5.5: Simulated input charge fitted with Landau distribution

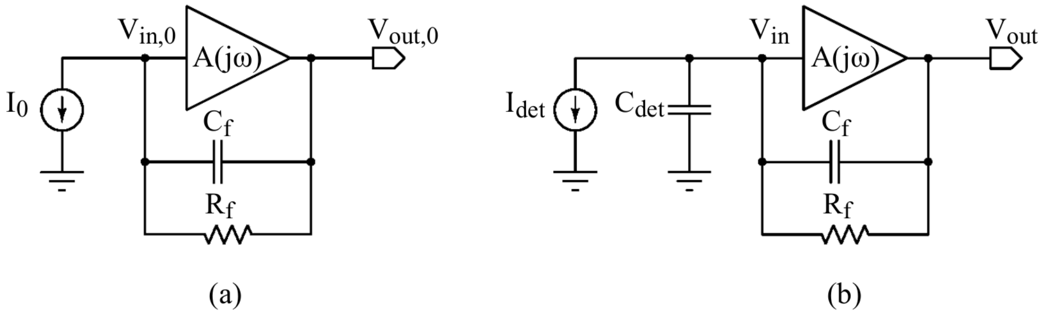


FIGURE 5.6: Amplifier model for calculation of a) input impedance and b) impulse response with an input capacitance.

is an exponential function, the time required to deliver the total charge to the amplifier is constrained by two factors: the characteristic collection time of the sensor t_c and the time constant of the low-pass filter formed by the detector capacitance and the amplifier input impedance, according to

$$I_{in} = \frac{V_{in}}{Z_{in}} = \frac{I_{det}(\jmath\omega)}{1 + j\omega C_{det} Z_{in}}, \tag{5.8}$$

which represents the current entering the amplifier. Considering Z_{in} purely real, the $\frac{1}{C_{det} Z_{in}}$ pole should be higher than the first pole of $I_{det}(\jmath\omega)$. In the specific case, the characteristic time of the low-pass filter has to be lower than the collection time, i.e.

$$C_{det}\frac{R_f}{A_0} + C_f R_f \leq t_c. \tag{5.9}$$

This sets a lower limit to the amplifier gain:

$$A_0 > \frac{R_f C_{det}}{t_c} = 120. \tag{5.10}$$

The effective gain value in the final circuit was set to 200. In this way, the low pass filter constant is one half of the collection time. Hence the feedback capacitance C_f is set to

$$C_f \leq \frac{t_c - C_{det} R_f / A_0}{R_f}. \tag{5.11}$$

From this equation follows that the optimum value for C_f is between 20 fF and 30 fF for a detector capacitance C_{det} between 4 pF and 6 pF. Simulations of the amplifier stability showed that a 80 degree phase margin is achieved with a value of 55 fF (Fig. 5.7), which was adopted in the final design. The effect of the feedback capacitance on the amplifier output can be visualized in Fig. 5.8, where the amplifier signal has been evaluated at different feedback capacitance values: 0, 20 fF and 55 fF. It can be seen that for 55 fF no undershoot and no oscillations are present. Finally, in order to keep the rise time of the signal at the maximum value achievable, the gain-bandwidth product of the amplifier $A_0\omega_0$ was maximized.

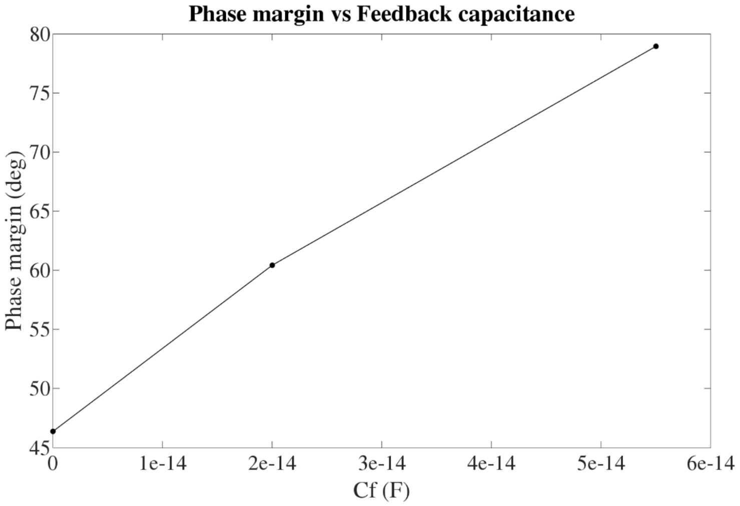


FIGURE 5.7: The amplifier stability as a function of the feedback capacitance C_f: the phase margin assumes an optimal value with C_f= 55 fC.

5.1.1 Amplifier architecture

The first amplification stage is based on a common source amplifier with split current source (Fig. 5.9). In order to improve the middle frequency gain, both the common source amplifier $M1$ and the current source $M4$ are cascoded with the transistors $M2$ and $M3$, respectively. This increases the output resistance of the transistors by the self gain of the cascoding transistor, which his given by the product of the transconductance and the drain-source resistance $g_m r_{ds}$. The bias current is provided by two independent branches: the left branch is powering the cascode transistors $M3$ and $M2$ with a current properly chosen to meet the desired specifications of output swing and slew rate. Taking into account that by reducing this current the output swing increases but the slew rate decreases, appears clear that a trade off is

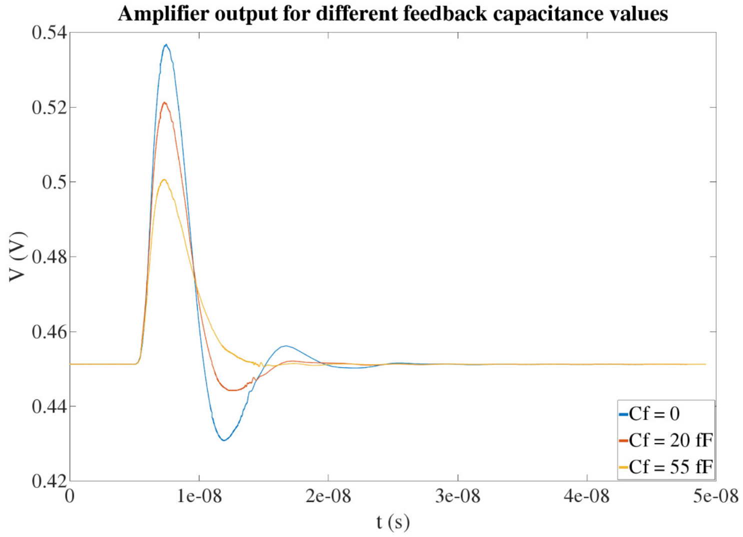


FIGURE 5.8: Simulation of the amplifier output for different feedback capacitance values.

TABLE 5.1: Characteristic values of amplifier transistors at operating point.

Transistor	$\mathbf{g_m}$ **(S)**	$\mathbf{r_{ds}}(\Omega)$	$\mathbf{I_{ds}}$ **(A)**
$M1$	165.6 m	51.6	12.4 m
$M2$	5.2 m	5 k	227 µ
$M3$	4.0 m	11.1 k	227 µ
$M4$	3.4 m	67.6 k	227 µ
$M5$	83.2 m	479	12.2 m
$M6$	119.5 m	69.18	12.2 m

necessary. To further improve the gain of the amplifier, a split current source is used ($M6$), providing additional current to $M1$ and thus fostering its transconductance. This current source is further cascoded with $M5$. Source degeneration resistors R_1 and R_2 are applied on the current sources $M4$ and $M6$ for noise reduction. These resistors have an effect on the total resistance of node **1**, whose two contributors are:

$$\begin{aligned} &g_{m3}r_{ds3} \cdot g_{m4}r_{ds4} \cdot R_1, \\ &g_{m2}r_{ds2} \cdot (r_{ds1} /\!/ g_{m5}r_{ds5} \cdot g_{m6}r_{ds6} \cdot R_2). \end{aligned} \tag{5.12}$$

The values of transconductance g_m, drain-source resistance r_{ds}, unity gain frequency f_{ug} and drain-source current I_{ds} for each device of the amplifier at the operating point are listed in Table 5.1: as shown, the current flowing in the right branch ($M5, M6$) determines most of the power consumption of the amplifier.

The middle-frequency gain of the telescopic cascode is thus given by the formula:

$$A_0 = -g_{m1} \cdot [g_{m2}r_{ds2}(r_{ds1} /\!/ g_{m5}r_{ds5}g_{m6}r_{ds6}R_2) /\!/ g_{m3}r_{ds3}g_{m4}r_{ds4}R_1] \simeq 213. \tag{5.13}$$

In first approximation, the circuit in Fig. 5.9 presents a pole at each node with at least one transistor. In the hypothesis of separated poles, the one with lowest frequency determines the ω_0 discussed in Eq. 5.2. From AC simulations, this pole is located at node **2**, between $M4$ and the resistor R_1 and its value is $\omega_0 = 3.4$ Grad/s, corresponding to a frequency of 540 MHz.

A buffer is used to decouple the output impedance of the telescopic cascode from the feedback resistance and the detector capacitance.
The value of the current in transistor $M1$, its transconductance g_{m1} and dimensions have been chosen to reach a high gain and a unity-gain-frequency of 14 GHz, to cope with the short collection time of the sensor. The power consumption is 14.4 mW, within the specification requirements of less than 20 mW [76].

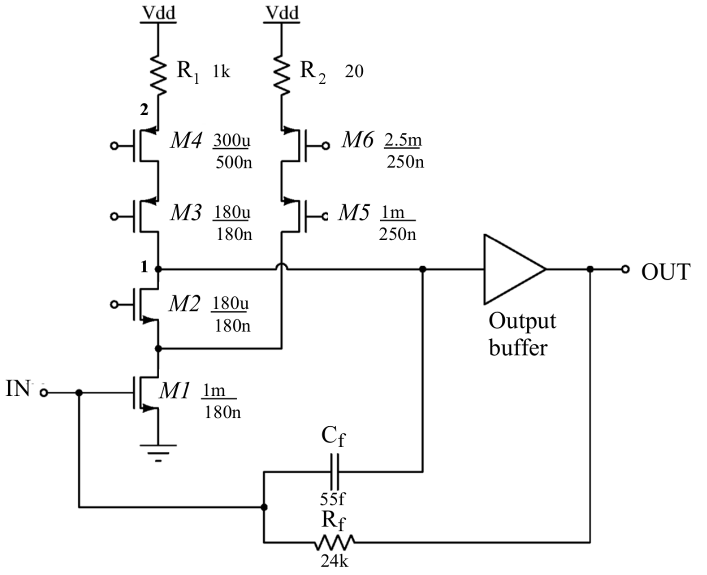


FIGURE 5.9: Amplifier architecture

5.1.1.1 Buffer

The decoupling buffer is a unitary gain amplifier with low output impedance realized with a source follower with active feedback. Its purpose is to increase the bandwidth of the transimpedance gain of the amplifier when connected to a capacitive detector. The effect of the buffer is shown in Fig. 5.11, where the transimpedance gain for different values of the detector capacitance is plotted with and without the buffer.

5.2 Discriminator

To achieve the highest speed, the discriminator is implemented with a cascade of two stages (Fig. 5.12). The total gain of the comparator is given by the product of the gains of the cascaded stages [55]. The first is a differential pair which further increases the gain of the amplifier, while the second is a single threshold discriminator. In the gain cell of Fig. 5.12.a the input transistors $M1+$ and $M1-$ are cascoded to reduce the Miller effect on their gate-drain capacitance. Two diode connected transistors $MS+$ and $MS-$ are added between the two arms of the differential pair to limit the maximum voltage swing, thus having a faster recovery when large input signals are applied. The load is implemented with two passive resistors, having a lower parasitic capacitance than active loads. The differential to single-ended conversion is performed with the stage depicted in Fig. 5.12.b.

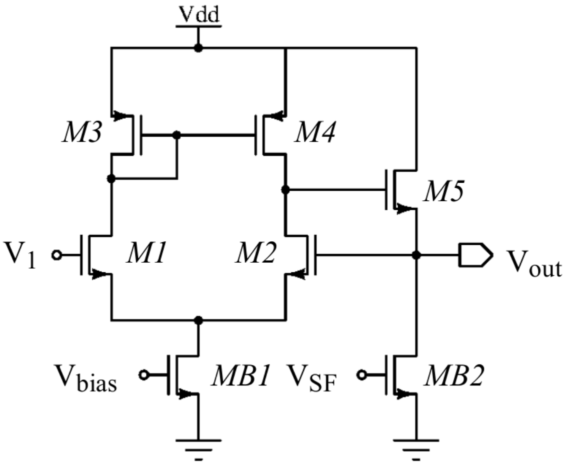


FIGURE 5.10: Amplifier output buffer

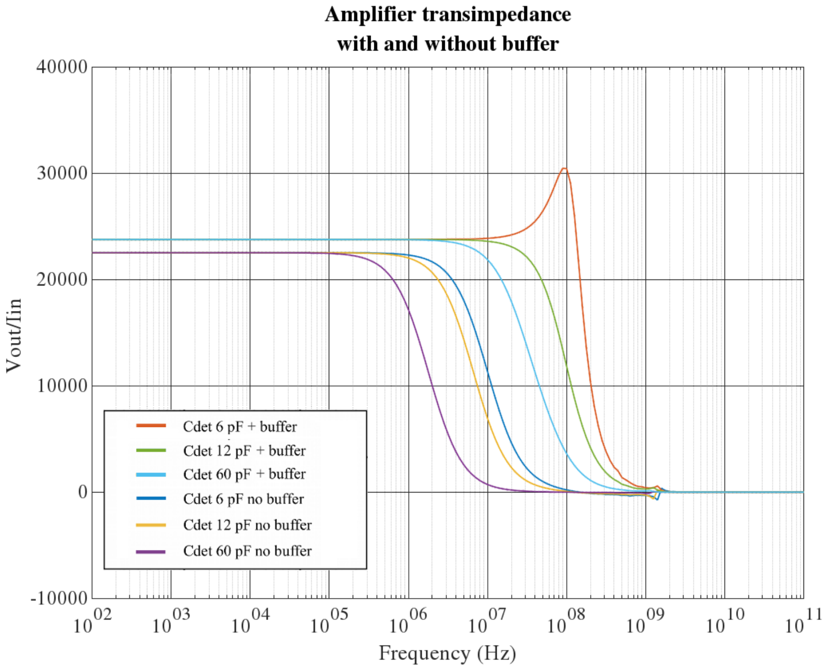


FIGURE 5.11: Amplifier transimpedance with an without output buffer for different detector capacitance values.

5.3 Stretcher

A delay line is used to add an offset to the discriminator output width, in order to match the minimum signal width required by HPTDC to detect both edges. A similar concept was already used in the NINO chip [31]. It is composed of ten consecutive digital delay blocks (Fig. 5.13), each formed by an inverter and two starving transistors: the amount of delay added to the signal is tuned by the gate voltage applied to them.

Indeed, the delay of an inverter can be controlled by adding two extra transistors as shown in Fig. 5.14 with a technique known as *current starving*: lowering V_n and increasing V_p increases the effective resistance of the inverter and thus the delay. The

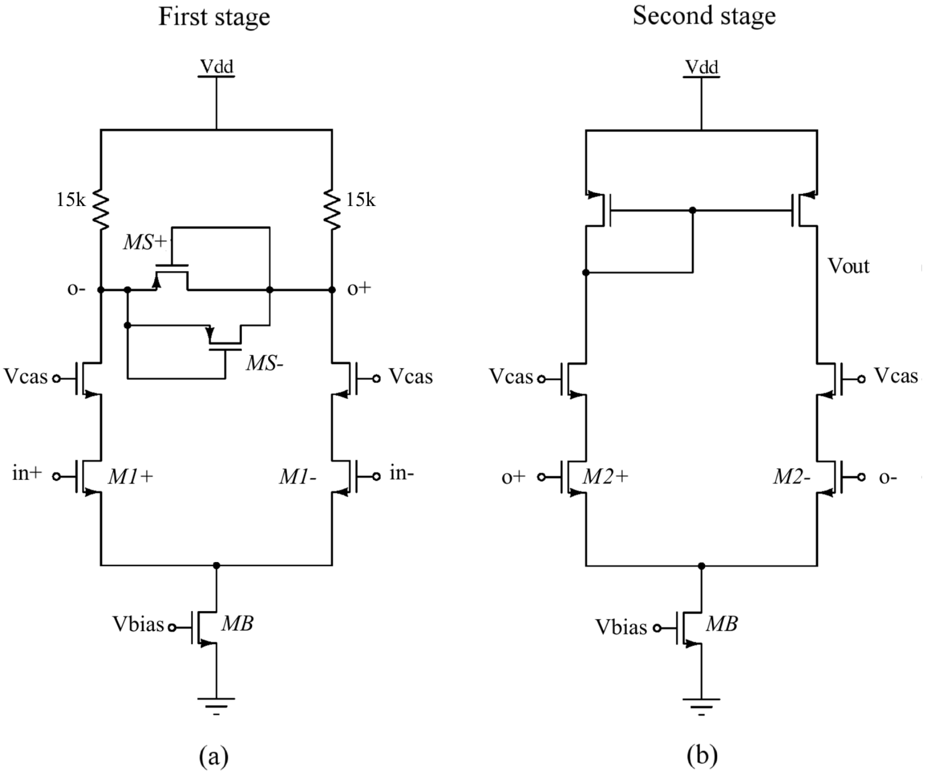


FIGURE 5.12: Discriminator

control voltages V_n and V_p are designed to be set from an external bias, allowing to remove differences in the resistance between nMOS and pMOS transistors due to process-related variations.

In simulations, the width of the offset added by the stretcher is set by varying the voltages applied to the inputs *Delay P* and *Delay N* of Fig. 5.13. An offset of $\sim$ 5 ns is added by putting 0.7 V to *Delay P* and 0.5 V to *Delay N*, while the minimum offset is obtained with *Delay P* set to 0 and *Delay N* set to 1.2 V: in this way, both the PMOS and NMOS starving transistors are always on and the delay added by them is minimum.

5.4 Schematic simulations

5.4.1 Small signal analysis

Through small signal analysis it is possible to verify most of the features of the amplifier as gain, input impedance and transimpedance. The transfer function of the amplifier $A_v = V_{out}/V_{in}$ has been evaluated as a function of the frequency (Fig. 5.15). The amplifier voltage gain is 218, corresponding to $\simeq$ 47 dB. The transfer function allows to derive the amplifier bandwidth (89.15 MHz) and the gain-bandwidth product (19.49 GHz).

The amplifier input impedance $Z_{in} = V_{in}/I_{in}$ as a function of the frequency is shown in Fig. 5.16: at lower frequencies it assumes the value of $\sim$ 109 Ω, increasing up to $\sim$ 120 Ω at a frequency of slightly less than 100 MHz. Recalling the formula of Eq. 5.2, the gain $A(\jmath\omega)$ has a first order pole in ω_0, which translates into a zero at ω_0 and a pole at $A_0\omega_0$ in the input impedance (Eq. 5.4). The zero at ω_0 explains the

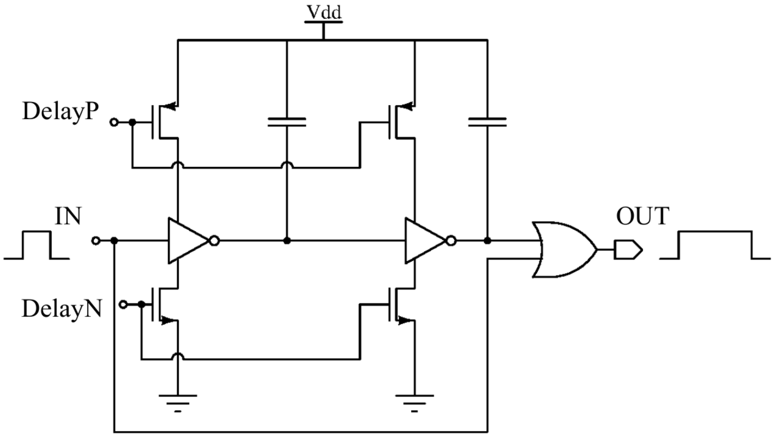


FIGURE 5.13: Stretcher

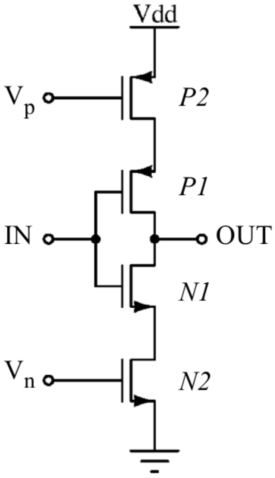


FIGURE 5.14: Inverter with starving transistors controlled by the voltages V_p and V_n

increase of the input impedance at a frequency of about 100 MHz, while its decrease has to be ascribed to the presence of a second pole in $A(\jmath\omega)$ that enters the feedback.

5.4.2 Transient simulations

Two kinds of input signals have been used for transient simulations. On one side, a realistic simulation of the actual signal shape obtained with *Weightfield 2* for a 50 µm thick UFSD sensor with a nominal gain of 15, delivering $\sim$ 8 fC to the amplifier input. Another quick and effective way to simulate the sensor signal for a given charge value is to use the trapezoidal signal explained in Sec. 5.0.1.

The amplifier output as a function of the signal charge is plotted in Fig. 5.17. The input charge ranges from 2 to 60 fC. A charge of 2 fC is at the very lower limit of the Landau distribution. The amplifier is linear up to 60 fC, while starts to saturate

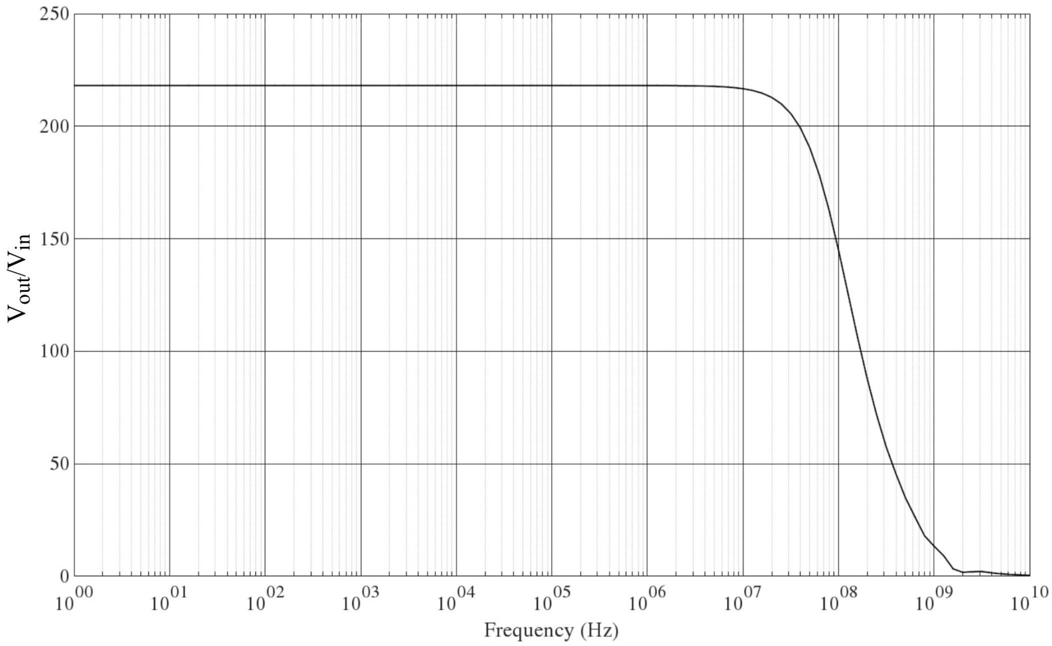


FIGURE 5.15: Amplifier AC gain vs. frequency

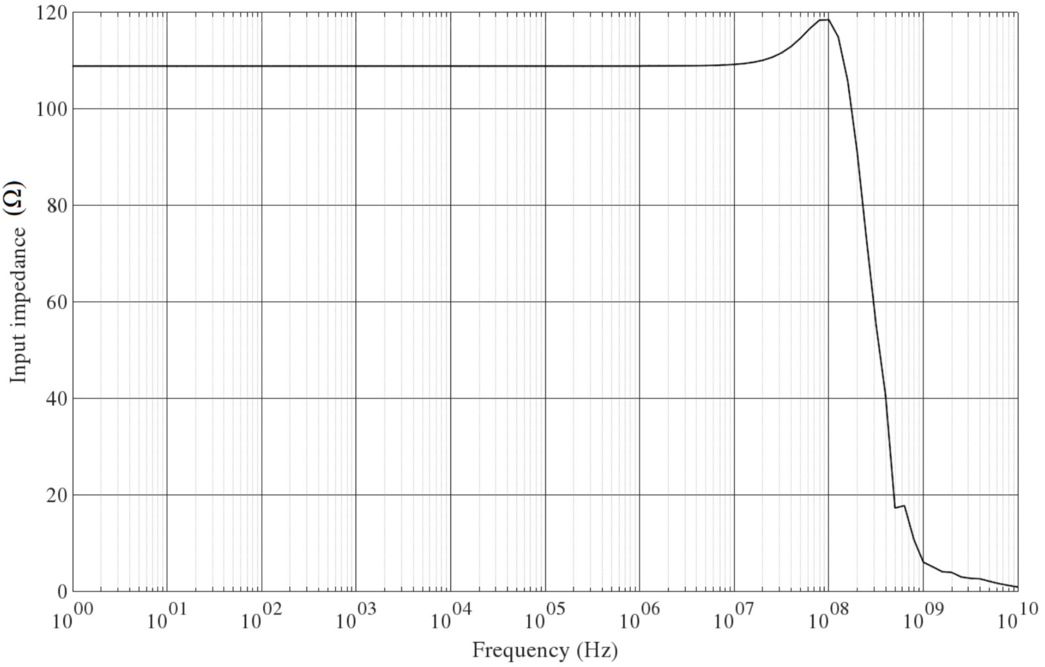


FIGURE 5.16: Amplifier input impedance $vs.$ frequency.

at $\sim$ 70 fC. This value is beyond the charge range predicted by the Landau distribution for a MIP hitting the 50 µm UFSD but can be easily achieved with a laser beam hitting the sensor or with a sensor delivering a higher charge, for example a device with higher gain (APD, SiPM).

The amplifier output in schematic simulations has a peaking time of $\sim$ 2 ns and a total duration of $\sim$ 10 ns. The the time during which the signal is above a fixed

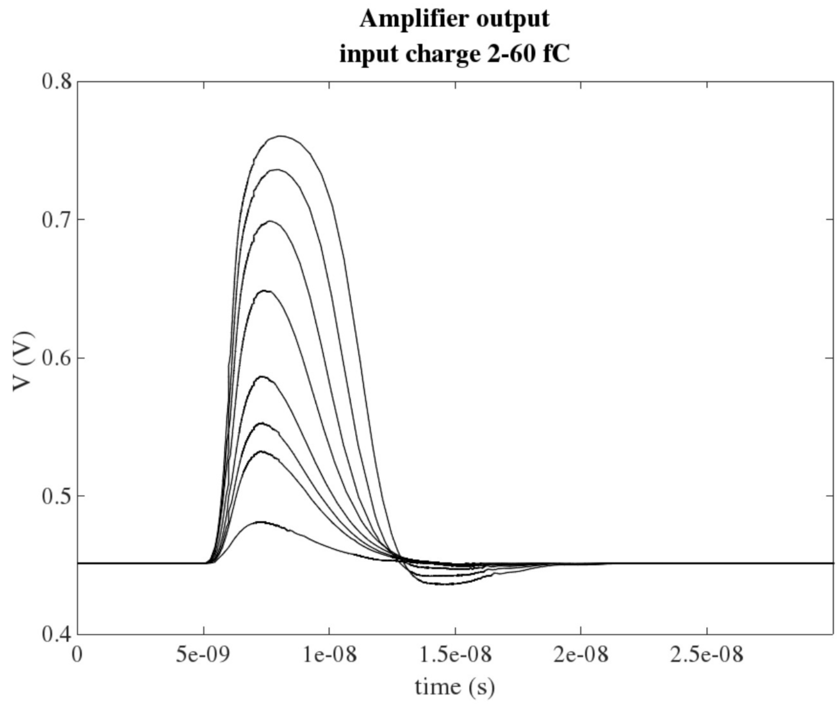


FIGURE 5.17: Shape of amplifier output for an input charge ranging from 2 to 60 fC and a detector capacitance of 6 pF.

threshold, i.e. the so called "Time Over Threshold"(ToT) has a logarithmic dependence on the input charge: before amplifier saturation, the ToT increases with an increase of the input charge, as shown in Fig. 5.18, where the ToT as a function of the input charge is plotted. The ToT is measured at half height of the LVDS output, with the stretcher set to the minimum starving value, which is achieved by applying 0 V to Delay P and 1.2 V to Delay N. The ToT expected for an input charge between 5 and 60 fC ranges from 6.8 ns to 9.2 ns for a threshold of 10 mV and from 4.5 ns to 8.9 ns for a threshold of 20 mV.
For the same charge range, the simulated Time of Arrival as a function of the Time Over Threshold is shown in Fig. 5.19.

5.4.2.1 Behavior for high charges

After amplifier saturation, the logarithmic relation between ToT and input charge no longer holds. In Fig 5.20 the amplifier output in a charge range between 2 and 200 fC is plotted for a threshold voltage of 7 mV. As shown in the picture, the ToT may loss monotonicity at high charge values and low threshold voltages, since the trailing edge of greater signals are crossing the threshold even before smaller ones. This effect can be visualized from Fig. 5.18 and Fig. 5.19. The decrease of the ToT with the input charge can be mitigated by setting a higher discriminator threshold: a threshold voltage of 20 mV ensures the monotonicity of the ToT. It is worth to notice that at the nominal working conditions, i.e. a charge ranging from 3 to 30 fC, the ToT is monotonic also with a threshold of 10 mV.

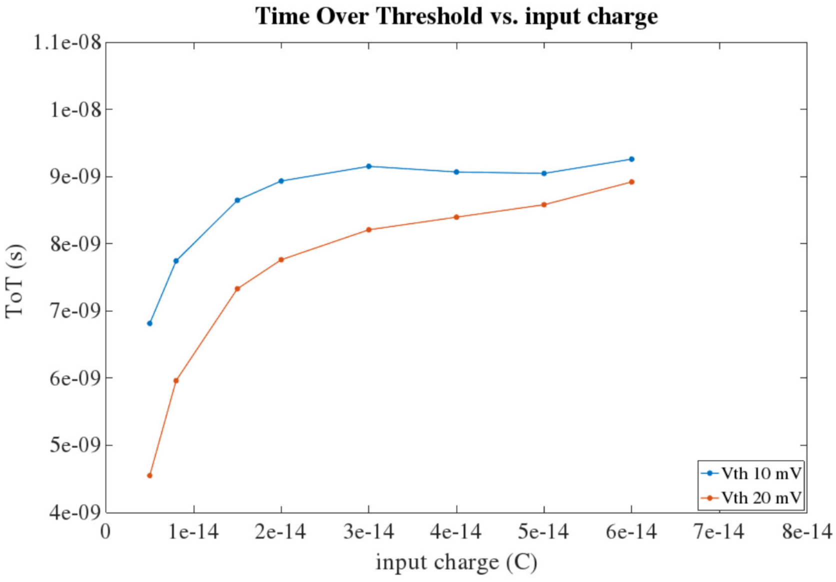


FIGURE 5.18: Time Over Threshold as a function of input charge for a discriminator threshold voltage of 10 mV and 20 mV.

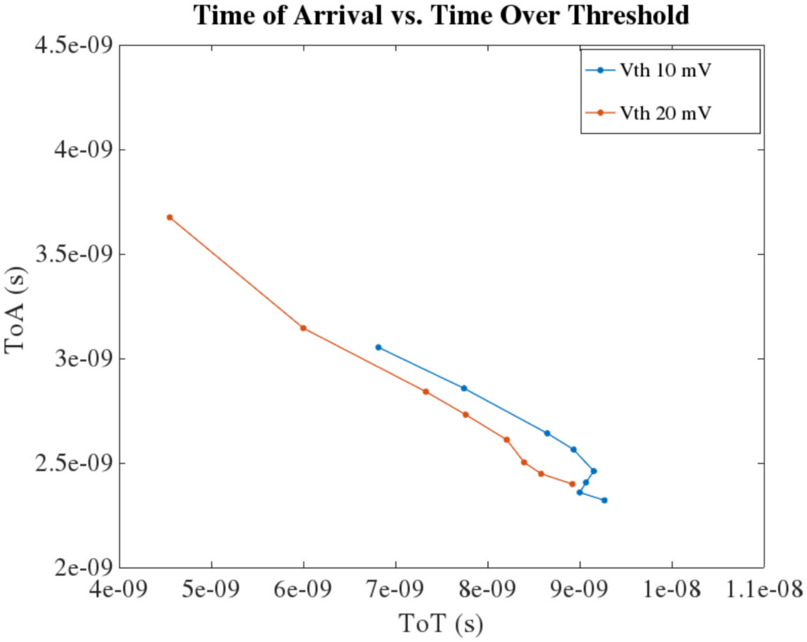


FIGURE 5.19: Time of Arrival as a function of Time Over Threshold for a discriminator threshold voltage of 10 mV and 20 mV for an input charge between 5 and 60 fC.

5.4.3 Noise simulations

The total output noise has been evaluated in a frequency range between 100 Hz and 100 GHz. The resulting value from schematic simulations is 0.6 mV. The effect of noise on the shape of the amplifier output can be visualized by means of transient noise simulations. The signal is distorted by the addition of noise as shown in

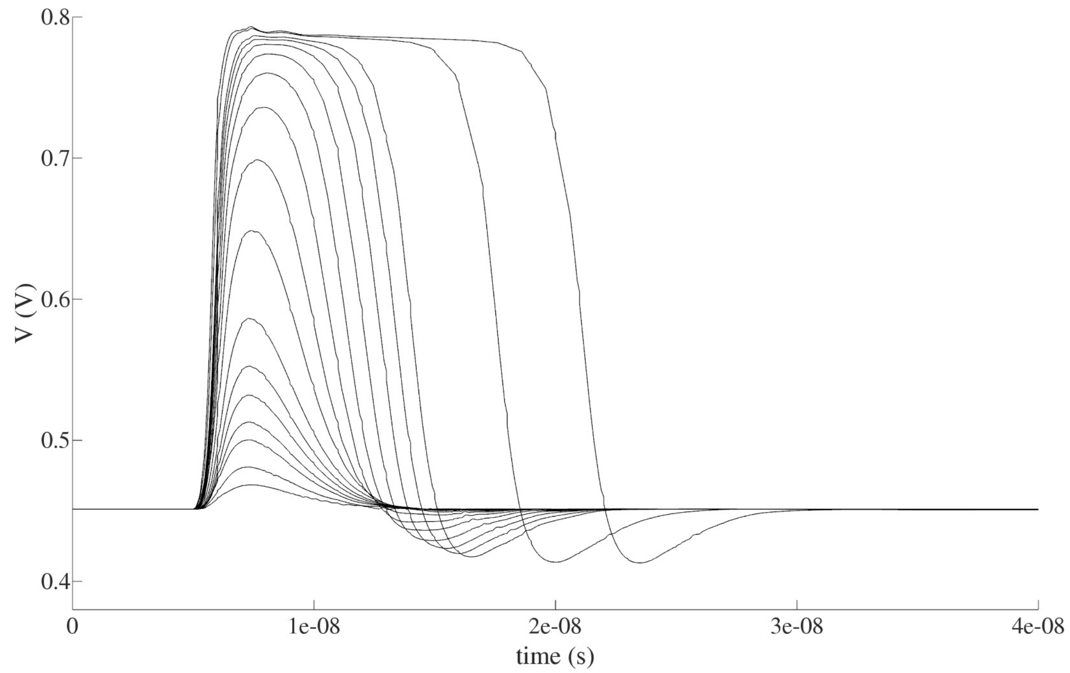


FIGURE 5.20: Amplifier output for different input charge values in the range between 2 and 200 fC and a detector capacitance of 6 pF.

TABLE 5.2: Noise contribution of the transistors of the telescopic cascode.

Transistor	**Weight (Simulation)**
$M1$	63%
$M2$	24%
$M3$	2.2%
$M4$	5.5%

Fig. 5.21 for an input charge of 8 fC and a detector capacitance of 6 pF.
The Equivalent Noise Charge as a function of the detector capacitance is shown in Fig. 5.22 and it increases as expected.

The noise contribution sources of the preamplifier is analyzed in a qualitative way. The three cascode transistors $M2$, $M3$ and $M5$ can be seen as three different source degenerated amplifiers where the degeneration role is played respectively by $M1$, $M4$ and $M6$. Therefore, their equivalent transconductances appear to be negligible if compared with the one of their degeneration transistors making the noise contribution of these devices small [55]. The same concept is applied also to transistors $M4$ and $M6$ where the degeneration has been obtained with the use of the of resistors $R1$ and $R2$ as depicted in Fig. 5.9. These considerations leads to an expected main contribution due to transistor $M1$.
This is confirmed by simulations results in table 5.2, in which it is shown that transistor $M1$ contributes for more than 50% to the total noise of the telescopic cascode.

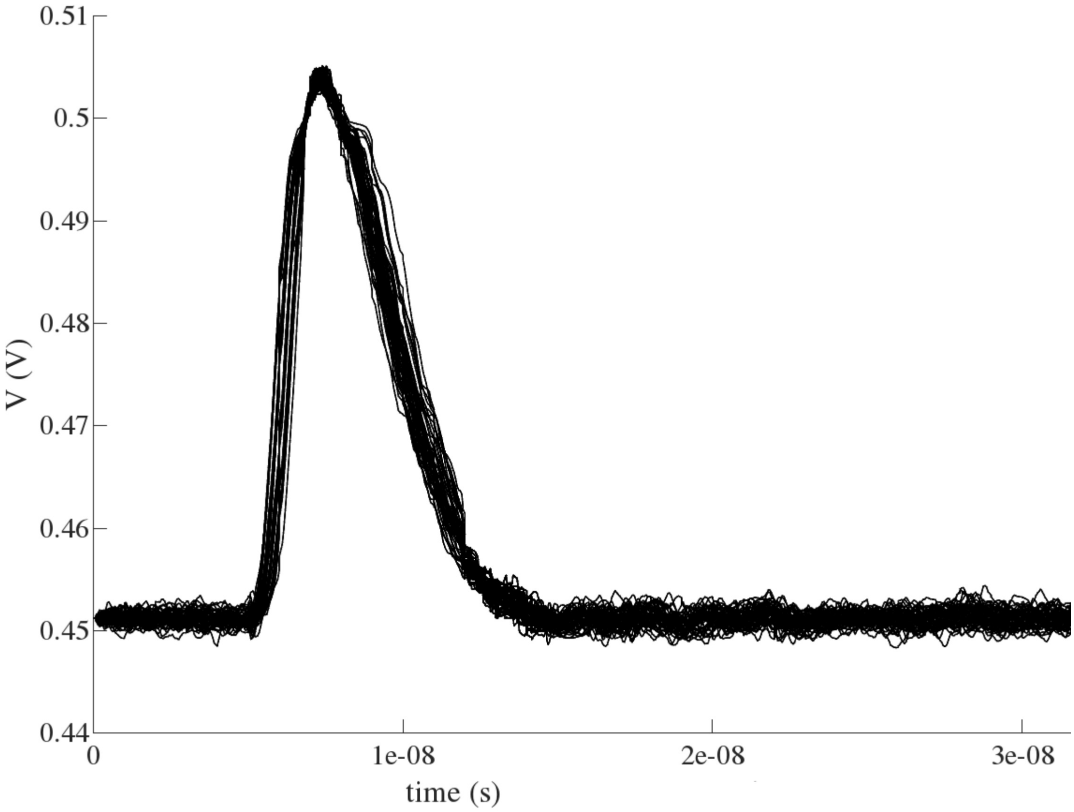


FIGURE 5.21: Transient simulation of the amplifier output for an input charge of 8 fC and a detector capacitance of 6 pF.

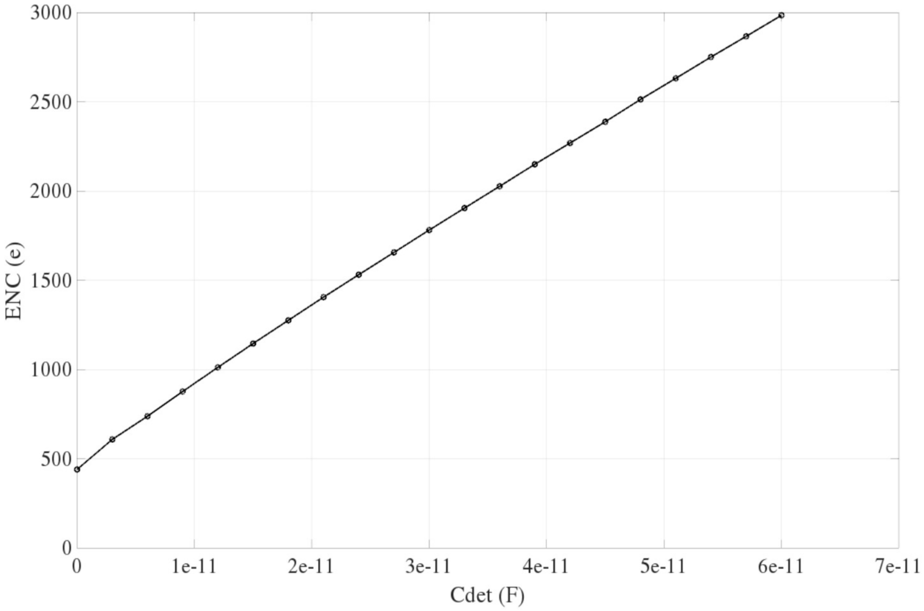


FIGURE 5.22: Equivalent noise charge as a function of detector capacitance.

TABLE 5.3: Jitter at the output of the first stage of the discriminator for different charge values and threshold voltages.

Input charge (fC)	Jitter (ps) V_{th}=9 mV	Jitter (ps) V_{th}=15 mV	Jitter (ps) V_{th}=20 mV	Jitter (ps) V_{th}=35 mV
3	80	571	n.a.	n.a.
6	25	30	37	141
9	16	17	18	16
12	12	13	14	17
15	11	12	13.5	16.5

5.4.3.1 Time resolution

The jitter at the output of the gain cell of the discriminator (Fig. 5.12.a) has been evaluated at different values of the discriminator threshold ranging from 9 mV to 35 mV, according to the formula of Eq. 3.21. The results are displayed in the Table 5.3.

5.5 Layout

Because of the high current requirements of the amplifier, the width of the metal lines used to connect the devices composing the amplifier has to be set high enough according to the electromigration rules provided by the foundry. These rules set an upper limit to the current density that can be delivered by each metal layer. A particular attention has been given to the reduction of parasitic capacitances rising from the overlap between metal lines and their proximity.

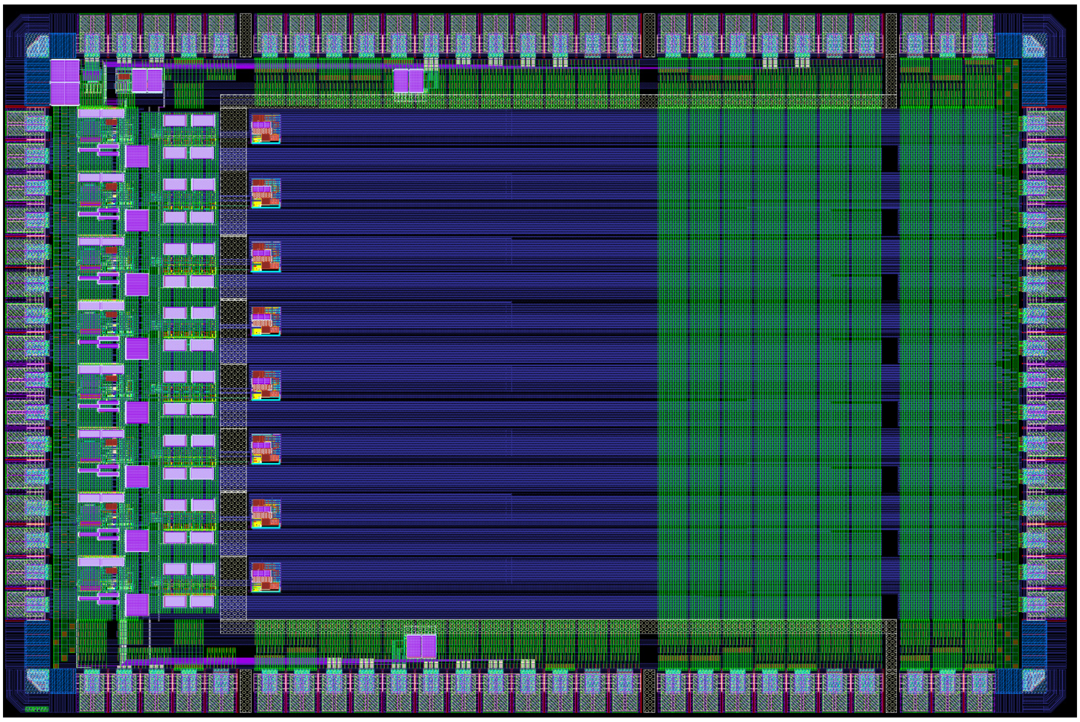


FIGURE 5.23: Final chip layout.

TABLE 5.4: Jitter at the first stage of the discriminator for different charge and threshold values.

Input charge (fC)	Jitter (ps) V_{th}=20 mV	Jitter (ps) V_{th}=35 mV
3	494	n.a.
6	100	227
9	54	62
12	39	39
15	31	30

5.5.1 Power domain splitting

The ASIC has four different power domains: three for the analog processing chain and one for the LVDS driver. The reason to separate the power domain of the amplifier from all the other power domains is that in this way the loop of the current signal from the detector is separated from everything else. This ensures that this loop follows the shortest possible path. This has also other benefits, as for example to separate the amplifier from the discriminator, which injects spikes into the power lines that might be re-injected in the amplifier. This is particularly helpful since the eight channels are detecting signals independently.

5.5.2 Post layout simulations

The effect of parasitic resistances and capacitances added by the physical implementation of the layout are simulated by means of post-layout simulations. The parasitic components are extracted from the layout and added to the schematic by the simulator.

The amplifier AC gain and input impedance simulated with post layout simulations and their comparison with schematic simulations are shown in Fig. 5.24 and Fig. 5.25, respectively. The gain obtained taking into account the parasitic components is higher but the bandwidth is reduced. However, the gain-bandwidth product $A_0\omega_0$ that appears in Eq. 5.6, dropping from 19.49 GHz to 14 GHz, is still acceptable for the application. A post layout transient simulation of the amplifier output for different values of the input charge is depicted in Fig. 5.26: the effect of parasitic components can be seen in an increase of the peaking time to 3 ns and in the presence of an undershoot at $\sim$ 10 ns, which is more pronounced when increasing the input charge.

To have a realistic estimate of the time resolution, post layout simulations have been used to evaluate the jitter at the output of the first stage of the discriminator, similarly to the schematic simulations of Sec. 5.4.3.1. The results for a threshold of 20 and 35 mV are shown in Table 5.4.

5.6 High Precision TDC

As previously explained, the TOFFEE ASIC is designed to produce a logic signal that can be read by the High Precision TDC. The HPTDC is based on the chip developed at CERN [77] for the LHC experiments and used in the CMS muon system. It is implemented in standard 250 nm CMOS technology and its architecture is shown

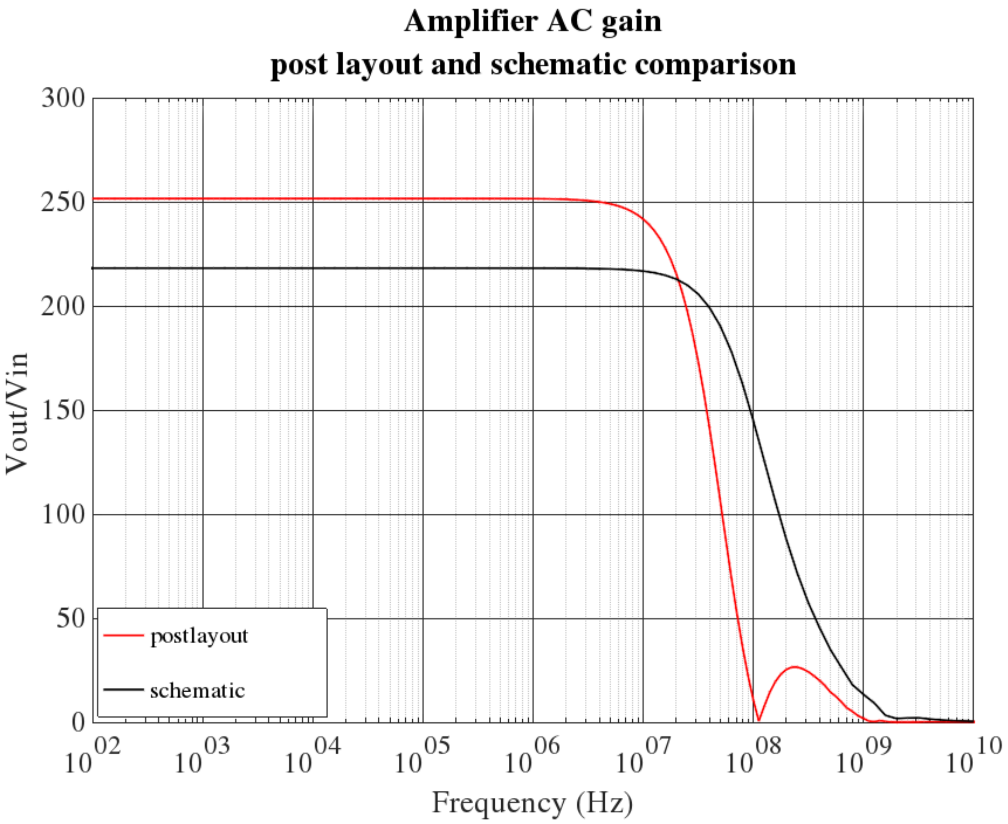


FIGURE 5.24: Post layout simulation of the amplifier AC gain (red line) compared with schematic simulation (black line).

in Fig. 5.27. The time binning and channel number of the HPTDC is configurable. The high-resolution mode provides 8 channels with 25 ps time binning. The time base for the TDC measurements is is a Delay Locked Loop (DLL) with 32 delay elements and a counter synchronized with the clock signal, which also drives the DLL. As clock reference, either the 40 MHz TDC clock or an on-chip Phase Locked Loop (PLL) can be employed [29]. This latter can be used as a filter to remove the jitter from the input clock or to increase the time resolution performing clock multiplication (160/320 MHz).

When a hit (leading and/or trailing edge) is detected, the state of the DLL and the coarse counter are stored. The TDC can be programmed to detect the leading and/or trailing edges of the signal, or alternatively can perform a measurement of the leading edge and the pulse width (pair mode).

4 measurements can be buffered by each channel before being written into the 256-word-deep buffer (L1 buffer) shared by 8 channels. Each channel has a small buffer acting as derandomizer, but the hit measurements are merged into the L1 buffer. If the channel buffer is full, any new hit will be ignored.

The HPTDC has two operation modes: the *Continuous Storage mode* and the *Trigger Matching mode*. When working in Continuous Storage mode, the measurements contained in the L1 buffer are sent directly to a readout FIFO, while in Trigger Matching mode provides hits related to a trigger are selected by a trigger-matching function. A 16 words deep trigger FIFO contains the temporary trigger information, composed by a trigger time tag (bunch id) and an event identifier (event id). Events related to a trigger are selected within a programmable window (trigger matching window).

The radiation tolerance of the chip technology is still not proofed.

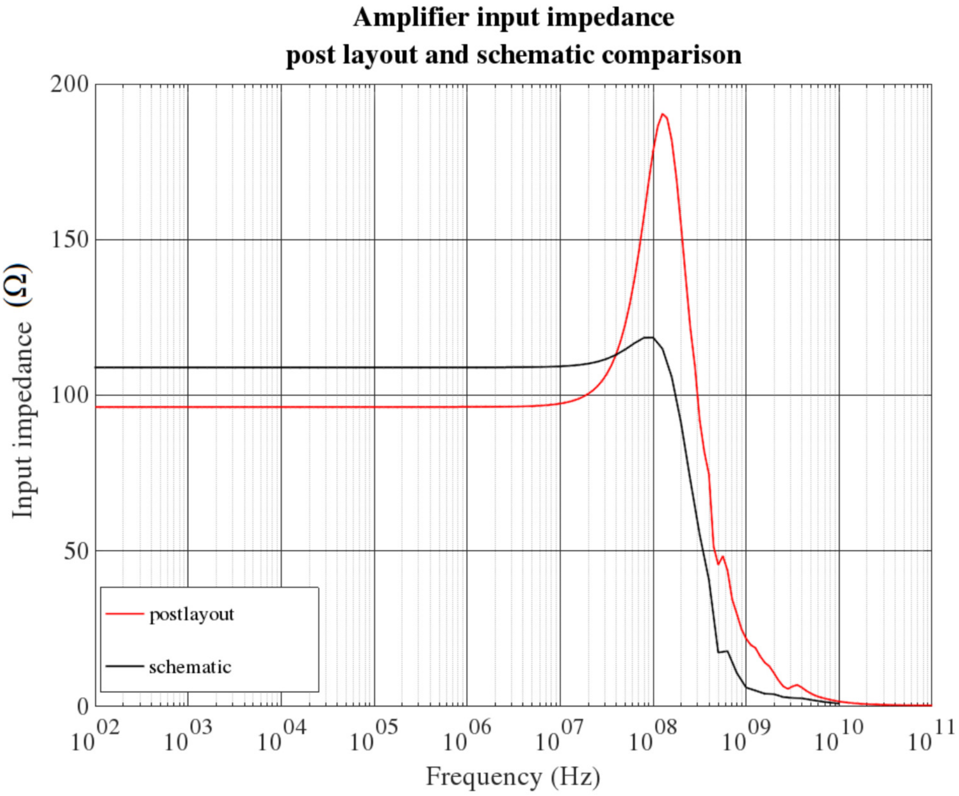


FIGURE 5.25: Post layout simulation of the amplifier input impedance (red line) compared with schematic simulation (black line).

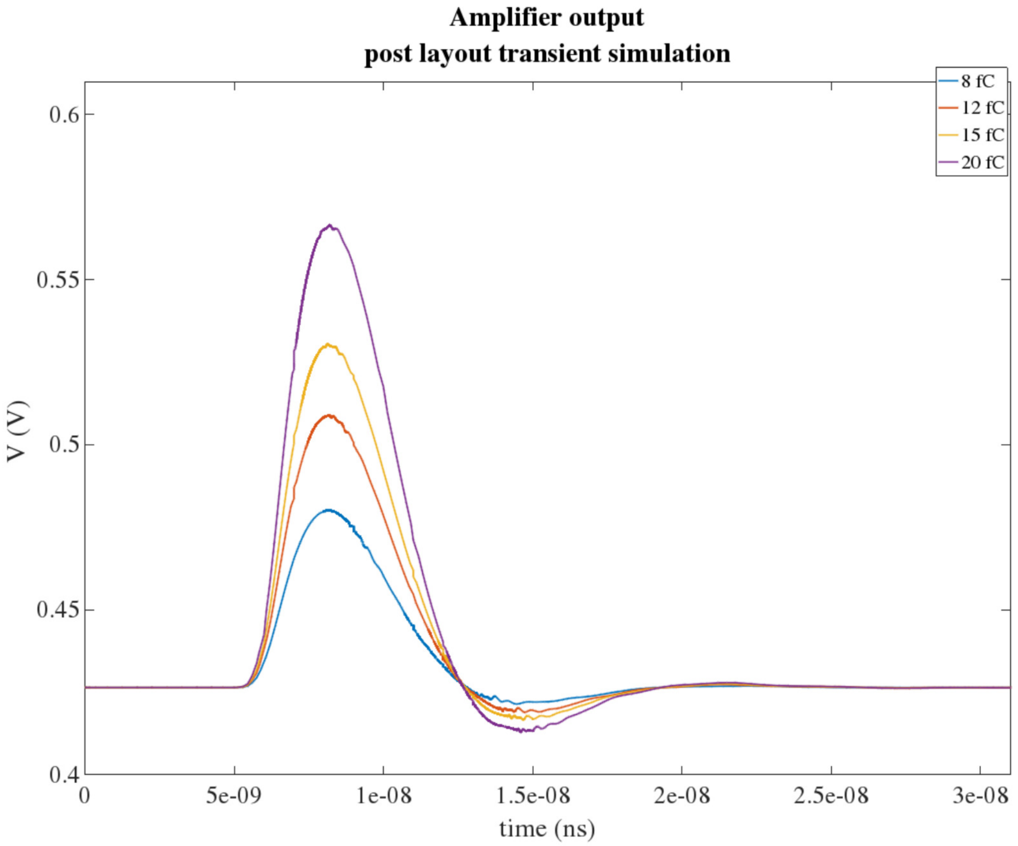


FIGURE 5.26: Post layout simulation of amplifier output at different values of the input charge for a detector capacitance of 6 pF.

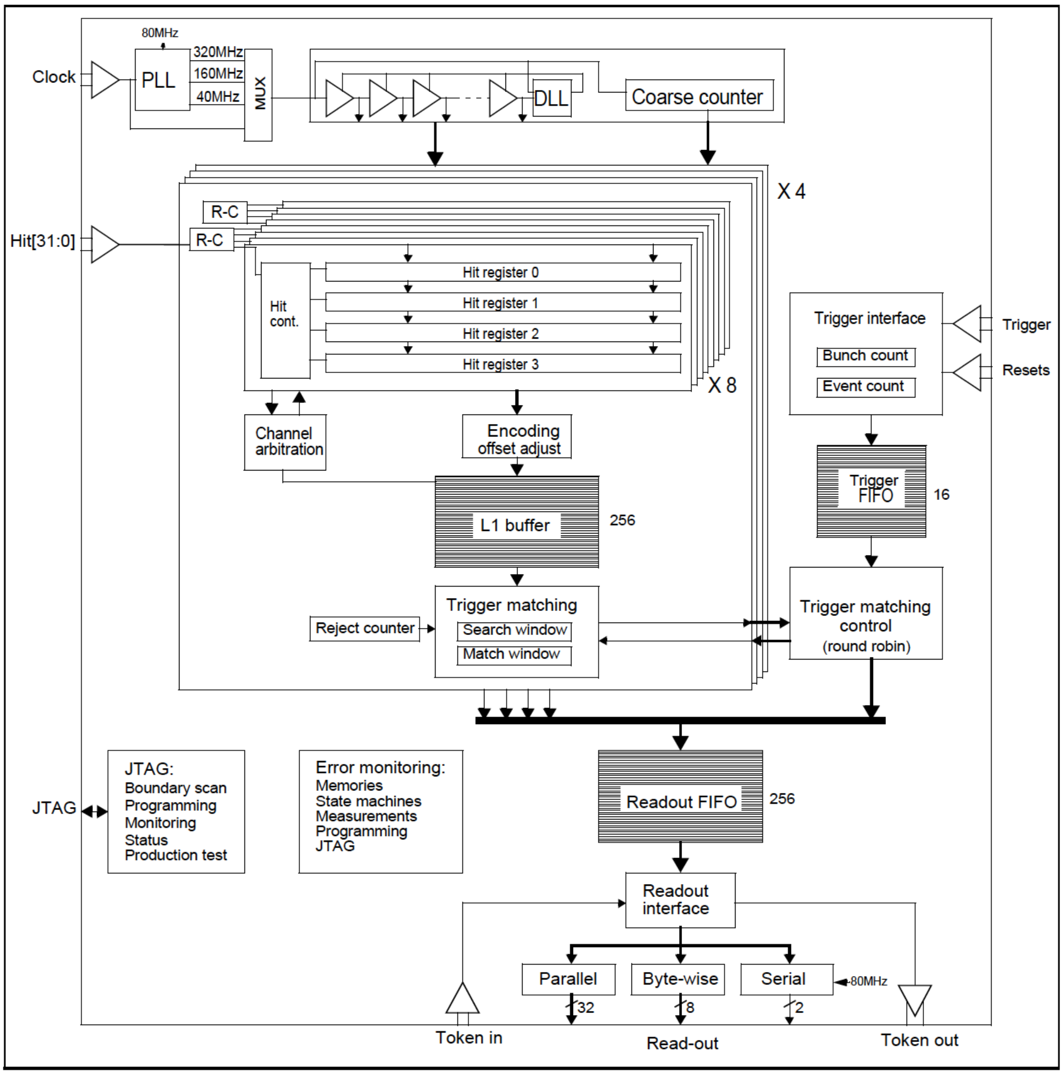


FIGURE 5.27: HTPDC architecture.

Chapter 6

Measurements on TOFFEE ASIC

The TOFFEE ASIC has been received from foundry in September 2016. Tests started in October using a first version of custom board, while tests with the ASIC combined with UFSD sensors started in January 2017 with a second version of boards. The setup has undergone successful beam tests since September 2017.
This chapter describes the experimental setups and the obtained results.

6.1 Laboratory tests with LIP testboard

LIP board has been intended to perform preliminary tests with TOFFEE and a test pulse and further tests with APDs developed by RMD. The board has been designed by R. Silva (LIP, PETsys Electronics) and allows to set the discriminator threshold with on-board trimmers or with an external voltage source for either the 8 channels independently or simultaneously.

The system has been tested in October 2016. The tests have been performed injecting a square input signal into the channels of the chip by using a pulse generator and a RC filter. Amplitude and width have been chosen to reproduce the expected sensor signal slew rate, whereas the values of R and C have been chosen to emulate the signal shape of UFSD sensors.

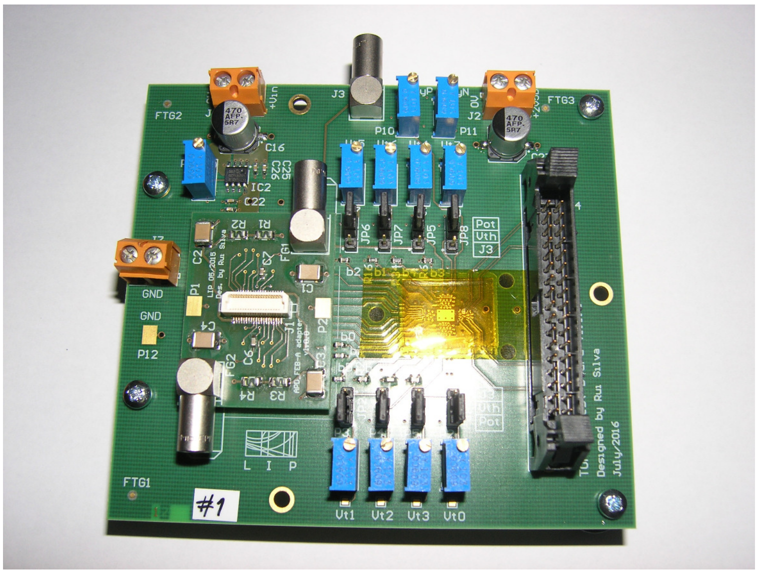


FIGURE 6.1: TOFFEE testboard by LIP

6.1.1 Experimental setup

The injected signal has an amplitude of 4.8 V, rise and fall times of 2.5 ns, a frequency of 100 kHz and a width of 1.38 µs, and is locked to a TTL trigger sent to an oscilloscope. The amplitude allows to maximize the pulser signal-to-noise ratio but this requires the use of a 50 dB attenuator to avoid the preamplifier saturation. The values chosen for the filter are R = 1 kΩ and C = 1 pF, to reproduce the charge collection time of the UFSD sensors. A DAC board is used to automatically set the discriminator threshold voltage with a remote computer. A 1 kΩ resistor in series with the DAC output ensures the stability of the DAC board driver.

The LVDS output of TOFFEE is read by a differential probe connected to the oscilloscope, which automatically measures the delay between the trigger signal and the LVDS signal, calculates the average and the standard deviation σ. The noise is evaluated from the slew rate and jitter of the output signal at various threshold values. The measured values, shown in (Fig. 6.2), demonstrate that most of them are compatible with simulations (predicting 0.6 mV). Higher noise values are due to an imprecise measurement of the slew rate. Moreover, the noise does not vary with the threshold value, and therefore it can be seen that the performances are not affected by the discriminator.

The pulse shape has been reconstructed by sweeping the threshold voltage from the baseline value to the one at which the threshold is too high to detect the pulse and measuring on the oscilloscope the time of arrival and the time over threshold for each threshold value. The resulting shape is shown in Fig. 6.3: it exhibits a signal of amplitude 90 mV with a peaking time of 3 ns, compatible with expectations for the signal of an UFSD delivering the same charge.

Tests with infrared laser and RMD APDs did not succeed due to discharges from the sensor injected into the ASIC and compromising its operation.

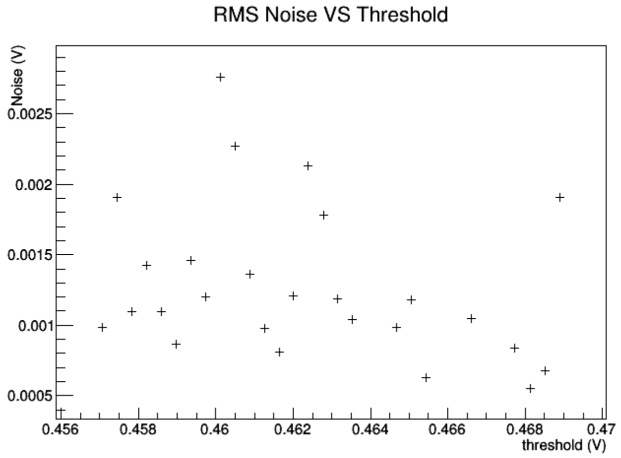


FIGURE 6.2: Measured noise as a function of the threshold voltage for 1 pF input capacitance.

By varying the discriminator threshold voltage, it has been possible to measure (i) the signal slew rate dV/dt

$$SR = \frac{V_{th2} - V_{th1}}{mean_2 - mean_1} = 25\ \mathrm{mV/ns}, \tag{6.1}$$

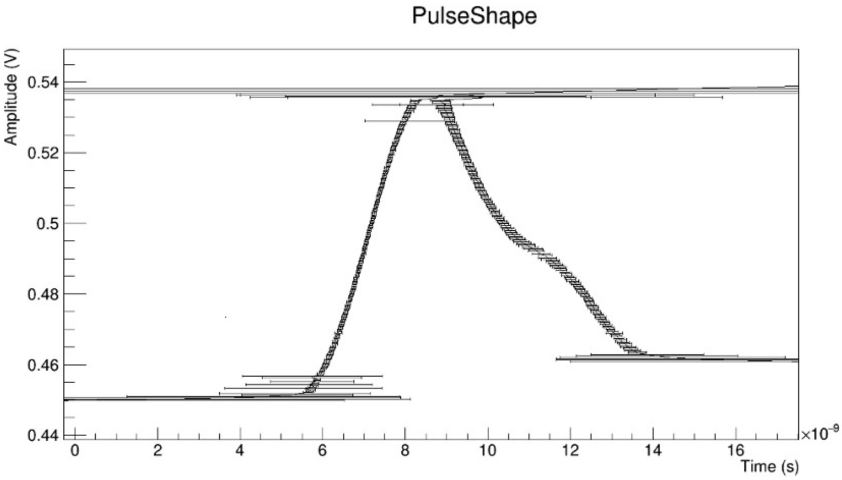


FIGURE 6.3: Reconstructed pulse shape from threshold voltage (amplitude), time of arrival and time over threshold (time).

(ii) the jitter of the TOFFEE + pulse generator system $\sigma = 44\ ps$ and (iii) to evaluate the noise

$$Noise = SR \cdot \sigma = 1.1 \text{ mV}. \tag{6.2}$$

These three values are consistent with simulations.
To measure the ASIC contribution to the total jitter as a function of the input charge, the previously described setup has been used. The input charge Q_{in} can be varied by sweeping the pulse amplitude V_{pulse} following the relation $Q_{in} = C{\cdot}V_{pulse}/attenuation$. The total jitter read at the oscilloscope (LVDS jitter) comprises the contributions of pulse generator jitter and the one from the ASIC, as shown in Fig. 6.4. The contribution due to the ASIC is thus given by

$$\sigma^2_{t,TOFFEE} = \sigma^2_{t,LVDS} - \sigma^2_{t,gen}. \tag{6.3}$$

For a charge of ~ 12 fC the expected resolution is ~ 36 ps.

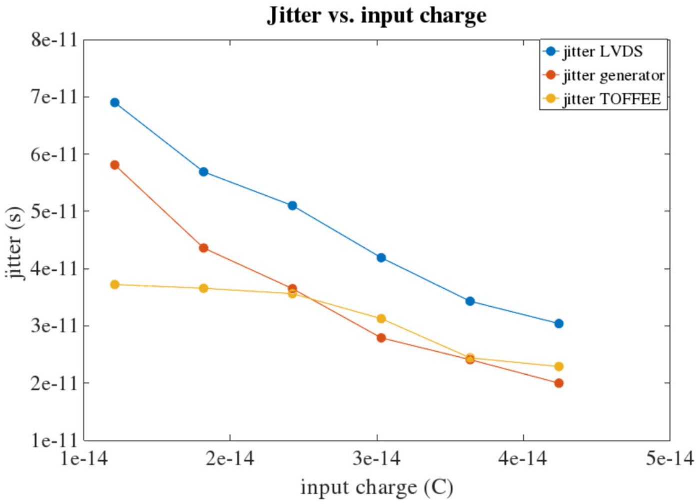


FIGURE 6.4: Measured jitter as a function of the input charge for an input capacitance of 1 pF.

6.2 Laboratory tests with INFN-Torino testboard

The second version of testboards has been designed by M. Mignone (INFN Torino). On this board, shown in Fig. 6.5, the sensor is glued on a dedicated plate providing the contact between the high voltage source and the sensor p backplane. This plate is surrounded by a ground reference frame for the sensor guardring. Eight n electrodes of the sensor are wire-bonded to the on-board pads connected to the input of each channel of the ASIC.
The discriminator threshold can be adjusted by means of on-board trimmers that allow to set the threshold for each channel independently.

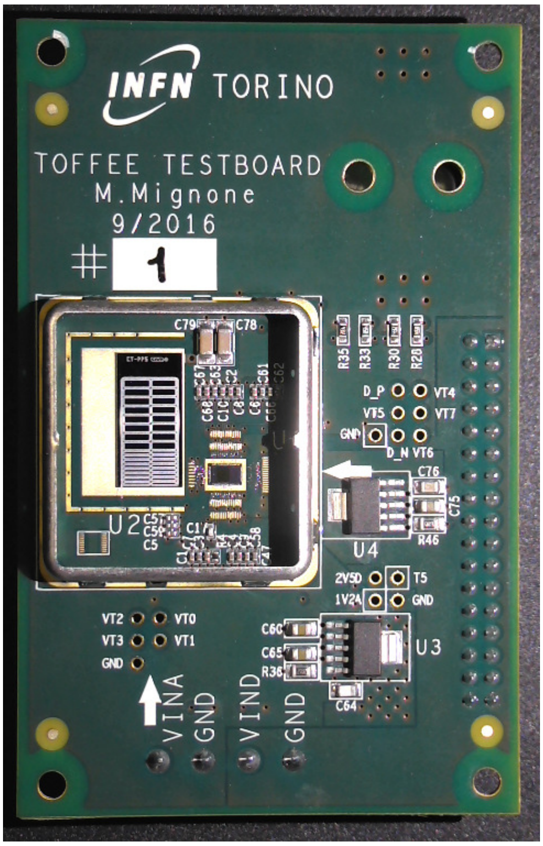


FIGURE 6.5: INFN-Torino testboard with TOFFEE wire-bonded to the CNM CT-PPS UFSD sensor.

6.2.1 Experimental setup

The most effective way to test the sensor with its readout electronics is with a particle beam (beam test). In absence of a particle beam, two other solutions can give analogous results with a rather simple laboratory setup, i.e. to produce the ionization inside the sensor with a source (a beta source can provide exactly a MIP signal) or with a laser. However, these two "simpler"methods have the drawback that both source and laser photons cannot produce a signal in more than one detector plane at the same time. For example, in a coincidence system, two or more detectors have to be crossed by the same particle, in order to derive the time of arrival of a detector using the other one as trigger. Moreover, the laser intensity has to be calibrated to know the actual amount of charge delivered to the sensor.

The charge deposition of a minimum ionizing particle in the sensor can be emulated with an infrared laser, whose wavelength crosses the whole sensor bulk. A focused 1064 nm infrared laser (ALD picosecond laser) consisting in (i) a laser controller that contains a pulser and allows to set the pulse frequency and light intensity and (ii) a laser head producing the light and terminated with an optic fiber. The fiber is mounted on a collimator, whose focal distance is 8.5 cm. A movable z stage allows to set the distance of the collimator in order to focus the laser spot.
The laser controller produces a TTL trigger with the selected frequency and is sent to the oscilloscope as trigger signal. The LVDS output is read by the oscilloscope with a differential probe connected to an IDC flat connector with a 100 Ω termination. The oscilloscope measures the time of arrival (ToA) of the LVDS signal with respect to the trigger and the duration of the LVDS pulse (ToT). The standard deviation of the ToA distribution is the time resolution of the system.
The board is fixed on a x, y movable plane that allows to move the board horizontally with respect to the laser beam, giving the possibility to change the position at which the laser beam hits the sensor.
The threshold can be set with a proper connector from the power supply to the onboard threshold pins. The amplitude of the amplifier signal is measured by scanning the threshold voltage from the baseline value up to the value at which the oscilloscope does not trigger anymore on the signal, which corresponds to the top of the amplifier signal.
The time resolution of the system is acquired while the laser intensity is changed to vary the amount of charge delivered to the sensor. The input charge is estimated by dividing the amplitude of the amplifier signal by the nominal gain of the amplifier, which corresponds to ~ 7 mV/fC. The setup allows remote control: the threshold

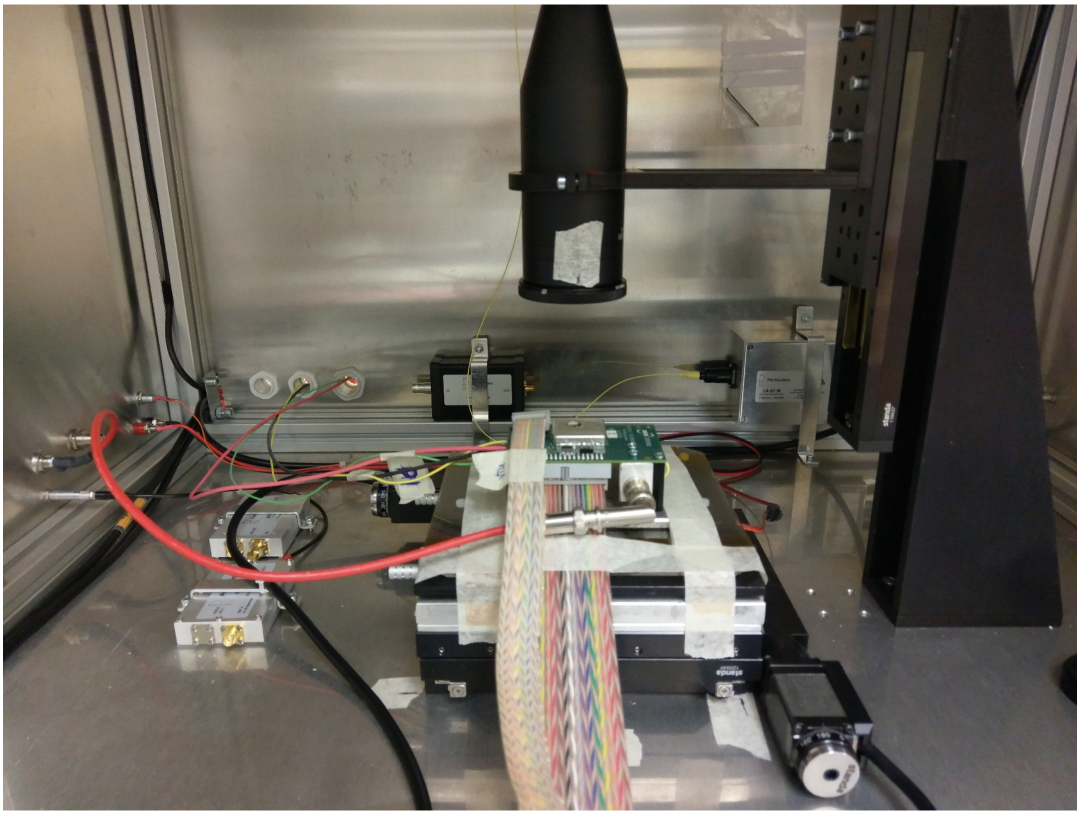


FIGURE 6.6: Laboratory setup for laser tests.

and laser intensity can be varied from a remote computer (running a RedHat or compatible distribution) by connecting the power supply and the laser controller through a GPIB cable (power supplies) and USB cable (laser). The oscilloscope can be optionally connected to the computer through a USB cable to acquire data automatically. Remote control is implemented with PyVISA, a Phython package that enables to control measurement devices over the National Instruments VISA protocol. Instrument configuration and data acquisition are then performed by running a Python script.

6.2.2 Tests with HPK 50 µm sensors

Hamamatsu 50D single pad sensors have been tested with the TOFFEE ASIC on INFN testboard. The sensor is a 50 µm thick Low Gain Avalanche Diode with a diameter of 1.3 mm^2, an active area of $\sim$ 1.7 mm^2 and a capacitance of about 3.5 pF. The sensor wire-bonded to the chip is shown in Fig. 6.7. Measurements results performed on two different boards reproducing with the laser an input charge ranging between 2 and 40 fC are displayed in Tables 6.1, 6.2 and plotted in Fig. 6.8. Laser tests conducted on this single pad sensor demonstrate that the measured standard deviation of the ToA for an input charge of $\sim$ 8 fC is between 40 and 50 ps. All the following tests have been performed with the stretcher set to the minimum starving value.

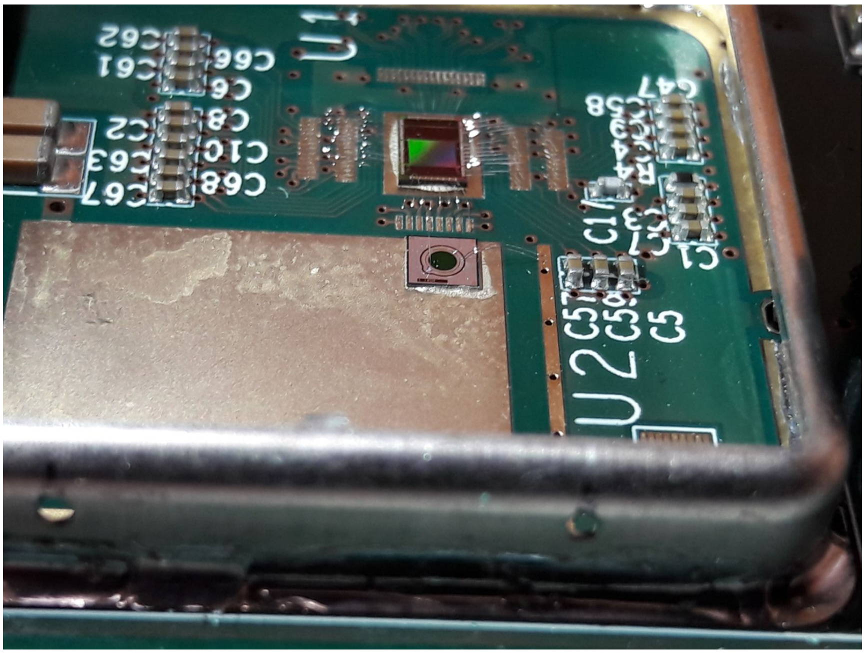


FIGURE 6.7: Hamamatsu 50D sensor wire-bonded to TOFFEE ASIC.

The dependence of the time of arrival on the time over threshold obtained with the previous laser tests is shown in Fig. 6.9 for a threshold voltage of 7 mV. Since up to 40 fC the ToT is monotonically increasing with the input charge (Fig. 5.18), this demonstrates that the spread of the ToA distribution is higher for lower charges and becomes progressively narrower with increasing ToT. Moreover, the value of the time of arrival is decreasing with increasing input charge as a result of a progressively steeper amplifier signal, in agreement with the expectations.

TABLE 6.1: Measurement results for laser tests conducted on TOFFEE + HPK 50D sensor (Board 1).

Input charge (fC)	ToT (ns)	σ ToT (ns)	ToA (ns)	σ ToA (ns)
2.3	5.2	0.55	29.8	0.28
2.7	6.5	0.3	29.3	0.14
3.2	7	0.23	29.05	0.107
5.0	8.16	0.17	28.55	0.05
6.5	8.31	0.14	28.44	0.044
8.3	8.72	0.12	28.28	0.037
9.7	9.15	0.12	28.15	0.034
11.3	9.28	0.108	28.08	0.03
12.7	9.37	0.102	28.03	0.029
14.3	9.45	0.1	28.98	0.028
16.0	9.53	0.092	28.93	0.027
18.2	9.58	0.089	28.89	0.024
20.7	9.59	0.082	28.87	0.024
23.5	9.69	0.078	28.83	0.021
27.0	9.7	0.073	28.8	0.022
30.3	9.69	0.067	28.77	0.021
32.8	9.64	0.06	28.77	0.021
40.0	9.58	0.052	28.71	0.02

TABLE 6.2: Measurement results for laser tests conducted on TOFFEE + HPK 50D sensor (Board 2).

Input charge (fC)	ToT (ns)	σ ToT (ns)	ToA (ns)	σ ToA (ns)
2.0	5.3	0.63	30.4	0.32
2.7	6.2	0.4	29.9	0.19
3.0	6.76	0.28	29.62	0.13
5.7	8.18	0.21	29.053	0.063
7.2	8.69	0.18	28.881	0.053
9.7	9.09	0.14	28.726	0.047
11.0	9.25	0.123	28.646	0.045
13.2	9.36	0.114	28.581	0.044
14.7	9.45	0.109	28.54	0.043
16.7	9.41	0.099	28.51	0.041
19.0	9.56	0.1	28.447	0.04
21.2	9.6	0.094	28.408	0.04
23.3	9.64	0.09	28.375	0.039
27.3	9.69	0.087	28.345	0.039
31.2	9.67	0.077	28.316	0.039
34.3	9.63	0.07	28.304	0.039
37.8	9.6	0.065	28.285	0.038
44.7	9.44	0.052	28.273	0.038

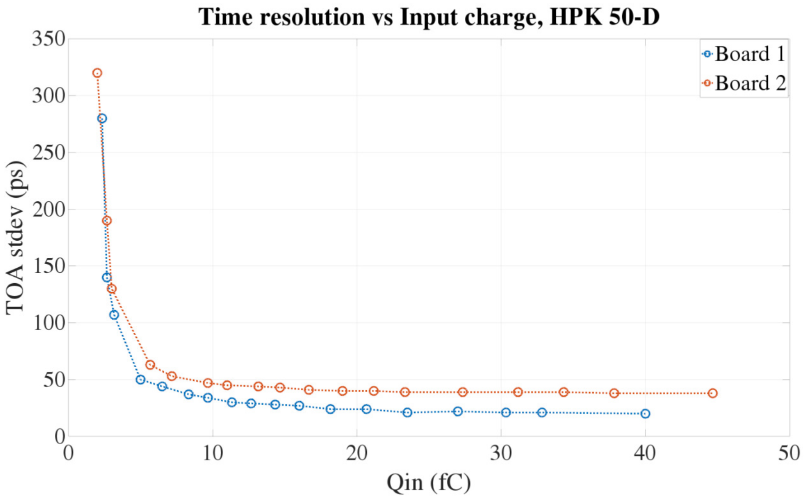


FIGURE 6.8: Measured time resolution as a function of the input charge for HPK 50 µm thick single pad sensors for a threshold voltage of 7 mV.

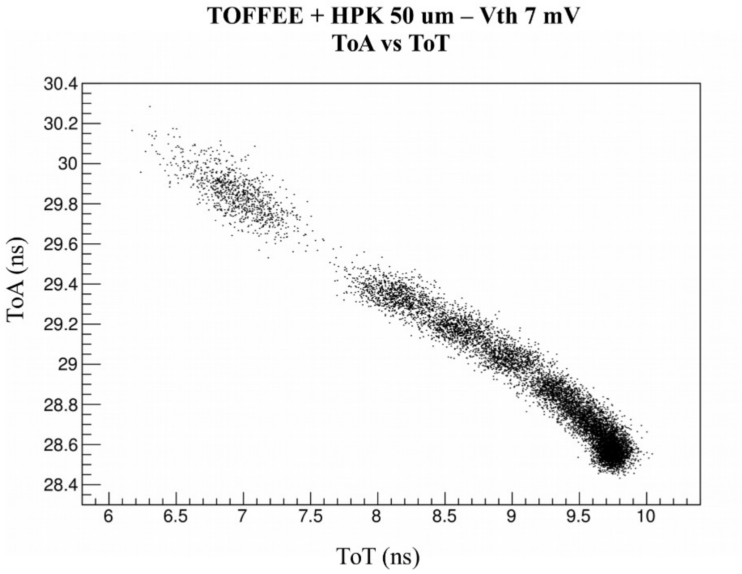


FIGURE 6.9: Time of Arrival as a function of Time Over Threshold for a discriminator threshold of 7 mV.

6.3 Data acquisition with HPTDC

The HPTDC chip (Sec. 5.6) can be found on the CAEN VME modules V1290A and V1290N: both are a 1-unit wide VME 6U modules and fit into standard VMEbus crates [78]. The V1290A houses 32 independent Multi-Hit/Multi-Event Time to Digital Conversion channels, namely 4 HPTDC chips developed by CERN. This module accepts both ECL and LVDS inputs. The V1290N contains 2 HPTDC chips (16 independent channels) and shares most of its features with the previous module, apart from requiring NIM inputs. The unit accepts ECL control signals that are common to all channels. The expected resolution in 21 bit dynamics is 25 ps (35 ps RMS) while the double hit resolution is 5 ns.

The TDC can be programmed by the user via a microcontroller that implements

TABLE 6.3: Measurement results for HPTDC resolution.

Measure	**Resolution (ps)**
Leading edge res.	25.01
Trailing edge res.	34.50
TOT res.	43.63

a high-level interface masking the TDC's hardware. The data acquisition can be performed in either "trigger matching mode", which foresees the configuration of a time window to match the event arrival, or in "continuous storage mode", as described in the previous chapter. In Trigger Matching mode, the V1290 module allows to configure the time offset of the matching window with respect to the trigger (either positive or negative), as the HPTDC itself does not allow matching windows straggling the trigger or delayed with respect to it.
Events are stored in a 32 kwords deep output buffer that can be readout by the acquisition in three ways: as single data, Block Transfer (BLT) or Chained Block Transfer (CBLT).

The data acquisition software is described in Appendix A.

6.3.1 Test of HPTDC board and signals timing

In order to evaluate the TDC contribution to total time resolution, the HPTDC time resolution for the channel Ch0 of module V1290N has been measured. The setup consists of:

1. VME module CAEN V1718, USB to VMEbus interface;
2. VME module CAEN V1290N, providing two HPTDC ASICs with NIM single ended inputs;
3. arbitrary waveform generator Agilent AFG3252;
4. computer with DAQ software.

The V1718 module is used to interface the computer for DAQ with the VME bus through a USB port. The computer configures the V1290N board and reads out the HPTDC data buffer. The waveform generator provides three signals for the V1290N, locked in phase:

1. external clock signal;
2. trigger signal;
3. input signal for Ch0.

The test has been conducted on 10001472 input signals and triggers at 400 kHz. The phase between the clock and the signal is constant. The time resolution of leading and trailing edge have been measured, as well as the TOT resolution to cross check the previous two. Results are shown in Table. 6.3, Fig. 6.10, 6.11, 6.12.

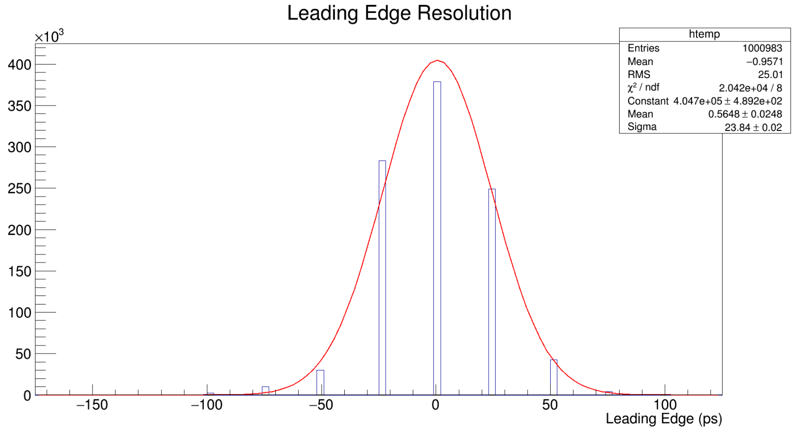


FIGURE 6.10: HPTDC leading edge resolution.

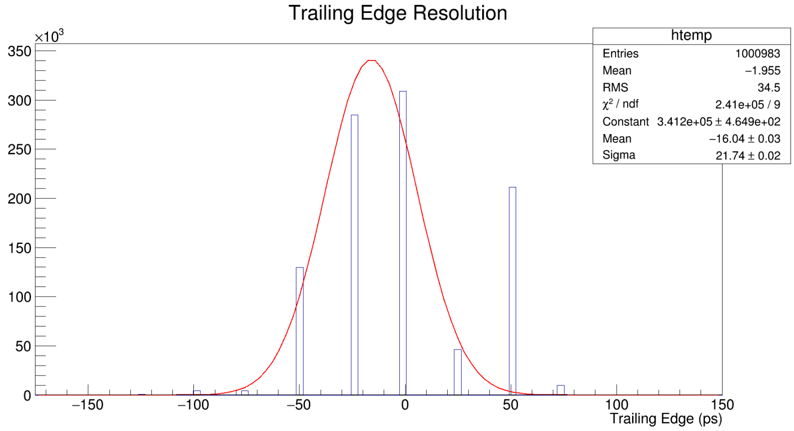


FIGURE 6.11: HPTDC trailing edge resolution.

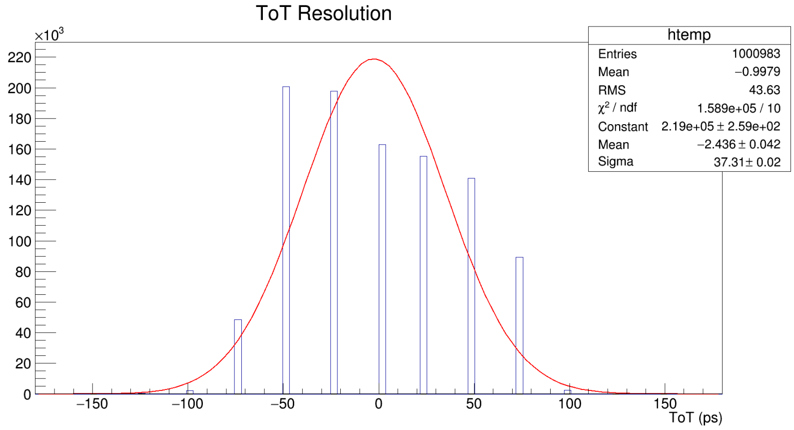


FIGURE 6.12: HPTDC TOT resolution.

6.4 Beam tests

In September 2017 the system UFSD + TOFFEE has been tested with 180 GeV pions at CERN (SPS H8 beam line). The telescope is composed of two boards, each containing a HPK 50 μm thick sensor read out by TOFFEE. The two sensors are aligned in order to produce a coincidence and can be considered identical. The difference of time of arrival between the two boards

$$\Delta ToA = ToA_2 - ToA_1 \tag{6.4}$$

is the time required by the impinging particle to cross the space between the two boards. The precision in resolving the difference of time of arrival, i.e. the root mean square of the ΔToA distribution gives the global time resolution of the two-board system, while the time resolution of a single board is given by

$$\sigma_{t,1board} = \frac{\sigma_{\Delta ToA}}{\sqrt{2}}. \tag{6.5}$$

This setup has undergone two successful beam tests with two different readout methods: differential probe connected to an oscilloscope, and HPTDC, and both led to consistent results that are described in the following.

6.4.1 Oscilloscope readout

The oscilloscope-based readout method, already employed in laboratory tests, allows to read out a single channel from each board by means of a differential probe. The two channels are chosen so as to produce a coincidence when hit by the same particle. Two differential probes connected to two aligned channels of the two boards are read by two different channels of the oscilloscope, this latter triggering when both boards are detecting a signal. The alignment of the two channels from Board 1 and Board 2 has been checked.
The obtained waveforms have been analyzed and led to the results presented in the following, which allowed to understand several features of the TOFFEE + UFSD system.

6.4.1.1 Estimation of amplifier gain

Since the value of the sensor gain as a function of the bias voltage is known from gain curves obtained with laboratory measurements as the ratio between the amplitude of a UFSD signal and the one of a sensor with unitary gain, the goal is to obtain the charge delivered by the sensor and to compare it with the amplitude of the amplifier signal in order to derive the gain of the amplifier. The gain G of the sensor as a function of the bias voltage has a mostly exponential behavior. The charge deposited by a minimum ionizing particle in a 50 µm sensor with unitary gain has been simulated and has the value Q_0 = 0.46 fC. Thus, the total charge Q_{in} delivered to the ASIC by the UFSD can be calculated according to

$$Q_{in} = Q_0 \cdot G, \tag{6.6}$$

and its increase with the sensor bias voltage is given by the relation between this latter and the gain.
Recalling that the value of the charge delivered by a MIP corresponds to the most probable value (MPV) of the Landau distribution, the probability to have a charge lower than the MPV is about 30%, while the probability to have a higher charge is 70%. With the lowest possible threshold, high enough not to trigger on noise pulses, a time interval is fixed to acquire 1000 events per spill. Then, the threshold is increased in such way that about the 70% of the events are acquired in the chosen time interval. In such way it is reasonable to assume that, by removing a portion of 30% of the total events, the threshold is close to the amplitude of the signal provided by a

TABLE 6.4: Measured amplitude of the amplifier signal for Board 1, along with the corresponding input charge and the extracted value of the amplifier gain for different bias voltages.

V_{bias} (V)	Q_{in} (fC)	Amplitude (mV)	Amplifier gain (mV/fC)
210	5.72	36	6.29
230	6.84	44	6.43
250	7.36	51	6.93
270	8.74	59	6.75
290	11.67	67	5.74
300	12.75	79	6.19
310	16.10	96	5.96

TABLE 6.5: Measured amplitude of the amplifier signal for Board 2, along with the corresponding input charge and the extracted value of the amplifier gain for different bias voltages.

V_{bias} (V)	Q_{in} (fC)	Amplitude (mV)	Amplifier gain (mV/fC)
210	5.72	36	6.29
230	6.84	44	6.43
250	7.36	52	7.07
270	8.74	63	7.21
290	11.67	73	6.26
300	12.75	86	6.74
310	16.10	113	7.02
320	18.86	133	7.05

MIP, and this value is progressively higher when increasing the sensor bias voltage. The amplifier gain is then extracted by dividing the signal amplitude by the charge Q_{in} delivered by the sensor for each value of the bias voltage. This procedure has been repeated for both boards, assuming that the sensors are identical. The measurements results are displayed in Table 6.4 for Board 1 and Table 6.5 for Board 2, while the amplifier gain is shown in Fig. 6.13 as a function of the bias voltage for the two boards, plotted alongside the gain curve of the sensor. From these results, it can be seen that the measured gain of the amplification stage of the TOFFEE ASIC is located between 6 and 7 mV/fC, being consistent with what expected from simulations.

6.4.1.2 ToT distribution

The Time Over Threshold is acquired by measuring the width of the LVDS signal at half of the pulse height. Data have been acquired sweeping the sensor bias voltage for three different threshold values: a low value, a medium one and a high one, namely 8 mV, 18 mV and 23 mV. The ToT distribution for a threshold of 8 mV is shown if Fig. 6.14.a for a bias voltage of 170 V (blue), 230 V (green) and 250 V (red). The same distribution for a threshold of 18 mV is shown in Fig. 6.14.b for a bias voltage of 230V (blue), 270 V (green) and 290 V (red), while for a threshold of 23 mV and bias voltages of 290 V (blue), 300 V (green) and 320 V (red) it is shown in Fig. 6.15.a.

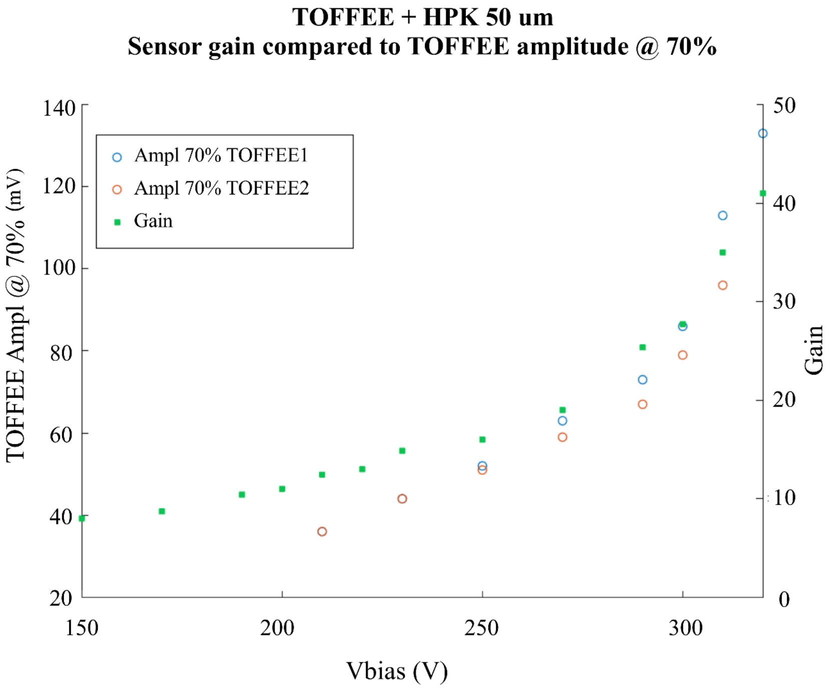


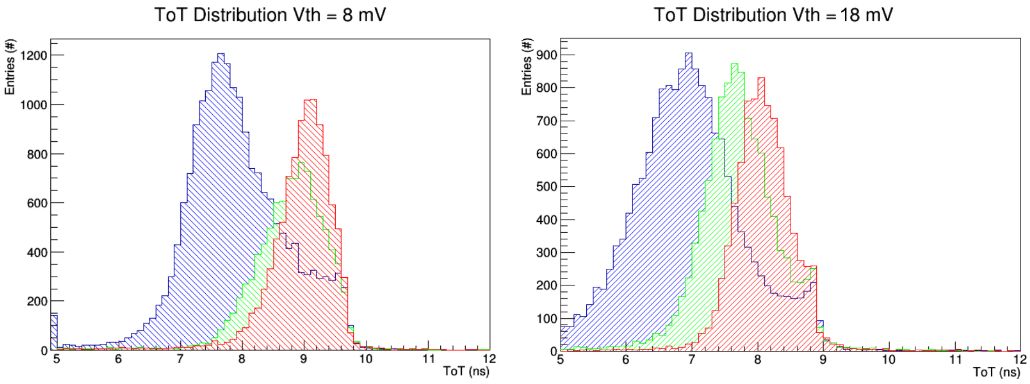
FIGURE 6.13: Extracted gain of Board 1 and Board 2 compared with the gain of the sensor as a function of the bias voltage.

Ideally, the ToT should follow a Landau distribution, but at higher thresholds and higher bias voltages, the information on the energy released by the impinging particle may be lost. All the previous distributions demonstrate the logarithmic behavior of the ToT as a function of the input charge, which increases up to the saturation value, occurring at $\sim$ 10 ns for a threshold of 8 mV and $\sim$ 9 ns for thresholds of 18 mV and 23 mV. In particular, in Fig. 6.14.a the distribution at 170 V (blue shape) shows a trend compatible with a Landau distribution until ToT saturation at $\sim$ 10 ns: the shape is rapidly increasing, reaches a peak at $\sim$ 7.5 ns and then starts to decrease with a lower slope. However, the ToT saturation does not allow the typical infinite decrease of the Landau distribution. An increase in the sensor bias voltage produces larger signals from the sensor that translate into larger amplifier pulses. In this way, however, the ToT is shifted towards its saturation value, leading to a narrower ToT distribution and a loss in energy resolution capability.

6.4.1.3 Coincidence time resolution

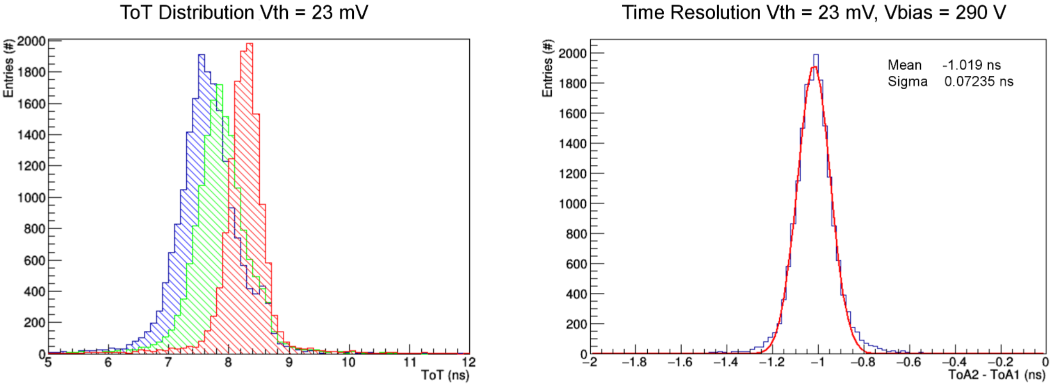
The behavior of the ToA as a function of the ToT has been evaluated, since due to time walk the time of arrival may have a dependence on the signal amplitude. By studying this relation and by applying the proper correction, the ToA can be made as independent as possible from the ToT. The $\Delta ToA\ vs.$ ToT for a threshold of 18 mV and a sensor bias voltage of 270 V before and after correction are displayed in Fig. 6.16. Fig. 6.16.a shows that the correction is not necessary after ToT = 9 ns, as no time walk is present. This is explained by an expected saturation of the slew rate increase. Data related to ToT higher than 9 ns have not been considered in the final analysis and in the presented data.

The time resolution of the two-board system for different sensor bias voltages has been measured for the three threshold values. It has been evaluated from data analysis by fitting the ΔToA distribution with a gaussian distribution and extracting the



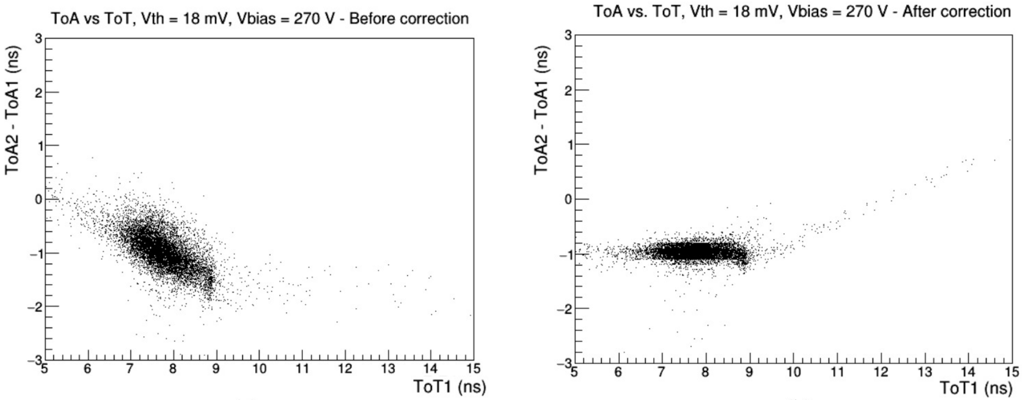
(a) (b)

FIGURE 6.14: a) Time Over Threshold distribution for a discriminator threshold voltage of 8 mV at $V_{bias} = 170$ V (blue), 230 V (green) and 250 V (red). b) ToT distribution for a threshold of 18 mV at $V_{bias} = 230$ V (blue), 270 V (green) and 290 V (red).



(a) (b)

FIGURE 6.15: a) ToT distribution for a threshold of 23 mV at $V_{bias} =$ 290 V, 300 V and 320 V. b)Time resolution for a sensor bias voltage of 290 V at a threshold of 23 mV.



(a) (b)

FIGURE 6.16: Difference in time of arrival for the two boards plotted as a function of time over threshold of Board 1 before correction (left) and after correction (right).

root mean square of the fit after time walk correction. The gaussian fit applied to the ΔToA distribution for a threshold of 23 mV is shown in Fig. 6.15.b. The time resolution of a single board is then derived by dividing the obtained root mean square by $\sqrt{2}$ (Eq. 6.5).
The time resolution as a function of time over threshold is plotted in Fig. 6.17. From this picture, it can be seen that higher thresholds have a better time resolution than lower ones. Moreover, the resolution is worsening for higher signals since due to the ToT saturation it is no longer possible to apply time walk corrections.
The time resolution of UFSD read out by TOFFEE has been compared with the one achieved by an UFSD read out by a discrete component board made by Santa Cruz Institute for Particle Physics. The resolution as a function of the sensor bias voltage is shown in Fig. 6.18.a, while the resolution as a function of the sensor gain is plotted in Fig. 6.18.b. Moreover, the time resolution measured at the beam test has been compared with a simulation prediction. In the simulation, performed with *Weightfield2*, a laser beam with varying intensity is producing a signal at into a 50 µm thick UFSD with the same features of the HPK 50D. In this way, a different amount of charge is collected by the sensor. Thus, a readout electronics composed by an amplifier and a discriminator and producing a signal similar to the TOFFEE ASIC is simulated. Fig. 6.19 shows the comparison between measurements and simulation results for different threshold voltages. The measured time resolution is slightly worse than the simulation prediction for compatible threshold values, while the overall trend seems to be respected.

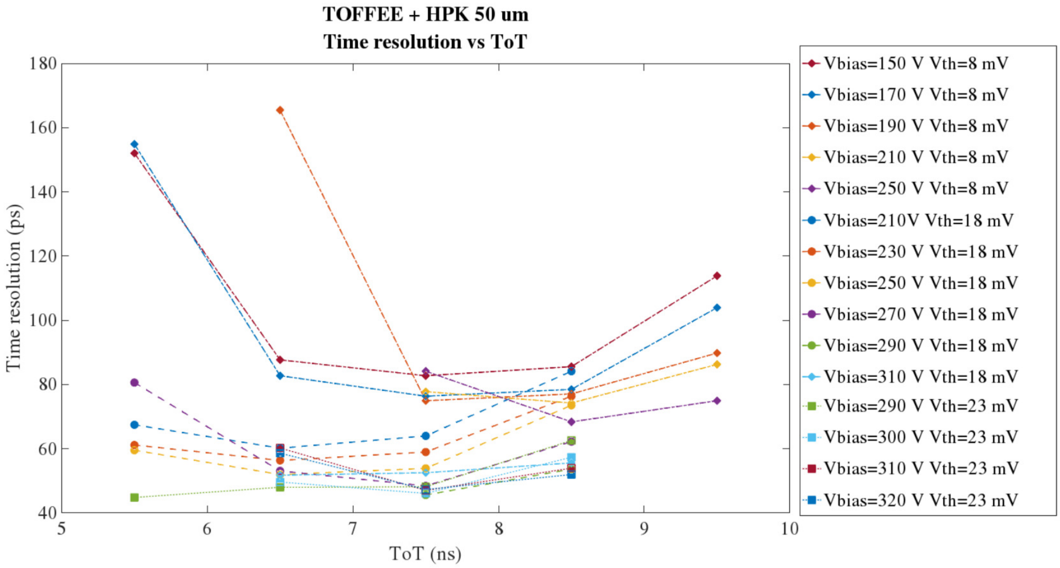
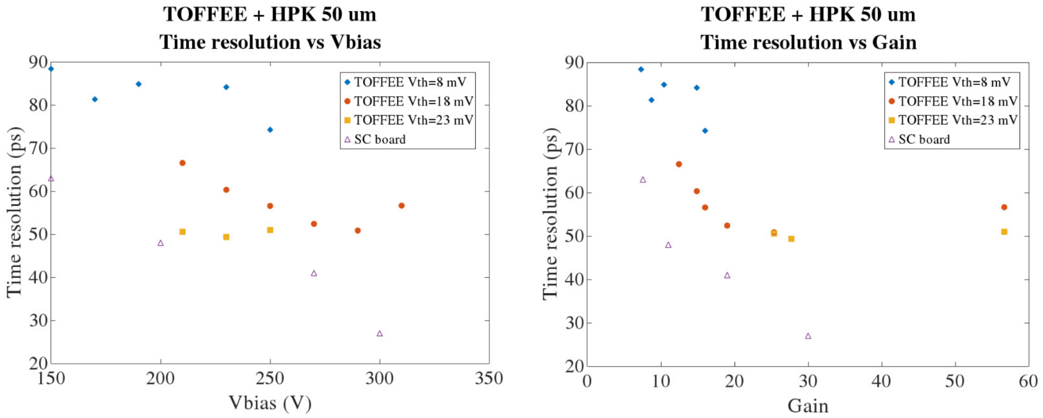


FIGURE 6.17: Time resolution as a function of Time Over Threshold for different sensor bias voltages and discriminator threshold voltages.

6.4.2 HPTDC readout

The coincidence of the two boards has been tested in a successive beam test in October 2017. In this beam test, the data acquisition has been performed by collecting digitized data from the HPTDC with the TOFFEE_DAQ software for threshold voltages of 15, 18, 21 and 27 mV.



(a) (b)

FIGURE 6.18: TOFFEE time resolution compared with the time resolution achieved by a discrete components board (SC board) as a function of the sensor bias voltage (left) and gain (right).

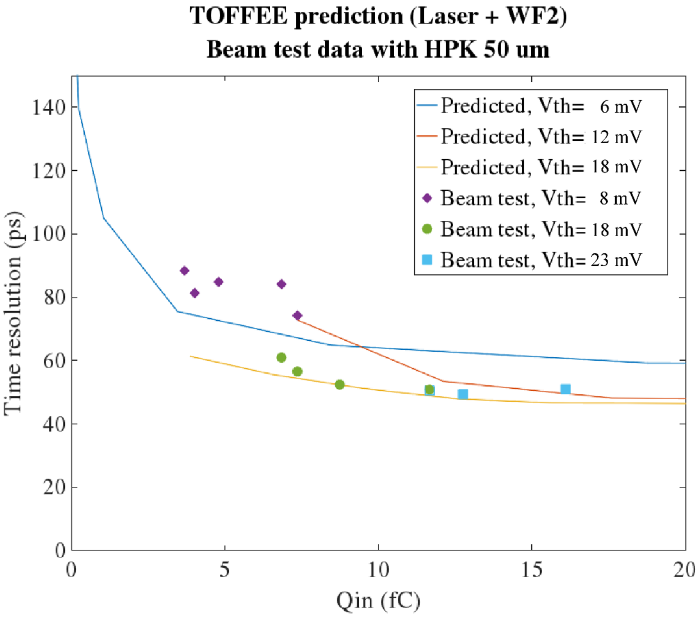
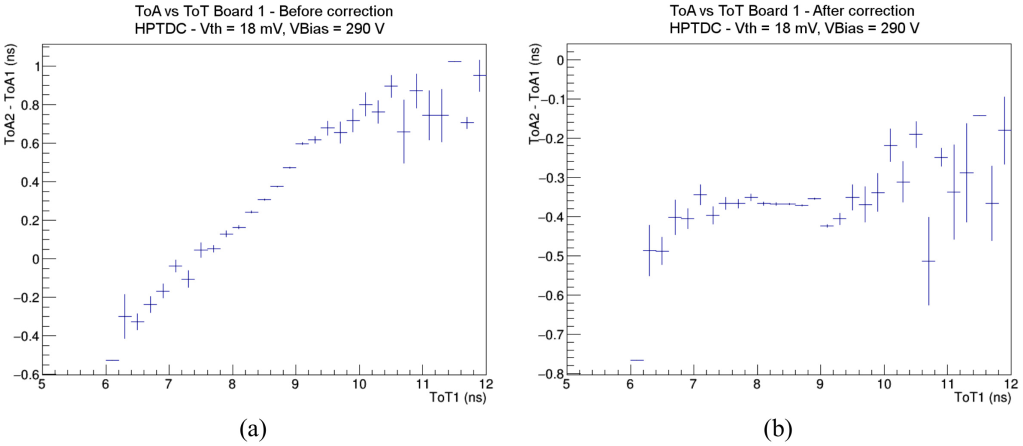


FIGURE 6.19: Comparison of TOFFEE time resolution as a function of the input charge simulated with *Weightfield2* (solid line) and beam test data (points).

The difference in time of arrival ΔToA has a dependence on the time over threshold, as shown in Fig. 6.20.a. This dependence has been evaluated and a correction has been made in order make the ToA as flat as possible with the ToT. The resulting ToA $vs.$ ToT after correction is shown in Fig. 6.20.b. The time resolution after ToT correction for the whole system and for a single board at a threshold of 18 mV for different bias voltages is displayed in Table 6.6, while the ΔToA distribution for the same threshold value and a bias voltage of 250 V is shown in Fig. 6.21: the root mean square of the distribution is $\sigma_{\Delta ToA} \sim$ 84 ps, and therefore the time resolution of the single board is $\sigma_{t,1board} \sim$ 60 ps. It has to be noted that all the previous values include the TDC contribution to the overall time resolution, which is about 30 ps. By removing the TDC contribution to the resolution, the actual timing capability of the



(a) (b)

FIGURE 6.20: Difference in time of arrival as a function of time over threshold for Board 1 before correction (left) and after correction (right).

TABLE 6.6: Time resolution for the 2-board system and for a single board at different bias voltages with HPTDC readout and a threshold of 18 mV.

V_{bias} (V)	Time res. 2 boards (ps)	Time res. single board (ps)
250	84	59.4
270	79	55.9
290	78	55.2
310	89	62.9

UFSD + TOFFEE system can be estimated as

$$\sigma_t^2 = \sigma_{t,1board}^2 - \sigma_{t,TDC}^2, \tag{6.7}$$

which leads to $\sim$ 52 ps in the case when $\sigma_{t,1board} \sim 60$ ps. For the previous ToA distribution, the comparison between different thresholds is shown in Fig. 6.22 for a bias voltage of 290 V and in Fig. 6.23 for a bias voltage of 310 V. It can be seen that increasing the threshold improves the time resolution.

Fig. 6.24 shows a summary of the time resolution as a function of the bias voltage acquired in beam tests. The resolution for different threshold values (8, 18 and 23 mV) obtained with oscilloscope readout is plotted alongside the one achieved with HPTDC and a threshold of 18 mV.

Both laboratory measurements and beam tests led to satisfactory and very promising results: the time resolution of the whole chain composed by UFSD sensor, TOFFEE ASIC and HPTDC has been proofed to be around 60 ps and can be considered in line with the expectations. With a time resolution of $\sim$ 50 ps, UFSD sensors coupled with TOFFEE demonstrate the feasibility of precise timing with silicon detectors. The system represents a valid alternative to the present CT-PPS timing detector and a promising start point for the development of sensors and front-end electronics matching the requirements of the CMS Endcap Timing Layer.

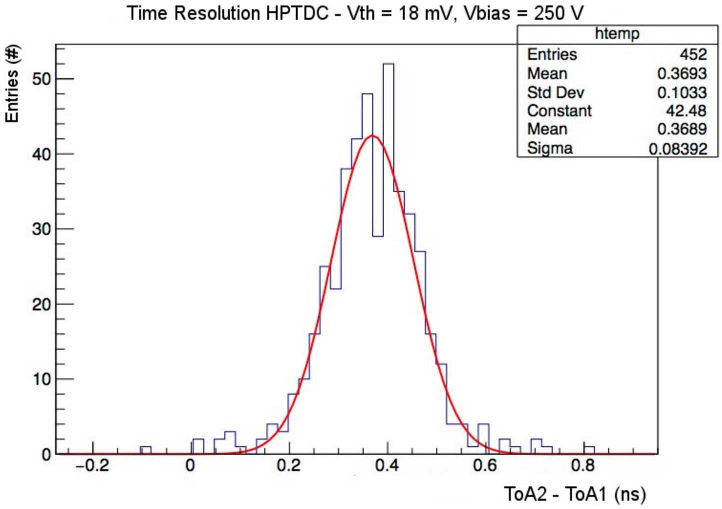
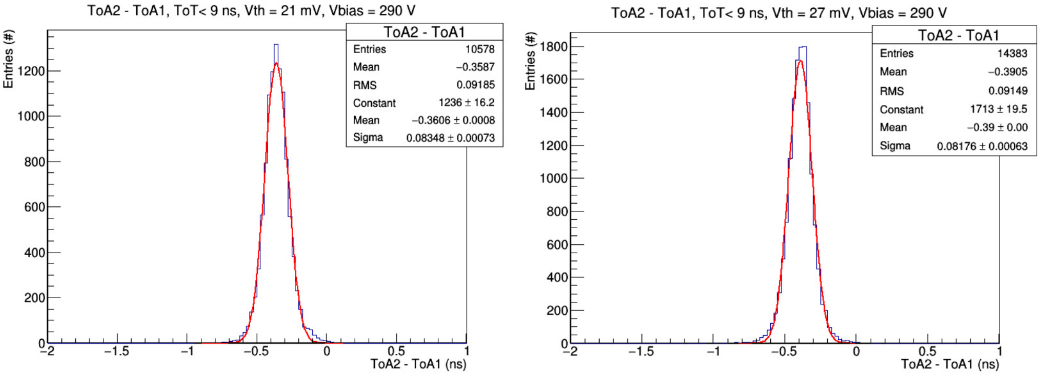


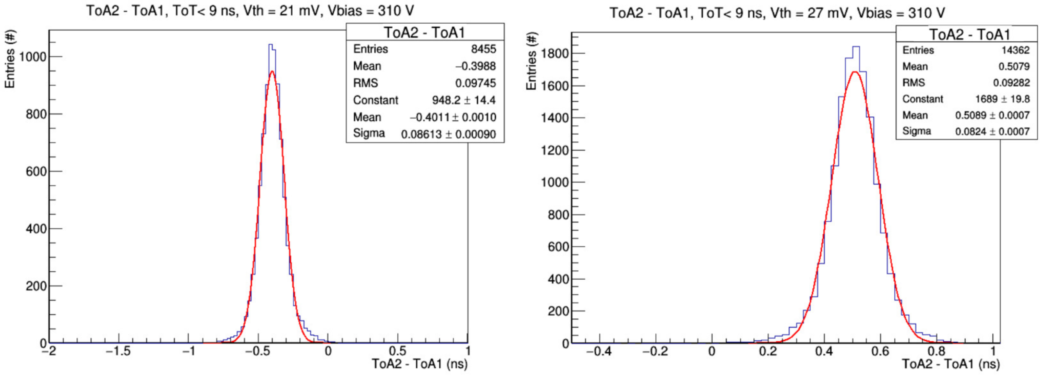
FIGURE 6.21: ToA distribution after ToT correction for the two-board system at a threshold of 18 mV and a bias voltage of 250 V.



(a)

(b)

FIGURE 6.22: ToA distribution after ToT correction for the two-board system at a bias voltage of 290 V for a threshold of 21 mV (left) and 27 mV (right).



(a)

(b)

FIGURE 6.23: ToA distribution after ToT correction for the two-board system at a bias voltage of 310 V for a threshold of 21 mV (left) and 27 mV (right).

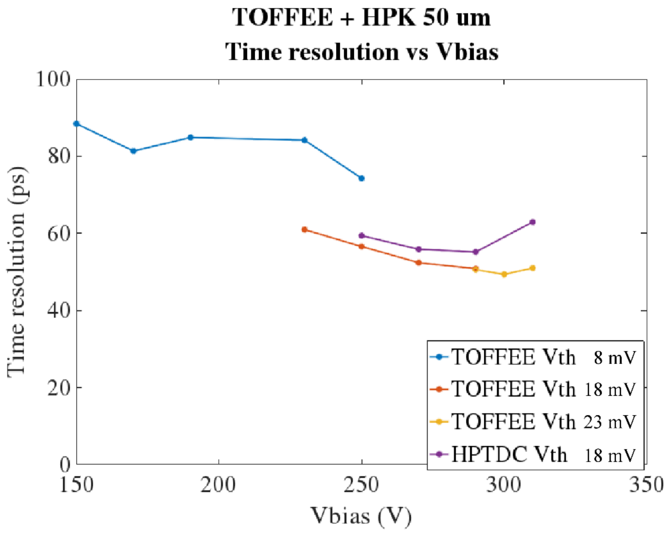


FIGURE 6.24: Comparison of the time resolution as a function of the sensor bias voltage obtained with oscilloscope readout and HPTDC readout.

Conclusions

The High Luminosity LHC will require an upgrade of the present state of the CMS detector. The option of implementing 4-dimensional tracking in order to improve the track-vertex association in presence of high pileup of events has been explored and will drastically increase the precision of the vertex reconstruction if a time resolution of $\sim$ 30 ps can be reached. In this thesis the Ultra-Fast Silicon Detectors (UFSD), i.e. silicon detectors optimized for both tracking and timing purposes have been presented. These detectors feature a thin bulk and an enlarged signal obtained by a moderate charge multiplication that can be achieved by a p^+ implant below the n electrode called "gain layer". The UFSD exploit the multiplication mechanism based on impact ionization that already characterizes devices like avalanche photodiodes (APDs) and silicon photo-multipliers (SiPMs), with the difference that the gain is constrained to some tens in order to maintain low levels of multiplication noise. Electrode segmentation is foreseen to achieve a 4-dimensional reconstruction.

The PhD activity focused on the development of UFSD in collaboration with FBK and of a front-end circuit for precise timing with UFSD. The design of the sensors started with TCAD simulations for the production of a 275 µm thick device with a gain layer doped with boron. The core of the simulation phase is the tuning the gain layer doping, i.e. finding a proper doping concentration for the gain layer. Once chosen the appropriate range of doping concentrations, studies have been performed to reduce the dead area between electrodes. The reduction of the inactive area between electrodes is fundamental as it can easily reach one half of the total area. AC coupled devices have been implemented in this production to achieve electrode segmentation without segmenting the gain layer. In this production, thirteen wafers have been fabricated with five different doses of the gain layer doping and two doses for the p electrode doping.

The second UFSD production is based on the reduction of the bulk thickness to 50 µm, which implied a rescaling of the multiplication layer to obtain a gain between 10 and 20 at a bias voltage of $\sim$ 200 V. Studies on the guardring width demonstrated that its size can be reduced to 10 µm without early breakdown. The UFSD2 production consists in ten wafers doped with boron and eight wafers doped with gallium, out of which six involve the use of carbonated dopant (boron or gallium). Gallium and carbonated gain layer dopants have been introduced in order to study their radiation hardness. In addition, boron-doped wafers come into two different configurations, namely high and low diffusion, corresponding to a different diffusion temperature of the gain layer after implantation.

The TOFFEE ASIC has been designed as readout circuit for UFSD within the CT-PPS experiment. The design is optimized for an input charge ranging between 2 and 60 fC and a detector capacitance of 3-6 pF. Landau distribution for a 50 µm thick UFSD with a nominal gain of 15. The most probable value is located at about 8 fC. The ASIC is composed of an amplifier, a single threshold discriminator, a stretcher and an LVDS driver. The circuit converts the sensor signal into a logic pulse with LVDS levels, which has been intended to be read out by the High Precision TDC.

The ASIC has been submitted to the foundry in May 2016 and received in September.

Laboratory tests have been performed on the TOFFEE ASIC with two types of custom boards: the first has been intended to investigate the operation of the ASIC by means of a test pulse, while the second foresees the readout of a UFSD by TOFFEE. Tests conducted on a 50 µm UFSD by HPK demonstrated that a time resolution of $\sim$ 50 ps for a MIP signal. This time resolution has been confirmed by beam tests performed at CERN SPS on a coincidence system composed by two aligned boards with identical sensors, when read out by a differential probe and an oscilloscope or by HPTDC.

Appendix A

Data acquisition software

A.1 TOFFEE DAQ

The data acquisition software, “TOFFEE DAQ”allows to acquire time measurements with the HPTDC, store them in the TDC data buffer, read the output buffer and write data into a text file.
The CAEN V1718 module, a USB to VMEbus interface is used to interface the computer for DAQ with the VME bus through a USB port and act as master. Communication to the modules requires the CAEN USB driver and the VME library to be installed on the computer. The computer configures the V1290 board (slave) and reads out the HPTDC data buffer.
The V1290 module programming is performed using an on-board microcontroller. In order to send and receive instructions and data to/from the TDC, a 16-bit word (OPCODE) has to be written into the Micro Register, which allows to send instructions to the microcontroller. The OPCODE is composed by an 8-bit command field specifying the operation to perform and an 8-bit object field defining the object over which the operation has to be performed (not all operations require and object specification). The communication with the microcontroller is realized with an “handshake”procedure checking the status of the Micro Handshake Register. This latter is a read-only register using only 2 bits: READ_OK and WRITE_OK. The handshake protocol between the VME and the microcontroller can be summarized into two main operations:

- **Write on Micro Register**:
  - The master module tests if the Micro Handshake Register is ready to be written by checking if the WRITE_OK bit is set to 1;
  - the WRITE_OK bit is automatically set to 0 after the datum is written and will set back to 1 when it can receive another one;
  - as soon as the WRITE_OK bit returns to 1, the VME can write another word into the register.

- **Read from Micro Register**:
  - The VME can read a valid datum if the READ_OK bit of the Micro Handshake Register is set to 1;
  - the READ_OK bit is automatically set to 0 after the datum is read and will set back to 1 when it can be read another one;
  - as soon as the READ_OK bit returns to 1, another word can be read.

Writing procedures often foresee to send two 16-bit words to the Micro Register, the first describing the type of operation and the second specifying the configuration. Once that the first word is written into the Micro Register address, the microcontroller remains in a wait status until the second word is written. Reading procedures also turn the microcontroller in a wait status until the configuration is read.
The data acquisition software is composed of a main program (`TOFFEE_DAQ.c`) and TDC library (`CAENHPTDClib.c`) containing all functions for the TDC configuration and readout used in the main program.
The main program executes these consecutive operations:

- initialization of the VME interface;
- initialization of the TDC board (either V1290A or V1290N);
- configuration of the TDC trigger matching mode;
- configuration of edge detection type and TDC resolution;
- configuration of the readout procedure;
- readout of the TDC output buffer and writing on text file.

The operations performed on the TDC are defined in the TDC library and explained hereafter.

`TOFFEE_HshkedMuCReadCycle` performs the polling of the Micro Handshake Register for READ_OK. If the register is ready to be read, the functions reads the content of this register;

`TOFFEE_HshkedMuCWriteCycle` performs the polling of the Micro Handshake Register for WRITE_OK. In the case of WRITE_OK, the functions writes the instruction into the register;

`TOFFEE_TDClogINFO` performs the first handshake to the TDC board, i.e. setting the board base address, the Micro Register address and the Micro Handshake Register address;

`TOFFEE_TRIGGERset` sets the acquisition mode of the TDC. The trigger matching mode is selected by writing the OPCODE 0x0000 in the Micro Register and detects only events within a time window around the arrival of the trigger signal. The trigger window can be further configured by setting its width with the OPCODE 0x1000 followed by the width value in 25 ps units and its offset with respect to the trigger (OPCODE 0x1100 and after the value). The sum of the window width and the offset should not exceed 1 µs (i.e. 40 clock cycles).
Trigger matching mode needs a trigger signal to be sent to the TDC: the trigger should be locked in phase with a clock signal and both have to be sent to the board from the external. The default procedure used in this DAQ is based on the continuous storage that makes the TDC self-triggering on its own clock;

`TOFFEE_DETECTIONset` configures the parameters for data acquisition:

edge detection the TDC can be programmed to to detect leading and/or trailing edges of the hit signal, or to acquire one leading edge and the corresponding pulse width (pair mode). In order to set the edge configuration, the OPCODE 0x2300 has to be written into the Micro Register address. After that, the edge type has to be specified, i.e. pair mode (0x0000), trailing edge (0x0001), leading edge (0x0010) or both (0x0011).

resolution sets the LSB of the leading/trailing. The first word is the OPCODE 0x2400 while the second is usually set to the default value 0x0011 corresponding to 25 ps, although coarser resolutions can be specified.

channel dead time the double hit resolution can be configured by sending 0x2800 as first word and a second word that can be either the default 5 ns value (0x0000), 10 ns (0x0001), 30 ns (0x0010) or 100 ns (0x0011). It is usually set to 5 ns.

`TOFFEE_DETECTIONread` reads the settings of the detection configuration described in `TOFFEE_DETECTIONset`, namely edge configuration (0x2300), resolution (0x2600) and dead time (0x2900), header and trailer status that has been set in `TOFFEE_ReadOutSET` (0x3200, only for trigger matching mode);

`TOFFEE_ReadOutSET` configures the TDC readout. When working it trigger matching mode, all data are written in the readout FIFO between an header and a trailer. The header contains the event id and a trigger time tag (bunch id), while the trailer contains the event id and the event word count. Header and trailer can be activated with the code 0x3000 and deactivated with 0x3100. In the default program the continuous storage is set and therefore header and trailer are disabled;

`TOFFEE_EventReadOut` checks first for the presence of bus errors and, if no errors are found, allocates memory for the event. The event will be stored into the struct `EVT_ptr` and will contain:

`n_meas` the measure number (a counter);

`tdc_id` the number of the TDC board;

`bunch_id`

`event_id` the

`channel_id` the number of the channel that has seen the event;

`edge` the edge type of the measurement (0 refers to leading, 1 to trailing);

`meas` the measured time value.

`TOFFEE_POLLstatRegReadCycle` performs the polling of the Status Register to check for data in the output buffer. The Status Register is a read-only 16-bit register containing information on the status of the module, such as the presence of data in the output buffer (DATA_READY: 0 if no data are available, 1 if yes), the status of the almost full/full levels of the output buffer (ALM_FULL, FULL: 0 if the level has not yet been reached, 1 if yes), the selected operating mode (TRG_MATCH: 0 for continuous storage, 1 for trigger matching), the status of the TDC header and trailer (HEADER_EN: 0 if disabled, 1 if enabled), presence of errors or bus errors, resolution (RES);

`TOFFEE_OutBufferReadOut` checks for data and writes the content of the output buffer into a file. Executes the polling of the Status Register by calling the function `TOFFEE_POLLstatRegReadCycle`. If data are available, the program allocates memory for them by calling `TOFFEE_EventReadOut`.

A.1.1 Threshold scan with TDC

On the basis of the main DAQ program, an automatic and remotely controlled threshold scan has been implemented. Inside the function `TOFFEE_POLLstatRegReadCycle`

, this program performs a loop on a selected range of threshold values. For each of these values, it executes the a ReadCycle for 1000 times and returns the number of times at which data are found. The threshold is set by a Phython script communicating to the power supply the voltage value to be applied by writing the proper SCPI instruction for “Voltage Set”on the GPIB bus.
In this way, it is possible to extract the band of threshold voltages corresponding to noise or to build an s-curve if a signal is applied.

Bibliography

[1] W. R. Leo. *Techniques for Nuclear and Particle Physics Experiments: A How-To Approach*. 2nd ed. Springer-Verlag, Berlin, Feb. 1994. ISBN: 3540572805.

[2] W. Hammer. In: *Zeitschrift für Physik* 12 (1911), p. 1077.

[3] G. Münzenberg. "Development of mass spectrometers from Thomson and Aston to present". In: *International Journal of Mass Spectrometry* 349-350 (Sept. 2013), pp. 9–18.

[4] J. Stachel. *Detectors in Nuclear and Particle Physics*. University Lecture. 2015.

[5] W. Klempt. "Review of particle identification by time of flight techniques". In: *Nucl. Instr. Meth. in Physics Research A* 443.1-2 (Aug. 1989), pp. 542–553.

[6] M. Bonesini. "A review of recent techniques of TOF detectors". In: *Astroparticle, Particle, Space Physics, Radiation Interaction, Detectors and Medical Physics Applications* 2 (2003), pp. 455–461.

[7] N.Harnew et al. "Vertex triggering using time difference measurements in the ZEUS Central Tracking Detector". In: *Nucl. Instr. Meth. in Physics Research A* 279.1-2 (July 1989), pp. 290–296.

[8] D. S. Bailey et al. "The design and performance of the ZEUS Central Tracking Detector z-by-timing system". In: *Nucl. Instr. Meth. in Physics Research A* 396.3 (Sept. 1997), pp. 329–349.

[9] C. Grupen and B. Shwartz. *Particle detectors*. 2nd ed. Cambridge University Press, 2008. ISBN: 978-0521840064.

[10] N. Cartiglia H.F.-W Sadrozinski A. Seiden. "4-Dimensional Tracking with Ultra-Fast Silicon Detectors". In: *Reports on Progess in Physics* (2017). DOI: `10.1088/1361-6633/aa94d3`.

[11] N. Cartiglia et al. "Tracking in 4 dimensions". In: *Nuclear Instr. Meth. Phys. A* 845 (May 2016), pp. 47–51. DOI: `10.1016/j.nima.2016.05.078`.

[12] H. Spieler. "Fast Timing Methods for Semiconductor Detectors". In: *IEEE Transactions on Nuclear Science* 29.3 (June 1982), pp. 1142–1158.

[13] E. Martin et al. "Review of results for the NA62 gigatracker read-out prototype". In: *Journal of Instrumentation* 7 (Mar. 2012). DOI: `10.1088/1748-0221/7/03/C03030`.

[14] M. Benoit et al. "100 ps time resolution with thin silicon pixel detectors and a SiGe HBT amplifier". In: *Journal of Instrumentation* 11 (Mar. 2016). DOI: `10.1088/1748-0221/11/03/P03011`.

[15] V. Sola et al. "Ultra-Fast Silicon Detectors for 4D tracking". In: *Journal of Instrumentation* 12 (Feb. 2017). DOI: `10.1088/1748-0221/12/02/C02072`.

[16] J. Butler et al. *Technical Proposal for the Phase-II Upgrade of the CMS Detector*. Tech. rep. Technical Report CERN-LHCC-2015-010. LHCC-P-008, CERN, Geneva, June 2015.

[17] S. Fartoukh. "Pile up management at the high-luminosity LHC and introduction to the crab-kissing concept". In: *Physical Review Special Topics-Accelerators and Beams* 17 (11 Nov. 2014), pp. 111001/1–13. DOI: `10.1103/PhysRevSTAB.17.111001`.

[18] G. Pellegrini et al. "Technology developments and first measurements of Low Gain Avalanche Detectors (LGAD) for high energy physics applications". In: *Nuclear Instr. Meth. Phys. A* 765 (2014), pp. 12–16. DOI: `10.1016/j.nima.2014.06.008`.

[19] N. Cartiglia et al. "Design opimization of Low Gain Avalanche Detectors". In: *Nuclear Instr. Meth. Phys. A* 796 (Oct. 2015), pp. 141–148. DOI: `10.1016/j.nima.2015.04.025`.

[20] H.F.-W Sadrozinski et al. "Ultra-fast silicon detectors (UFSD)". In: *Nuclear Instr. Meth. Phys. A* 831 (2016), pp. 18–23. DOI: `10.1016/j.nima.2016.03.093`.

[21] ATLAS Collaboration. *ATLAS Phase-II Upgrade Scoping Document*. Tech. rep. Technical Report CERN-LHCC-2015-020. LHCC-G-166, CERN, Geneva, Sept. 2015.

[22] S. White. "R&D for a Dedicated Fast Timing Layer in the CMS Endcap Upgrade". In: *Acta Phys. Pol. B Proc. Suppl.* 7 (4 Sept. 2014), pp. 743–750. DOI: `10.1016/j.nima.2016.05.114`.

[23] Y. Musienko et al. "Radiation damage studies of silicon photomultipliers for the CMS HCAL phase I upgrade". In: *Nucl. Instr. Meth. in Physics Research A* 787 (2015). New Developments in Photodetection (NDIP14), pp. 319–322. DOI: `10.1016/j.nima.2015.01.012`.

[24] Y. Musienko and A. A. Karneyeu. "Studies of SiPMs for the CMS HCAL Upgrade". In: *PoS (PhotoDet2015)* 073 (2015). International Conference on New Photo-detectors PhotoDet2015 6-9 July 2015.

[25] A. Heering et al. "Effects of very high radiation on SiPMs". In: *Nucl. Instr. Meth. in Physics Research A* 824 (2016). Frontier Detectors for Frontier Physics: Proceedings of the 13th Pisa Meeting on Advanced Detectors, pp. 111–114. DOI: `10.1016/j.nima.2015.11.037`.

[26] M. D. Rolo et al. "TOFPET ASIC for PET applications". In: *Journal of Instrumentation* 8 (Feb. 2013). DOI: `10.1088/1748-0221/8/02/C02050`.

[27] M. D. Rolo et al. "A low-noise CMOS front-end for TOF-PET". In: *Journal of Instrumentation* 6 (Sept. 2011). DOI: `10.1088/1748-0221/6/09/P09003`.

[28] A. Di Francesco et al. "TOFPET2: a high-performance ASIC for time and amplitude measurements of SiPM signals in time-of-flight applications". In: *Journal of Instrumentation* 11 (Mar. 2016). DOI: `10.1088/1748-0221/11/03/C03042`.

[29] Peter Lambert. *CMS-TOTEM Precision Proton Spectrometer Technical Design Report*. Tech. rep. The CMS and TOTEM Collaboration, Sept. 2014.

[30] R. Arcidiacono et al. "A new timing detector for the CT-PPS project". In: *Nucl. Instr. Meth. in Physics Research A* 845 (Feb. 2017), pp. 16–19. DOI: `10.1016/j.nima.2016.05.114`.

[31] F. Anghinolfi et al. "NINO: an ultra-fast and low-power front-end amplifier/discriminator ASIC designed for the multigap resistive plate chamber". In: *Nucl. Instr. Meth. in Physics Research A* 533.1-2 (Nov. 2004), pp. 183–187. DOI: `10.1016/j.nima.2004.07.024`.

[32] E. Garutti. "EndoTOFPET-US a novel multimodal toolfor Endoscopy and positron emission tomography". In: *2012 IEEE NSS/MIC Conf. Record* (2012).

[33] W. Shen et al. "STiC - a Mixed Mode Chip for SiPM ToF Applications". In: *2012 IEEE NSS/MIC Conf. Record* (2012).

[34] A. Di Francesco et al. "Experimental results with TOFPET2 ASIC for time-of-flight applications". In: *Nucl. Instr. Meth. in Physics Research A* (2017). Article in press. DOI: `https://doi.org/10.1016/j.nima.2017.11.034`.

[35] S.M. Sze. *Semiconductor Devices, Physics and Technology*. Wiley, New York, 1985.

[36] H. Spieler. *Semiconductor Detector Systems*. Oxford University Press, 2005.

[37] G. Lutz. *Semiconductor Radiation Detectors*. Springer-Verlag, Berlin, 1999. ISBN: 3-5406-4859-3.

[38] L. Landau. "On the energy loss of fast particles by ionization". In: *J.Phys.(USSR)* 8 (1944), pp. 201–205.

[39] Sergey Vainshtein Michael Levinshtein Juha Kostamovaara. *Breakdown Phenomena in Semiconductors and Semiconductor Devices*. World Scientific, 2005.

[40] B. Jayant Baliga. *Fundamentals of Power Semiconductor Devices*. Springer, 2008.

[41] L. Rossi. *Pixel Detectors: From Fundamentals to Applications*. Springer, 2006.

[42] W. Schockley. "Currents to Conductors Induced by a Moving Point Charge". In: *Journal of Applied Physics* 10.9 (1938), pp. 635–636.

[43] S. Ramo. "Currents Induced by Electron Motion". In: *Proceedings of the IRE* 9.27 (Sept. 1939), pp. 584–585.

[44] *Opto-semiconductor handbook*. Hamamatsu Photonics K.K.

[45] *Avalanche photodiode – A user guide*. PerkinElmer Optoelectronics. 2010.

[46] *Characteristics and use of Si APD (Avalanche Photodiode)*. Hamamatsu Photonics K.K. 2004.

[47] R.J. Mclntyre. "Multiplication Noise in Uniform Avalanche Diodes". In: *IEEE Trans. Electron Devices* 13.1 (Jan. 1966), pp. 164–168.

[48] V. A. J. van Lint. *Mechanisms of Radiation Effects in Electronic Materials*. Wiley, 1980.

[49] M. Moll. "Radiation Damage in Silicon Particle Detectors - microscopic defects and macroscopic properties -". PhD thesis. Hamburg: University of Hamburg, 1999.

[50] E. A. Burke. "Energy Dependence of Proton-Induced Displacement Damage in Silicon". In: *IEEE Transactions on Nuclear Science* 6.33 (Dec. 1986), pp. 1276 –1281.

[51] P. A. Aarnio M. Huhtinen. "Estimation of pion induced displacement damage in silicon". In: *HU-SEFT R-C02* (1993).

[52] G. P. Summers et al. "Damage correlations in semiconductors exposed to gamma, electron and proton radiations". In: *IEEE Transactions on Nuclear Science* 6.40 (Dec. 1993), pp. 1372 –1379.

[53] P. A. Aarnio M. Huhtinen. “Pion induced displacement damage in silicon devices”. In: *Nucl. Instr. Meth. in Physics Research A* 3.335 (Nov. 1993), pp. 580–582.

[54] Konobeyev et al. A. Yu. “Neutron displacement cross-sections for structural materials below 800 MeV”. In: *Journal of Nuclear Materials* 2.186 (Jan. 1992), pp. 117–130.

[55] A. Rivetti. *CMOS: Front-End Electronics for Radiation Sensors*. CRC Press, 2015. ISBN: 9781466563100.

[56] L. Pacher. “Development of Integrated Pixel Front-End Electronics in 65 nm CMOS Technology for Extreme Rate and Radiation at HL-LHC”. PhD thesis. University of Torino, 2015.

[57] S. Garbolino. “Integrated Front-End Electronics for High Precision Timing Measurements with Radiation Detectors”. PhD thesis. Università degli Studi di Torino, 2011.

[58] D. R. Holberg P. E. Allen. *Analog Circuit Design*. Oxford University Press, 1987.

[59] J. Gal. “Investigation of the jitter of the constant fraction timing method based on the comparison of the original signal and the stretched and attenuated one”. In: *Nuclear Instruments and Methods* 133.2 (Mar. 1976), pp. 341–346.

[60] S. Henzler. *Time-to-Digital Converters*. Springer, 2010. ISBN: 9048186277.

[61] P. F. Manfredi. E. Gatti. “Processing the signals from solid-state detectors in elementary-particle physics”. In: *La Rivista del Nuovo Cimento* 9.1 (Jan. 1986), pp. 1–146. DOI: `10.1007/BF02822156`.

[62] G. Lutz et al. “Low noise monolothic CMOS front-end electronics”. In: *Nucl. Instr. Meth. in Physics Research A* 263.1 (1988), pp. 163–173.

[63] E. Gatti et al. “Timing of pulses of any shape with arbitrary constraints and noises: optimum filters synthesis method”. In: *Nucl. Instr. Meth. in Physics Research A* 457.1-2 (2001), pp. 347–355.

[64] G. F. Dalla Betta et al. “Design and TCAD simulation of double-sided pixelated low gain avalanche detectors”. In: *Nuclear Instr. Meth. Phys. A* 796 (2015), pp. 154–157. DOI: `10.1016/j.nima.2015.03.039`.

[65] G. Paternoster et al. “Developments and first measurements of Ultra-Fast Silicon Detectors produced at FBK”. In: *Journal of Instrumentation* 12 (2017). DOI: `10.1088/1748-0221/12/02/C02077`.

[66] P. Fernandez et al. “Simulation of new p-type strip detectors with trench to enhance the charge multiplication effect in the n-type electrodes”. In: *Nuclear Instr. Meth. Phys. A* 658.1 (Dec. 2011). DOI: `10.1016/j.nima.2011.04.056`.

[67] N. Cartiglia et al. “Beam test results of a 16 ps timing system based on ultra-fast silicon detectors”. In: *Nuclear Instr. Meth. Phys. A* 850 (Apr. 2017), pp. 83–88. DOI: `10.1016/j.nima.2017.01.021`.

[68] F. Cenna et al. “Weightfield2: A fast simulator for silicon and diamond solid state detector”. In: *Nuclear Instr. Meth. Phys. A* 796 (Oct. 2015), pp. 149–153. DOI: `10.1016/j.nima.2015.04.015`.

[69] S. Meroli et al. “Energy loss measurement for charged particles in very thin silicon layers”. In: *Journal of Instrumentation* 6 (June 2011). DOI: `10.1088/1748-0221/6/06/P06013`.

[70] N. Cartiglia et al. "The 4D pixel challenge". In: *Journal of Instrumentation* 11 (Dec. 2016). DOI: `10.1088/1748-0221/11/12/C12016`.

[71] R. E. Bank et al. "Transient Simulation of Silicon Devices and Circuits". In: *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 4.4 (Oct. 1985), pp. 436–451.

[72] R. van Overstraeten and H. de Man. "Measurement of the Ionization Rates in Diffused Silicon p-n Junctions". In: *Solid State Electronics* 13.1 (1970), pp. 583–608.

[73] Y. Okuto and C. R. Crowell. "Threshold Energy Effect on Avalanche Breakdown Voltage in Semiconductor Junctions". In: *Solid State Electronics* 18.2 (1975), pp. 161–168.

[74] M. Mandurrino et al. *Numerical simulation of charge multiplication in ultra-fast silicon detectors and comparison with experimental data*. poster. IEEE NSS-MIC 2017. Oct. 2017.

[75] M. Ferrero et al. "Developments in the FBK Production of Ultra-Fast Silicon Detectors". In: *IEEE NSS-MIC 2017 Conference Record* (Oct. 2017).

[76] F. Cenna et al. "TOFFEE: a full custom amplifier-comparator chip for timing applications with silicon detectors". In: *Journal of Instrumentation* 12 (2017). DOI: `10.1088/1748-0221/12/03/C03031`.

[77] J. Christiansen. *High Performance Time to Digital Converter version 2.2*. CERN-EP/MIC. 2004.

[78] *Mod. V1290-VX1290 A/N, 32/16 Ch. Multihit TDC - Technical information manual*. CAEN S.p.A. Nov. 2016.