A study of the effect of parasitic elements on the timing performance of SiPM readout electronics

(Article begins on next page)

09 May 2024

POLITECNICO DI BARI

Department of Electrical and Information Engineering
ELECTRICAL AND INFORMATION ENGINEERING PH.D. PROGRAM

ING-INF/01 - ELECTRONICS

Final Dissertation

_____

A study of the effect of parasitic elements on the timing performance of SiPM readout electronics

_____

by
Pietro Antonio Paolo Calò

Supervisor:
Prof. Cristoforo Marzocca

Coordinator of Ph.D. Program:
Prof. Luigi Alfredo Grieco

Course XXXII, 2016-2019

POLITECNICO DI BARI

Department of Electrical and Information Engineering
ELECTRICAL AND INFORMATION ENGINEERING PH.D. PROGRAM

ING-INF/01 - ELECTRONICS

Final Dissertation

_____

A study of the effect of parasitic elements on the timing performance of SiPM readout electronics

_____

by
Pietro Antonio Paolo Calò

Supervisor:
Prof. Cristoforo Marzocca

Referees:
Prof. Stefano D'Amico
Dr. Andrea Brogna

Coordinator of Ph.D. Program:
Prof. Luigi Alfredo Grieco

Course XXXII, 2016-2019

To my parents

CONTENTS

PREFACE

Technology breakthroughs in the development of devices and systems for light detection and measurement have traditionally supported scientific advance. Today there exist many emerging healthcare, bio-imaging, and life-science applications that require highly specialized detection solutions.

The main aim of this dissertation to investigate the effect of parasitic elements on the timing performance of readout electronics for low light detection system based on Silicon Photomultipliers (SiPMs). SiPMs represent a well consolidated and cost-effective technology, featuring inherent fast timing, sensitivity and high dynamic range that prove beneficial for recent applications demanding excellent timing resolution.

To date, relatively little modelling has been done to understand the impact of nonidealities associated with the interconnections used to couple the photodetector with the front-end electronics.

A systematic approach to design high performance SiPM readout electronics is developed and the results are critically analysed.

Starting from a consolidated electrical model of the complete detection chain, the first part introduces a mathematical study of either a complete and approximate analytical characterization of the most crucial parameters involved in the determination of the timing performance of the system, with special emphasis on the effects of parasitic interconnection inductance, often underestimated.

The second part proposes the complete design of a front-end implemented in a standard CMOS 130 nm process from TSMC, aimed at achieving state-of-art timing accuracy performance when coupled with large area SiPMs, intended for applications such as ToF-PET.

Lastly, the results of the evaluation tests carried out with an experimental electronic board implementing two embodiments of circuital architecture for SiPM interface are discussed.

1 CHAPTER 1

1.1 INTRODUCTION

In classical physics, electromagnetic radiation (EMR) is described as a wave of the electromagnetic field that propagates through space carrying electromagnetic energy. Assuming a sinusoidal plane wave travelling at a fixed wave speed, that corresponds to the speed of light, c, if in vacuum, its frequency is inversely proportional to its wavelength. The wavelength is expressed in meters (m) and the frequency in hertz (Hz), though for the energy is preferable to use a unit named electron volt (eV), which is equal to $1.602 \cdot 10^{-19}$ J and corresponds to the amount of energy gained by one electron when the electrical potential at the electron increases by 1 V. The electromagnetic spectrum (Fig. 1-1) encompasses and groups all types of waves based on their wavelengths.
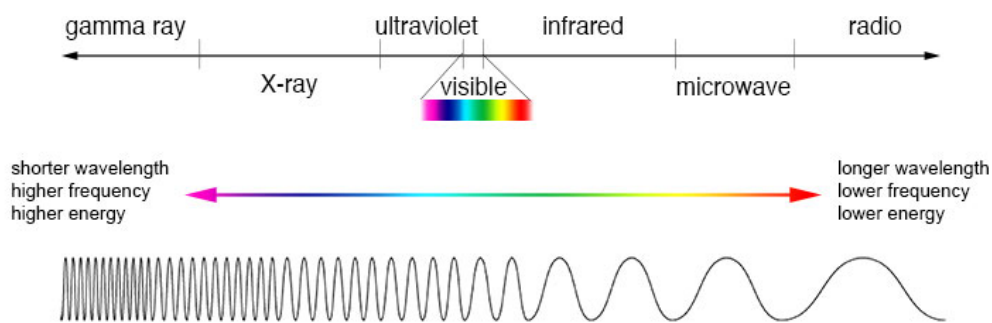
Fig. 1-1: The most commonly used types of EMR (Credit: NASA's Imagine the Universe).

In 1905, Albert Einstein explained the photoelectric effect using the idea that electromagnetic radiation was associated to particles [1], named *photons* by G. N. Lewis in 1926, whose energy can be calculated using Planck's equation:

$$E_{photon} = h\,\nu \tag{1.1}$$

where E_{photon} is the energy of a photon in joule, h is Planck's constant equal to $6.626 \cdot 10^{-34}$ joule per second (J·s) and ν is the radiation frequency in hertz. The photoelectric effect (Fig. 1-2) consists in a metal kicking out electrons when a radiation of enough frequency (regardless its intensity) is incident on its surface.
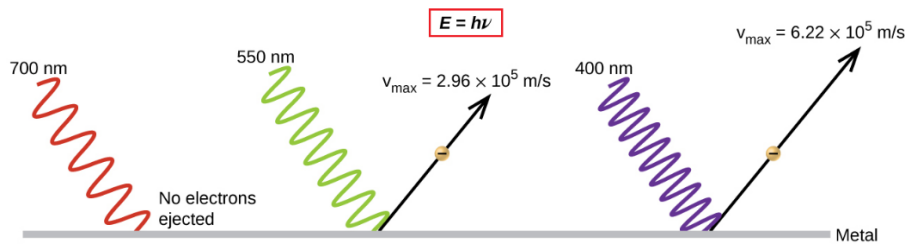
Fig. 1-2: A sketch of the photoelectric effect due to light quantization.

He also demonstrated that the product of the energy of a photon by its wavelength is a constant, meaning that photons with higher energies have shorter wavelengths and vice versa. Radio waves are photons with low energies, microwave photons with a little more energy than radio waves, infrared photons with even more, then come visible, ultraviolet, X-rays, and, the most energetic of all, gamma-rays (γ) with energies that can be as high as millions of electron volts (MeV). X-rays and gamma-rays are preferably characterized through their energy.

On a par of photonic particles, which, however, are electrically neutral and have no weight, tiny fast-moving massive particles such as *alpha* or other heavy charged particles, electrons, and neutrons, also behave as waves. In accordance with wave-particle duality, a fundamental tenet of quantum mechanics, they give off energy in the physical form of particle radiation, the counterpart of electromagnetic radiation associated to photons. Depending on their energy levels, radiation can be categorized in ionizing and non-ionizing. Ionizing radiation (such as X-rays and gamma rays) is more energetic than non-ionizing radiation; consequently, when ionizing radiation passes through materials, it deposits enough energy to break molecular bonds and displace (or remove) electrons from atoms. Other than ionization, interactions of radiation with matter induce photon scattering, absorption and attenuation; new particles can be created whereas existing ones annihilate one with another: when radiation emerges from matter, if it is even going to, its characteristics and properties may not be the same. How it is affected by matter depends on the intensity and wavelength, and the matter itself, that may also undergo deep modifications.

An example of interaction between radiation and matter is **Compton effect**.
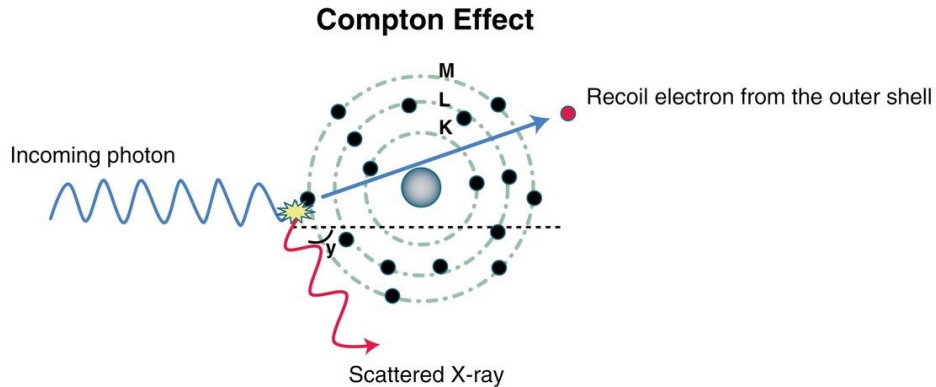
Compton Effect

Fig. 1-3: A sketch of the Compton effect with a photon kicking off a valence electron.

It is a partial absorption process where an original photon interacts with a "free" electron of the outer shell of an atom and loses energy causing the electron to recoil, as shown Fig. 1-3.

The fraction of the photon energy that is transferred depends on the scattering angle. When the incoming photon is only slightly deflected, little energy is transferred to the electron. Instead, maximum energy transfer occurs when the incoming photon is backscattered from the electron and its original direction is reversed. Since all angles of scattering will typically occur, the recoil electrons are produced with a continuum of energies, ranging from near zero to a maximum represented by the backscattering edge.

Another example of interaction is the **electron-positron annihilation** in Fig. 1-4.
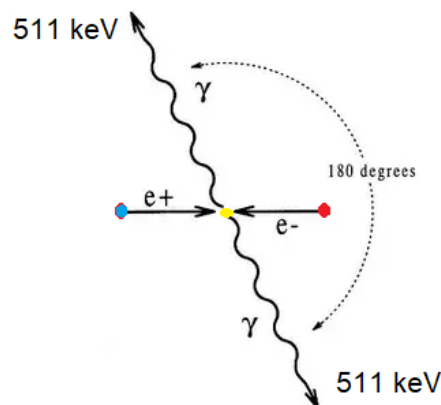
Fig. 1-4: A positron annihilates with an electron producing two anti-parallel γ's.

This is the process in which a positron[1] (or β^+ particle) collides with an electron (or β^- particle) resulting in the annihilation of both particles. They have equal masses that are converted in two annihilation gamma ray photons with an energy of about 511 keV and moving in opposite directions (for the conservation of energy and momentum).

Electron-positron annihilation is the basis of positron emission tomography (PET), a non-invasive medical imaging technique that will be revised in the next sections.

1.2 MATERIALS AND SYSTEMS

The interaction of radiation with matter may be undesirable, but it may also be exploited to construct radiation detectors used for radiation measurement.

Different applications and settings call for different types of detectors, each having various ways it can be specialized to fit a role.

Depending on the specific role, the three categories of detectors most commonly used for research and medical purposes are:

- Gas-filled detectors

- Scintillators

- Solid State detectors and Photomultiplier tubes (PMT).

The latter two are often organized in modules forming the sensors of the scintillation-based radiation detection systems.

The operation of a **gas-filled detector** is based on the ionization of gaseous molecules by incident radiation. A typical example of gas-filled detector is the ionization chamber. This device, shown in Fig. 1-5, resembles a cylindrical capacitor, with a central anode for collecting electrons and an outer cathode for collecting positive ions.

---

[1] The positron is the antiparticle of the electron, coming from β^+-decay.
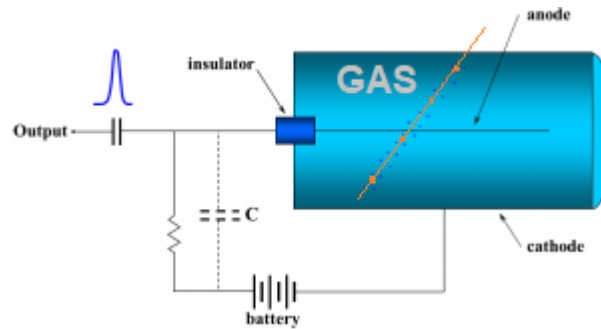
Fig. 1-5: Schematic apparatus of a gas-filled detector.

The ionising particle passes through the gas that fills the capacitor, creating positive ions and electrons that are swept to the electrodes by the electric field. Under different conditions of bias voltage and gas pressure they can be used either as counters (Geiger-Muller mode) or dosimeters; serving as a dosimeter, the target measurement is the net amount of released energy.

A **scintillator** is a crystal that exhibits scintillation, the property of luminescence, when excited by ionizing radiation. Luminescent materials can absorb ionizing radiation energy and scintillate, re-emitting a fraction of the absorbed energy in the form of light photons. The number of photons produced is proportional to the energy of the absorbed primary photon and their emission is governed by the Poisson distribution. The statistical uncertainty that provokes fluctuations in energy measures should scale as the reciprocal of root square of N, where N is the average number of light photons. Therefore, detectors that re-emit the largest number of photons show the best energy resolution, which is defined as the full width at half maximum (FWHM) of the photopeak at a certain energy. Energy can be measured using a scintillation detector that is obtained coupling a transparent scintillator to other types of photodetectors such as a Photo-Multiplier Tube (PMT), a photodiode, or a Silicon Photo-Multiplier (SiPM); being sensitive to visible radiation, they absorb the light emitted by the scintillator and re-emit it in the form of electrons via the photoelectric effect. The subsequent multiplication of those electrons (also called photoelectrons) results in an electrical charge pulse that can then be processed to give meaningful information about the particle that originally struck against the scintillator.

There exist many types of scintillators that differ in materials (e.g. inorganic crystals, organic liquids, noble gases and liquids, plastic scintillators) and exhibit different luminescent characteristics; however, a useful scintillator detector should fulfil the following requirements [2]:

- high light yield, i.e. high efficiency for converting ionization energy to light output [photons/MeV] and proportional response to radiation intensity to achieve the highest energy resolution

- an emission spectrum that matches to the spectral sensitivity of the coupling photodetector

- short decay time of induced luminescence so that fast signal pulses can be generated, and timing resolution improved.

Examples of scintillators used in high-energy physics, spectrometry, radiometry of gamma-radiation, and positron tomography are *LaBr$_3$(Ce)*, *LYSO(Ce)*, and *BGO*.

LaBr$_3$(Ce) - Lanthanum Bromide - is a salt compound of Lanthanum and Bromine and one of a new generation of inorganic scintillator-based gamma radiation detectors. LaBr$_3$(Ce) scintillators have fast light output decay times and provide excellent energy resolution performance.

LYSO(Ce) - Lutetium Yttrium Orthosilicate (Lu$^{1.8}$Y$^{.2}$SiO^5:Ce) - is a relatively new scintillator crystal with a high density, high light output, short decay time and good radiation hardness characteristics.

BGO - Bismuth Germanate (Bi$_4$Ge$_3$O$_{12}$) - is a highly effective gamma-ray absorber because of its high density.

Solid State detectors and **PMT** are suitable for applications that require high speed, low noise, and high gain in the visible light spectrum. They can be used stand-alone or coupled to scintillators to meet today's increasingly stringent demands in many diverse applications, such as medical imaging, nuclear and high-energy physics including the latest cosmic-ray research.

A PMT consists of a photocathode and a series of dynodes in an evacuated glass enclosure, as shown in Fig. 1-6.
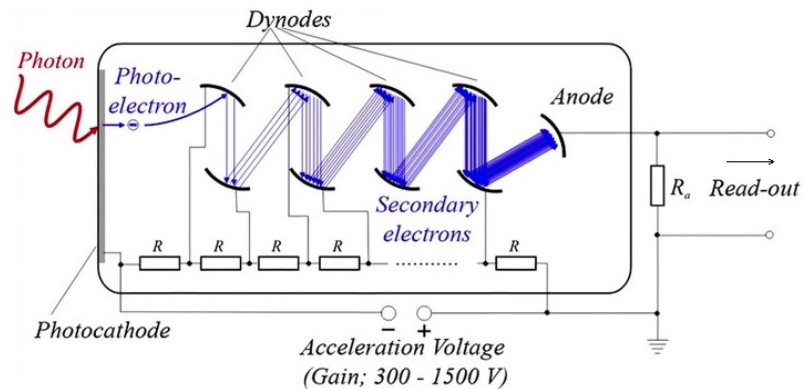
Fig. 1-6: A schematic representation of a PMT.

When a photon of high enough energy strikes the photocathode, an electron (photoelectron) is kicked off due to the photoelectric effect. The photocathode is at a high negative voltage, in the range from 300 to 1500 V. The photoelectron is accelerated towards a series of additional electrodes (dynodes). The dynodes are each maintained at successively lower negative potentials and can generate additional electrons. This cascading effect creates from $1 \cdot 10^5$ to $1 \cdot 10^7$ electrons for each photon hitting the first cathode depending on the number of dynodes and the accelerating voltage. This amplified electrical signal is collected at the anode where it can be processed using an electronic readout equipment.

A solid-state detector exploits the property of a semiconductor crystal to create electron-hole pairs upon radiation interaction [3]. To collect the free charges created before they recombine an electric field must be present throughout the active volume of the device and the subsequent drift of the electrons and holes toward electrodes on the surface of the semiconductor material will generate a current pulse in much the same manner as the motion of ion pairs in a gas-filled sensor. However, using an intrinsic semiconductor realizes a "bulk semiconductor detector" that is far from being an ideal radiation detector because of its leakage current in the absence of radiation that would eventually mask the impulse current level produced by radiation interaction (some tenth of microamperes).

A great improvement occurs with a p-n junction that forms a semiconductor diode. Doping with other materials is often used to change the electrical characteristics of semiconductor materials. For example, Silicon (Si) is a tetravalent element forming four covalent bonds in the crystal structure. When it is doped with a pentavalent element, Group V in the periodic table, it becomes an n-type semiconductor and an energetic *donor* level (site) is established just underneath the conduction band, with the conduction electrons that are completely dominated by the number of *donor* sites. On the contrary, when silicon is doped with a trivalent element such as Boron, it becomes an p-type semiconductor and an energetic acceptor level is established just above the valence band: holes are created in the valence band for approximately every acceptor impurity added and the number of holes in the valence band is increased by the same number of *acceptor* sites.

The diffusion of a given concentration of *donor* impurities on a p-type Silicon creates the p-n junction and a space charge region is established within the conductive, doped semiconductor material that is mobile charge carriers free. There exists an electric field that is normally increased by external reverse biasing. Clearly, if an electron-ion pair is created in that depletion zone, the negative charges are accelerated towards the n-type region and the holes (moving positive charges) towards the p-type, where they can be efficiently collected thus giving rise to a current pulse in the external circuit.

Semiconductor detector are designed to have high detection efficiency and internal signal amplification that guarantee even a very weak radiation to be, first, detected with a high probability, then, suitably recorded.

1.2.1 FAST TIMING APPLICATIONS

In many applications, the goal of a radiation detection system is about either recording impact time and energy of individual interacting photons, or measuring the total amount of energy released by a bunch of particles hitting the detector over a known measurement time, e.g. the scintillation time in a scintillator detector, as in spectroscopy. In the latter case, accurate time of arrival recording of first particles can also be of interest for the scope, whereas in the former case the objective may be to indicate just the rate at which the events are occurring, as in counting applications. The time of arrival of an individual photon is closely related to the time of appearance of a pulse at the detector output: the more stable the measurement of occurred delay, the more likely is that two photons arriving within an extremely short time interval are resolved, i.e. detected as separate entities.

In recent years, emerging techniques in life sciences have turned the spotlight on counting and time measurement of ultra-fast evolving and low-intensity luminous signals with picosecond accuracy [4], while simultaneously being able to perform energy measurement with higher light fluxes. The Time-Correlated Single Photon Counting (TCSPC) is one of the leading techniques which many others rely on: Fluorescence Lifetime Imaging Microscopy (FLIM), Förster Resonance Energy Transfer (FRET), and Fluorescence Correlation Spectroscopy (FCS) are just a non-exhaustive list of them.
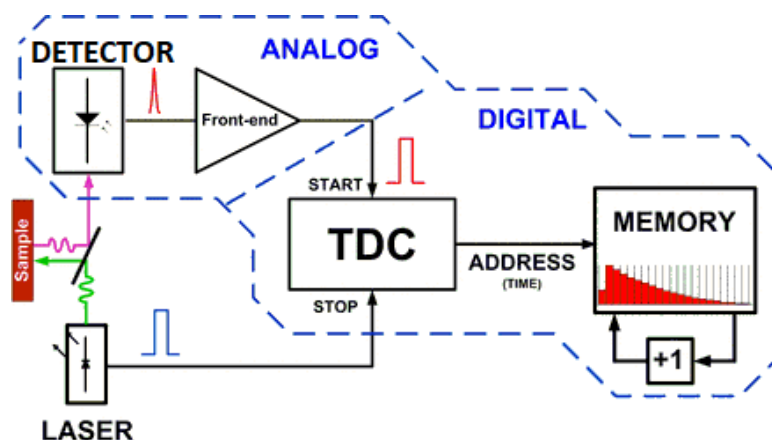
Fig. 1-7: Detection chain for fluorescence decay reconstruction.

In the TCSPC technique, the fluorescence time trace of a sample is recorded by measuring the times of arrival of single photons after excitation by a high repetitive light source. Each event is periodically stored in a memory by adding a '1' at an address proportional to the detection time. After many periods, enough photon events for a required statistical data precision have been collected, and the distribution of the photons over the time in the signal period has been built up. The result represents the waveform of the fluorescence decay pulse. The complete detection chain used to reconstruct the fluorescence decay is shown in Fig. 1-7.

In many nuclear measurements, it is advantageous to be able to determine that two photons are emitted in the same nuclear process and therefore may be sensed by two separate detectors in virtual time coincidence.

Time of Flight (ToF) is one of those technologies that can take medical imaging a step forward. It is used in ToF-PET for better image quality and shorter scanning times than conventional positron emission tomography to the benefit of patients under treatment.
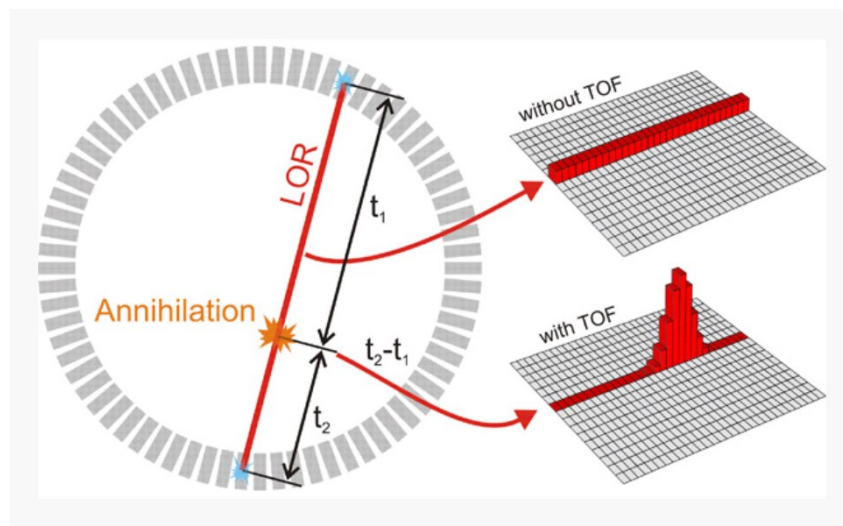
Fig. 1-8: Conventional PET and TOF-PET reconstruction.

In PET application, radionuclides are administered to the patient to localize disease processes inside the body. The molecules of these radioactive substances (tracers) concentrates at specific organs with a certain biological function and decay emitting a positron (β^+). After its release, the positron quickly annihilates with an electron and, preferentially, two 511 keV gamma photons are produced in nearly opposite directions (see Fig. 1-4). The pair of gamma photons is detected by radiation detectors in a cylindrical configuration around the body, thus defining a line, the line of response (LOR), as shown in Fig. 1-8. From many LORs, a map of the tracer concentration is generated using a tomographic reconstruction algorithm; however, as the tridimensional pixels (voxel) along each of the lines are all given the same probability of being the place of annihilation, the reconstructed image will have statistical blurring.

In PET with TOF option, the position of the annihilation event along the LOR is determined by accurately measuring the delay between the flight times of two photons stopped in two detectors of the PET scanner that are supposed in "time coincidence": if the difference of arrival times is a shorter time than a coincidence window (traditionally from 4 to 10 ns), the photons are considered as physically associated to the same annihilation event and the measured time difference is stored. The coincidence timing resolution (CTR) is the resolution with which the TOF difference between the two gamma photons of an annihilation pair is measured by two PET detectors. It is referred to as a FWHM parameter of the Gaussian distribution of time differences and poses the major limitation to the improvement of the spatial resolution in TOF-PET. In fact, to narrow annihilation position down to 1.5 mm of confinement along the LOR, a CTR of 10 ps would be required, which is unimaginable at state of art. Indeed, a CTR of 100 ps ($\sim$ 1.5 cm) has been achieved in TOF-PET using LYSO and LaBr$_3$:Ce scintillators coupled to Silicon Photo-Multipliers (SiPMs) [5],[6].

1.2.2 SiPM AS FAST PHOTODETECTOR

The Silicon Photo-Multiplier (SiPM) is a solid-state photodetector that combines low noise, high gain and fast timing. It represents a rugged, reliable and cost-effective technology for many applications requiring the detection of low light levels. In the last years, remarkable research efforts have been devoted, on the one hand, to improve the basic performance of this kind of detectors, for instance increasing the Photon Detection Efficiency (PDE), and, on the other hand, to reduce the impact of their main drawbacks, such as dark count rate, afterpulsing and optical crosstalk [7]. As a result, the possible application spectrum of SiPM detectors becomes wider and wider [8], covering fields where traditionally they have been considered a valid replacement for PMTs, such as Time of Flight Positron Emission Tomography (ToF-PET) [9] and calorimetry [10], but also more recently emerging areas, such as Light Detection and Ranging (LiDAR) [11] besides Time Correlated Single Photon Counting (TCSPC) [12].

An SiPM sensor is a single-photon-sensitive device that integrates a dense array of Single-Photon Avalanche Diode (SPAD) on common Silicon substrate. It has a high gain and high detection efficiency so that even a single photon can be detected producing an output current pulse. SiPMs can also be arranged to form matrices and tiles as in Fig. 1-9.
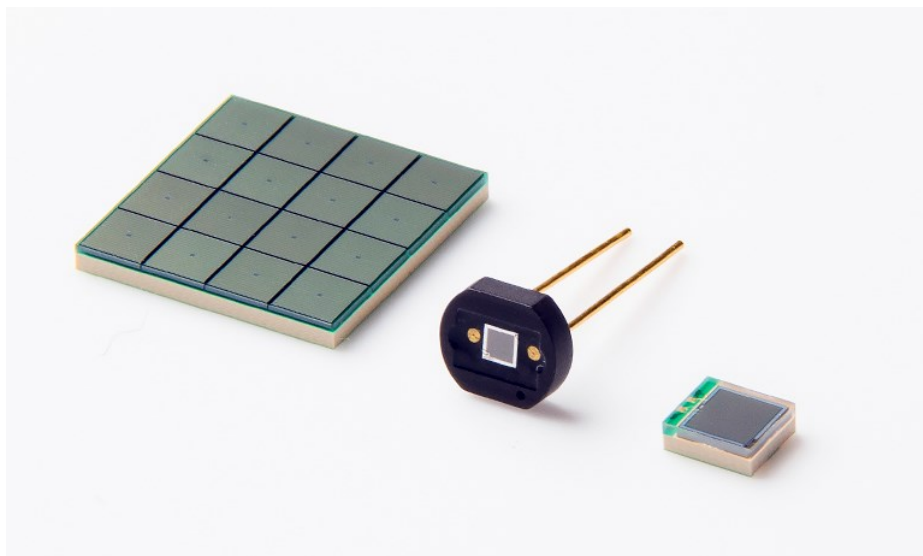

Fig. 1-9: Commercial types of SiPMs (Credit: Hamamatsu Photonics K. K.)

The SPAD is the smallest sensitive element of an SiPM, called micro-cell (or pixel). A typical SiPM has micro-cell densities of between a hundred and several thousand per mm^2, depending upon the size of the micro-cell. Each SPAD sensor is a p-n junction that functions as a photodiode operated in *Geiger Mode*, i.e. working beyond the breakdown voltage of the junction, and integrates a series quenching resistor R_q. When a micro-cell fires in response to an absorbed photon, an avalanche breakdown event is initiated causing a photocurrent to flow through the micro-cell. This results in a voltage drop across the quenching resistor, which in turn reduces the bias across the diode to a value below the breakdown, thus quenching the photocurrent, i.e. stopping the avalanches from going on. Once the photocurrent has been quenched, the voltage across the diode recharges to the nominal bias value. The time it takes for the micro-cell to recharge to the full operating voltage is called the 'recovery' time.

The single SPAD can respond to just one photon a time and, during recovery, it cannot detect other photons. For this reason, a single SPAD cannot count the number of incoming photons unless the photon rate is lower than the inverse of the recharge time that depends approximately on the time constant $\tau_d = R_q C_d$, where C_d is the diode capacitance in reverse bias.

The SiPMs overcome this limitation thanks to the parallel arrangement of several micro-cells, each working independently from another; in fact, the Geiger avalanche will be confined to the single micro-cell it was initiated in, and, during the avalanche process, all other micro-cells of the SiPM will remain fully charged and ready to detect upcoming photons.

When N photons are detected (which means that N photons arrive on N different micro-cells producing N single-cell signals) the SiPM output pulse is N-times larger than the single-cell response, in that, the N independent current pulses just add up at the SiPM terminals.

A circuital arrangement of the SiPM is shown in Fig. 1-10.
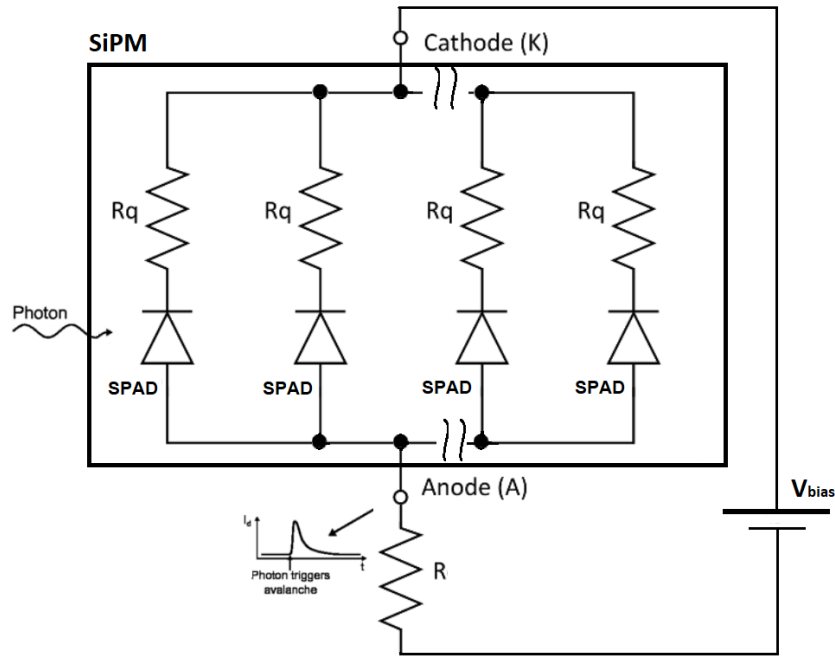
Fig. 1-10: The parallel connection of micro-cells in an SiPM.

Note that both the amplitude and the area of each SiPM pulse, which is the total charge delivered by the detector, are proportional to the number of detected photons (neglecting for the time being intrinsic sources of errors). However, if the number of incoming photons is comparable to the number of micro-cells in the SiPM, the probability that more than one photon hits the same micro-cell becomes relevant and saturation occurs. In this case, the amplitude and the area of the output pulse become a non-linear function of the number of incoming photons.

Leaving aside the open issues regarding the structural and technological optimization efforts to improve the performance of SiPMs, there are many electrical and statistical aspects that should be considered when effective single photon timing is the demanding requirement of an SiPM-based application [13], [14]. Each of them contributes to Single Photon Time Resolution (SPTR), a characteristic parameter of a single-photon detection system which measures the accuracy of the estimate of the arrival time of the single photon. SPTR is also referred to as the time jitter of the detection system.

The same elements that influence fast timing may also have an impact on charge spectrum and, thus, on the achievable energy resolution [15]. Indeed, the charge spectrum represents the probability that the system responds releasing an expected charge amount when stimulated by a bunch of photons carrying a fixed energy, said probability being reported as a function of energy, dosed and swept over a limited range (spectrum).

The most important performance parameter for an SiPM are:

- Gain

- Photon Detection Efficiency

- Primary noise (Dark Count Rate)

- Correlated noise (Afterpulsing, Optical Crosstalk).

Gain.

Every time an absorbed photon triggers an avalanche in the active volume of a micro-cell, a quantized amount of charge is generated by the SiPM. The gain G of a micro-cell, and hence of the sensor, is then defined as:

$$G = \frac{C_d \cdot \Delta V}{q} = \frac{Q_{\mu cell}}{q} \tag{1.2}$$

where C_d is the capacitance of the micro-cell, $Q_{\mu cell}$ is the charge of the micro-cell, q is the electron charge, and ΔV is the excess voltage bias of the SiPM beyond the breakdown voltage of the p-n junction. The sensor output is a photocurrent, and the total charge Q_{tot} generated from an event is given by:

$$Q_{tot} = N_{fired} \cdot G \cdot q \tag{1.3}$$

where N_{fired} is the number of fired micro-cells, G is the gain of the SiPM defined in and q is the electron charge. The total charge Q_{tot} is also equal to the integral of the photocurrent pulse.

Photon Detection Efficiency.

PDE is a measure of the sensitivity of the detector defined as the ratio between the average number of detected and impinging photons. A 100% of PDE is an ideal situation in which all the impinging photons succeed in firing a micro-cell.

The PDE is the product of Quantum Efficiency (QE), triggering probability (Pt), and Fill Factor (FF), as expressed by the following equation:

$$PDE = QE \cdot Pt \cdot FF \qquad (1.4)$$

QE represents the probability that a photon impinging on the SiPM is transmitted to the Silicon, absorbed in the Silicon and finally converted in an electron/hole pair. QE is a function of the wavelength and angular incidence of the incoming photons. The triggering probability (Pt) is the probability that the generated electron/hole pair successfully initiates a self-sustaining avalanche process and thus produces an output current pulse. Pt is a strong function of the overvoltage ΔV, and it increases with ΔV. FF accounts for the fact that each micro-cell in the SiPM has necessarily some dead area on its periphery to make room for the quenching resistor, isolating structures and metal lines for signal routing, which necessarily limit active area. Thus, it is defined as the ratio between the active and the total areas of the device.

The main limiting factor in the PDE is the fill-factor: in fact, for sufficiently high overvoltage, both QE and Pt are close to 1 and PDE saturate to FF.

Primary noise.

In the absence of light, if an electron, or a hole, originates inside the active region of a micro-cell in absence of light, an avalanche is initiated (with probability Pt) and an output photocurrent pulse is observed. This is called a *dark event*. The number of dark events per unit time is the dark count rate (DCR) and represents the main source of noise in an SiPM. In SiPMs, the thermal generation of carriers doubles approximately every 10 °C, and so does the DCR. Moreover, the DCR

scales according to the area of the SiPM and it is an increasing function of the overvoltage ΔV.
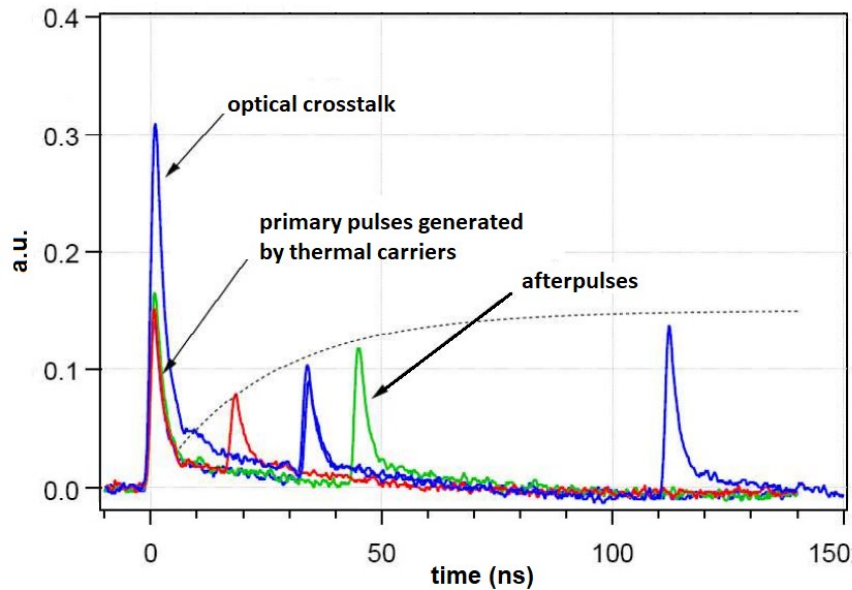
Fig. 1-11: Typical noisy signal shapes of an SiPM.

In an SiPM a dark event is indistinguishable from a photo-generated one. In fact, a dark pulse has the same shape and amplitude of one light triggered photoelectron (pe⁻), thus their integrals are equal to the charge released by the fired micro-cell. In applications where the useful signal produced is always greater than 1 pe⁻, a threshold can be set to, e.g., 1.5 pe⁻ to get rid of the dark events.

Correlated noise.

Both Afterpulse (AP) and Optical Crosstalk (OC) events originate from a primary event (which can be either a photon event or a dark event) and this is the reason for which they are referred to as correlated noise (see Fig. 1-11).

During avalanche, carriers can be trapped in energy states produced by defects in the Silicon. After a delay that can take as many as tens of nanoseconds, the trapped carriers are released, potentially initiating an avalanche and creating an afterpulse in the same micro-cell.

The crosstalk is defined as the probability that photons emitted by a micro-cell undergoing avalanche breakdown cause an avalanche event in other adjacent micro-cells. The process happens instantaneously and therefore, single incident photons may occasionally generate signals equivalent to 2 or 3 photons, or even higher. This effect can be seen in Fig. 1-11, which shows the output of a sensor in the dark. As for AP, in first approximation, OC probability increases more than linearly with the overvoltage and quadratically with the cell size.

Recently, a growing number of low-light applications is demanding SiPMs with large sensitive areas to accommodate for high input dynamic range. This requires either more eventually smaller-sized micro-cells on the same sensor surface or larger sensor size with larger capacitances. This leads to unavoidable degradation of the intrinsic high timing performance of the sensor, because larger capacitances coupled with parasitic interconnections and finite input impedance of the front-end electronics slow down the signals and cause the noise of the whole detection system to increase.

1.3 DETECTION SYSTEMS FOR SiPMs

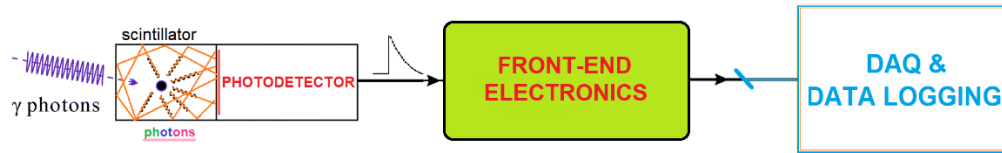A functional block diagram of a typical readout system for light detection is shown in Fig. 1-12.

Fig. 1-12: Functional diagram of the standard readout chain for photodetectors.

Either directly impinging or coming from a scintillator, the light stimulates the photodetector to produce a current signal proportional to the absorbed energy, while the front-end electronics will provide analog signal processing and retrieve suitable timing and pulse height information for digital data processing and offline analysis.

As a photodetector in the visible region of EM spectrum, the bulky PMT has been around for a long time and is still the choice when it comes to detect individual photons, thanks to its extremely high gain and well-established readout schemes. SiPMs are at the forefront of technology advance, though. They feature ultra-fast response and high gain to enable innovation; nonetheless, when large detection area is required it comes at cost of huge additional capacitance, meaning that the SiPM cannot be yet considered a pick-and-replace device. Indeed, to preserve its intrinsic excellent timing performance either a choice of niche solutions commercially available or a clever design of customized front-end electronics must be pursued.

1.3.1 FRONT-END TOPOLOGIES

The earliest stage of the front-end electronics is the preamplifier that interfaces directly with the photodetector to read out its signal. The circuital configuration of the preamplifier is the distinguishing mark of the whole front-end electronics and influences the choice of the remaining blocks of the readout chain to such an extent that, the denominations 'preamplifier' and 'front-end' are often equally

adopted. Typically, an Application Specific Integrated Circuit (ASIC), with much more complex circuitry than the preamplifier alone, is used for amplification and digitization of both energy and timing information. To comply with large SiPM matrices, multi-channel architectures are implemented on a single chip, each channel being devoted to handle a specific macro-pixel (SiPM) of the monolithic detector to which it belongs (up to 256 macro-pixels on a single device). Eventually, whenever too many channels are involved, the organization and management of the information provided by each channel within the detection system may play a key role to attain the desired performance and a dedicated digital macro is embedded on chip for control and communication purpose.

Among the different front-end topologies that can be used to read out a photodetector, the Charge Sensitive Amplifier (CSA) represents the most straightforward solution. The basic CSA principle is shown in Fig. 1-13.
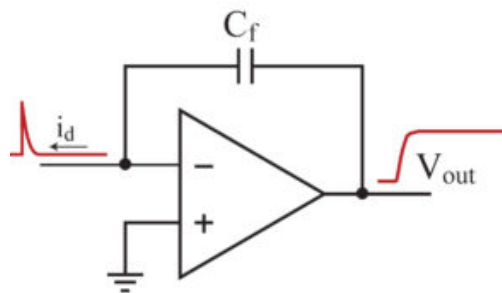
Fig. 1-13: Concept scheme of a CSA.

It is designed to convert an input charge into a voltage output signal, whose peak amplitude is proportional to the charge at the input through the gain factor $1/C_f$, where C_f is the feedback capacitance between the input and output. Despite its potentially good noise performance, it is not necessarily an optimal arrangement for an SiPM. In fact, the charge delivered by the detector can be very large because of its high gain, which often makes direct charge integration on the feedback capacitance impractical. When deep-submicron CMOS technologies are used and the voltage headroom is limited by the low supply voltage, large integration capacitances are needed, to accommodate for an extended dynamic range. For instance, in a typical PET application based on a LYSO scintillator, the number of

photoelectrons contributing to the SiPM signal can be as large as 2000 that, given the gain of the SiPM equals $1 \cdot 10^6$, corresponds to a total input charge of 320 pC. With a 1 V maximum output voltage swing for the amplifier in Fig. 1-13, there would be the necessity to integrate a feedback capacitance of 320 pF, which is not feasible in standard CMOS technologies usually employed to realize multi-channel readout ASICs. Even with more relaxed requirements in terms of maximum input charge, the integration capacitance would still be quite large. This means that, to preserve the fast rising-edge of the signal produced by the detector and to achieve accurate time measurements, the amplifier should be able to drive large values of load capacitance while featuring large bandwidth, which is possible only with ever increasing power consumption. Other issues could also arise related to stability, once again due to large capacitive loads.

The most commonly used approaches to the design of a front-end for SiPM are the voltage mode approach and the current mode approach whose general embodiments are shown in Fig. 1-14.
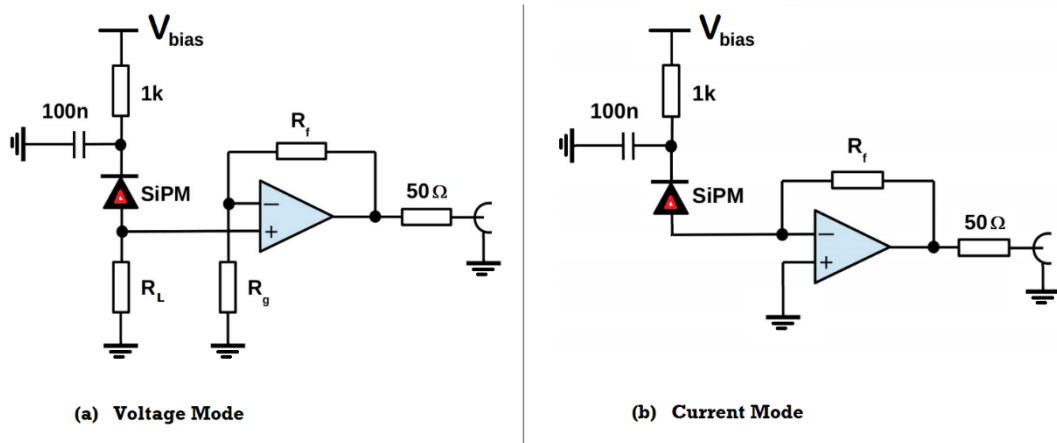
Fig. 1-14: (a) Voltage mode approach; (b) Current mode approach.

In a voltage mode approach, the current pulse of the detector is converted into a voltage pulse by means of the resistor R_L and a voltage preamplifier is used to amplify the signal, whereas in a current mode approach either a current amplifier or a transimpedance amplifier (TIA), as in Fig. 1-14, can be used, the TIA acting as a current buffer implementing a current-to-voltage conversion at the output. The

final decision about the most suitable configuration for optimal timing is influenced by several parameters; however, few simple remarks are now exposed that can successfully drive the choice.

Concerning time measurements, in the great majority of the SiPM applications Leading Edge Discrimination (LED) is adopted as time pick-off method, where the ratio of the rms output noise of the front-end electronics, σ_n, to the slope of the output signal of the front-end around the chosen threshold is evaluated, according to the well-known equation [16]

$$\sigma_t = \frac{\sigma_n}{\frac{dV}{dt}\Big| V_{OUT} = V_{TH}} \tag{1.5}$$

Considering an accurate model of the SiPM [17] coupled to a resistance R_{IN} that mimics the input resistance of the preamplifier, the effect of an avalanche in the SiPM can be read out either in terms of voltage across this resistor or current flowing through it. In the former case a voltage amplifier is intended as input stage of the front-end whereas a current amplifier (or a current buffer that converts the input current into a voltage signal with a resistor at the output) is behind in the latter case.

Fig. 1-15 shows the simplified circuit when a single micro-cell is fired.
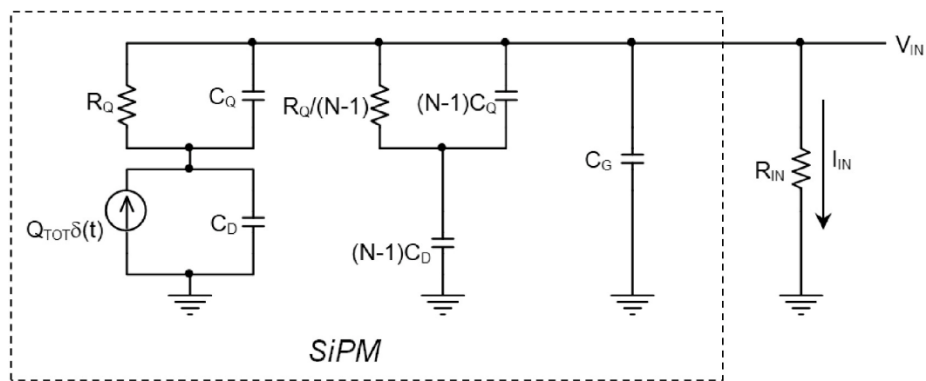
Fig. 1-15: SiPM model coupled to a generic preamplifier with input impedance R_{IN}.

Considering an approximate analysis of the circuit, it can be demonstrated [18] that the contribution to the voltage across R_{IN} can be split up into a fast component

$V_{INF}(t)$ and a slow component $V_{INS}(t)$. The 'fast' charge Q_F is a fraction of the total charge delivered during an avalanche that almost instantaneously flows through C_Q and rapidly reaches the input node of the preamplifier. It is responsible for

$$V_{INF}(t) \cong \frac{Q_F}{C_{HF}} e^{-\frac{t}{\tau_F}} \tag{1.6}$$

where C_{HF} is the high frequency equivalent capacitance of the detector and τ_F, for low values of R_{IN}, is the product of C_{HF} by R_{IN}. Instead, the slow component $V_{INS}(t)$ exhibits a long tail dominated by the slow time constant $t_S = t_R + R_{IN} (C_G + N \cdot C_D)$, where $t_R = R_Q (C_D + C_Q)$.

It goes without saying that only the fast component of the input voltage developed across R_{IN} contributes to achieve a good timing, whereas the slow component has poor relevance in this respect.

What holds true is that a very low input resistance is always preferable for the front-end electronics of a SiPM detector. In Fig. 1-16 the effects of the variation of R_{IN} on the fast and slow components of the signal $V_{IN}(t)$ are highlighted.

Fig. 1-16: Simulation of the fast and slow components of $V_{IN}(t)$ for two different values of R_{IN}.

28

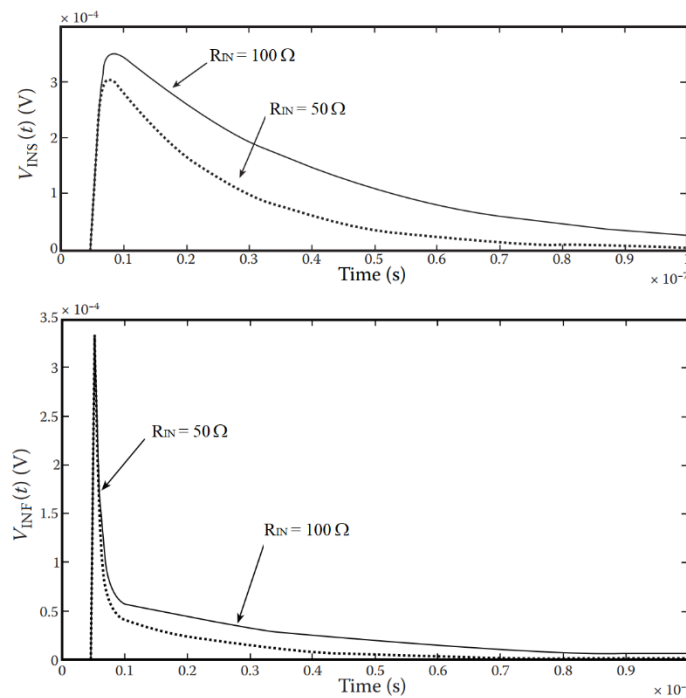If R_{IN} decreases, the contribution of the slow component of the voltage transient response becomes less relevant and the fall time of the fast component gets faster. This means that the charge released by the detector is collected more quickly if the input resistance of the front-end is reduced; in fact, due to a larger discharge current flowing in R_{IN}, the total charge on the equivalent capacitance C_{HF} that is seen in parallel to Rin at higher frequencies, is discharged more rapidly. Moreover, the tail of the signal is apparently slower for larger values of R_{IN}. Consequently, if R_{IN} increases, the rate of the events that the detection system can sustain is reduced, due to possible pile-up effects. In addition, the timing performance deteriorates, especially in case the LED time pick-off method is applied, as an effect of the baseline fluctuations that occur when the signal overcomes the threshold.

The analysis done hitherto refers to mostly ideal conditions. Some important parasitic components may significantly affect, for instance, the rise time of the fast component of the transient voltage that, in the previous analysis, was limited only by the very fast time constants of the avalanche breakdown. A remarkable contribution in this sense comes from the parasitic inductance L_{par} associated to the interconnection between the detector and the front-end electronics [19]. The simulations for Fig. 1-17 are carried out with an inductance L_{par} = 10 nH.
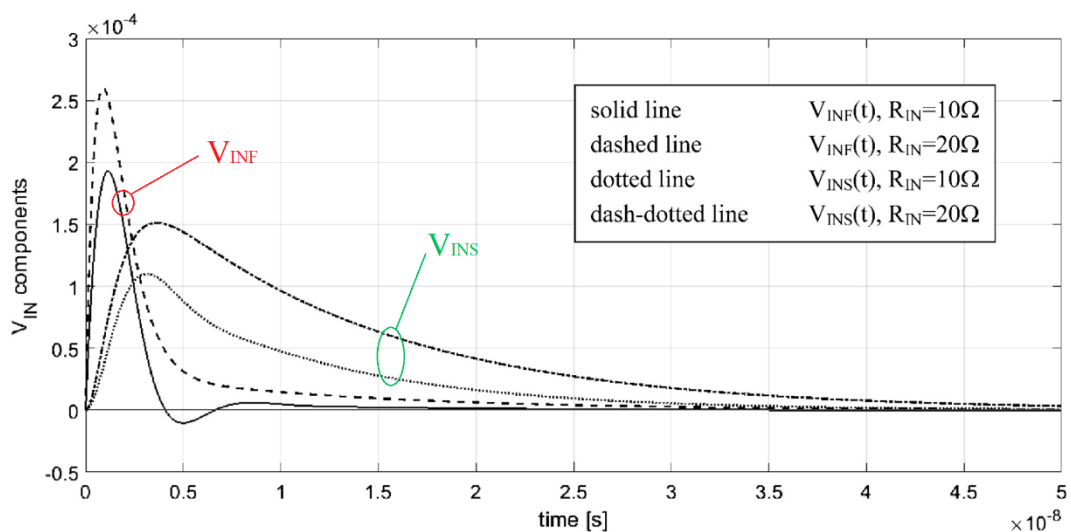

Fig. 1-17: Simulation of the fast and slow components of the voltage signal $V_{IN}(t)$ for two different values of R_{IN}, in presence of a parasitic interconnection inductance L_{par} = 10 nH.

In case of presence of an inductance associated to the interconnections, it is apparent from Fig. 1-17 that the slope of the fast component of the voltage signal, in a voltage mode approach, is larger when R_{IN} is increased, which is good for timing accuracy. On the other hand, this feature cannot be fully exploited because once again increasing the resistance causes slower collection of the charge and longer signal tail, in presence of the parasitic L_{par} as well as in the ideal case, leading to increased pile-up probability.

As far as the current I_{IN} which flows into the input resistance of the preamplifier is concerned, some interesting conclusions can be drawn. The behaviour of the fast and the slow components of this current, $I_{INF}(t)$ and $I_{INS}(t)$ respectively, are shown in Fig. 1-18, obtained with simulations carried out under the same conditions reported for the curves of Fig. 1-17.
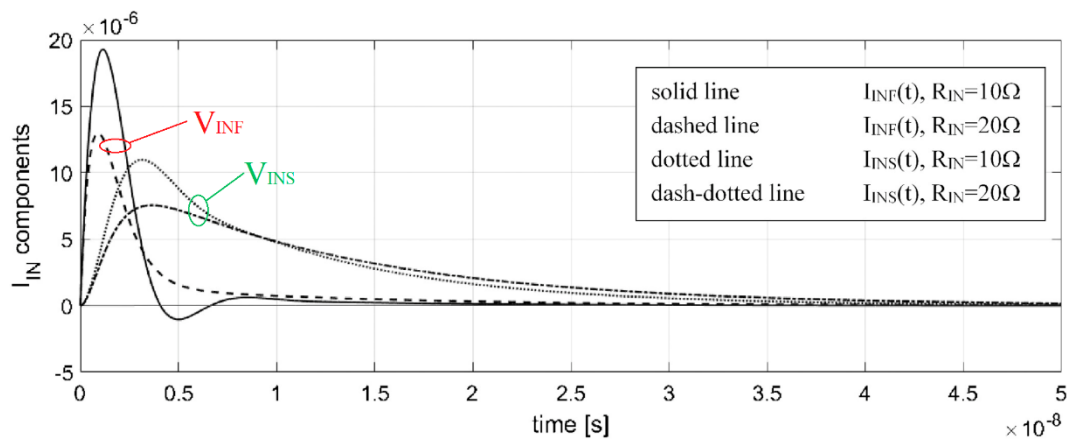
Fig. 1-18: Simulation of the fast and slow components of the current signal $I_{IN}(t)$ for two different values of R_{IN}, in presence of a parasitic interconnection inductance L_{par} = 10 nH.

Lower values of the input resistance correspond apparently to both larger values of the slope of the fast component $I_{INF}(t)$ and shorter tails for both components. This suggests that a very effective read-out approach can be based on a current mode preamplifier which reads the current pulse generated by the SiPM at very low impedance levels, discharging quickly the large equivalent capacitance of the detector, C_{HF}, and reproduces this current on a high impedance node, so that it can be further processed for the extraction of the time and energy information. This is a very common approach for the front-end electronics used for SiPM and

different implementations of this scheme have been applied in several realizations of read-out circuits.

The issues related to the different approaches to the readout electronics for SiPMs, in the light of the characteristics of the detector, for both energy and time measurements will be now discussed [17].

Voltage mode readout approach.

According to Eq. (1.5), to achieve good timing performance, the preamplifier must be able to preserve as much as possible the fast rise time of the input signal at its output, thus it must feature large bandwidth. In case a voltage amplifier is used to read-out the detector, since, as discussed above, R_{IN} cannot be too large, the amplifier should also have sufficient gain, to reproduce an output signal of suitable amplitude, so that it can be conveniently processed by the next blocks, e.g. an integrator and a comparator for energy or time measurements respectively, as schematically depicted in Fig. 1-19.
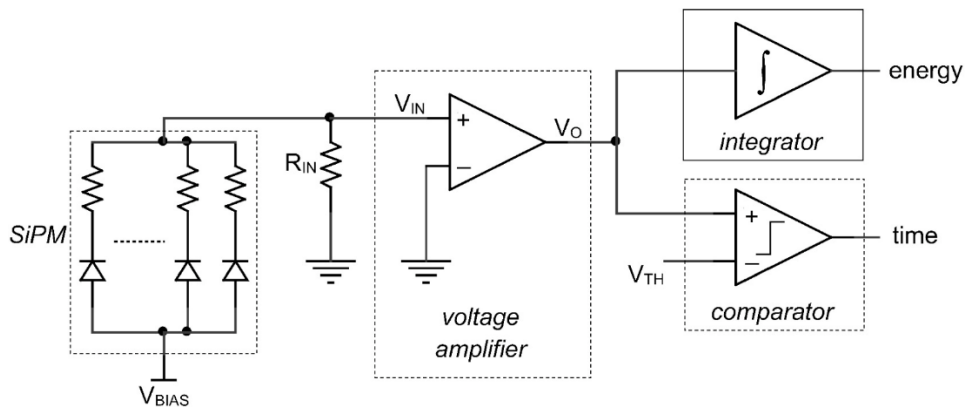

Fig. 1-19: SiPM readout with voltage mode approach.

High gain–bandwidth product is usually a difficult specification to meet if large power consumptions are not put up with, making this approach not effective in applications where very low levels of lights must be detected, timing accuracy is a relevant specification and the number of readout channels is large.

On the other hand, when the dynamic range of the input signal is large, thus low voltage gain values are needed, and the specifications on the time accuracy are relaxed, the voltage mode approach can be conveniently applied.

As far as the noise performance of the circuit in Fig. 1-19 is concerned, the total equivalent input voltage noise of the voltage amplifier, which is typically associated to a common source input transistor, is directly summed to the voltage across R_{IN}, causing a degradation of the Signal to Noise Ratio (SNR) and limiting the maximum timing resolution that is possible to achieve, according to Eq. (1.5).

Current mode readout approach.

As already pointed out, current mode preamplifiers are commonly used to read out SiPM detectors and several implementations with discrete components have been proposed in the literature [20]. Fig. 1-20 shows the basic principle of this approach: a current buffer with very low input impedance is coupled to the detector and exploits the advantages of small R_{IN} values, illustrated at the beginning of the present section.
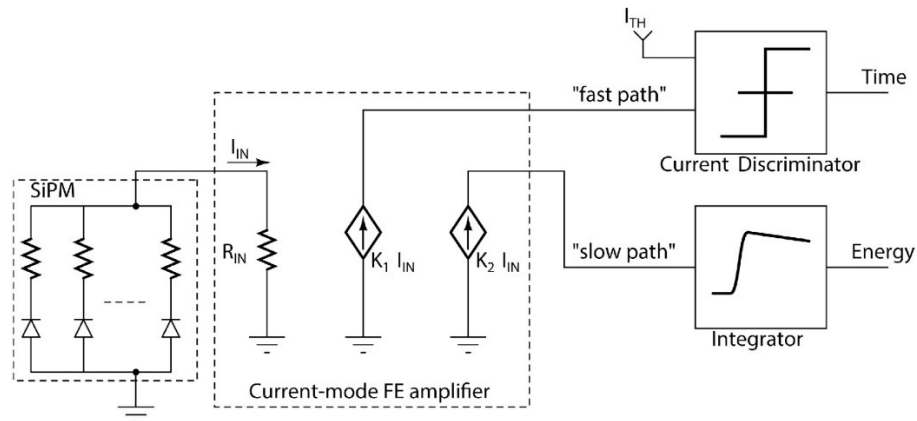

Fig. 1-20: Basic structure of a current mode analog channel for SiPM.

The output signal of the buffer is a high impedance replica of the current pulse generated by the detector that can be easily reproduced with different scaling factors (K1 and K2 in Fig. 1-20) and used to establish different "fast" and "slow" signal paths, optimized for charge or time measurements. Typically, large

bandwidths are easier to be achieved with current mode amplifiers, because of the absence of high impedance nodes, thus the output signal can follow the very fast leading edge of the current pulse generated by an SiPM, resulting in good performance in terms of time resolution. As shown above (see Fig. 1-18) low values of R_{IN} contribute to improve the timing performance of the current-mode readout approach and make it faster the time constant that rules the temporal evolution of the tail of the output pulse. This implication represents an advantage of the current-mode approach over the voltage-mode approach.

1.3.2 TIMING METHODS

The timing information is carried by the leading edge or rising portion of the detector output pulse. To obtain this information, the pulses from the detector are handled very differently than they would be, for example, in a pulse-height analysis system. Several analog standard methods exist that can be profitably exploited to attain accurate measures of time. New trends are coming up that assume waveform sampling and interpolation [21]: these ones revolve around mixed-mode and full digital techniques.

LED is the favourite time pick-off method with SiPMs and in several readout circuits, based on both voltage and current mode, a fast voltage comparator is used to form an output signal with a very sharp transition when the detector pulse overcomes the threshold. This trigger signal marks the arrival of the event and can be time-stamped by means of a Time to Digital Converter (TDC).

In some cases [22], the discriminator is composed by the cascade of low gain, large bandwidth voltage amplifiers, whose overall gain is high enough to generate a fast, full swing output pulse in response to the signal generated by a single micro-cell of the SiPM under avalanche breakdown. Hysteresis can be added to the discriminator by means of a small amount of positive feedback, to avoid undesired output transitions due to the noise.

In current mode front-end circuits, a current discriminator is often used in the fast signal path for time pick-off. A common structure for the fast discriminator is the one proposed in [23], schematically represented in Fig. 1-21.
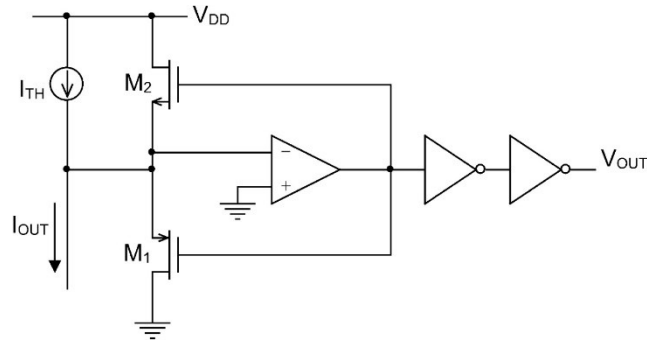
Fig. 1-21: Structure of a current discriminator.

The output current pulse of the front-end is compared with a threshold current, thus avoiding current-to-voltage conversion required with a voltage discriminator. With no signal applied, the PMOS M_1 is switched on and carries the difference between the threshold I_{TH} and the output current of the front-end I_{OUT}, so that the output voltage of the inverting amplifier is grounded. As soon as I_{OUT} overcomes the threshold, due to the arrival of a valid event, M_1 turns off and the NMOS M_2 turns on, thus the output of the amplifier makes a fast transition from low to high voltage level that is sensed and amplified by the cascaded inverters.

Time walk, i.e. the dependence of the time when the threshold is overcome on the amplitude of the signal, is a typical issue of the LED time pick-off technique, and the classic circuit solution for this problem is Constant Fraction Discrimination (CFD). A constant fraction discriminator is a circuit that is triggered at a fixed time after the leading edge of the input pulse has reached a constant fraction of its final amplitude. Therefore, provided the pulses have all the same shape, the time of occurrence of the output pulse is independent of the pulse amplitude and the 'walk' is virtually eliminated. However, the shape or rise time of the preamp output pulses can vary, for example, when a scintillator is used. In this case, the statistical emission of light photons may give rise to wide rise time and shape

fluctuations of the signal and the CFD proves inadequate as a time pick-off method.

Very good time resolution can be obtained with more sophisticated time pick-off techniques, based on digital processing of many samples of the SiPM pulse. The basic circuit structure exploited in multichannel fast digitizer ASICs is represented by the Switched Capacitor Array (SCA), whose simple example of architecture is illustrated in Fig. 1-22 [24].
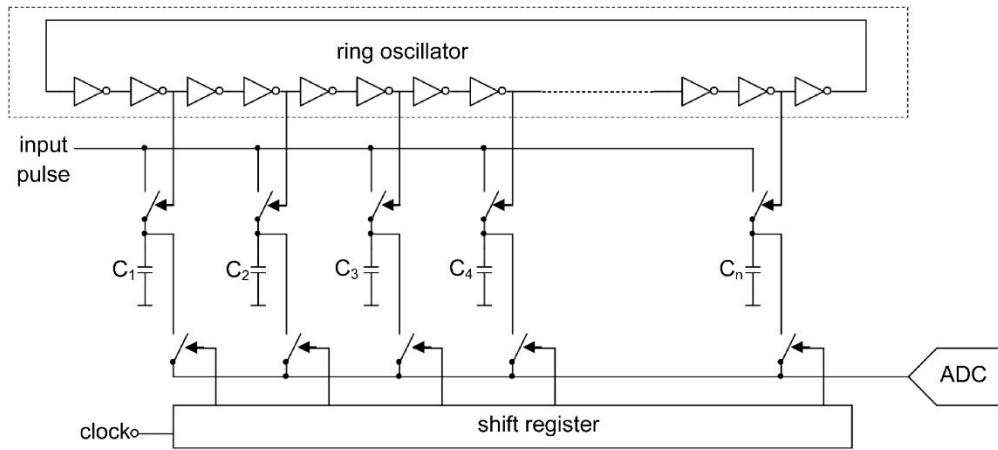
Fig. 1-22: An example of fast sampler: Domino structure.

The sampling signal propagates through the inverter chain, which forms a ring oscillator, and the capacitors are used as analog memories to store the samples of the signal. The depth of the capacitor array allows the storage of the SiPM pulse and, after the sampling phase, the shift register allows the readout of the array towards an ADC, which can either be on-board or off-chip. As an example of performance of this kind of circuits, the DRS4 ASIC hosts 8+1 channels composed of an array with 1024 storage cells; the sampling rate can be varied from 700 MS/s to 6 GS/s, the analog bandwidth is equal to 950 MHz and the power consumption is comprised between 10 and 40 mW/channel, depending on the sampling speed and the selected mode of operation.

Concerning the digital time pick-off methods that can be used if several samples of the raising edge of the detector pulse are made available, there is a broad range of solutions. A large class of the techniques are an extension of the corresponding

analog ones and exploit the samples of the signal to reduce the effects of noise and make more accurate the evaluation of the time when the threshold is crossed. Interpolation of the available samples around the threshold and normalization of the pulse amplitude is often used to increase the total number of samples and improve the accuracy in the determination of the threshold crossing time.

Another class of methods is based on true digital algorithms. To reconstruct the start time of the event from the start time of a 'golden' reference signal, a possible approach tries to find a matching between a reference pulse, evaluated by means of real data or by theoretical analysis, and the measured samples.

1.4 STATE-OF-ART IN FRONT-END ELECTRONICS FOR SiPMs

Several examples and architectures of front-end circuits for SiPMs have been proposed over the recent years. One of the most appreciated implementations based on the current mode approach is the front-end of the NINO ASIC [22], schematically depicted in Fig. 1-23.
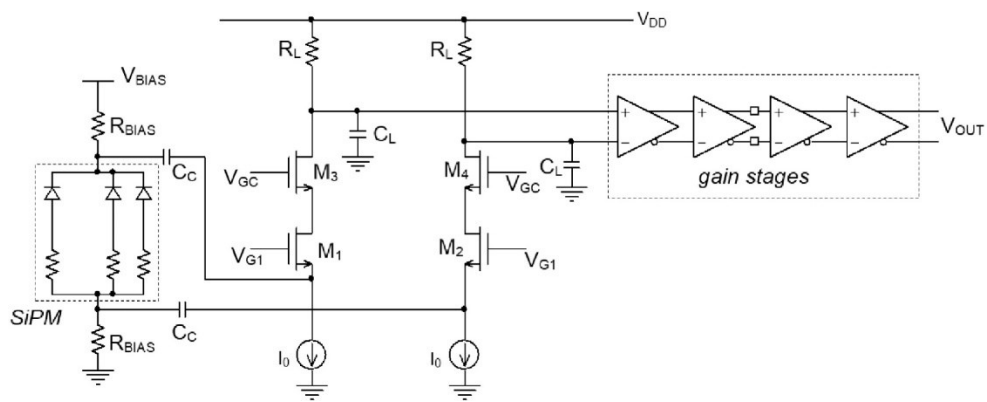
Fig. 1-23: Analog channel of the ASIC NINO.

A fully differential configuration is employed to increase the immunity of the circuit against power supply and ground noises. Cascode transistors, M_3, M_4, are used to decouple the drains of the input common gate MOSFETs, M_1, M_2, from the output nodes and to increase the output resistance of the current buffer. The current signal of the detector is converted into a voltage by means of the load resistors R_L, which form the dominant time constant of the circuit (about 760 ps) with the load capacitance C_L. The input resistance of the stage is set by the

transconductance g_m = 50 mA/V of the common gate MOSFETs, corresponding to a total differential resistance seen by the SiPM of 40 Ω. The open loop configuration of the front-end makes the circuit very fast and free from any stability concerns. When coupled with a commercial 3 × 3 mm^2 SiPM by Hamamatsu with an overvoltage of 1.5 V, the rise time of the preamplifier output signal is equal to 1.5 ns for a single fired micro-cell and the estimated time jitter, with a signal-to-noise ratio of 15, is 100 ps rms [20]. The front-end is followed by four differential gain stages, each with a voltage gain of 6 and 500 MHz bandwidth, which form the discriminator, while the power consumption of the preamplifier in combination with the discriminator is 20 mW. Recent measurements carried out by coupling the detector to a Hamamatsu S13360-3050CS SiPM biased at 62 V, achieve a remarkable Single Photon Time Resolution (SPTR) of 64 ps rms [25].

Another dedicated readout chip for fast timing applications in time-of-flight medical imaging and particle physics experiments is STiC3 [26]. It also uses a differential front-end based as well on an open loop input common gate stage and a load resistor to form a voltage signal that is compared to a threshold by means of a fast comparator for the generation of the trigger signal. In the last version of the circuit, the load resistor has been implemented by means of a diode connected MOSFET as depicted in Fig. 1-24 (only one branch of the differential structure is shown).
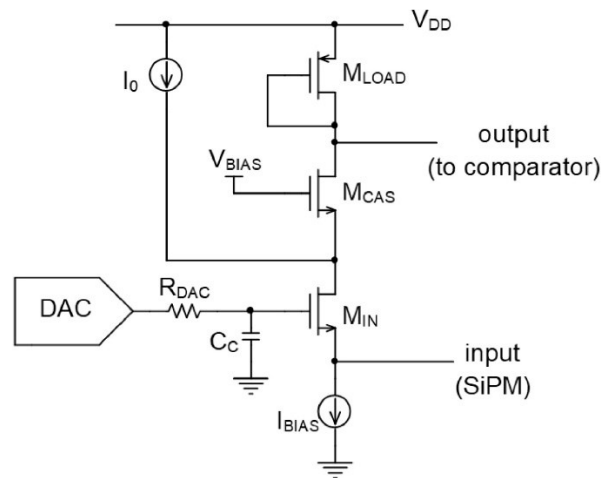
Fig. 1-24: Front-end of the ASIC STiC (half side of the differential current mode preamplifier).

37

The capacitance C_C allows keeping the value of the input resistance R_{IN} close to the reciprocal of the input transconductance also at high frequencies, mitigating the effects of the output resistance of the Digital-to Analog Converter (DAC) used to fine-tune the bias voltage of the SiPM. The results in terms of SPTR are comparable to the ones reported for the NINO ASIC: using the same detector (Hamamatsu S13360-3050CS) an SPTR of 67.1 ps rms has been achieved, with a total power consumption of 25 mW/channel [27].

A further example of read-out ASIC intended for SiPM detectors and based on a current buffer (current-mode approach) is the TOFPET2 circuit [28]. The front-end exploits the feedback on a Regulated Common Gate (RCG) input stage to lower the input impedance and can be coupled to both n-on-p and p-on-n devices: Fig. 1-25 shows the basic circuit used for one of the two SiPM polarities.
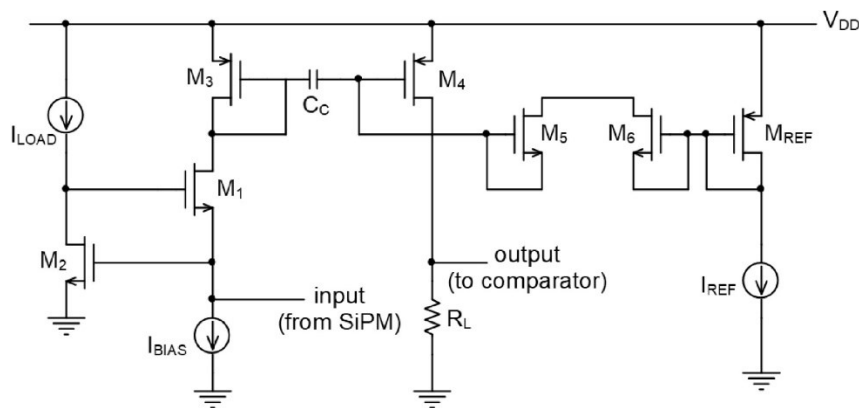
Fig. 1-25: The RCG preamplifier of the ASIC TOFPET2.

The load of the regulated common gate is the diode-connected PMOS M_3, which is AC coupled to a common source amplifier with passive load, to convert the current pulse of the detector into the input voltage of the discriminator. The operating point of the common source M_4 is settled by means of the voltage reference formed by I_{REF} and M_{REF}, connected to the gate of M_4 via a large resistor, in the order of giga-ohms, realized by the back-to-back cut-off MOSFETs M_5 and M_6. The same arrangement, basically consisting in an AC coupled current mirror, is replicated to obtain different signal paths with suitable scaling factors, to be used for energy and time measurements. As an example of the timing

performance of the TOFPET2 ASIC, an SPTR of 95 ps *rms* has been measured using a fast laser source and one of the SiPM of the Hamamatsu 4 × 4 array HPK S13361-3050AE-04, biased at 7.5 V overvoltage. The overall power consumption of the ASIC has resulted in less than 10 mW per channel.

Lastly, an example of voltage mode preamplifier for SiPM detectors is the front-end of the first version of the ASIC PETA [29] designed as an evolution of a previous circuit intended for photomultiplier tubes. In this circuit, the fast signal path used for time measurements is based on a fully differential voltage amplifier composed of the cascade of 5 low-gain, high speed differential gain stages with diode loads, to maximize the bandwidth (see Fig. 1-26).
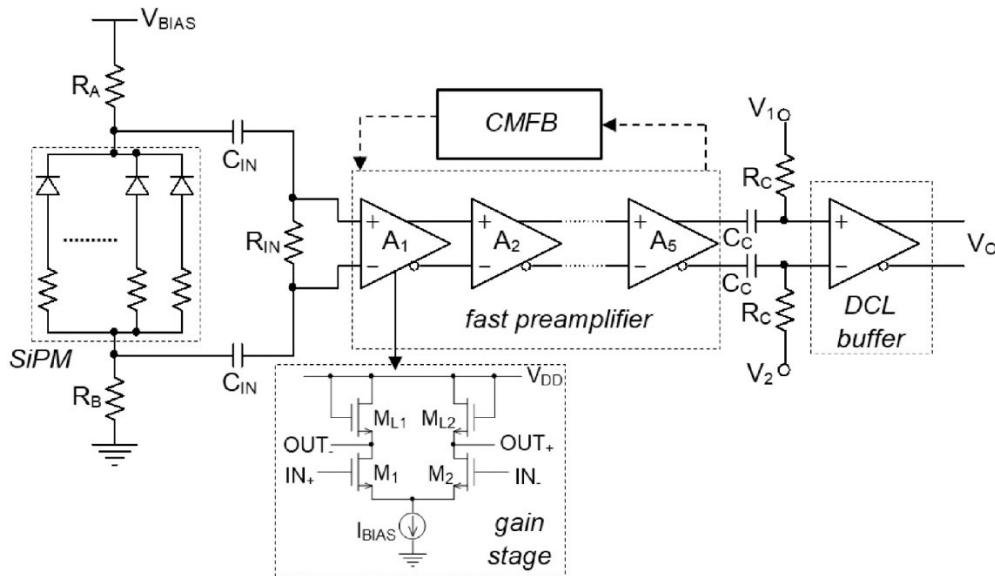
Fig. 1-26: Fast path of the ASIC PETA.

The fully differential structure guarantees immunity from common mode noise and is less sensitive to ground bounce and noise coming from the switching of digital parts, but more off-chip passive components are needed (AC coupling) and the input pad number of the ASIC is doubled. The preamplifier reads out the signal across an internal, adjustable termination resistance with nominal value of 50 Ω. The gain of the preamplifier is 20 V/V and its maximum bandwidth is 900 MHz, which can be limited in two steps by means of a low pass filter. A slow common mode feedback block is also used to stabilize the common mode of the

preamplifier and a differential current mode logic (DCL) buffer, AC coupled to the preamplifier, acts as a fast discriminator.

The PETA3 version of the ASIC provides very interesting results in terms of time resolution. In fact, a Coincidence Time Resolution (CTR) of 190 ps FWHM can be obtained by coupling two channels of the ASIC to an FBK 3×3 mm^2 SiPM, used to read out $3 \times 3 \times 5$ mm^3 LYSO scintillators exposed to 511 keV γ-rays; however, the power consumption of 32 mW/channel marks adversely the overall outstanding timing performance of the ASIC.

Table 1-1 is a short data collection which covers some readout ASICs that have been developed over the last decade, regarding their main features.

Table 1-1: Some examples of the readout electronic chips with different approaches.

ASIC	Preamplifier (approach)	Timing meas. (accuracy)	Energy meas. (Dynamic Range)	Power Consumption
NINO	Current	64 ps *rms* (SPTR)	2000 pe$^-$	20 mW/ch
STiC3	Current	67 ps *rms* (SPTR)	2500 pe$^-$	25 mW/ch
TOFPET2	**Current**	**95 ps rms (SPTR)**	**7500 pe$^-$**	**10 mW/ch**
PETA	Voltage	190 ps *rms* (CTR)	8 bit ($\log_2$(MAX/MIN))	32 mW/ch

It can be concluded that the current mode approach is the most viable solution for the readout of SiPMs, especially in applications where timing resolution is of interest, as, for instance, ToF-PET. TOFPET2 is the benchmark for next generation ASICs.

1.4.1 BASIC64

BASIC64 is 64-channels front-end ASIC designed to detect $511 \; \mathrm{keV}$ gamma rays in Positron Emission Therapy (PET) systems when coupled to an array of SiPMs fitting to pixelated Lutetium Fine Silicate (LFS) 16x16 scintillator matrix [30].

Each channel (see Fig. 1-27) implements a simple current buffer in common gate configuration and can perform both time and energy measurements with adequate timing accuracy (about 300 ps FWHM) and extended dynamic range of about 400 pC (2500 pe$^-$ for an SiPM gain of $1 \cdot 10^6$), as guaranteed by simulations.
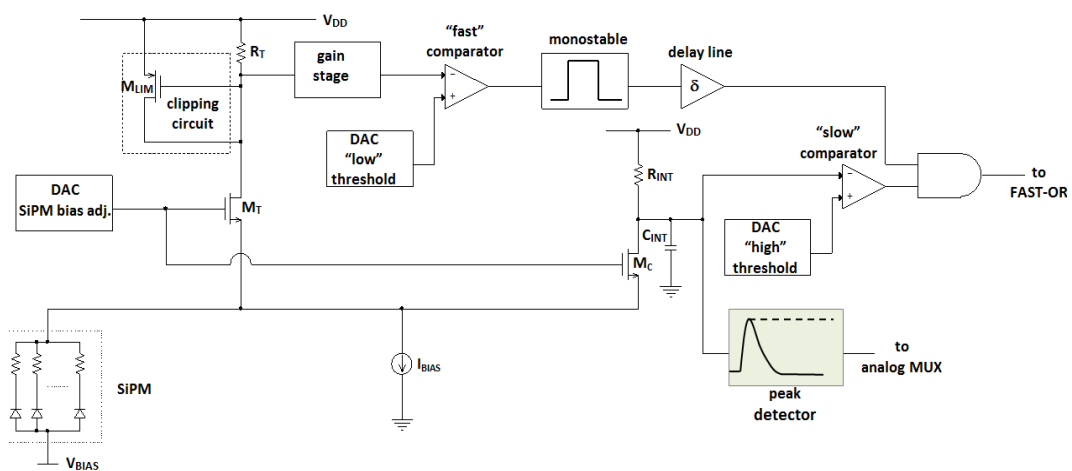
Fig. 1-27: Structure of the analog channel of BASIC64.

It features a double-threshold technique for dark pulse rejection, an internal 8-bit Wilkinson ADC for digital data of energy information and is capable to manage three different triggering schemes and to eventually issue an external trigger for timestamping the earliest arrival time of a scintillation event.

It was considered eligible as a case study to understand the reliability of the simulations in presence of a parasitic inductance as large as 100 nH.

Being originally intended for PET, it was proven by measurements that it is well suited to timing, so far [31]. In fact, it features about 40 ps rms time jitter for charge greater than 5 pC (roughly 30 pe$^-$ piled-up), regardless of a preamplifier input resistance that can be made as large as 50 Ω.

TEST SETUP AND MEASUREMENTS

The ASIC has been characterized in terms of time jitter and charge dynamic range. The experimental test setup in Fig. 1-28 includes a prototype test board for the front-end ASIC, a fast step voltage generator board (PULSER) for charge injection and a digital oscilloscope for signals visualization and data acquisition.
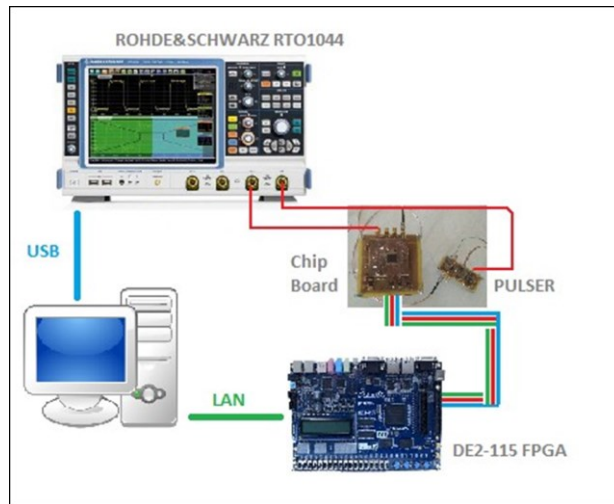
Fig. 1-28: Experimental test setup.

The ASIC configuration, the external control signals required by the readout procedure and data acquisition are managed by means of an FPGA development board. The PULSER, which reproduces the effect of charge injection from the SiPM using a 1000 pF series capacitor, makes it possible to vary the injected charge in the range from 16 pC to 576 pC with the resolution of a 16 bit DAC. A charge injection test has also been carried out for timing characterization using a series capacitor equal to 330 pF to inject low levels of charge. The results of experimental test and characterization for timing accuracy and gain linearity are presented and confirm that BASIC64 is suitable for PET applications and can be further developed to comply with ToF-PET requiring moderate levels of timing accuracy.

To better understand the influence of the parasitic interconnection components on the circuital parameters that influence time jitter a theoretical analysis will be developed, and the design of a new front-end architecture will be discussed.

2 CHAPTER 2

Moving from the electrical modelling of a simple photo-receiver based on SiPM, the essential mathematical expressions that make a good estimate of the system timing performance have been derived and validated by comparison with numerical simulations.

2.1 INTRODUCTION

Accurate time measurement is becoming the design challenge for a growing number of photo-detection systems in applications such as time-of-flight positron emission tomography (ToF-PET) [1], γ-ray spectroscopy [2], time-correlated single-photon counting (TCSPC) [3] and distance measurements (LIDAR) [4], where the resolution required to estimate the time of occurrence of a valid event can be in the order of 100 ps or even less. Silicon Photomultipliers (SiPMs), are becoming the detectors of choice for this kind of applications, due to their intrinsically fast response, characterized by sub-nanosecond rise times and single-photon sensitivity. Unfortunately, several sources of uncertainty can undermine the theoretical timing performance, mostly associated with the presence of both the front-end electronics and the parasitic interconnection components.

Electronic noise and bandwidth limitation introduced by the front-end electronics cause an inevitable increase of the time jitter of the detection system. To prevent this effect from being emphasized, it is indeed mandatory to consider the real-world effects as part of the design process. For instance, it is widely demonstrated that enlarging the front-end bandwidth and making extremely low input impedance are good design practices to neutralize the effects of the large equivalent detector capacitance; nonetheless, it is also well known that stray components associated to the interconnections between the photodetector and the front-end electronics can cause unrecoverable harmful effects. When timing performance is demanded to approach its theoretical limit, the effects of parasitic components can never be neglected but need to be reviewed as a system specification and accounted for as well as the intrinsic electrical parameters of the SiPM and the characteristics of the front-end.

The purpose of the present theoretical study is to derive an essential set of simple mathematical expressions connecting the measurable characteristics of the single photon time response of a SiPM-based electronic readout system to some of the most important circuit parameters such as input resistance, gain, bandwidth, at the presence of the interconnection parasitic inductance.

A factorized expression of the system transfer function is proposed, in such a form that the individual weight of the electrical parameters participating in the formation of the output response can be easily identified and calculated. Thanks to this novel analytical study it is possible to draw some useful conclusions about the explicit influence of the parasitic electrical components on the timing performance of a SiPM-based detection system and to revise some practical guidelines for the choice and the design of the front-end architecture that is well-suited to the application. Eventually, to accomplish the task, a comparative noise analysis for two basic preamplifier configurations has been carried out and the timing jitter for both voltage and current mode approaches has been estimated as a function of the SiPM parameters and the parasitic components, for different bias points of the input transistor. The results of the study are discussed and validated by circuit simulations and experimental tests.

2.2   THE TRANSFER FUNCTION OF A SiPM-BASED DETECTION SYSTEM

The complete circuit model of a SiPM detector with N micro-cells coupled to an electronic front-end based on the classic current-mode approach [5], [6] is shown in Fig. 2-1. The preamplifier is a current buffer with gain A_i, which exhibits a low input resistance R_{in}. The load resistance R_L converts the output current of the buffer into the output voltage V_{out} and the overall transfer function is dominated by a single output time-constant $\tau_L = R_L C_L$, that sets the system bandwidth. For

simplicity, in Fig. 2-1, only one microcell of the detector is supposed to undergo a Geiger discharge, but this does not affect the generality of the conclusions that will be drawn once provided that the sensor linearity is ensured.
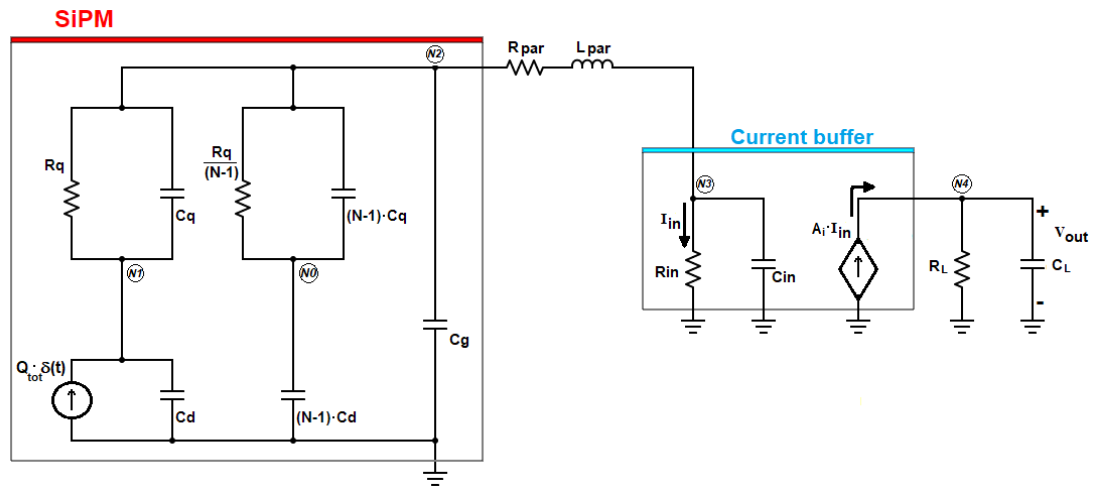
Fig. 2-1: Equivalent circuit model of the SiPM coupled to a current-mode front-end.

The electrical model of the SiPM developed in [8] has been exploited in [8], [9] to derive the transfer function from the input charge Dirac's delta to the current I_{in} flowing through the input resistance of the preamplifier in Fig. 2-1. The model of the SiPM includes the delta-like current source $I_{\mu cell}(t)=Q_{tot}\cdot\delta(t)$ that accounts for the Geiger discharge, the quenching resistor R_q, the parasitic capacitance across it, C_q, the photodiode capacitance C_d and the capacitance C_g, the latter due to the routing metal grid used to connect in parallel all the micro-cells. The element L_{par} represents the parasitic inductance associated with the interconnection wire between the SiPM and the front-end. Compared to the analysis reported in [8], the only additional element that has been introduced in Fig. 2-1 is the resistor R_{par}, that accounts for both the series substrate resistance of the SiPM, R_{sub}, as reported in [10], and other series parasitic resistances; finally, it improves the accuracy of the SiPM model fit. To carry out the analysis in the s-domain with the aim of determining a suitable analytical expression for the current $I_{in}(s)$, the input part of the circuit in Fig. 2-1 has been redrawn. After applying the Norton equivalent to the SiPM model network and reducing to the parallel connection of two admittances $Y_{det}(s)$ and $Y_{par, in}(s)$, the circuit can be drawn as in Fig. 2-2.
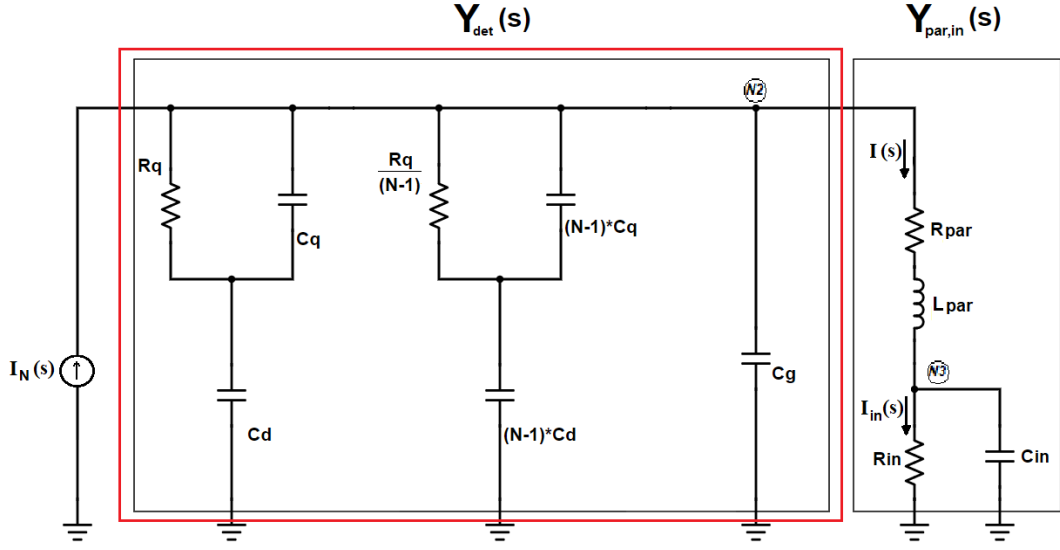
The expression of the Norton equivalent current $I_N(s)$ in Fig. 2-2 is the following:

$$I_N(s) = Q_{tot} \frac{1 + \tau_q s}{1 + \tau_r s} \tag{1}$$

where the time constants $\tau_q = R_q C_q$ and $\tau_r = R_q (C_d + C_q)$ appear.

After expressing the parallel admittances $Y_{det}(s), Y_{par,in}(s)$ respectively as

$$Y_{det}(s) = N C_d \frac{s(1 + \tau_q s)}{1 + \tau_r s} + s C_g \tag{2}$$

$$Y_{par,in}(s) = \frac{1}{R_s} \cdot \frac{\omega_n^2}{s^2 + 2 \zeta \omega_n s + \omega_n^2} (1 + \tau_{in} s) \tag{3}$$

with $\tau_{in} = R_{in} C_{in}$, $R_s = R_{in} + R_{par}$, $\omega_n^2 = \frac{1}{L_{par} C_{in}} \cdot \frac{R_s}{R_{in}} = \frac{1}{\tau_{in}} \cdot \frac{R_s}{L_{par}}$, and

$2 \zeta \omega_n = \frac{1}{\tau_{in}} + \frac{R_{par}}{L_{par}}$, the application of the current divider rule leads to

$$I_{in}(s) = \frac{1}{1 + \tau_{in} s} \cdot \frac{Y_{par,in}(s)}{Y_{par,in}(s) + Y_{det}(s)} \cdot I_N(s) \tag{4}$$

47

By substituting expressions from (1) to (3) in (4), the expression of the Laplace transform of the current flowing into the input resistance of the preamplifier R_{in} can be rearranged as:

$$I_{in}(s) = Q_{tot}\omega_n^2 \frac{(1+\tau_q s)}{(1+\tau_{in}s)\cdot(1+\tau_r s)\omega_n^2 + R_s \cdot s \cdot (s^2 + 2\zeta\omega_n s + \omega_n^2)\cdot[C_g(1+\tau_r s) + NC_d(1+\tau_q s)]} \quad (5)$$

The trigger signal for timing is formed either converting the output current of the buffer, $A_i \cdot I_{in}(s)$, into a voltage, as illustrated in Fig. 2-1, and using a voltage comparator, or directly using a leading-edge current discriminator. Of course, when a current discriminator is used for time pickoff, the output current signal can more accurately follow the very fast leading edge of the current pulse generated by the SiPM; in fact, fully current-mode amplifiers easily achieve large bandwidths, because of the absence of high impedance nodes. The overall current amplification factor to be used depends on the characteristics of the current discriminator.

In Fig. 2-1, the output current of the current buffer is converted into the output voltage

$$V_{out}(s) = \frac{K_R}{(1 + \tau_L s)} \cdot I_{in}(s), \quad (6)$$

where $\tau_L = R_L C_L$ is the time constant associated to the output pole and $K_R = A_i \cdot R_L$ is the overall transimpedance gain of the front-end.

Eq. (5) and Eq. (6) can be rearranged in the following way:

$$I_{in}(s) = Q_{tot} \frac{(1 + \tau_q s)}{1 + a_1 s + a_2 s^2 + a_3 s^3 + a_4 s^4}, \quad (7)$$

$$V_{out}(s) = Q_{tot} \frac{(1 + \tau_q s)}{1 + a_1 s + a_2 s^2 + a_3 s^3 + a_4 s^4} \cdot \frac{K_R}{(1 + \tau_L s)}, \quad (8)$$

48

where

$$a_1 = \tau_{in} + \tau_r + R_s(C_g + NC_d),$$

$$a_2 = \tau_{in}\tau_r + C_g(R_s\tau_r + L_{par} + R_{par}\tau_{in}) + NC_d(R_s\tau_q + L_{par} + R_{par}\tau_{in}),$$

$$a_3 = L_{par}(C_g + NC_d)\tau_{in} + (L_{par} + R_{par}\tau_{in})(C_g\tau_r + NC_d\tau_q),$$

$$a_4 = L_{par}(C_g\tau_r + NC_d\tau_q)\tau_{in}.$$

Eq. (8) describes the relationship between the total charge released by a single micro-cell undergoing a Geiger discharge and the output voltage of the preamplifier for the current-mode front-end in Fig. 2-1. Resorting to a voltage-mode approach, the current pulse of the detector is early converted into a voltage by R_{in}, while the resulting input voltage signal is now amplified by a high input impedance voltage amplifier with gain A_V. In this case, the output voltage $V_{out}(s)$ can still be expressed by Eq. (8), with the unique formal exception of the transimpedance gain $K_R=A_i{\cdot}R_L$ to be replaced with the factor $K'_R=A_V{\cdot}R_{in}$. Eq. (8) is characterized by one zero and five poles that make the corresponding closed-form expression in the time-domain rather complicated to be worked out. Indeed, it is true that it can be either easily calculated as the inverse Laplace transform of (8) running a MATLAB® script or plotted by solving the electrical network shown in Fig. 2-1 using a SPICE-based program; nonetheless, any information about the influence of each model parameter on the dynamic behaviour of the system would get lost. Furthermore, it would be rather complicated to establish any direct relation between those parameters and any of the relevant specifications involved in timing performance, such as the slope of the output signal. To do this, it can be advantageous to try and simplify the complex mathematical expressions derived from the complete analysis that has been carried out so far.

2.3 ANALYTIC APPROXIMATION OF THE MODEL FOR THE STUDY OF FAST TRANSIENT

As far as the design of front-end electronics for SiPM is concerned, numerical simulations are rarely effective at earlier stages of the design flow to put the designers on the right track for meeting the required final specifications.

A wise action to take is referring to simplified analytical expressions reproducing with enough mathematical accuracy the characteristics of the system response as a function of model parameters, such as the parasitic coupling inductance L_{par}, the input resistance of the current buffer, R_{in}, and the amplifier bandwidth, and focusing on the most relevant parts of the output waveform that may have an impact on the timing performance.

The idea behind the approximation method proposed in the following is to correctly reproduce just the leading edge of the SiPM response, i.e. the portion involved in time pickoff process.

First, since the approximation needs to be valid only at sufficiently high frequencies, the second order term $(1 + \tau_{in}s)(1 + \tau_r s)$ in Eq. (5), which affects mostly the slow trailing edge of the response, can be disregarded with minor impact on the shape factor of the fast and steep edge. Consequently, Eq. (5) can be rewritten and factorized as follows:

$$I_{in}(s) \cong \frac{Q_{tot}}{C_g + NC_d} \cdot \frac{(1 + \tau_q s)}{s \cdot (1 + \tau_p s)} \cdot \frac{1}{R_s} \frac{\omega_n^2}{(s^2 + 2\zeta\omega_n s + \omega_n^2)}, \qquad (9)$$

where $\tau_p = \frac{C_g \tau_r + NC_d \tau_q}{C_g + NC_d}$.

By substituting the expression (9) in (6), the following approximation for the voltage $V_{out}(s)$ is obtained:

$$V_{out}(s) \cong \frac{Q_{tot}}{C_g + NC_d} \cdot \frac{(1 + \tau_q s)}{s \cdot (1 + \tau_p s)} \cdot \frac{1}{R_s} \frac{\omega_n^2}{(s^2 + 2\zeta\omega_n s + \omega_n^2)} \cdot \frac{K_R}{(1 + \tau_L s)}, \qquad (10)$$

Eq. (10) shows that, in the s-domain, the system can be broken up into three submodules, each representing one out of the three blocks in Fig. 2-3. Each of them is characterized by the transfer function of the corresponding factor in Eq. (10): $G_1(s)$ accounts for the SiPM model with its electrical parameters; $G_2(s)$ represents the effects of the interaction between the SiPM and the preamplifier that largely depends on both the parasitic interconnections and the input impedance of the current buffer; $G_3(s)$ represents the current-to-voltage transfer function of the preamplifier.

$$I_{\mu cell}(s) = Q_{tot} \xrightarrow{\qquad} \boxed{\underbrace{\frac{1}{C_g + N C_d} \cdot \frac{(1+\tau_q s)}{s \cdot (1+\tau_p s)}}_{G_1(s)}} \rightarrow \boxed{\underbrace{\frac{1}{R_S} \cdot \frac{\omega_n^2}{(s^2 + 2\zeta\omega_n s + \omega_n^2)}}_{G_2(s)}} \rightarrow \boxed{\underbrace{\frac{K_R}{(1+\tau_L s)}}_{G_3(s)}} \xrightarrow{\quad} V_{out}(s) = Q_{tot} \cdot G_1(s) \cdot G_2(s) \cdot G_3(s)$$
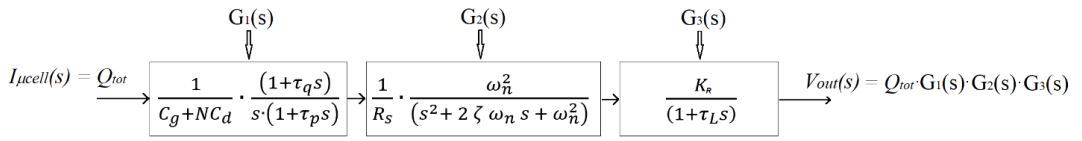
Fig. 2-3: Block diagram of the factorized transfer function.

The contribution to the final shaping of each of the main blocks of the detection chain on the initial charge pulse can be properly isolated by visual inspection of the transfer function of the complete system represented by Eq. (10).

The electrical parameters of the SiPM used for numerical simulation and calculation are reported in the following Table 2-1 and will be adopted throughout the analysis.

Table 2-1: Electrical parameters of the SiPM used for the study.

R_q	182.75 kΩ
C_q	17.72 fF
C_d	75.17 fF
C_g	36.85 pF
R_{par}	22.9 Ω
Ncell	3600

The inverse Laplace transforms of expressions (8) and (10) are plotted in Fig. 2-4. They represent, respectively, the response of the complete model and the

response of the approximate model when just one micro-cell of the SiPM is fired, with the following set of parameters: L_{par} = 10 nH, R_{in} = 10 Ω, C_{in} = 1 pF, K_R = 1.2 kΩ and BW = $1/2\pi\tau_L$ = 1 GHz.
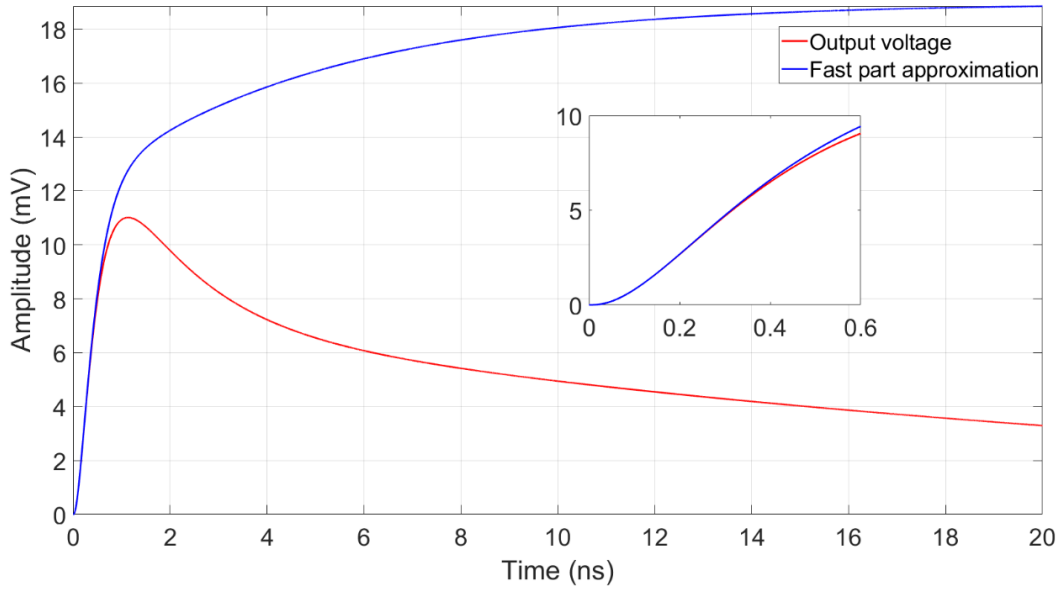
Fig. 2-4: Comparison between the output voltage behaviour predicted by Eq. (8) and Eq. (10). The inset shows a very good matching in the rising edge of the response, relevant for applications where timing accuracy is of interest.

It is apparent from Fig. 2-4 that the approximate expression of $V_{out}(s)$ and the fast part of the complete response fit perfectly together with marginal deviation around the region that is of interest for time pickoff, whereas the two plots start to diverge in proximity of the peak of the exact model (red curve). It is worthy to notice that the fitting is valid irrespective of any choice of the parameters used to derive either the transfer function and its approximation.

When the leading-edge discrimination technique is used for time pickoff, the slope of the response at the instant when the chosen threshold V_{TH} is reached is a key parameter in the definition of time resolution, according to the well-known expression [11]:

$$\sigma_t = \frac{\sigma_n}{\left.\dfrac{dV_{out}(t)}{dt}\right|_{V_{out}=V_{TH}}} \tag{11}$$

in which σ_n is the *rms* electronic noise at the preamplifier output and σ_t is the time jitter, i.e. the measurement uncertainty in the identification of the occurrence time of the detected event. According to Eq. (11), once established the noise level, the steepest the waveform at the threshold crossing point above the noise level, the lower the jitter.

The time derivatives of the waveforms in Fig. 2-4 have been calculated as well, and their plots are reported in Fig. 2-5 for the first part of the transient response, which is relevant for time measurements. purposely limited to the initial part of the transient responses, which is relevant for time measurement. The good agreement between the complete expression (8) and the approximate expression (10) is showcased below.
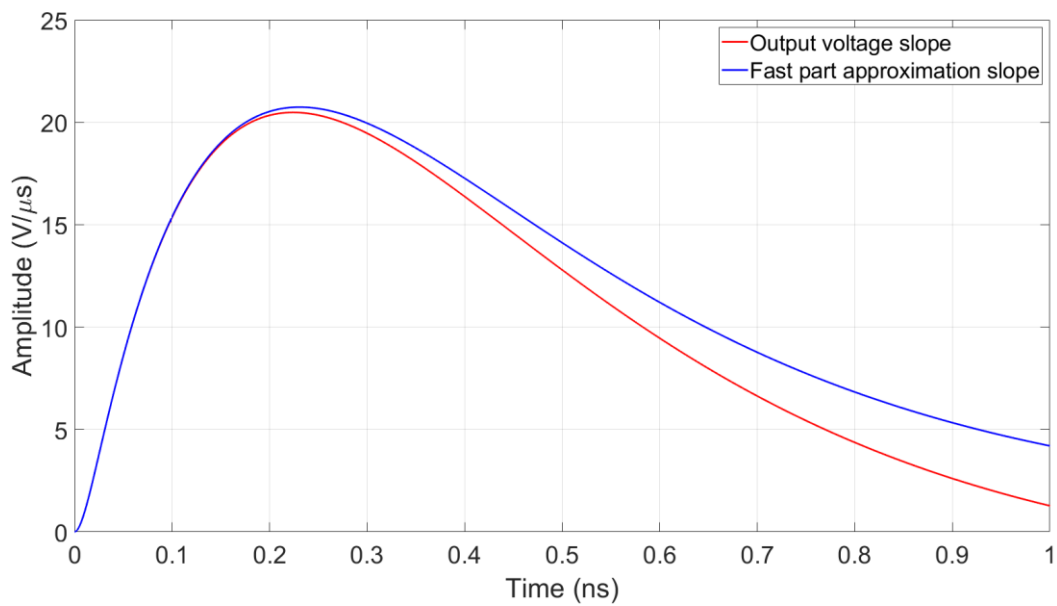
Fig. 2-5: Time derivatives of the response V_{out} (t) and its approximation (10).

The maximum amplitudes of the time derivatives for both the responses of the complete model and its approximation (10) have been reported in Fig. 2-6 as a function of R_{in}, swept in the range from 1 Ω to 100 Ω. The blue curve refers to a voltage-mode approach, where a voltage amplifier with a single pole roll-off and 1GHz bandwidth is used to amplify the voltage across R_{in}. On the other hand, the red curve is related to the current-mode approach (see the scheme illustrated in Fig. 2-1), and τ_L is the same time constant used for the voltage-mode case.

The slopes of both responses shown in Fig. 2-6 and computed at their inflection points, i.e. the most convenient threshold level for timing according to (11), perform differently. For comparison purpose, the magnitudes on the graph refer to different scales. After fixing *A_i, R_L, and A_v*, the slopes assume the same value just for that particular value of R_{in} satisfying the condition *$A_i \cdot R_L = A_v \cdot R_{in}$*.
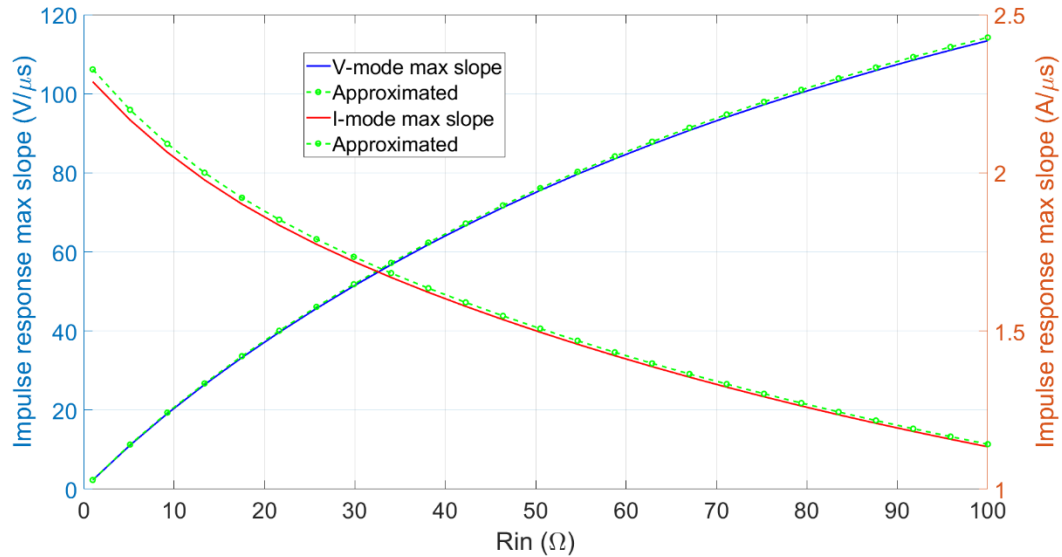
Fig. 2-6: Maximum amplitudes of the derivatives as a function of Rin for both the exact response and its approximation, for both current-mode and voltage-mode approach with arbitrary amplification factors A_v and A_i.

It is apparent that the maximum slope increases for increasing values of R_{in} for the voltage-mode approach, whereas it decreases for the current-mode approach, thus confirming that the behaviours of the complete model, represented by equations (8), and the approximate expressions (9) and (10) match with good accuracy.

2.4 THE INFLUENCE OF THE MODEL PARAMETERS ON THE FAST SYSTEM RESPONSE

As already pointed out, deriving a comprehensive analysis of a fifth order system in the time domain is a complicated task. Thus, a simplified approach may help gain insight into the problem. The inverse Laplace transform of $I_{in}(s)$ from Eq. (9) can be used to approximate the initial part of the current waveform $I_{in}(t)$. If the preamplifier bandwidth is wide enough not to distort the current signal of the

SiPM, the first order roll-off function can be dropped, and a fourth order expression can be assumed to represent the output response of the detection system. The time derivative can be accomplished in the s-domain by multiplying the transfer function by *s* and evaluating the inverse Laplace transform of the resulting expression. Eventually, considering the factorized form of Eq. (9), that can be derived from Eq. (10) after dropping the preamplifier transfer function, the time derivative of the current pulse, i.e. its slope, can be obtained as the convolution between the inverse Laplace transforms of $F_1(s) = sQ_{tot}G_1(s)$ and $G_2(s)$:

$$\mathcal{L}^{-1}\{s \cdot I_{in}(s)\} = Slope_I(t) = \frac{Q_{tot}}{C_g + NC_d} \int_0^t f(t-v) \cdot g_2(v) \cdot dv$$

$$= \frac{\alpha \cdot Q_{tot}}{C_g + NC_d} \cdot g_2(t) + \frac{(1-\alpha) \cdot Q_{tot}}{C_g + NC_d} \frac{e^{-\frac{t}{\tau_p}}}{\tau_p} \int_0^t e^{\frac{v}{\tau_p}} \cdot g_2(v) \cdot dv$$

$$\tag{12}$$

where $f(\theta) = \mathcal{L}^{-1}\{F(s)\} = \mathcal{L}^{-1}\left\{\frac{(1+\tau_q s)}{(1+\tau_p s)}\right\}$,

$$g_2(\theta) = \mathcal{L}^{-1}\{G_2(s)\} = \mathcal{L}^{-1}\left\{\frac{1}{R_s} \cdot \frac{\omega_n^2}{(s^2 + 2\,\zeta\,\omega_n\,s + \omega_n^2)}\right\}, \ \alpha = \frac{\tau_q}{\tau_p}.$$

The poles of $G_2(s)$ can either be real or conjugate complex depending on the coefficients of the denominator. To avoid oscillatory responses, the attention is turned to solving for underdamped responses only, characterized by the fact that the expected poles are real and located in the left half of the s-plane. In the overdamped case, the definition of the second order system $G_2(s)$ is restricted to the solutions of the denominator with $\zeta > 1$, therefore $\tilde{g}_2(\theta)$, the inverse Laplace transform of the restricted function, takes the familiar form

$$\tilde{g}_2(\theta) = \frac{1}{R_s} \cdot \frac{\omega_n}{2 \cdot \sqrt{\zeta^2 - 1}} \cdot \left(e^{-\omega_n \cdot \left(\zeta - \sqrt{\zeta^2 - 1}\right) \cdot \theta} - e^{-\omega_n \cdot \left(\zeta + \sqrt{\zeta^2 - 1}\right) \cdot \theta}\right) \tag{13}$$

Since $\tau_{in} = R_{in}C_{in}$, $R_s = R_{in} + R_{par}$, $\omega_n^2 = \frac{1}{L_{par}C_{in}} \cdot \frac{R_s}{R_{in}} = \frac{1}{\tau_{in}} \cdot \frac{R_s}{L_{par}}$,

$$2\,\zeta\,\omega_n = \frac{1}{\tau_{in}} + \frac{R_{par}}{L_{par}},$$

Eq. (13) can be reformulated as

$$\tilde{g}_2(\theta) = \frac{1}{\sqrt{L_{par}^2 + 2L_{par}R_{par}\tau_{in} + \tau_{in}^2 R_{par}^2 - 4R_s\tau_{in}L_{par}}} \cdot$$

$$e^{-\frac{1}{2L_{par}\tau_{in}}\left[(L_{par}+R_{par}\tau_{in}) - \sqrt{L_{par}^2 + 2L_{par}R_{par}\tau_{in} + \tau_{in}^2 R_{par}^2 - 4R_s\tau_{in}L_{par}}\right]\cdot\theta} -$$

$$\frac{1}{\sqrt{L_{par}^2 + 2L_{par}R_{par}\tau_{in} + \tau_{in}^2 R_{par}^2 - 4R_s\tau_{in}L_{par}}} \cdot$$

$$e^{-\frac{1}{2L_{par}\tau_{in}}\left[(L_{par}+R_{par}\tau_{in}) + \sqrt{L_{par}^2 + 2L_{par}R_{par}\tau_{in} + \tau_{in}^2 R_{par}^2 - 4R_s\tau_{in}L_{par}}\right]\cdot\theta}$$

$$(14)$$

Since, considering typical values of the parameters involved, it is true that

$$\tau_{in} = R_{in}C_{in} \ll \frac{L_{par}}{R_{par}}, \tag{15}$$

then, as a practical matter

$$\sqrt{L_{par}^2 + 2L_{par}R_{par}\tau_{in} + \tau_{in}^2 R_{par}^2 - 4R_s\tau_{in}L_{par}} \cong L_{par}\sqrt{1 - 4\tau_{in}\frac{R_s}{L_{par}}} =$$

$$L_{par} \cdot K,$$

where

$$K = \sqrt{1 - 4\tau_{in}\frac{R_s}{L_{par}}}$$

and, finally, from equation (14), a more compact form can be written as

$$\tilde{g}_2(\theta) = \mathcal{L}^{-1}\{\tilde{G}_2(s)\} = \frac{1}{L_{par} \cdot K} \cdot \left[e^{-\frac{(1-K)}{2\tau_{in}}\theta} - e^{-\frac{(1+K)}{2\tau_{in}}\theta}\right] \tag{16}$$

To ensure that K is a real number, the following inequality shall be satisfied:

$$\tau_{in} = R_{in}C_{in} < \frac{L_{par}}{4R_s}, \tag{17}$$

To replace $g_2(\theta)$ with $\tilde{g}_2(\theta)$ in Eq. (12) and simplify the forthcoming computational effort, both the conditions (15) and (17) set on τ_{in} must be checked out, and the resulting damping ratio must be greater than 1, as well. The method provided hereafter serve to wrap up the constraints given in (15) and (16) in a mathematical easy-to-use package. A real number factor m, tied to be smaller than 1/10, is introduced to express τ_{in} as a fraction m of (L_{par}/R_{par}). By replacing the time constant $\tau_{in} = R_{in}C_{in}$ with m·(L_{par}/R_{par}) in (17), the following constraints for R_{in} and C_{in} are derived:

$$R_{in} < \left(\frac{1-4m}{4m}\right) \cdot R_{par} , \quad C_{in} < \left(\frac{4m^2}{1-4m}\right) \cdot \frac{L_{par}}{R_{par}^2} \quad (18)$$

where the constraint for C_{in} is obtained after substituting the constraint for R_{in} in (15) and solving for C_{in}. Fig. 2-7 shows the plots of R_{in_max} and C_{in_max}, the upper limits of disequality conditions (18), as functions of m, for R_{par}=22.9 Ω. The plots of C_{in_max} correspond to as many as three values of L_{par}. For instance, if L_{par}= 11 nH, the corresponding C_{in_max} curve is highlighted. After a value for C_{in} is chosen within the range of possible values on that curve (from few fF to 1.3 pF in this case), the value for m can be retrieved first and, subsequently, after jumping onto the deep blue curve, the corresponding value of R_{in_max} (smaller ones are also permitted).
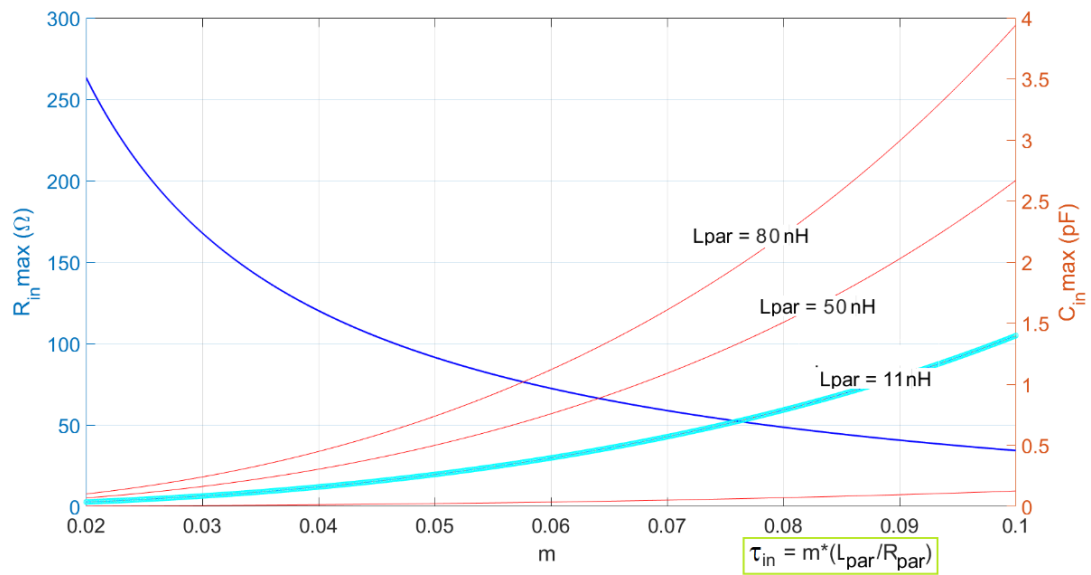

Fig. 2-7: Rin_max and Cin_max as a function of the coefficient m. Cin_max is plotted for different values of Lpar.

When L_{par} increases, the range of permitted values for C_{in} becomes larger. The larger the value chosen for C_{in}, the smaller the maximum permitted value for R_{in}, which certainly makes sense because τ_{in} is linearly dependent on both L_{par} and m and, once L_{par} and m are fixed, from (15), the product $R_{in}C_{in}$ is determined as well. It is worthy to notice that after R_{in_max} is pinpointed on the blue curve, all the values that are smaller than the pinpointed value are admitted down to somewhere just above zero. When the described checking procedure is successfully applied on (15) and (17), the resulting damping ratio is necessarily greater than 1 and the equations become consistent with the initial assumptions.

After replacing in (12) $g_2(\theta)$ with $\tilde{g}_2(\theta)$ as reported in (16) and after few calculations, the expression of the slope in the time domain can be written as:

$$
\begin{aligned}
Slope_I(t) \; = \; & \frac{(1-\alpha)\cdot V_0}{L_{par}K\tau_p} \\
& \cdot \left\{ \left[\frac{1}{\frac{1}{\tau_p} - \frac{(1+K)}{2\tau_{in}}} - \frac{1}{\frac{1}{\tau_p} - \frac{(1-K)}{2\tau_{in}}} \right] \cdot e^{-\frac{t}{\tau_p}} \right. \\
& + \left[\frac{\alpha\tau_p}{(1-\alpha)} + \frac{1}{\frac{1}{\tau_p} - \frac{(1-K)}{2\tau_{in}}} \right] \cdot e^{-\frac{(1-K)}{2\tau_{in}}\cdot t} \\
& \left. - \left[\frac{\alpha\tau_p}{(1-\alpha)} - \frac{1}{\frac{1}{\tau_p} - \frac{(1+K)}{2\tau_{in}}} \right] \cdot e^{-\frac{(1+K)}{2\tau_{in}}\cdot t} \right\} \\
= \; & A0 \cdot e^{-\frac{t}{\tau_p}} + A1 \cdot e^{-\frac{(1-K)}{2\tau_{in}}\cdot t} - A2 \cdot e^{-\frac{(1+K)}{2\tau_{in}}\cdot t}
\end{aligned}
$$

(19)

$$
A0 = \frac{(1-\alpha)\cdot V_0}{L_{par}K\tau_p} \cdot \left[\frac{1}{\frac{1}{\tau_p} - \frac{(1+K)}{2\tau_{in}}} - \frac{1}{\frac{1}{\tau_p} - \frac{(1-K)}{2\tau_{in}}} \right], \quad A1 = \frac{(1-\alpha)\cdot V_0}{L_{par}K\tau_p} \cdot \left[\frac{\alpha\tau_p}{(1-\alpha)} + \frac{1}{\frac{1}{\tau_p} - \frac{(1-K)}{2\tau_{in}}} \right], \quad A2 =
$$

$$
\frac{(1-\alpha)\cdot V_0}{L_{par}K\tau_p} \cdot \left[\frac{\alpha\tau_p}{(1-\alpha)} - \frac{1}{\frac{1}{\tau_p} - \frac{(1+K)}{2\tau_{in}}} \right] \text{ and } V_0 = \frac{Q_{tot}}{C_g + NC_d}.
$$

Since the term $A0 \cdot e^{-\frac{t}{\tau_p}}$ decays very slowly when compared to the terms whose coefficients are A1 and A2, it can be considered almost constant and dropped while calculating the derivative of (19). The expression of the time when the slope of the output waveform reaches its maximum amplitude is then

$$t_{MAX_S} \cong R_{in} C_{in} \cdot \ln\left(\frac{1+K}{1-K}\right), \qquad (20)$$

where the approximation $\ln\left(\frac{1+K}{1-K}\right) \cong \ln\left(\frac{1+K}{1-K} \cdot \frac{A1}{A2}\right)$ has been taken.

In the following Fig. 2-8 and Fig. 2-9, expression (20) is compared to the expression of maximum slope occurring time obtained after solving for time the inverse Laplace transform of the factorized expression (9) of I$_{in}$(s). Two edge cases have been considered. In Fig. 2-8 an ideal situation is shown in which the parasitic elements L$_{par}$ and C$_{in}$ can barely affect the ultra-fast intrinsic response of the SiPM.
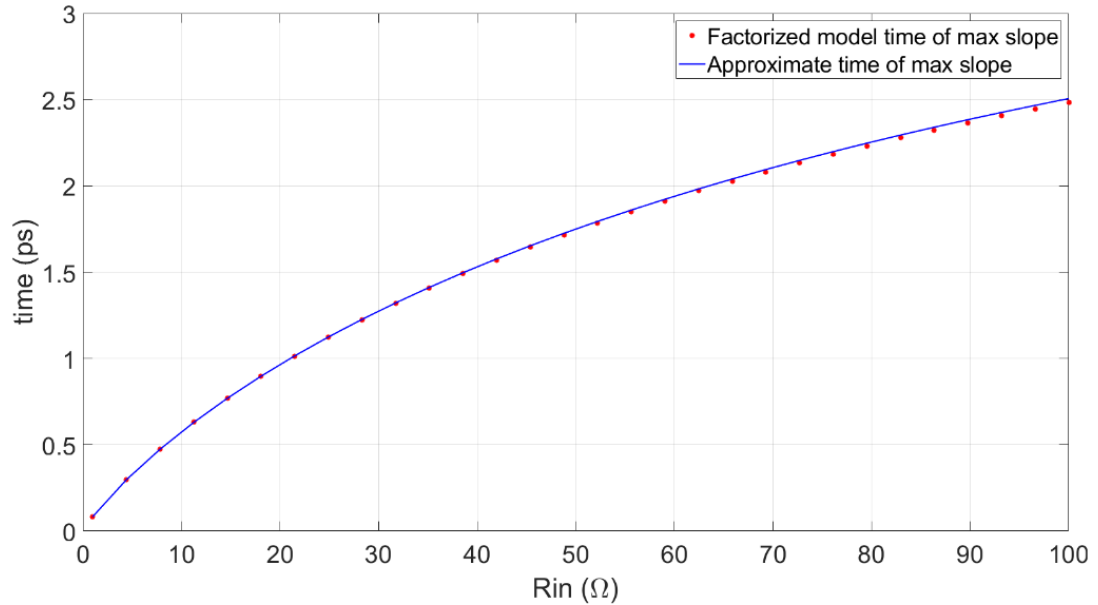

Fig. 2-8: The curve of maximum slope time and its approximate Eq. (20) for an almost ideal situation of the amplifier bandwidth approaching infinity, with Lpar = 1 nH, Cin = 10 fF and m = 0.045.

Indeed, Fig. 2-9 shows that when larger values of the parameters are considered, the response becomes slower. In both situations, increasing the input resistance

leads to increase the time when the maximum slope of the current response occurs. Once more, a very good matching between the approximation (20) and the results given by expression (9) is achieved.
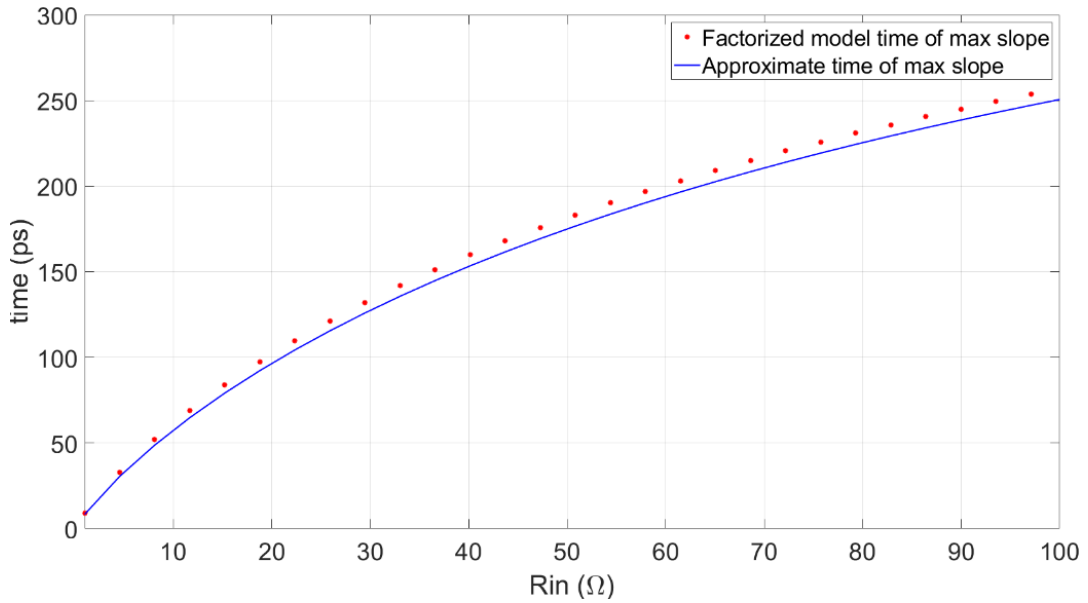

Fig. 2-9: The curve of maximum slope time and its approximation by Eq. (20) where parasitic inductance and input capacitance are quite large: L_{par} = 100 nH, C_{in} = 1000 fF and m = 0.045.

After replacing variable t with expression (20) in equation (19), an expression of the maximum slope of the response is obtained in a concise form as a function of the previously defined parameter K and the coefficients A0, A1, A2:

$$SL\_max_I = Slope_I(t_{MAX\_S}) = A0 + A1 \cdot \left(\frac{1-K}{1+K}\right)^{\frac{1-K}{2K}} - A2 \cdot \left(\frac{1-K}{1+K}\right)^{\frac{1+K}{2K}}$$

(21)

In Fig. 2-10 and Fig. 2-11 a comparison between the maximum slope of the current pulse $I_{in}(t)$ derived from Eq. (9) and its approximation (21) is reported as a function of the input resistance R_{in}. Both figures show that Eq. (21) can provide a very good approximation of the behaviour of the maximum slope. Moreover, it is apparent that the slope amplitude decreases as R_{in} increases.
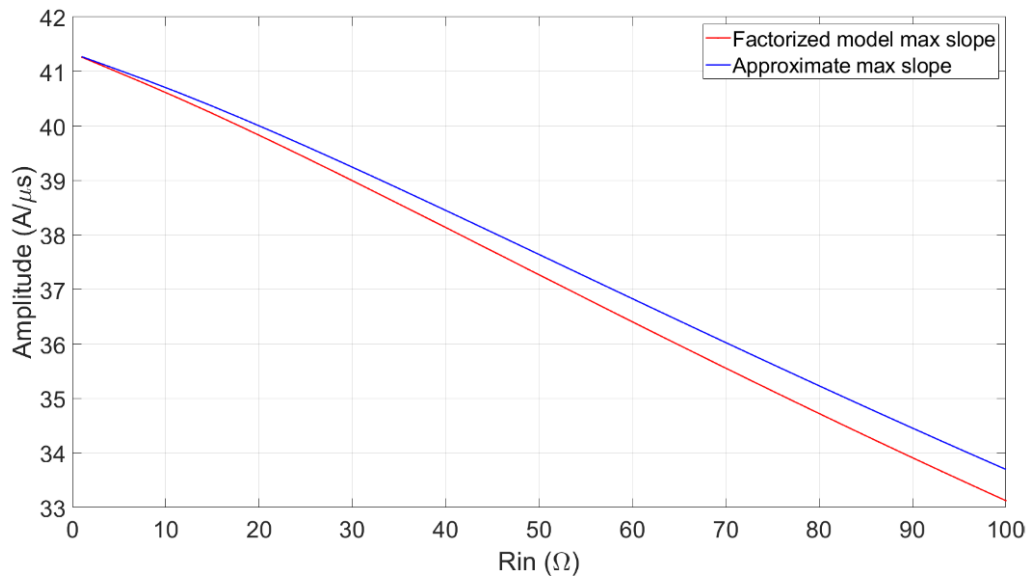
Fig. 2-10: The curve of maximum slope and its approximation by Eq. (21) for an almost ideal situation of the amplifier bandwidth approaching infinity, where Lpar = 1 nH, Cin = 10 fF, m = 0.045.

In Fig. 2-11, the parasitic inductance and the input capacitance are quite large, determining a decrease of the maximum slope of the signal if referred to Fig. 2-10.
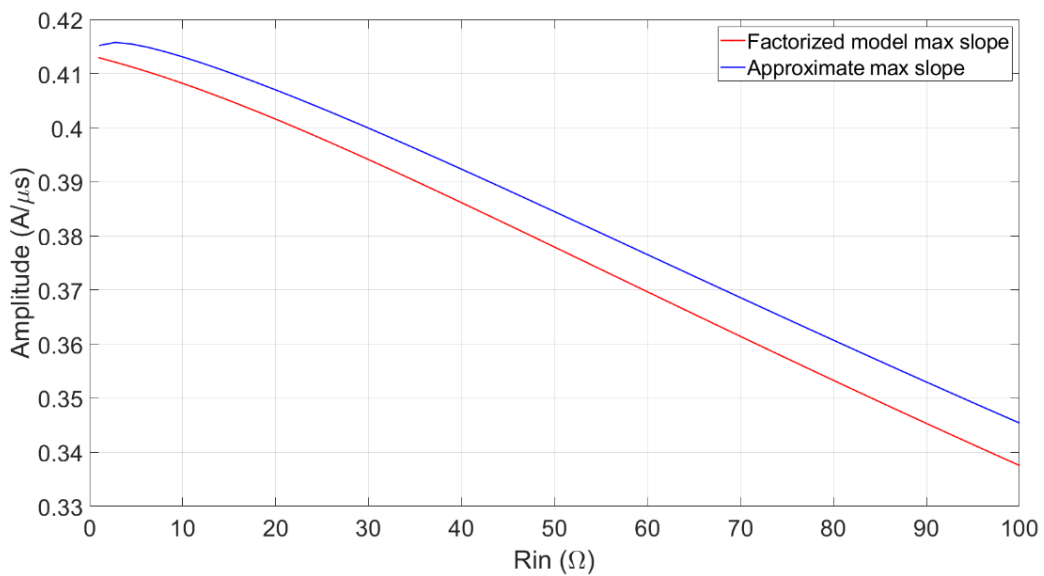
Fig. 2-11: The curve of maximum slope and its approximation with a different set of parameters, where Lpar = 100 nH, Cin = 1000 fF, m = 0.045.

On the contrary, as far as the voltage signal across the input resistor is concerned, its maximum slope would increase with increasing input resistance because its approximate equation is now obtained after multiplying Eq. (21) by R_{in}.

2.5 A COMPREHENSIVE ANALYSIS INCLUDING THE FRONT-END BANDWIDTH

In real life, a preamplifier has a finite bandwidth which contributes to change pulse shape of the response in a decisive manner. To model the effect of the finite bandwidth, the preamplifier can be described as a single pole transfer function, so that the impact of this pole on the dynamic behaviour of the detection system can be investigated. An approximate transfer function of the output voltage with the bandwidth constraint of the amplification stage can be written as Eq. (10), which is reported hereafter again for convenience,

$$V_{I,out}(s) \cong \frac{Q_{tot}}{C_g + NC_d} \cdot \frac{(1 + \tau_q s)}{s \cdot (1 + \tau_p s)} \cdot \frac{1}{R_S} \cdot \frac{\omega_n^2}{(s^2 + 2\,\zeta\,\omega_n\,s + \omega_n^2)} \cdot \frac{K_R}{(1 + \tau_L s)}. (22)$$

In a current-mode approach (marked by the subscript I in Eq. 22), a transimpedance amplifier with gain K_R and a cut-off frequency set by the pole $1/\tau_L$ is used to convert the input current signal into a proportional voltage signal (see Fig. 2-1).

Recalling what already pointed out, the gain K_R can be replaced with the product $K'_R = A_v \cdot R_{in}$ to switch from a current-mode to a voltage-mode readout approach.

Just multiplying by R_{in} turns the expression of $V_{I,out}(s)$ in Eq. (22) into

$$V_{V,out}(s) \cong \frac{Q_{tot}}{C_g + NC_d} \cdot \frac{(1 + \tau_q s)}{s \cdot (1 + \tau_p s)} \cdot \frac{R_{in}}{R_S} \cdot \frac{\omega_n^2}{(s^2 + 2\,\zeta\,\omega_n\,s + \omega_n^2)} \cdot \frac{A_v}{(1 + \tau_L s)}. (23)$$

A simplified analytical approach, which can lead to practical guidelines for optimum circuit design, is possible also in this case, as follows.

Assuming $\tau_{in} = R_{in}C_{in} \ll \frac{L_{par}}{R_S}$, which is a stronger assumption than (15), because $R_S = R_{par} + R_{in}$ is greater than R_{par}, entails a further simplification $2\,\zeta\,\omega_n \cong \frac{1}{\tau_{in}}$ and then the second order transfer function included in Eq. (22), can be approximated as a first order transfer function:

$$G_2(s) = \frac{1}{R_s} \frac{\omega_n^2}{(s^2 + 2\,\zeta\,\omega_n\,s + \omega_n^2)} \cong \frac{1}{R_s} \frac{\frac{1}{\tau_{in}\cdot\tau_A}}{\left(s^2 + \frac{s}{\tau_{in}} + \frac{1}{\tau_{in}\cdot\tau_A}\right)} \cong \frac{1}{R_s} \frac{\frac{1}{\tau_{in}\cdot\tau_A}}{\left(s + \frac{1}{\tau_{in}}\right)\left(s + \frac{1}{\tau_A}\right)} \cong$$

$$\frac{1}{R_s} \frac{1}{(1 + \tau_A s)}, \qquad \tau_A = \frac{L_{par}}{R_s}$$

$$(24)$$

In fact, this dominant pole approximation is valid because the time constant τ_{in} is very small compared with τ_A (typically $\tau_A/\tau_{in} > 10$), thus the two poles of the system can be considered respectively equal to $-1/\tau_A$ and $-1/\tau_{in}$. Therefore, considering the contributions to the overall response, the faster exponential associated to τ_{in} comes to equilibrium, i.e. decays to zero, almost instantaneously, compared to the slower exponential associated with τ_A. Consequently, it is admissible to hold just the contribution of the slower time constant.

Another approximation that can be done at high frequencies, valid for the very first part of the transient response, is the following:

$$\frac{1 + \tau_q s}{1 + \tau_p s} \cong \frac{\tau_q}{\tau_p}$$

This further position leads to further simplification of the expression (22), as in the following

$$V_{I,out}(s) \cong \frac{Qtot}{Cg + N \cdot Cd} \cdot \frac{\tau_q}{\tau_p} \cdot \frac{1}{R_s} \cdot \frac{1}{s} \cdot \frac{1}{(1 + \tau_A s)} \cdot \frac{K_R}{(1 + \tau_L s)}. \qquad (25)$$

The Laplace transform of the slope of the output pulse, $Slope_{V,I}(s)$ is obtained after multiplying expression (25) by the variable s:

$$Slope_{V,I}(s) = \frac{Qtot \cdot K_R}{Cg + N \cdot Cd} \cdot \frac{\tau_q}{\tau_p} \cdot \frac{1}{R_s} \cdot \frac{1}{\tau_A \cdot \tau_L} \cdot \left(\frac{a}{s + 1/\tau_A} + \frac{b}{s + 1/\tau_L}\right) \qquad (26)$$

where $a = -\frac{\tau_A \cdot \tau_L}{\tau_L - \tau_A}$ and $b = \frac{\tau_A \cdot \tau_L}{\tau_L - \tau_A}$.

63

Thus, just to summarize, Eq. (26) has been figured out after completing the factorized model with the additional pole of the amplifier, associated with the time constant τ_L, considering a first order approximation for the transfer function $G_2(s)$ and, lastly, assuming $\frac{1+\tau_q s}{1+\tau_p s} \cong \frac{\tau_q}{\tau_p}$.

Eq. (26) can be conveniently used to derive easy-to-handle equations for the time of maximum slope and the maximum slope value as a function of the most relevant parameters involved in the dynamic behaviour of the detection system.

The inverse Laplace Transform of Eq. (26) leads to the final expression for the slope of the output voltage signal in the time domain

$$Slope_{V,I}(t) = \mathcal{L}^{-1}\{Slope_{V,I}(s)\} = \frac{Qtot \cdot K_R}{Cg + N \cdot Cd} \frac{\tau_q}{\tau_p} \cdot \frac{1}{R_s} \cdot \frac{1}{\tau_A \cdot \tau_L} \cdot \left(a \cdot e^{-\frac{t}{\tau_A}} + b \cdot e^{-\frac{t}{\tau_L}} \right)$$

(27)

Taking the time derivative of Slope $_{V,\,I}$(t), equating it to zero and solving for time variable t, the time t_{MAX_S} that conveys the steepest slope of the output pulse, as a function of τ_A, τ_L, L_{par}, and R_s is

$$t_{MAX_S} = \frac{\tau_A \cdot \tau_L}{\tau_A - \tau_L} \cdot \ln\left(\frac{\tau_A}{\tau_L}\right) = \frac{L_{par}}{R_s} \cdot \frac{1}{\theta - 1} \cdot \ln\theta \qquad (28)$$

where $\theta = {\tau_A}/{\tau_L}$ is a normalization variable depending on L_{par}, R_s and τ_L while t_{MAX_S} is a continuous positive function of θ. It may happen that using larger values of C_{in} and ultra-wideband amplifiers ($\tau_L < 0.5$ns), the approximate expression (28) of t_{MAX_S} proves faulty, therefore a corrective term $\tau_{in} = R_{in}C_{in}$ can be added to (28) to keep on ensuring the validity of the approximation with the following Eq. (29)

$$t_{MAX_S} = \frac{\tau_A \cdot \tau_L}{\tau_A - \tau_L} \cdot \ln\left(\frac{\tau_A}{\tau_L}\right) = \frac{L_{par}}{R_s} \cdot \frac{1}{\theta - 1} \cdot \ln\theta + \tau_{in} \qquad (29)$$

Fig. 2-12 shows t_{MAX_S} as a function of R_{in} for different values of L_{par}, according to Eq. (28), with Cin = 0.5 pF and BW = 0.5 GHz. The results of the approximation are compared with simulations of the complete model, showing a very good matching.
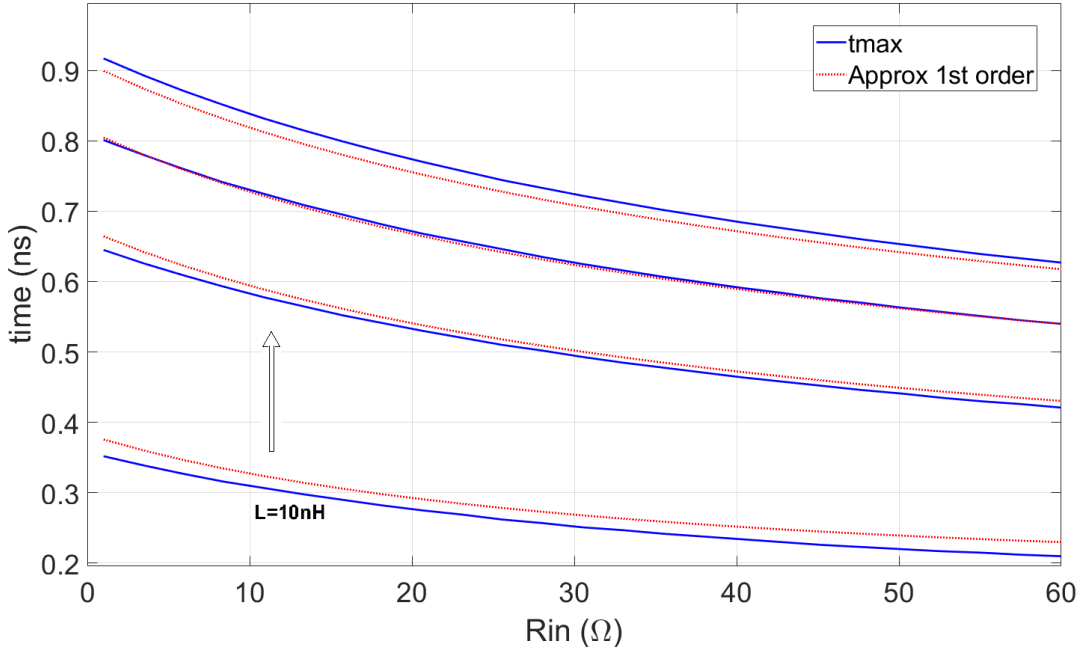

Fig. 2-12: The curves represent the location of the time derivative peak of the response $V_{I, out}$ (t) as a function of four increasing values of parasitic inductance, L = 10, 40, 70, 100 nH.

While input resistance decreases the time when the peak is reached increases and the same occurs with increasing L_{par}.

By substituting the (28) into (27), the value of the maximum slope is

$$Slope_{V,I}(t_{MAX_S}) = \frac{Qtot \cdot K_R}{Cg + N \cdot Cd} \cdot \frac{\tau_q}{\tau_p} \cdot \frac{1}{R_s} \cdot \frac{1}{\tau_L} \cdot e^{-\frac{t_{MAX_S}}{\tau_L}} \tag{30}$$

and, proceeding with θ substitution, the following expression is obtained:

$$Slope_{V,I}(t_{MAX_S}) = \frac{Qtot \cdot K_R}{Cg + N \cdot Cd} \cdot \frac{\tau_q}{\tau_p} \cdot \frac{1}{R_s} \cdot \frac{1}{\tau_L} \cdot \theta^{\frac{\theta}{1-\theta}} \tag{31}$$

The expression for a voltage-mode approach is

$$Slope_{V,V}(t_{MAX_S}) = \frac{Qtot \cdot A_v}{Cg + N \cdot Cd} \cdot \frac{\tau_q}{\tau_p} \cdot \frac{R_{in}}{R_s} \cdot \frac{1}{\tau_L} \cdot \theta^{\frac{\theta}{1-\theta}} \tag{32}$$

Therefore, the time of maximum slope and the maximum response slope itself have been expressed with simple mathematical relationship as a function of the key parameters of SiPM, the preamplifier bandwidth and parasitic components. Fig. 2-13 and Fig. 2-14 make a direct comparison of the exact slopes and their approximate functions. The curves reported in the following Fig. 2-13 prove that the proposed lower order approximation of the complete high order system gives very good results in terms of accuracy.

The selected inductor values are 10 nH, 40 nH, 70 nH, 100 nH and an intermediate value of 0.5 pF has been chosen for C_{in}. When sweeping R_{in} from 1 Ω to 60 Ω the condition $\tau_{in} = R_{in}C_{in} \ll \frac{L_{par}}{R_S}$ is fully satisfied.
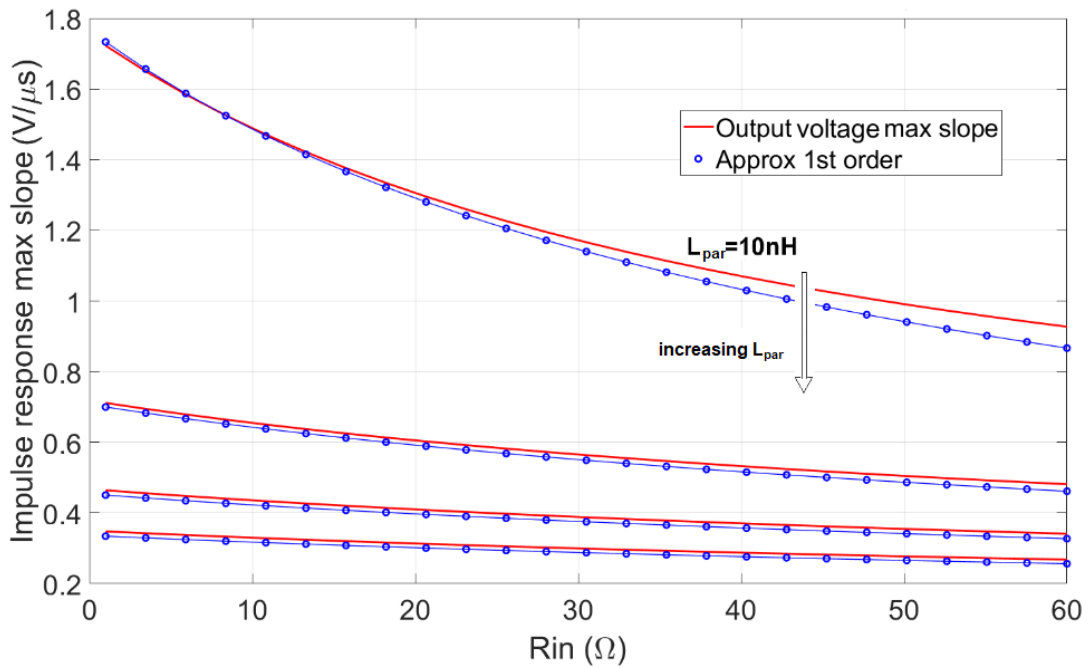

Fig. 2-13: The curve of maximum slope for the current-mode approach and its approximation by first order Eq. (31), with Lpar = 10nH, 40 nH, 70 nH, 100 nH, Cin = 0.5 pF and BW = 0.5 GHz.

Fig. 2-14 shows some plots of Eq. (32) as a function of R_{in} for a specific set of values of the inductance L_{par}.
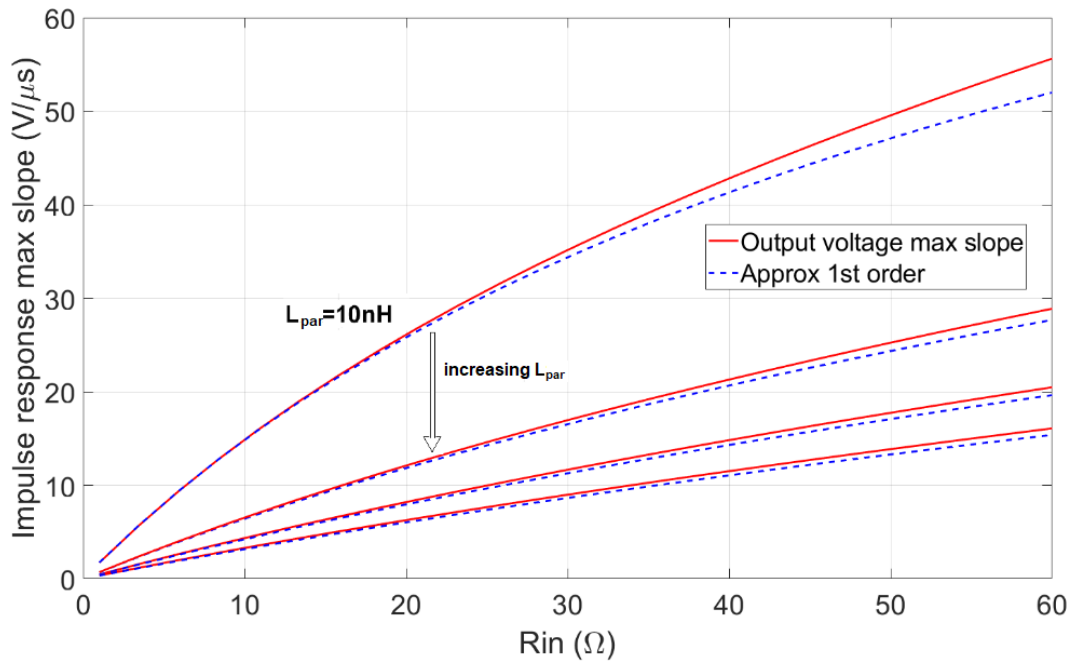
Fig. 2-14: The curve of maximum slope for voltage-mode approach and its approximation by first order Eq. (31), with Lpar = 10nH, 40 nH, 70 nH, 100 nH, Cin = 0.5 pF and BW = 0.5 GHz.

It is well worth noticing that increasing values of the input resistance R_{in} cause an increase of the maximum slope for the voltage-mode approach; the opposite occurs for the current-mode approach.

2.6 QUALITATIVE ANALYSIS OF THE FAST TRANSIENT AND PRACTICAL DESIGN CRITERIA

The set of approximate analytical equations (29), (31) and (32) allows for a straightforward estimation of the timing characteristics of a SiPM-based detection system. They have been worked out referring to the exact point of maximum slope of the response where an ideal discriminator threshold should be set to achieve the best time resolution. Nonetheless, the approximate equation of the slope can work quite well even if the threshold of the discriminator falls around the optimum. In fact, to prevent false triggers due to electronic noise with standard deviation equal to σ_n, as a rule of thumb the threshold is set at a level around five times σ_n and, typically, a good compromise between noise and steepness is found when the threshold is chosen to lay somewhere in the bounded region from 1/3 to 1/5 of the pulse peak.

Fig. 2-15 shows the slope of the current pulse calculated with the exact model and with the approximate model as a function of input resistance. The threshold for time pickoff is equal to 1/5 of the pulse peak. Even if the threshold is not exactly positioned where the slope of the pulse reaches the maximum, the approximated model is still valid and can reproduce the slope with an accuracy that is close to that foreseen by the exact model.
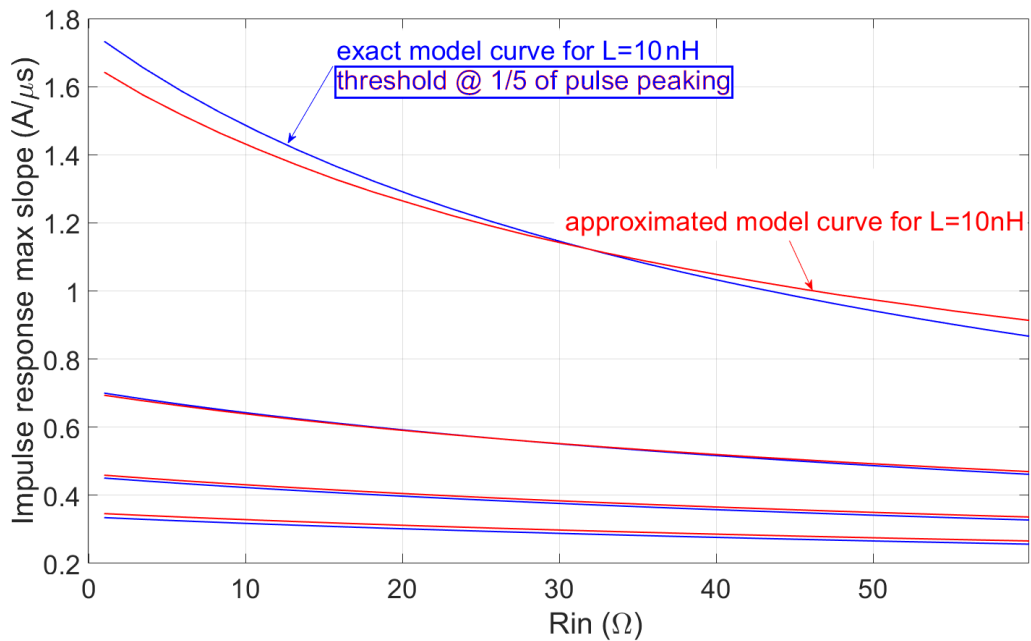
Fig. 2-15: Curves comparing the trends of the slope of the current pulse as a function of R_{in} for the approximate model with optimal time pickoff and the exact model with a suboptimal threshold value. The circuital parameters are equal to the parameter set used for the curves in Fig. 2-14.

Fig. 2-16 shows the output waveform and its slope for C_{in} = 0.5 pF, BW = 0.5 GHz R_{in} = 30 Ω, and L = 100 nH. If the threshold for time pickoff is chosen to be as large 1/5 of the pulse peak, the value of the slope is not much different from its maximum value, that would be retrieved if just an ideal threshold were set at the optimal point (maximum of the first order derivative of the output pulse).

In fact, the slope is rather flat around the inflection point of the response and it cannot change abruptly around the peak; this is also because of low-pass filtering and parasitic inductance that smooth the leading edge of the signal.
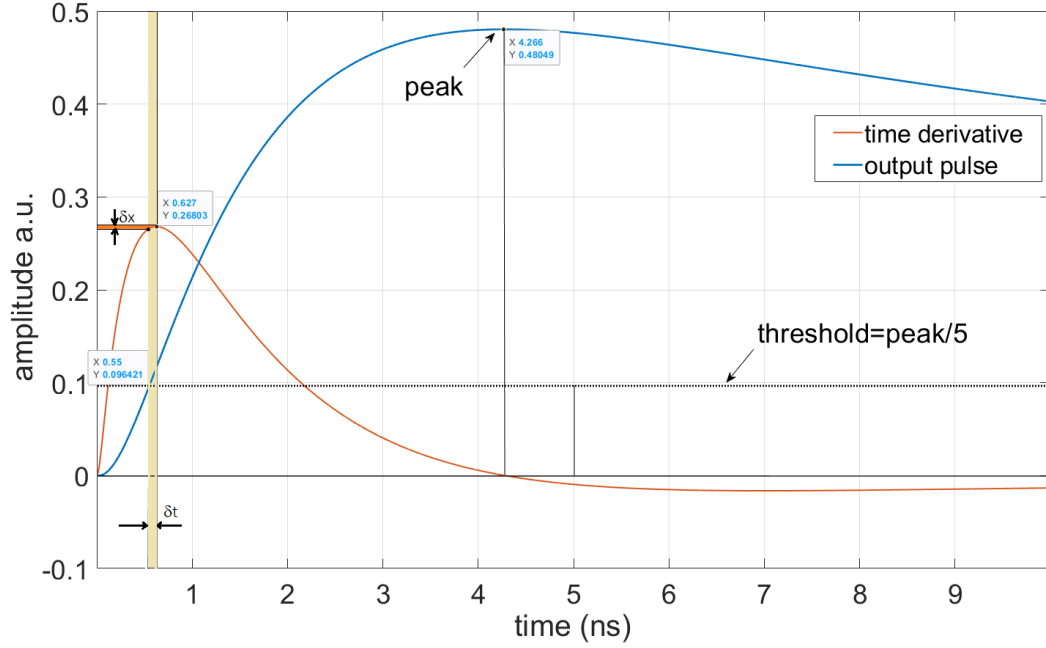
Fig. 2-16: The pulse and its time derivative: the difference between the maximum slope and the slope for the threshold at 1/5 of the peak is very small.

To take a further step towards the full comprehension of the influence that the parameters of the SiPM, the parasitic interconnection elements and the front-end electronics exert on the dynamic behaviour of the detection system, a deeper analysis of the approximate expression of the response (25) is essential.

After rewriting (25) as

$$V_{I,out}(s) \cong M \cdot \frac{1}{R_s} \cdot \frac{1}{(1 + \tau_A s) \cdot (1 + \tau_L s)} \cdot \frac{1}{s} = H_I(s) \cdot \frac{1}{s} = Y_I(s),$$

$$M = \frac{Qtot \cdot K_R}{Cg + N \cdot Cd} \cdot \frac{\tau_q}{\tau_p} \tag{33}$$

the inverse Laplace transform of (33), that very accurately approximates the leading edge of the voltage response for a current-mode approach, can also be regarded as the unity step response $y_I(t)$ of the second order low-pass system whose transfer function is

$$H_I(s) = \frac{1}{R_s} \cdot \frac{M}{(1 + \tau_A s) \cdot (1 + \tau_L s)} \tag{34}$$

69

Instead, for a voltage-mode approach, K_R in (33) must be replaced with K_R', thus the transfer function (34) becomes:

$$H_V(s) = \frac{1}{R_s} \cdot \frac{M'}{(1 + \tau_A\,s\,) \cdot (1 + \tau_L s)} \qquad (35)$$

where $M' = \frac{Qtot \cdot K_R'}{Cg + N \cdot Cd} \cdot \frac{\tau_q}{\tau_p}$.

Since the denominator has two real and negative roots associated to the time constants τ_A and τ_L, the second order low-pass system is overdamped and may possibly become critically damped if $\tau_A = \tau_L$. It shapes the leading edge of the output pulse, thus influencing the timing performance of the detection system. The time constants τ_A and τ_L contribute to shape both the rise time (hence the slope) of the response and the system bandwidth (hence the noise).

Introducing again the normalization variable $\theta = {^{\tau_A}}/_{\tau_L}$, and using expression (29) for t_{MAX_S}, Eq. (31) and (32) become:

$$\boldsymbol{Slope_{V,I}}(t_{MAX_S}) = \frac{Qtot}{Cg + N \cdot Cd} \cdot \frac{\tau_q}{\tau_p} \cdot \frac{A_I R_L}{L_{par}} \cdot \theta^{\frac{1}{1-\theta}} \qquad (36)$$

$$\boldsymbol{Slope_{V,V}}(t_{MAX_S}) = \frac{Qtot}{Cg + N \cdot Cd} \cdot \frac{\tau_q}{\tau_p} \cdot \frac{A_V R_{in}}{L_{par}} \cdot \theta^{\frac{1}{1-\theta}} \qquad (37)$$

These equations show that parasitic interconnection elements have a significant impact on the achievable slope, thus influencing timing performance.

After the transimpedance gain has been chosen and L_{par} has already been estimated, Eq. (36) states that the slope just depends on θ and that, by fixing θ, the slope shall be fixed as well. On the other hand, it is apparent from Eq. (37) that the trendline of the slope as a function of θ retraces the same trendline of Eq. (36), though, now, its magnitude is decided also by the input resistance R_{in}. This explains why, once the bandwidth and the parasitic inductance are known, in a voltage mode approach the slope tends to increase when the input resistance increases, adversely to what happens in a current mode approach. This will become clearer

with some practical examples. First, a bandwidth BW = 0.5 GHz has been chosen and the corresponding τ_L has been calculated; secondly, R_{in} has been swept from 0.1 to 100 Ω and the corresponding array of values has been calculated for θ. It was found in few tries that taking the gain A_iR_L in (36) equal to 25 times the gain A_v in (37) and dropping all the common parameters would make it possible to compare the slopes within the same range of values on the same picture, as apparent from Fig. 2-17. With this intentional choice, the plots of the slopes intersect for R_{IN} = 25 Ω. Eventually, equations (36) and (37) have been rearranged to obtain the following *normalized* expressions:

$$Slope_{V,I}(t_{MAX_S}) = \frac{25}{L_{par}} \cdot \theta^{\frac{1}{1-\theta}} \qquad Slope_{V,I}(t_{MAX_S}) = \frac{R_{in}}{L_{par}} \cdot \theta^{\frac{1}{1-\theta}} \qquad (38)$$

Fig. 2-17 shows the plots from expressions (38) as functions of θ when varying R_{in} with L_{par} = 70 nH.
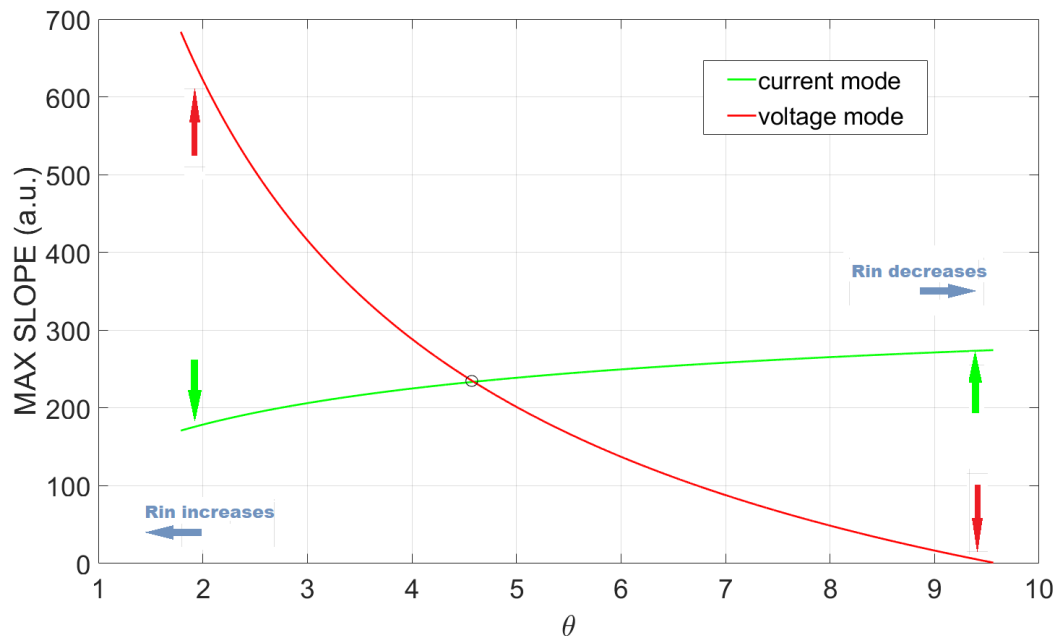

Fig. 2-17: The slopes as a function of ϑ when varying R_{in} with L_{par} = 70 nH.

The intersection point of the curves of Fig. 2-17 is the point where the equality condition is satisfied and corresponds to R_{in} = 25 Ω.

It makes a big difference moving from that point either to the left or to the right, depending on whether the green curve or the red curve is being tracked. The trend of the maximum achievable slope for the current-mode is flatter than that for the voltage-mode and it weakly depends on R_{in} when $R_{in} > 25\ \Omega$. On the contrary, the slope for the voltage-mode is strongly dependent on the value of θ, that is R_{in}; in fact, the steepness rolls off rapidly moving sideways to the right of the equality point, whereas it increases moving to the left.

Fig. 2-17bis shows the plots corresponding to expressions in (38) when varying the bandwidth, i.e. τ_L, with L_{par} constant and R_{in} equal to 5 Ω, 25 Ω, 50 Ω.
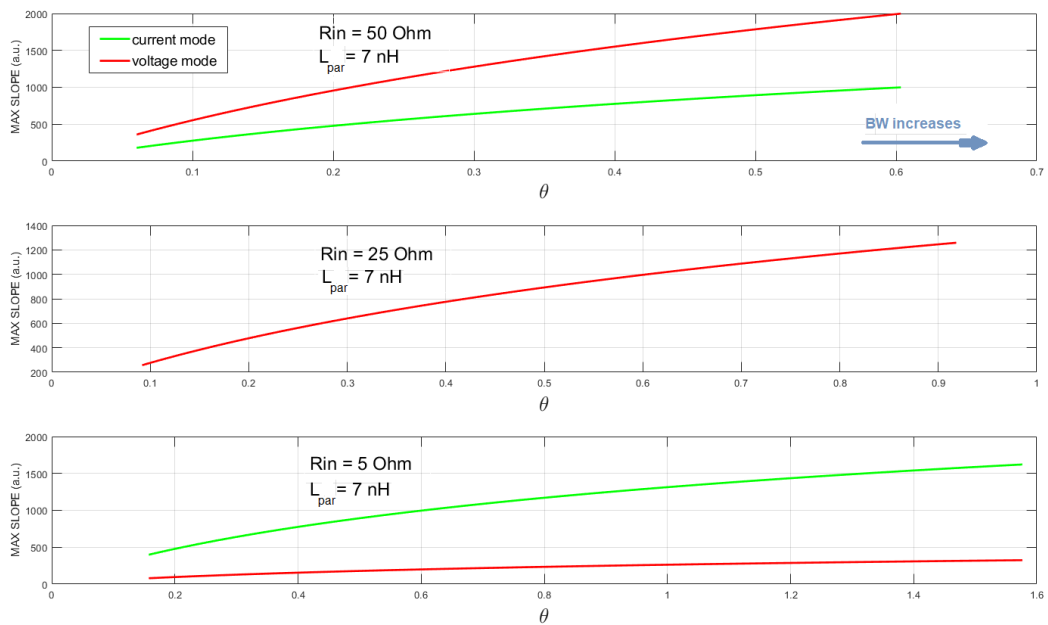
Fig. 2-17bis: The slopes as a function of the normalization variable ϑ, varying BW for three values of R_{in} and L_{par} = 7 nH.

It is apparent that for R_{in} = 25 Ω the curves are coincident and that while the magnitude of the slope is larger with the current-mode rather than with the voltage-mode for R_{in} smaller than 25 Ω, the opposite holds true whenever R_{in} is greater than 25 Ω.

Fig. 2-17ter shows the plots corresponding to expressions in (38) when varying the inductance while keeping bandwidth BW constant and equal to 1 GHz.
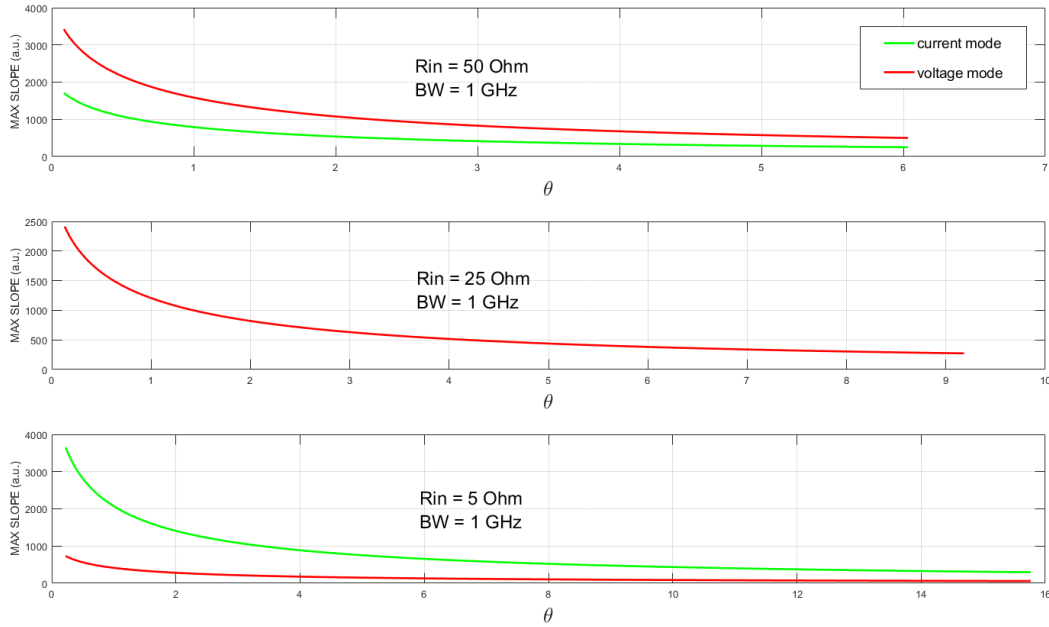
Fig. 2-17ter: The slopes as a function of the normalization variable ϑ, varying Lpar for three values of R_{in} and BW = 1GHz.

Here again for R_{in} = 25 Ω the curves are coincident and it is apparent that, while the slope of the current mode is better than voltage mode when a lower value is chosen for Rin, the opposite occurs when R_{in} is greater than 25 Ω, though the trendline is different with respect to the previous plot, because now the steepness of the slope for both cases diminishes with increasing L_{par}.

To analyse the trend of the slopes against the bandwidth, three values have been considered, namely 0.5 GHz, 1 GHz, 3 GHz. The curves are shown in Fig. 2-18.
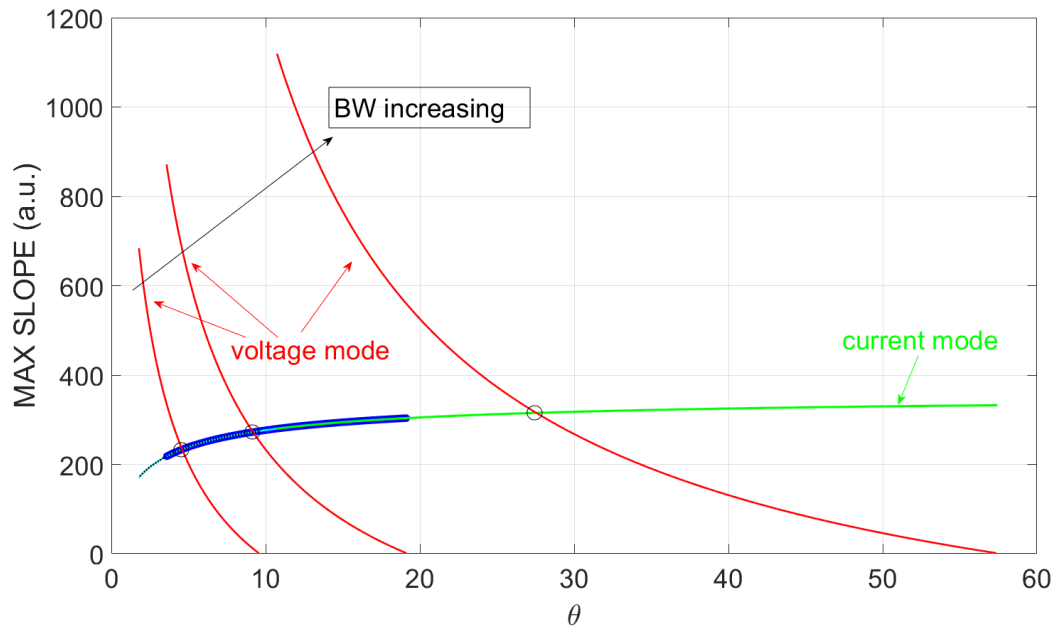
Fig. 2-18: Slopes for three values of bandwidth.

The curve of the slope for the voltage-mode shifts to the right if the bandwidth is increased.

This confirms the fact that for the voltage-mode, when the input resistance varied holding a fixed value for the bandwidth and the parasitic inductance, the input resistance reverses the trend that can be observed for the current mode because it directly contributes to increase the slope acting as a multiplying linear function.

One could get to the same conclusion observing the plot in Fig. 2-19, where the evolution of the output waveforms for a current and a voltage mode in the time domain are shown.

As R_{in} increases the pole of the system transfer function associated with τ_A is pushed away to higher frequencies for both the current-mode and the voltage-mode. However, while the slope for current-mode decreases because the amplitude of the current signal decreases as well, for the voltage-mode it increases with R_{in}, as the slope is being multiplied by R_{in} itself.
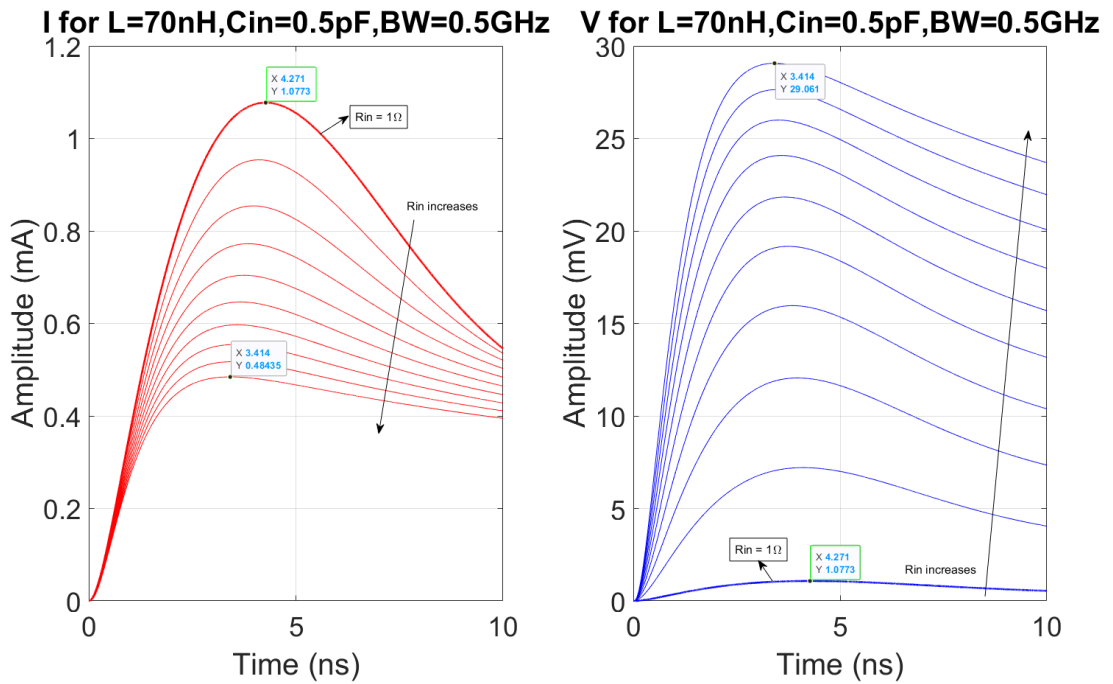
Fig. 2-19: The current and voltage responses as functions of input resistance.

On the other hand, it is relevant to note that R_{in} influences the effective recovery time of the detector τ_{reff}, i.e. the time constant which rules the tail of the output current pulse as pointed out in [12] and its value must be kept reasonably low.

The concepts exposed hitherto are simple and effective, even though obtained through approximations. The first order approximation of Eq. (13) as reported in (24) is useful to carry out a qualitative study of the timing performance of the detection system once the SiPM model is assigned and a preliminary estimation of the parasitic components is available. The approximate transfer function of the whole detection system formed by the term $G_1(s)$, accounting for the SiPM parameters, the first order term coming from Eq. (24) and by the first order amplification factor proves to be effective in describing the behaviour of the fast raising edge of the signal.

2.7 TWO EXAMPLES WITH NOISE INCLUDED IN THE MODEL

Any signal processing that requires amplification and filtering may add noise to the measurement. The output noise power depends on the gain and the bandwidth of the measurement system, and it is also influenced by the impedance of the driving source; in our case, those parameters need to be fine-tuned to minimize noise and maximize the signal slope targeting the best achievable jitter, the figure of merit that characterizes the timing performance of the detection system.

Jitter prediction in our system is not a trivial task and different results can be obtained according to the specific approach, either voltage or current mode. The generalized set of simplified equations devised for slope can dramatically reduce the calculation effort but does not account for noise, whose impact on the jitter is a function of the same parameters that determine the signal slope. Moreover, the design choices that lead to reduce noise and increase signal slope are often conflicting, which suggests that a speed-noise trade-off must be necessarily found.

THE BASIC NOISE MODEL FOR AMPLIFIERS

To calculate the sensitivity limits of a noisy amplifier, the noise performance is usually expressed in terms of input referred noise, which would give the same output noise as the circuit under consideration after inherent noise sources have been removed. Such consideration allows one to compare directly the equivalent input noise with the amplitude of the input signals, thus assessing directly the effect of noise on those signals.

A general and useful representation of noise for network analysis is based on two equivalent noise sources, the input equivalent voltage noise source and the input equivalent current noise source, as shown in the Fig. 2-20 below.
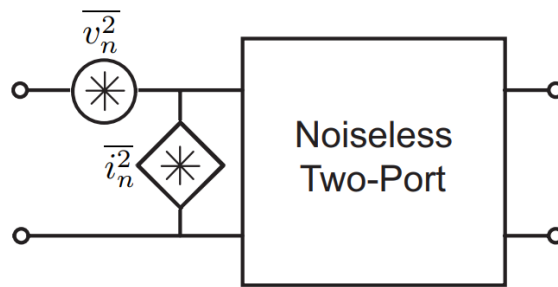
Fig. 2-20: Noise model of a two-port network.

The voltage and current noise sources are described in terms of power spectral densities, $\overline{v_n^2}$, $\overline{i_n^2}$ and may vary with frequency, operating point and the amplifier components and architecture. In the case of amplifiers, it will primarily be the input element (typically a transistor) that has the greatest impact on its noise performance. The model in Fig. 2-20 is valid for any source impedance, provided the correlation between the two noise sources is considered. When the correlation becomes large, it may be convenient to go back to the original circuit for calculation; on the contrary, when the correlation is negligible, noise power spectra can directly be measured and characterized. This is done by first short circuiting the input of the noisy circuit and measuring the output noise to get to $\overline{v_n^2}$ and after by open circuiting the input to get to $\overline{i_n^2}$.

The internal noise sources in operational amplifiers are normally uncorrelated. That is, they are randomly related to each other in time and there is no systematic phase relationship. Uncorrelated noise quantities can be combined as root sum-squares. Thus, if $\overline{n_1^2}$, $\overline{n_1^2}$ and $\overline{n_3^2}$ are uncorrelated noise power densities, then their combined value is

$$N_{rms} = \sqrt{\overline{n_1^2} + \overline{n_1^2} + \overline{n_1^2}}. \tag{39}$$

The basic approach in noise error calculations, then, is to identify the noise sources, segment them into groups conveniently defined (for instance in terms of the shape of their noise spectral densities), compute the root mean square (rms) value of each group, and then combine them by root-sum-squares, to get the total

noise. When dealing with a typical application that requires the optimization of the signal-to-noise ratio (SNR), the technique for noise calculation is applied at the input terminals of the amplifier where the highest signal to noise ratio can be found and where all the noise sources should be referred to.

The approach is different with the design of readout electronics for SiPM when the performance of interest is the timing accuracy and leading-edge discrimination is the chosen time pick-off technique. The total jitter is computed as the ratio between the *rms* electronic noise V_{no} at the output of the preamplifier and the slope of the pulse $v_o(t)$ at the input of the discriminator around the chosen threshold V_{th} as in Eq. (11). Consequently, the noise sources that corrupt the signal (causing jitter) must be referred to the preamplifier output after each of them are being shaped by a specific transfer function (noise gain) that ultimately depends on the specific circuit solution adopted to read out the SiPM signal. An overview of the main architectures commonly applied for the readout electronics dedicated to SiPM detectors has been proposed in [5]. As already pointed out, the current mode approach and the voltage mode approach share the preferences for single photon timing measurements applications, each having peculiar advantages and drawbacks.

To make noise analysis easier, it is appropriate to focus on the characteristics of the SiPM as driving source impedance connected to an amplifier with finite input impedance including the effects of parasitic electrical components. For a current mode and a voltage mode approach, respectively, a suitable choice of the amplification device also helps to simplify the analysis.

Once the driving source impedance is known and properly characterized, an analytical study of the effects of noise can be carried out after some reasonable assumption and the results of this analysis can be easily validated.

The purpose of the following analysis is to compare two typical preamplifier topologies that represent the basic configurations used when either a current mode or a voltage mode is adopted.

Usually, as well-known, R_{in} is kept low enough to preserve the SiPM linearity [13]. In case of voltage-mode approach, R_{in} is a linear resistor that transforms the sensor current pulse into a voltage, thus, as already mentioned, the preamplifier must be a suitable voltage amplifier, typically a bipolar that has a high transition frequency and guarantees high bandwidth (limitation is only τ_L). Conversely, for a current mode signal processing, a MOSFET is more suitable because it is used as a current buffer and current noise contribution associated to the base of bipolar is missing.

Based on the preceding short dissertation on the basic noise model of an amplifier, either equivalent input voltage noise or current noise can be neglected when the source impedance is characterized by extreme values. From guidelines of front-end design, rules of thumb for the selection of a noise-optimized amplifier are as follows:

- dealing with a relatively small source resistance forces the designer to select a voltage amplifier with a low-noise input voltage, since the specification on input noise current is not so important, as the input resistance makes its contribution to voltage noise negligible. This may steer the designer toward a bipolar-input op-amp;
- if, on the other hand, the source resistance is large, it is advisable to look at an amplification solution with a very low input current noise, thus considering solutions based on JFET or CMOS input op-amps.

To meet the requirements for low noise design and without losing generality but gaining in simplicity, good candidates to explore the noise performance of both current and voltage mode approaches as a function of parasitic inductance, input resistance and bandwidth are:

1) the Common Emitter amplifier employing an RF BJT;
2) the Common Gate amplifier employing an RF MOSFET.

In Fig. 2-21 the symbols of the BJT and the MOSFET are represented along with the respective noise sources.
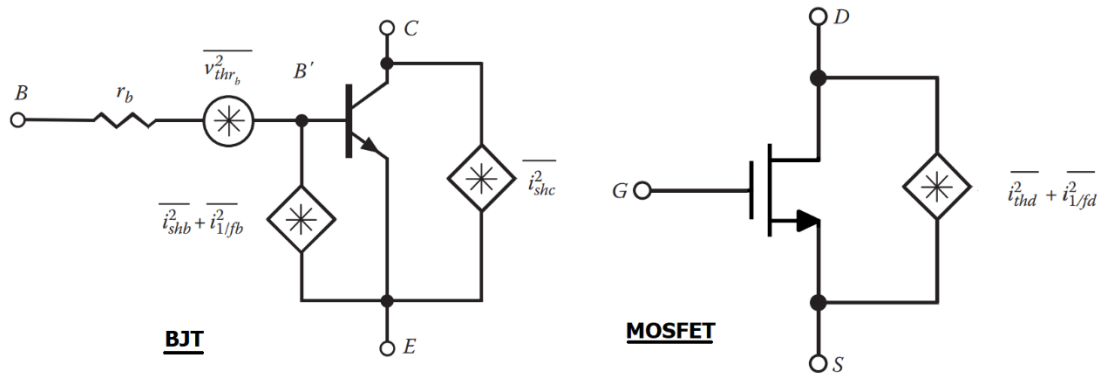
Fig. 2-21: Diagrams of BJT and MOSFET with noise sources.

For the BJT the thermal noise of the base-spreading resistance, the shot noise and the 1/f noise of the base bias current I_B and the shot noise of collector current I_C are reported. The power spectral densities (mean-square values) of these noise sources are given respectively by:

$$\overline{v_{r_b}^2} = 4kT\, r_b \tag{40}$$

$$\overline{i_{shb}^2} = 2q\, I_B \tag{41}$$

$$\overline{i_{1/f_b}^2} = \frac{K_f\, I_B^{\gamma}}{f^{\gamma}} \tag{42}$$

$$\overline{i_{shc}^2} = 2q\, I_C \tag{43}$$

In the next steps involving noise calculation, flicker noise (42) and base bias current noise (41) will be neglected since they give only second order contributions to the overall noise.

For the MOSFET the main noise sources are the thermal noise and the 1/f noise of the drain current I_D. Their power spectral densities are given by:

$$\overline{i_{thd}^2} = 4kT\left(\frac{2g_{mCG}}{3}\right) \tag{44}$$

$$\overline{i_{1/f_d}^2} = \frac{K_f\, I_B^{\gamma}}{L^2 C_{ox} f^{\gamma}}. \tag{45}$$

In the next steps involving noise calculation, flicker noise (45) will be neglected.

For all the given equations k = 1.38 · 10⁻²³ (JK⁻¹) is Boltzmann's constant, T is the absolute temperature in Kelvin degrees (K), q = 1.602 · 10⁻¹⁹ (C) is the electronic charge, V_T = kT/q, is the thermal voltage (V).

ASSUMPTIONS

To comply with the analysis developed in previous sections where simplified equations for slope calculation were devised, both CE and CG amplifiers are supposed to have a single pole transfer function and to exhibit the same bandwidth; the bandwidth of the single pole amplifier is a design parameter that can be varied by properly choosing the output load resistor and the output capacitance to form the time constant $\tau_L = R_L C_L$ that sets the dominant pole. To validate this assumption, it is necessary to demonstrate that, for CG and CE amplifiers, the input pole frequency is essentially the higher of the two pole frequencies that shape the frequency response.

- COMMON GATE AMPLIFIER

Common-Gate is a broadband amplifier that does not suffer from Miller effect. In fact, almost no input-output capacitive coupling exists, since the gate-drain capacitance C_{GD} is grounded; moreover, C_{GD} and the load capacitance C_L are in parallel and can be lumped together. If standard small-signal high frequency model is used, the low-pass filters at the input and at the output of the amplifier will have the following 3-dB cut-off frequencies, respectively:

$$f_{H\ input} \cong \frac{g_m}{2\pi C_{GS}}, \qquad f_{H\ output} \cong \frac{1}{2\pi R_L C_L} \qquad (46)$$

Here R_L is the total resistance that loads the drain, while C_L includes C_{GD}. The input impedance is very small, as the transconductance sets the input resistance R_{in} ≅1/g_m to be few tenths of ohm (i.e. 25 Ω if I_D=1 mA and V_{eff}=100 mV) and the C_{GS}, that represents the C_{in}, can be made smaller than 1 pF as the transition frequency can be very high. Hence the output pole is dominant.

- COMMON-EMITTER AMPLIFIER

The frequency behavior of a Common-Emitter amplifier is plagued by the base-collector capacitance that complicates the evaluation of the frequency response and affects the input impedance. Given the capacitance is placed across inverting voltage gain nodes as is the case with the CE amplifier, the Miller effect is used to relieve a lot of computational effort. As pointed out by Miller's theorem, the bridging capacitor can be replaced with a larger capacitor across the input and a marginally bigger one across the output. The capacitor across the input is magnified by voltage gain and a trade-off is often necessary between gain and bandwidth. Nonetheless, under reasonable approximations, typical expressions for the poles of a Common-Emitter amplifier are

$$f_{H\ input} \cong \frac{1}{2\pi R_{in}\{C_\pi+(1+g_m R_L)C_\mu\}}, \ f_{H\ output} \cong \frac{g_m}{2\pi\{(1+g_m R_L)C_L\}} \cong \frac{1}{2\pi R_L C_L}$$

(47)

where R_{in} is the physical resistor that, for a voltage mode approach, converts the sensor current into a voltage signal, while C_{in} is represented by the sum of C_π, the base-emitter capacitance, and the Miller capacitance, namely $C_{in}=C_\pi+(1+g_m\cdot R_L)\cdot C_\mu$. In nearly all Common-Emitter amplifiers, one pole is dominant, and the high 3-dB frequency is essentially the lower of the two pole frequencies.

It is important to notice that the Miller effect pole is rarely considered the non-dominant pole, even though the Miller effect pole is not definitively the dominant pole, as demonstrated in [14]. In fact, the input pole f_{Hinput} is highly dependent on R_{in}, whereas the output pole $f_{Houtput}$ is independent of R_{in}. That dependence-independence relationship of the poles is such that pole at frequency f_{Hinput} is dominant for large R_{in} and pole $f_{Houtput}$ is dominant for small R_{in}. In our case, since R_{in} is assumed to be very small and, provided C_{in} is around 1 pF or less, the high 3-dB frequency associated to the input can be high enough not to dominate. *A cascode amplifier, the cascaded arrangement of a Common-Emitter stage with a Common-Base stage, represents an effective solution for the Miller effect issue,*

using a second transistor, configured as a Common-Base, which decouples the input and the output of the amplifier.

Lastly, both CE and CG amplifiers are required to exhibit the same input resistance and capacitance equal to R_{in} and C_{in}, thus sharing the same value of the input time constant $\tau_{in} = R_{in}C_{in}$. To accomplish the task

$$v_{o_{CE}}(t) = v_{o_{CG}}(t) \tag{48}$$

$$R_{in_{CE}} = R_{in_{CG}} = R_{in} \tag{49}$$

implying, in the low and midrange frequency band

$$g_{m_{CE}}R_L v_{in_{CE}}(t) = g_{m_{CE}}R_L R_{in}i_{in_{CE}}(t) = R_L i_{in_{CG}}(t) \tag{50}$$

$$i_{in_{CE}}(t) = i_{in_{CG}}(t) = i_{in}(t), \tag{51}$$

where $i_{in}(t)$ is the sensor current flowing into the equalized input impedances of both CE and CG amplifiers.

Exploiting the relations (50) and (51) it can be written the following

$$g_{m_{CE}}R_{in} = 1 \tag{52}$$

and, considering that the input resistance of CG amplifier is the reciprocal value of the MOS transistor transconductance, then

$$g_{m_{CE}} = g_{m_{CG}} \tag{53}$$

To fit to the constraint expressed by (53), considering the bipolar transistor is operated in forward active mode and the MOS transistor in the saturation region on the edge of strong inversion with $V_{eff} \cong 100\ mV$, the bias current I_D of the latter should at least be two times the bias current I_C of the former (*the bias current is to be doubled to keep the input resistance low unless feedback is used*).

As far as the choice of R_L is concerned, if I_c = 1 mA and I_D = 2 mA with R_{in}=25 Ω and I_{in} is roughly 10 µA, to have an output voltage signal of around 10mV then R_L is to be 1 kΩ. This means that, to have 1GHz bandwidth, C_L must be 250 fF which is feasible.

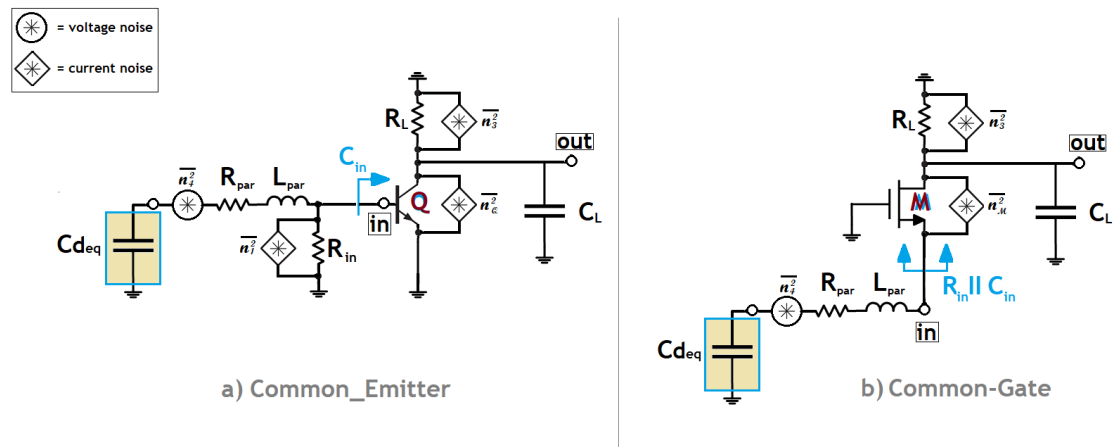Fig. 2-22 shows the basic schemes of both the CE and the CG amplifier with source impedance and noise sources.

Fig. 2-22: Diagrams of a) CE and b) CG amplifiers with noise sources.

In the CE scheme, the resistor Rin is supposed to prevail in the parallel connection against r_π, the small signal input resistance of the BJT, provided the current gain of the transistor is very large.

The circuit topologies for small signals only are presented, i.e. power supplies and biasing are omitted but ideally assumed to make the transistors operate correctly. The transfer functions from the voltage and current input of the amplifiers to their output terminals are:

$$a) \ \ H_{CE}(s) = \frac{V_{OUT}(s)}{V_{IN}(s)} = g_{mCE}R_L \cdot \frac{1}{(1+\tau_L s)}, \ \ b) \ \ H_{CG}(s) = \frac{V_{OUT}(s)}{I_{IN}(s)} = R_L \cdot \frac{1}{(1+\tau_L s)}$$

$$(54)$$

The total output voltage noise is a superposition of inherent sources of noise in semiconductor devices along with every physical component in the circuit generating noise.

The noise power spectral densities of both shot and thermal noises are independent of frequency (white noise) and, in the frequency band of interest for the present analysis, the mean squared noise fluctuations are expected to be proportional to the measurement bandwidth. The complex-valued transfer

function H(f) describes exhaustively the effect of a linear system on the amplitude and phase of an input signal at frequency f. The power gain of such a system at that frequency would then be $|H(f)|^2$. If a signal with power spectral density p(f) passes through the system, then the signal power spectral density at the system's output will be

$$p'(f) = |H(f)|^2 p(f) \qquad (55)$$

Thus generating an average output power (noise variance σ^2)

$$\langle P \rangle = \int_0^{+\infty} |H(f)|^2 p(f) df \qquad (56)$$

With this idea of the noise power spectral density in mind, the problem of determining the total output rms noise for both CE and CG configurations can be addressed.

2.7.1 RMS OUTPUT NOISE OF COMMON-EMITTER CONFIGURATION

Starting from scheme a) in Fig. 2-22, the power noise densities of all the noise sources will be referred to the output by computing their transfer functions one at a time to get the filtered power densities that will be integrated over all frequencies from zero to infinity and the mean square contributions will be summed up together; the square root of the result will lead to the rms output voltage noise, σ_{n_CE}.

Looking back to Fig. 2-1 , it is- clear that, to relief the burden of noise analysis, at high frequencies the equivalent capacitance of the detector (about 100 pF for large area SiPM) can be regarded as a short circuit thus grounding the parasitic resistance R_{par} as reported in the picture below.
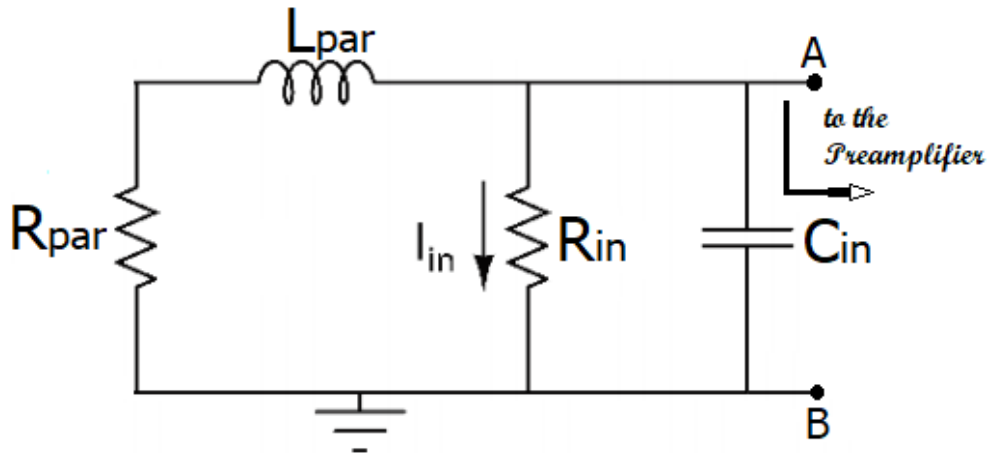
Fig. 2-23: Simplified model of the driving impedance of the amplifier used for noise purpose calculation only.

1) NOISE FROM RESISTOR R_{in}

Only the noise current from R_{in} is included, the equivalent capacitance of the detector is short-circuited, and the following scheme is used to calculate the noise power spectral density of the voltage signal that feeds the CE amplifier.

Fig. 2-24: Scheme for the calculation of the power spectral density of the input voltage signal generated by the current noise source associated to Rin that will be shaped by the CE amplifier.

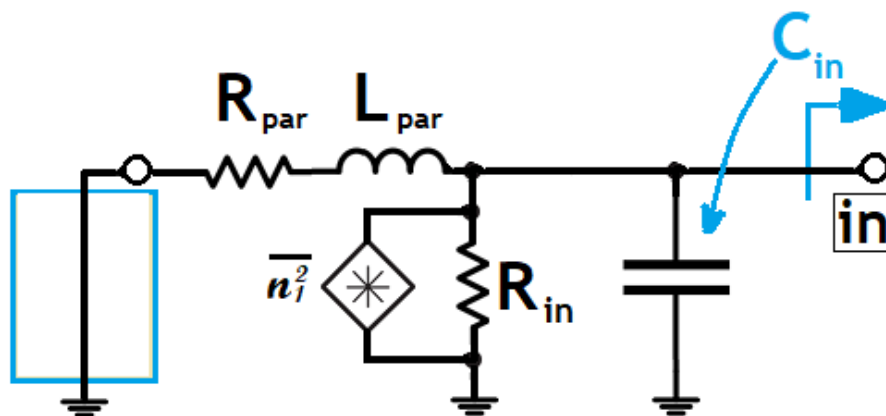The transfer function between the voltage at the input of the CE amplifier and the noise current from resistor R_{in} is

$$H_{R_{in}}(s) = \frac{v_{in,R_{in}}(s)}{i_{n_1}(s)} = R_{in} \cdot (1 + \tau_B s) \cdot \frac{\omega_0^2}{(s^2 + 2\,\delta\,\omega_0\,s\,+\,\omega_0^2)} \qquad (57)$$

where, after assuming $\tau_{in} \ll \tau_B = \frac{L_{par}}{R_{par}}$, it results

$$\omega_0^2 = \frac{1}{b\tau_{in}\tau_B}, \quad \delta \cong \sqrt{\frac{b\tau_B}{4\tau_{in}}}, \quad b = \frac{R_{par}}{R_{par} + R_{in}} = \frac{R_{par}}{R_s} \qquad (58)$$

The poles of (57) are the roots of the denominator, namely s_1 and s_2

$$s_{1/2} = -\delta\,\omega_0 \left(1 \pm \sqrt{1 - \frac{1}{\delta^2}} \right) = -\frac{1}{2\tau_{in}} \left(1 \pm \sqrt{1 - \frac{4\tau_{in}}{\tau_A}} \right), \qquad (59)$$

where $\tau_A = \frac{L_{par}}{R_s}$ is the time constant already defined in (24). The constraints (17) still apply for τ_{in} and τ_A and ensure that the two solutions are real numbers. Furthermore, for typical values of τ_{in} and τ_A

$$0.01 < \frac{\tau_{in}}{\tau_A} < 0.1 \qquad (60)$$

thus, the argument of the square root contained in (59) will range from approximately 0.5 (for the upper limit of (60)) to roughly unity. A linear approximation of the square root function between 0.5 and 1 leads to replace the square root term in (59) with the following $y \cong \left(2 - \sqrt{2} \right) \cdot x + \left(\sqrt{2} - 1 \right)$.

After some manipulations, two real poles are derived: if τ_L, the time constant associated with the dominant pole of the system, is supposed to be larger than τ_{in} (i.e. the input pole isn't dominating the response), the higher frequency pole being close to $1/\tau_{in}$ has negligible filtering effect over the resulting noise, whereas the lower frequency pole does. Therefore, the (57) can be written as

$$H_{R_{in}}(s) = \frac{v_{in,R_{in}}(s)}{i_{n_1}(s)} \cong R_{in} \cdot \frac{R_{par}}{R_s} \cdot \frac{(1 + \tau_B s)}{(1 + r\tau_A s)}, \quad with \ r = \frac{2 + \sqrt{2}}{4}. \quad (61)$$

Therefore, considering expression a) in (54), the power noise voltage density due to R_{in} as a function of frequency at the amplifier output is

$$\frac{\overline{v_{o,R_{in}}^2(\omega)}}{\Delta\omega} = \frac{4kT}{R_{in}} \left| H_{R_{in}}(j\omega) \cdot H_{CE}(j\omega) \right|^2. \quad (62)$$

After substituting (61) and (54) in (62) and integrating over frequency, the average noise power coming from R_{in} becomes

$$V_{o,R_{in}}^2 = 4kTR_{in} \left(\frac{R_{par}}{R_s} \right)^2 Gain_{CE}^2 \cdot \frac{1}{4(r\tau_A + \tau_L)} \cdot \left(1 + \frac{\tau_B^2}{r\tau_A\tau_L} \right), \quad (63)$$

where $Gain_{CE} = g_{mCE}R_L$ is the voltage gain of the CE amplifier.

It is worthy noticing that this analysis is conservative for noise, because a larger amount of noise is filtered out in real life due to the filtering action of the pole that has been neglected. Furthermore, being r < 1, the presence of the parasitic inductance causes a peak in the noise power gain, whatever its magnitude.

2) NOISE FROM RESISTOR R_{par}

In the present case, only the noise current from R_{par} is included, the equivalent capacitance of the detector is short-circuited, and the following scheme reported in Fig. 2-25 is used to calculate the noise power spectral density of the voltage signal that feeds the CE amplifier.
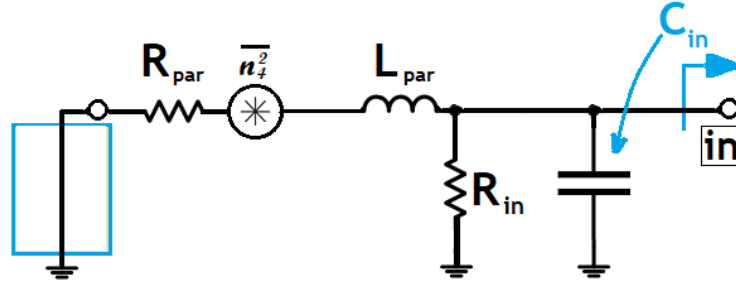
Fig. 2-25: Scheme for the calculation of the power spectral density of the input voltage signal generated by the voltage noise source associated to Rpar. It will be shaped by the CE amplifier.

The transfer function between the voltage at the CE amplifier input and the noise voltage from resistor R_{par} is

$$H_{Rpar}(s) = \frac{v_{in,R_{par}}(s)}{v_{n_4}(s)} = \frac{R_{in}}{R_s} \cdot \frac{1}{\left(\frac{s^2}{\omega_0^2} + \frac{2\,\delta}{\omega_0}\,s + 1\right)} \qquad (64)$$

where the parameters ω_o and δ are the same as those for Eq. (58) and consequently, after assuming $\tau_{in} \ll \tau_B = \frac{L_{par}}{R_{par}}$

$$\omega_0^2 = \frac{1}{b\tau_{in}\tau_B}, \quad \delta \cong \sqrt{\frac{b\tau_B}{4\tau_{in}}}, \quad b = \frac{R_{par}}{R_{par} + R_{in}} = \frac{R_{par}}{R_s} \qquad (65)$$

The poles of (64) are the roots of the denominator, and after retracing the analysis accomplished for the noise calculation from R_{in}, the power noise voltage density from R_{par} as a function of frequency at the amplifier output is

$$\frac{\overline{v_{o,R_{par}}^2(\omega)}}{\Delta\omega} = 4kTR_{par} \left|H_{Rpar}(j\omega) \cdot H_{CE}(j\omega)\right|^2. \qquad (66)$$

After substituting (64) and (54) in (66) and integrating over frequency, the average noise power coming from R_{par} is

$$V_{o,R_{par}}^2 = 4kTR_{par} \left(\frac{R_{in}}{R_S}\right)^2 \cdot Gain_{CE}^2 \cdot \frac{1}{4(r\tau_A + \tau_L)}, \tag{67}$$

where the parameters involved have already been defined.

3) NOISE FROM BJT

If it is assumed that the base-spreading resistance can be neglected along with base current shot noise and that the flicker noise can either be neglected or its effects minimized, the mean square output voltage of the BJT from its collector current shot noise source that contributes to the overall output noise power is

$$\frac{\overline{v_{o,BJT}^2(\omega)}}{\Delta\omega} = 2qI_C \left|\frac{R_L}{(1 + \tau_L j\omega)}\right|^2 = 2kTR_L \cdot Gain_{CE} \left|\frac{1}{(1 + \tau_L j\omega)}\right|^2 \tag{68}$$

After integrating over frequency, the average noise power coming from the BJT is

$$V_{o,BJT}^2 = 2kTR_L \cdot Gain_{CE} \cdot \frac{1}{4\tau_L}. \tag{69}$$

4) NOISE FROM RESISTOR R_L

The thermal current noise from R_L simply sees the parallel combination of R_L and C_L and the mean square output voltage from this resistor is

$$\frac{\overline{v_{o,R_L}^2(\omega)}}{\Delta\omega} = \frac{4kT}{R_L} \cdot \left|\frac{R_L}{(1 + \tau_L j\omega)}\right|^2. \tag{70}$$

After integrating over frequency, the average noise power coming from R_L is

$$V_{o,R_L}^2 = 4kTR_L \cdot \frac{1}{4\tau_L}. \tag{71}$$

After some manipulations involving the average noise power terms from 1) to 4), the rms output voltage of the CE amplifier is obtained and can be written as

$$V_{o,CE} = \sqrt{V_{o,R_{in}}^2 + V_{o,R_{par}}^2 + V_{o,BJT}^2 + V_{o,R_L}^2} \tag{72}$$

2.7.2 RMS OUTPUT NOISE OF COMMON-GATE CONFIGURATION

Starting from scheme b) in Fig. 2-22, the power noise densities of all the noise sources will be referred to the output by computing their transfer functions one at a time to get the filtered power densities that will be integrated over all frequencies from zero to infinity and the mean square contributions will be summed up together; the square root of the result will lead to the rms output voltage noise (σ_{n_CG}).

1) NOISE FROM RESISTOR R_{in}

The incremental input resistance does not generate noise.

2) NOISE FROM RESISTOR R_{par}

Only the noise current from R_{par} is now included, the equivalent capacitance of the detector is short-circuited and the following scheme is used to calculate the noise power spectral density of the current signal that flows into the source terminal of the CG amplifier and is conveyed to the output.
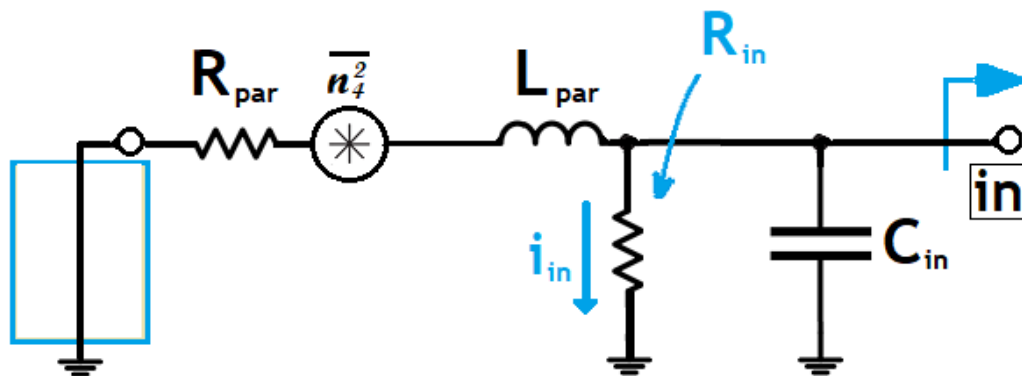

Fig. 2-26: Scheme for the calculation of the power spectral density of the input current signal generated by the voltage noise source associated to Rpar. The CG is a current buffer and the input current will be transformed into an output voltage shaped by the load impedance.

The transfer function between the current through the incremental resistance R_{in} and the noise voltage from resistor R_{par} is

$$H_{R_{par}}(s) = \frac{i_{in,R_{par}}(s)}{v_{n_4}(s)} = \frac{1}{R_s} \cdot \frac{1}{\left(\frac{s^2}{\omega_0^2} + \frac{2\,\delta}{\omega_0}\, s + 1\right)}, \tag{73}$$

where the parameters ω_o and δ are the same as those for Eq. (58) and (65) and consequently, after assuming $\tau_{in} \ll \tau_B = \frac{L_{par}}{R_{par}}$, it can be written

$$\omega_0^2 = \frac{1}{b\tau_{in}\tau_B}, \quad \delta \cong \sqrt{\frac{b\tau_B}{4\tau_{in}}}, \quad b = \frac{R_{par}}{R_{par} + R_{in}} = \frac{R_{par}}{R_s}. \tag{74}$$

The poles of (73) are the roots of the denominator, and after retracing the earlier analysis and approximations, the power noise voltage density from R_{par} as a function of frequency at the amplifier output can be written as:

$$\frac{\overline{v_{o,R_{par}}^2}(\omega)}{\Delta\omega} = 4kTR_{par}\,\left|H_{R_{par}}(j\omega) \cdot H_{CG}(j\omega)\right|^2. \tag{75}$$

After substituting (73) and (54) in (75) and integrating over frequency the average noise power coming from R_{par} is

$$V_{o,R_{par}}^2 = 4kTR_{par}\left(\frac{R_L}{R_s}\right)^2 \cdot \frac{1}{4(r\tau_A + \tau_L)}, \tag{76}$$

where the involved parameters have already been defined.

3) NOISE FROM MOSFET

The inherent thermal noise of the MOSFET is represented by a current generator in parallel to the device whose value depends on the transconductance and on a parameter that is a function of the device operating region and short channel effects, to name the major contributor. It is assumed to be equal to 2/3, as in Eq.

(44). Only the noise current from MOSFET is included, the equivalent capacitance of the detector is short-circuited, and the following scheme is used to calculate the noise power spectral density of the voltage signal at the output of the CE amplifier.
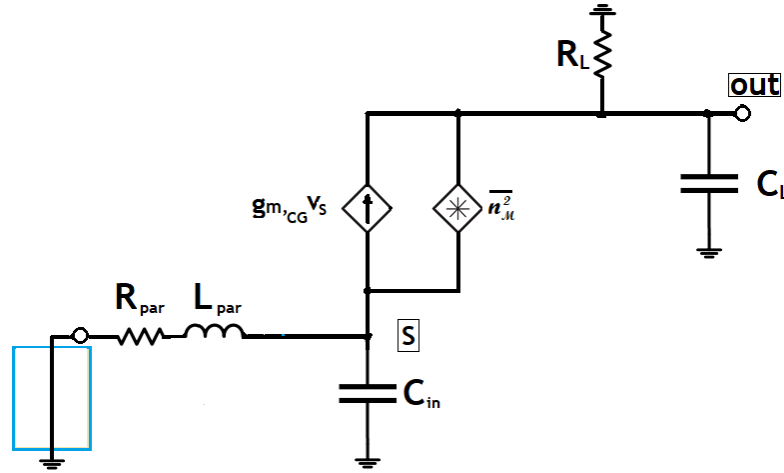

Fig. 2-27: Scheme for the calculation of the power spectral density of the output voltage signal from MOSFET thermal noise, using the small signal model of the transistor.

The transfer function between the output current through the load impedance R_L and C_L and the noise current from MOSFET is

$$H_M(s) = \frac{i_{o,M}(s)}{i_{n_M}(s)} = \frac{R_{in}}{R_s} \cdot \frac{1 + R_{par}C_{in}s + L_{par}C_{in}s^2}{\left(\frac{s^2}{\omega_0^2} + \frac{2\delta}{\omega_0}s + 1\right)} \qquad (77)$$

where, the parameters ω_o and δ are the same as those for Eq. (58) and consequently, after assuming $\tau_{in} \ll \tau_B = \frac{L_{par}}{R_{par}}$, it can be written

$$\omega_0^2 = \frac{1}{b\tau_{in}\tau_B}, \quad \delta \cong \sqrt{\frac{b\tau_B}{4\tau_{in}}}, \quad b = \frac{R_{par}}{R_{par} + R_{in}} = \frac{R_{par}}{R_s}, \qquad (78)$$

while the second order polynomial at the numerator is characterized by

$$\omega_{0Z}^2 = \frac{1}{L_{par}C_{in}}, \quad \delta_Z = \frac{R_{par}}{2}\sqrt{\frac{C_{in}}{L_{par}}} \qquad (79)$$

The roots of the numerator are two complex and conjugate zeroes for the transfer function. Interestingly, the asymptotic behaviour is the same as it would be for a repeated simple pole at ω_{0Z}, except for the peak that is not accounted; the slope decreases by 40dB/decade at this location. So, the uncorrected Bode plot for the complex conjugate zeroes is the same as it would be for a simple repeated pole, with ω_{0Z} behaving as the break point in this case. To speed up the analysis, just one of the two zeroes is kept for computing the average power noise. This approximation is conservative for noise and is in part balanced by the dominant pole approximation.

The zero is equal to ω_{0Z} and is placed at a very low frequency, playing as a zero in the origin. The poles of (77) are the roots of the characteristic equation and have already been calculated leading to the dominant pole approximation. The power noise voltage density from the MOSFET as a function of frequency at the amplifier output is

$$\frac{\overline{v_{o,M}^2(\omega)}}{\Delta\omega} = 4\text{kT}\left(\frac{2g_{m_{CG}}}{3}\right) \cdot |H_M(j\omega) \cdot H_{CG}(j\omega)|^2. \qquad (80)$$

After substituting (77) and (54) in (80) and integrating over frequency the average noise power coming from the MOSFET is

$$V_{o,M}^2 = 4\text{kT}\left(\frac{2g_{m_{CG}}}{3}\right) \cdot \left(\frac{R_{in}}{R_s}\right)^2 \cdot Gain_{CG}^2 \cdot \frac{1}{4(r\tau_A + \tau_L)} \cdot \left(1 + \frac{L_{par}C_{in}}{r\tau_A\tau_L}\right), \qquad (81)$$

where the parameters have already been defined and $Gain_{CG} = R_L$.

4) NOISE FROM RESISTOR R_L

The thermal current noise from R_L simply sees the parallel combination of R_L and C_L and the mean square output voltage from this resistor is

$$\frac{\overline{v_{o,R_L}^2(\omega)}}{\Delta\omega} = \frac{4kT}{R_L} \cdot \left|\frac{R_L}{(1 + \tau_L j\omega)}\right|^2 \qquad (82)$$

After integrating over frequency, the average noise power coming from R_L is

$$V_{o,R_L}^2 = 4kTR_L \cdot \frac{1}{4\tau_L}. \qquad (83)$$

The sum of all the average noise power terms is

$$V_{o,CG}^2 = V_{o,R_{par}}^2 + V_{o,MOS}^2 + V_{o,R_L}^2 \qquad (84)$$

After some manipulation, the rms output voltage of the CG amplifier is obtained:

$$V_{o,CG} = \sqrt{\frac{kT}{C_L} \left\{ \left[\frac{R_{in}\tau_L^2}{4R_s^2 C_L(r\tau_A + \tau_L)}\right] \cdot \left[\frac{2}{3}\left(1 + \frac{L_{par}C_{in}}{r\tau_A\tau_L}\right) + 4\right] + 1\right\}} \qquad (85)$$

As an example, for the CG configuration (current-mode approach), the expression of noise from (85) and the expression of slope from (36) can be merged to form the total output time jitter that is calculated as the ratio between rms noise and maximum slope. Assuming L_{par} = 40 nH, C_{in} = 400 fF, R_L = 50 Ω, and $A_I R_L$ = 25, R_{in} and BW are varied to get a 3D plot of the results, as reported in the following figure. In the same figure, the time jitter as a function of bandwidth is also reported with R_{in} = 25 Ω.

It is evident that there exists a minimum value for jitter for a given set of circuital parameters, and that, depending on the value of the parasitic inductance, once the input resistance is chosen, an optimal bandwidth can be identified that doesn't match with the as large as possible one.
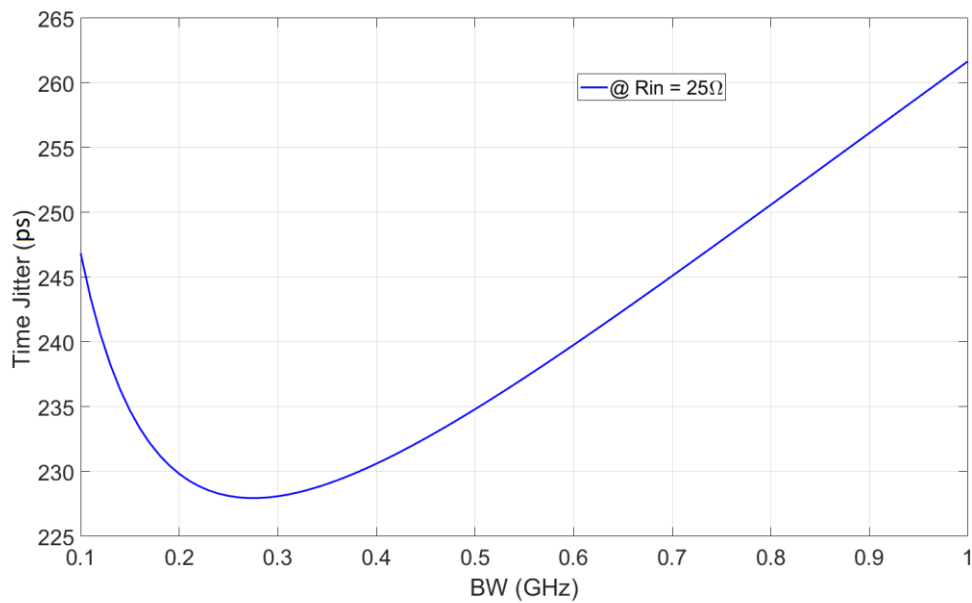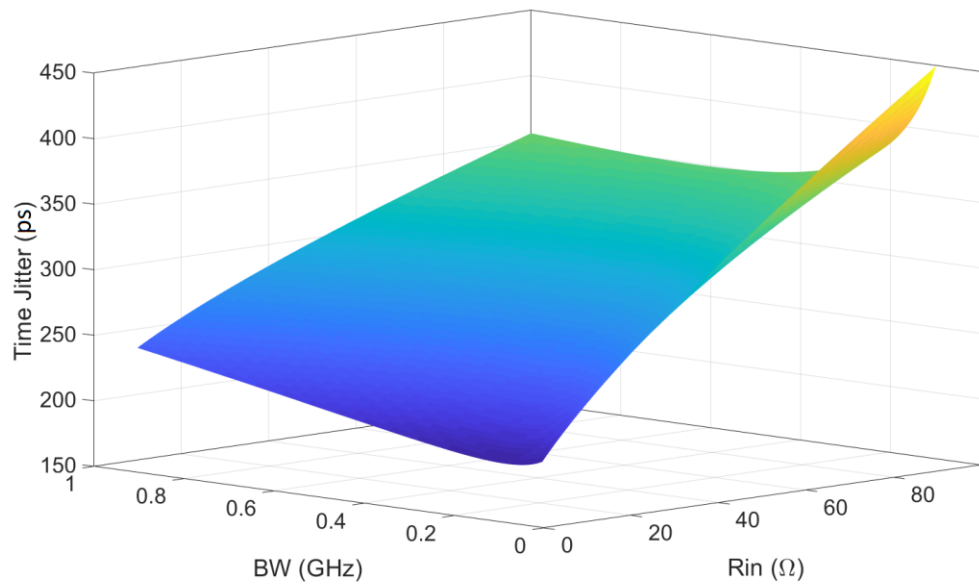

Fig. 2-28: Time Jitter as a function of input resistance and bandwidth.

2.8 CONCLUSION

The problem of approximating the transfer function of a SiPM-based detection system model has been faced in the present work. A new methodology has been developed to determine the slope of a detection system based on SiPM. By means of mathematical simplifications in the waveform region of fast transient involved in timing extraction, the contributions to the shaping of the output waveform, respectively coming from the SiPM, the interconnect series parasitic impedance and the front-end, have been decoupled in the s-domain. Electronic noise and bandwidth limitation of the front-end electronics cause an inevitable increase of the time jitter. As far as the design of front-end electronics for SiPM is concerned, at earlier stages of the design flow numerical simulations are rarely effective to provide designers with useful indications for achieving the desired specifications. Whenever it is required to push for timing performance, first it is highly desirable to refer to simplified analytical expressions reproducing with enough mathematical accuracy the characteristics of the system response as a function of model parameters, such as the parasitic coupling inductance L_{par}, the input resistance of the current buffer R_{in} and the amplifier bandwidth, and focusing on the most relevant parts of the output waveform that have an impact on the jitter. If accurate timing is required, the parasitic components must be reviewed as an inherent part of the system that influences the performance in the same way as the intrinsic electrical parameters of the SiPM and the characteristics of the front-end do. The role of the series parasitic inductance against the slope of the timing signal has been eventually showcased by simple mathematical relationships. Increasing L_{par} introduces a strong penalty on speed; furthermore, contrasting the common feeling, even decreasing the input resistance causes the system to slow down. However, broadly speaking, the choice of the current mode approach for the design of the front-end amplifier appears to be the best solution for fast timing as it allows the designer to properly trade input resistance with both bandwidth and gain, in a more convenient way with respect to the voltage mode approach.

3 CHAPTER 3

3.1 INTRODUCTION

The front-end electronics may represent the major limitation to meeting the specifications of an arbitrary detection system based on SiPMs with large area. Typically, when the greatest expectation of the design is placed on timing measurement to improve Single Photon Time Resolution (SPTR), high gain-bandwidth product, very low input impedance and low noise must be addressed. The gain should be kept high enough to detect weak current signals and the bandwidth large enough to preserve the fast timing as accurately as possible. On the other hand, increasing large bandwidth could degrade jitter, because noise level would also increase. In the measurement of energy, inaccuracies and non-linearity usually arise from the intrinsic noise of the detector itself, associated to afterpulsing, optical crosstalk, dark pulses and gain fluctuations from cell to cell; moreover, in those applications using a scintillator coupled to the SiPM, the energy resolution is affected by the statistics of the photons delivered by the crystal. Being optimized for single photon events, the preamplifier would unavoidably saturate, while trying to accommodate for a very large number of photons, thus imposing a penalty to the detector linearity.

Eventually, the effects of parasitic passive components due to the interconnections should be taken care of from a design perspective, accepting the idea that they will never disappear altogether. It will be demonstrated that, in presence of the parasitic interconnection inductance, the design criteria of maximizing the bandwidth and minimizing the input impedance may not represent optimal design choices to approach the inherent best theoretical timing accuracy of the SiPM.

3.2 A PROOF OF CONCEPT PROTOTYPE

The design of a front-end channel suitable for interfacing large area SiPMs in ToF-PET and single photon counting applications will be presented, as the result of a research project supported by *SLAC National Accelerator Laboratory*, and in part developed with the collaboration of *SLAC TID-AIR research design team*

The design is intended for the readout and digitization of signals from fast photon detectors in applications where a fast timing is required. It is also conceived to allow for high data rate accommodation and wide range charge measurements.

The design workflow has been created within Cadence® Virtuoso® ADE analog design framework. The Analog Design Environment (ADE) has been customized with the Process Design Kit (PDK) licensed by TSMC for its CMOS 130 nm process.

A schematic view of the functional blocks of the SiPM readout system is shown in Fig. 3-1. The Front-End subset hosts the SiPM Analog Front-End that has been designed and validated by simulation and is being presented hereafter.
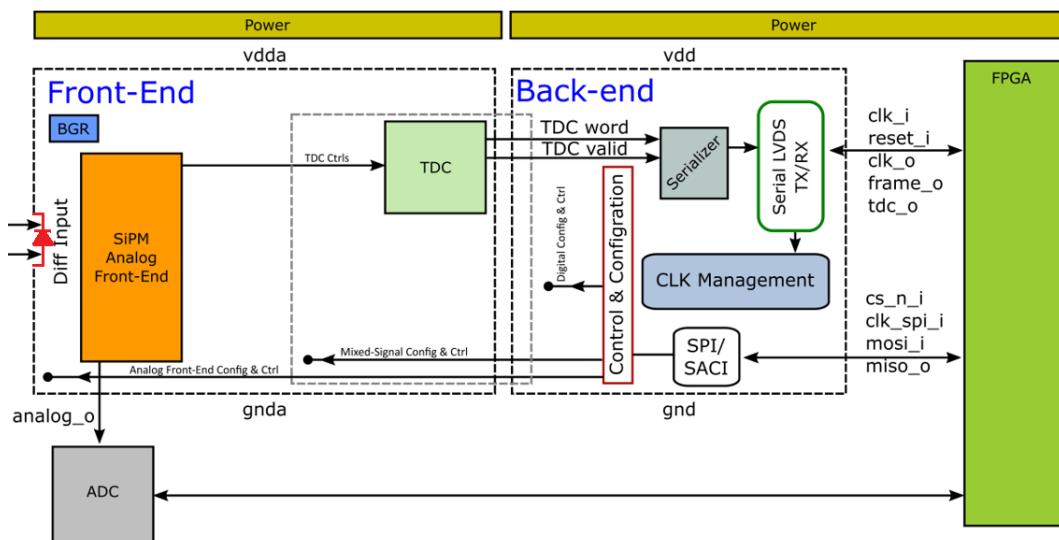

Fig. 3-1: System diagram.

The signal current delivered by the SiPM is converted into a voltage signal and further processed before triggering a timestamp for a valid event. Each event trigger is managed by internal control logic that eventually enables the digitization of both time and charge measurements using an embedded Time-to-Digital

Converter (TDC) and an off-chip Analog-to-Digital Converter (ADC). The back-end circuitry interfaces with the FPGA for data communication and storage, circuit parameter settings and clock distribution.

3.3 PRELIMINARY REQUIREMENTS

The front-end channel is required to be sensible to the single photon with a time jitter specification around 20 ns *rms* (inherent electronics jitter only), while enabling linear and accurate measurement of the charge, from the single photon level up to the relevant number of photons involved in a scintillation event. In this case, an SiPM of the kind described in Table 3-1, that has already been characterized in [1] and tested in SLAC facilities, is supposed to be coupled with a fast, high yield scintillator, e.g. LSO/LYSO.

Table 3-1: Comparison between physical and electrical parameters of three SiPMs.

PARAMETER	FBK NUV-HD	SensL J 3x3 TSV	Hamamatsu S13360
Area	4x4 mm^2	3x3 mm^2	3x3 mm^2
Cell size	25 μm	35 μm	50 μm
Number of cells	25600	5676	3584
Fill Factor *	73 %	75 %	74 %
Max PDE *	45 %	40 %	45 %
Max number of photoelectrons (before saturation) *	15360 [0.6 times # of cells]	3406 [0.6 times # of cells]	2151 [0.6 times # of cells]
Quenching resistor [kΩ] *	2000	240	310
SPAD capacitance [fF] *	37.5	160	110
Recovery time [ns] *	75	38.4	34.1

Room temperature data.

For example, a wrapped and polished Cerium doped LYSO scintillator can give off as much as 13800 photons under a 511 keV gamma ray scintillation event. Considering an ideal optical coupling of the scintillator to the SiPM and a PDE of about 40%, the theoretical number of photons which would contribute to the SiPM signal of a 511 keV event is equal to 5,500; this could result in some linearity

issue for both SiPMs from SensL and Hamamatsu, because the ratio between the maximum number of fired micro-cells and the total no of available micro-cells must be less than 0.6 to avoid saturation.

What has been demonstrated so far is that a realistic estimate of the effective photoelectrons involved is somewhere between 1100 and 2000, thus the maximum charge that the preamplifier is required to handle has been calculated considering the worst case of 2000.

As already pointed out in the first chapter of this work, the total charge released by the SiPM in response to impinging photons depends on the number of firing micro-cells, N_f, on the overvoltage and on the total pixel capacitance. Looking up the data in the previous table, under an overvoltage of 4 V, when N_f is equal to 2000 the maximum total charge is found to be 1280 pC.

The design of the channel prototype is the first step for a highly integrated multi-channel low power ASIC SiPM signal processing to be applied, for instance, in medical PET and ToF-PET applications. Therefore, the channel must feature low power consumption, with average value below 10 mW. The input stage must also provide a differential readout scheme to be interfaced with the SiPM using AC coupling capacitors in series, thus preventing the DC polarization of the front-end from interfering with the SiPM High Voltage biasing.

In the Table 3-2 the main challenging features for the design of the front-end channel have been summarized.

Table 3-2: Main features from preliminary requirements.

TARGET FEATURE	Minimum value	Maximum value
Time Accuracy *	-	**20 ps *rms*** (1 pe⁻)
Charge Measurement	1 pe⁻ → **150 fC**	8000 pe⁻ →**1280 pC**
Power consumption	-	**10 mW**
Supply voltage (typ. **1.2 V**)	**1.08 V**	**1.32 V**

*inherent performance of analog front-end to be guaranteed by design

3.4 DEFINITION OF TARGET SPECIFICATIONS

To provide a better understanding of the design challenges posed by the present project, the major issues are clarified and the decisional process that has led to define the final target specifications based on the preliminary requirements will be explained.

Balanced pseudo-differential inputs.

Single-ended inputs are more susceptible than fully differential inputs to coupled-noise and DC offsets. Pseudo-differential inputs are like fully differential inputs in that they isolate signal ground from the ADC ground, and unlike single-ended inputs, DC common-mode voltages can be cancelled out. Its symmetry results in a good power supply and substrate noise rejection, according to the solution proposed, for instance, in the NINO ASIC [2]. The differential structure will come with two *legs*. Therefore, throughout the design process described in the following, only one of the two 'legs' will be addressed, and all the choices will be extended to the other one for symmetry.

Power Budget.

Typically, an array of SiPM is read out by a multi-channel ASIC whose high channel count demands power dissipation of each channel to be as low as possible. A power budget equal to 10 mW per channel cannot be broken to make the design of ASICS attractive for compact and cost-effective acquisition systems with a very large number of readout channels. Aiming for optimal timing, this is challenging because:

- The differential structure requires double biasing current than single;
- The differential input impedance is doubled
- The bandwidth trades with power.

Low-power design criteria have driven the choice of the new front-end architecture. Feedback techniques and subthreshold operation for the devices are explored to meet the required specification about power consumption while

keeping up expectations about timing accuracy, charge resolution, dynamic range and linearity.

Charge measurement and gain of the analog channel.

An RC active integrator can guarantee a charge-to-voltage gain with a good linearity over a wide range of input charge. From the preliminary requirements, the front-end is required to accommodate for input pulses carrying a charge which spans from 1 pe⁻ to roughly 8000 pe⁻ (considering a charge of about 150 fC for the single pe⁻, i.e. a gain of slightly less than $1 \cdot 10^6$, this corresponds to a maximum charge of about 1200 pC). The integrator cannot manage the full maximum charge without a suitable attenuation (consider that with a 1.2 V supply voltage, to integrate a charge of 1 200 pC without any attenuation an integration capacitance of 1200/1.2 pF, i.e. 1 nF, is needed, which is impossible to be made on-chip). Thus, when progressively large values of the charge are expected, the current signal produced by the front-end must be adequately attenuated and the circuit must be configured in different ways according to the expected dynamic range of the charge to be integrated. For single photon energy resolution, a high sensitivity is required to integrate the minimum charge of 150 fC released in a single photon event. Since we have also severe requirements on the dynamic range of the total charge, we decided to foresee four different configurations for the charge-to-voltage gain of the integrator. The largest gain range can be applied to accommodate for up to 150 piled-up photoelectrons and is suitable also for single photon events. Three more ranges are required to cover the full span of charge up to 8000 pe⁻. This solution has been adopted, for instance, also in the TOFPET2 ASIC, that has four possible ranges: 7.5 - 750 pe⁻, 15 - 1500 pe⁻, 30 - 3000 pe⁻, 60 - 6000 pe⁻, considering a gain of the SiPM equal to $1.25 \cdot 10^6$, i.e. 200 fC for the single photon.

Bandwidth, input impedance, noise.

A half-side input impedance of 10 Ω to ground (20 Ω in differential mode) is the design target value. This value must be guaranteed for each 'leg' almost up to 200 MHz thus allowing for a less than 1ns rise time of the differential output voltage

pulse. After assuming a single photon peak current of 10 μA, a current-to-voltage amplification factor of 1250 would lead the half-side voltage to peak after 2 ns with 12.5 mV height: the amplitude of the output differential signal would be equal to 25 mV, with an average slope of 25 V/μs.

Finally, to get to a time jitter of 20 ps *rms* the output differential noise integrated over the whole frequency bandwidth cannot be lower than 800 μV *rms*, with a crest factor of 5 that gives a peak-to-peak Signal-to-Noise Ratio (SNR) of 22 dB. This is also suitable to perform time measurement with a Leading-Edge Discrimination (LED) technique, as the comparator threshold can be moved as low as possible to seek the maximum slope point and the minimum time jitter.

3.5 DESIGN CRITERIA FOR THE CHOICE OF THE INPUT STAGE

The typical large active area of a SiPM results in a large equivalent capacitance presented at the input stage of the front-end that unavoidably calls for a pre-amplifier solution which exploits the advantage of a current buffer in terms of low input impedance [3], [4]. If a fast response circuit is required, it is important to design with initial emphasis on speed performance, trying to minimize first the adverse impact of the detector capacitance on the bandwidth, which can be done only by proper choice of the input impedance. Nonetheless, as time jitter depends on both speed and noise, the latter cannot be overlooked when aiming at time performance optimization. Indeed, once the system is decided in terms of power consumption, gain and bandwidth, it is not possible to significantly reduce noise anymore: a design strategy is needed to trade speed and gain and trying to possibly find the best trade-off with moderate effort. To gain wide-ranging tunability of the input impedance with minimum impact on power dissipation, the benefits of a feedback mechanism have been exploited; at the same time, the instability issue has been avoided by watching the risk to go after extremely low values of input resistance. In fact, as already proven, the interconnect parasitic inductance along with input resistance and sensor capacitance result in gain

peaking at high frequencies and instability of the circuit with ringing transient response [5].

Another way to increase gain and decrease noise is to change the inherent noise figure of the transistors along with their transconductance by means of proper sizing in weak and moderate inversion regions [6].

Despite intended for PET and spectroscopy, BASIC64 has demonstrated to feature good timing performance, so far. The biasing current of the input CG stage is equal to 2.4mA and the input impedance is equal to 17 Ω. Further decrease of the input impedance is still possible, tough at expense of dramatic increase of power dissipation which would make noise increase with little chance to achieve a time jitter in the picoseconds range.

Feedback can be conveniently exploited to reduce the input resistance of the front-end without increasing too much the power consumption, thus the configuration that has been deemed useful for this task is the Regulated Common Gate (RGC). In Fig. 3-2 the schematic of an RCG-TIA circuit, with a basic model of the SiPM is depicted.
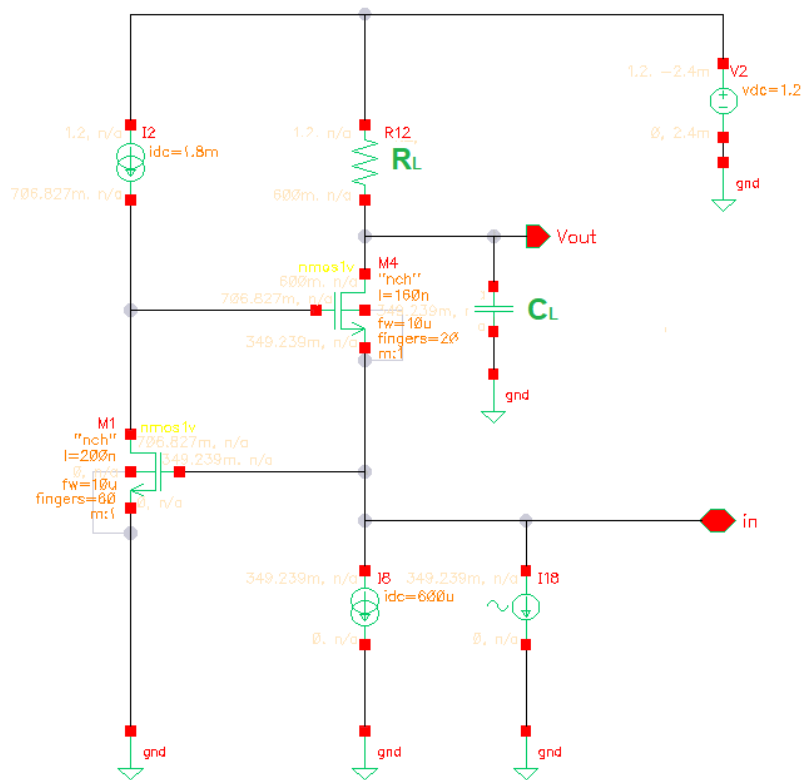
Fig. 3-2: Regulated common gate circuit, with an R_L-C_L load at output node.

The current sources for transistor biasing are ideal and the only parasitic capacitances left are the intrinsic capacitances of the transistors, whereas, to change the bandwidth the load impedance at the output can be varied by changing the capacitance C_L.

The circuit in Fig. 3-2 has been simulated in Cadence Virtuoso Analog Design Environment (ADE). The simulations have been carried out using a basic SiPM model coupled to the current buffer. The model is composed of a short current pulse that accounts for the fast component of the charge $Q_0 \cdot \delta(t)$ and a high frequency equivalent capacitance of the detector C_{Eq}. At a preliminary design stage, this simplified model can describe with sufficient accuracy the very first part of the output pulse and it is suitable to study the behaviour of the system when the threshold for time pickoff is set very close to the signal baseline.

With a proper choice of the transistor size and posing moderate constraint on the biasing current budget, the inherent bandwidth can be made extremely large:

increasing biasing current leads to a bandwidth as large as 5 GHz with an input resistance as low as 1 Ω.

This is demonstrated in Fig. 3-3, where the input resistance is far below 5Ω up to 500 MHz.
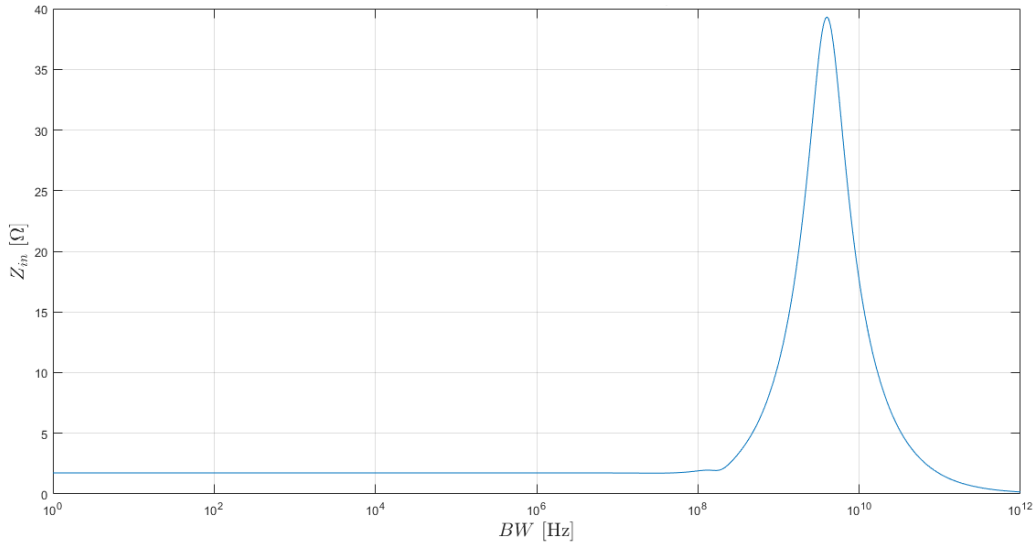

Fig. 3-3: Input impedance of the RGC as a function of the frequency.

Indeed, the peaking module of the input impedance needs to be closely monitored as it occurs in the high frequency band of the spectrum where instead a small value is expected to match with the fast signal of the SiPM. The lower the floor value at midrange, the higher the peak that might trigger oscillations in the transient response. This behaviour might be emphasized by the input network realized by the large equivalent capacitance of the detector C_{eq} (hundreds of pF), by the series inductances and by the input resistance itself, thus leaving open issue on the fact that extremely low resistance is the best choice for timing.

The expression used to calculate the jitter of a front-end for SiPM has already been introduced (first chapter (1.5)). It depends on electronic noise and signal slope as:

$$\sigma_t = \frac{\sigma_n}{Slope} \tag{1}$$

where *Slope* is the magnitude of the time derivative of the signal waveform calculated at the time when the signal overcomes the comparator threshold.

Fig. 3-4 shows a plot which summarizes the jitter simulations with the effects of electronic noise included in the model, according to the expression (1).
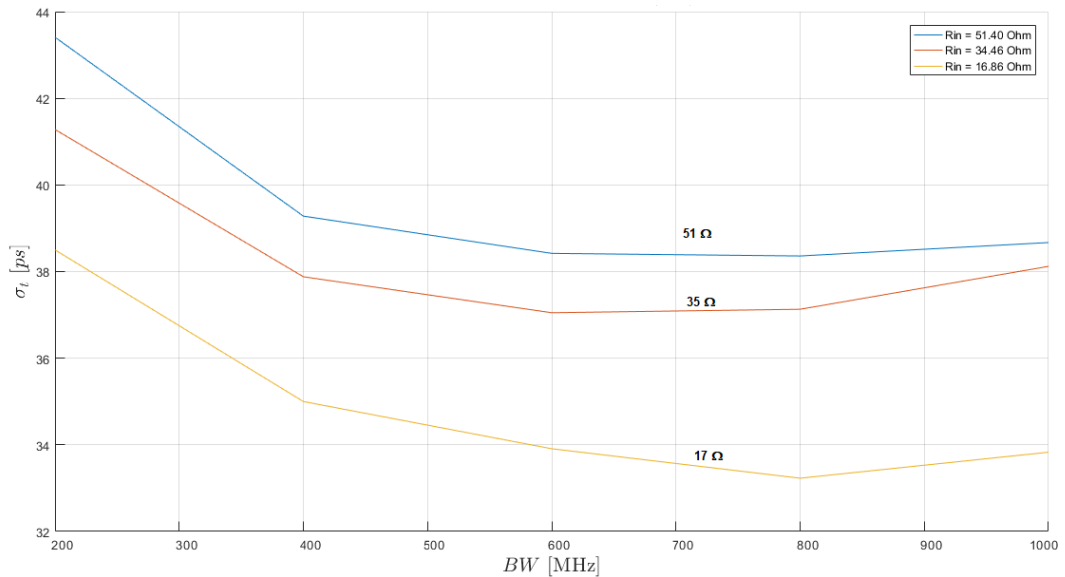

Fig. 3-4: Time jitter as a function of the bandwidth for three values of the input resistance Rin.

The plot describes the jitter as a function of the bandwidth for three different values of the input resistance with a parasitic inductance L_{par} = 10nH that mimics the interconnection wire between the detector and the front-end electronics. Smaller input resistances allow to increase timing accuracy better than larger, while decreasing L_{par}, lower absolute values for the jitter σ_t can be achieved. The bandwidth has been varied by sweeping the value of the load capacitance at the output of the circuit while keeping the load resistor equal to 50 Ω.

It can be concluded what follows:

1. Decreasing the input resistance, the timing performance of the circuit improves. The minimum value that considered here is around 17 Ω and lower values correspond to curves characterized by lower values of σ_t. The drawback is the increase of the power consumption for decreasing values of R_{in}.

2. An optimum value for the bandwidth exists, for each value of the input resistance, due to the increase of the effects of noise and to the limited

improvement of the slope caused by the parasitic inductance. On the other hand, this minimum is not abrupt and the constraint on the bandwidth can be relaxed.

3. The power consumption of the simple circuit adopted for this preliminary analysis is too high so far and it is supposed to increase trying to achieve the required levels of timing accuracy (not lower than 20 ps rms) with a differential structure.

3.5.1   HANDMADE RCG PARAMETER CALCULATIONS

The Regulated Common Gate (RCG) has much in common with Common Gate (CG) and additional benefits emerge.

Fig. 3-5 shows the schematic typology of the RCG and its small signal model for gain, input impedance and noise calculation.
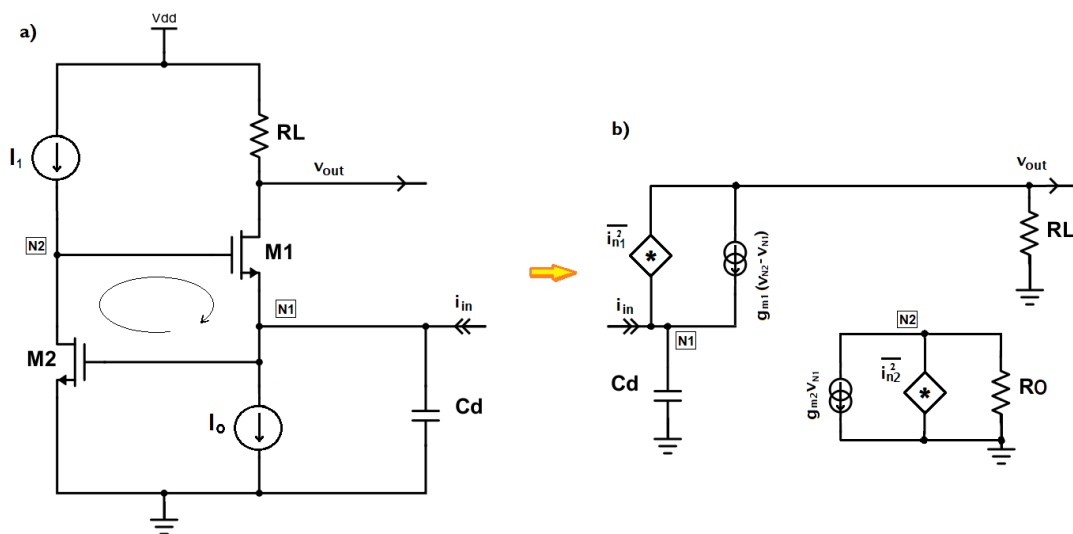

Fig. 3-5: a) The schematic of RCG amplifier and b) its small signal model.

Considering the limited current budget and the emphasis on timing for this project, the RCG is best candidate because it grants much more freedom degrees to optimize the noise and to achieve the desired input impedance, whereas noise contribution of the input transistor in a CG topology depends on its transconductance, which defines, at the same time, the input impedance of the preamplifier.

If the bias current sources I_0 and I_1 are ideal and the C_d is the equivalent detector capacitance, including the noise current sources in the calculation, the following set of equations can be written:

$$v_{N2} = -(g_{m2}v_{N1} + i_{n2}) \cdot R_o \qquad (2.a)$$

$$v_{N1} = (i_{in} + i_{n1} + g_{m1}(v_{N2} - v_{N1}) \cdot \frac{1}{sC_d} \qquad (2.b)$$

$$v_{out} = -[i_{n1} + g_{m1}(v_{N2} - v_{N1})] \cdot R_L \qquad (2.c)$$

where the load resistance R_L is supposed to dominate the output resistance of M1 in the parallel connection. Assuming $C_d = 0$, $i_{n1} = i_{n2} = 0$ and solving (2) for v_{N1}/i_{in}, the equivalent input impedance of the RCG amplifier is

$$R_{in} \cong \frac{1}{(g_{m2}R_o) \cdot g_{m1}} = \frac{1}{A_{v2}g_{m1}}, \qquad (3)$$

where $A_{v2} = g_{m2}R_o$ represents the open loop gain of the feedback loop that is equal to the voltage DC gain of Common Source (CS) transistor M_2. The input impedance of a simple CG TIA is $1/g_m$. This means that using RCG TIA can lower the input impedance by A_{v2} times that of the regular CG TIA. Signal gain calculation is performed with the assumption that the noise current generators associated to M1 and M2 are open ($i_{n1} = i_{n2} = 0$). Solving (1) for v_{out} and under the assumption that $A_{v2} \gg 1$, one obtains:

$$v_{out}(s) \cong i_{in}R_L \frac{1}{1 + s\tau_{in}} \qquad (4)$$

where $\tau_{in} = R_{in}C_d$ and R_L is the transimpedance gain. The signal transfer function of the RCG TIA and the signal transfer function of the CG TIA are formally identical but the dominant pole in the former is A_{v2} times that of the latter.

The spectral density of power thermal noise current is described by:

$$\overline{i_n^2}/\Delta f = 4KT\gamma n g_m \tag{5}$$

where g_m is the transconductance of the transistor, T is the absolute temperature, k is the Boltzmann constant, n is the slope factor and γ is a bias dependent parameter taking values from 1/2 to 2/3 for an ideal transistor operating from weak to strong inversion.

It can be easily demonstrated that, at first approximation, both the active devices and the load resistor R_L contribute to the output noise. Assuming $i_{in} = i_{n2} = 0$ and solving (1) for v_{no1} one obtains:

$$v_{no1}(s) = i_{n1}R_L\frac{s\tau_{in}}{1+\ s\tau_{in}} \tag{6}$$

After transferring (6) in the frequency domain, calculating the module, squaring and integrating over the whole frequency range, the expression of the output voltage noise will be a function of M1 current noise times the load resistor. Therefore, M1 contributes to the output noise with its g_{m1} and can be referred to the input as a parallel noise source.

On the contrary, assuming $i_{in} = i_{n1} = 0$ and solving (2) for v_{no2} one obtains:

$$v_{no2}(s) = \frac{i_{n2}}{g_{m2}}R_LC_d\frac{s}{1+\ s\tau_{in}} \tag{7}$$

that shows how output voltage noise decreases when g_{m2} is increased and that M2 contributes as a series noise source at the input.

Noise optimization can be achieved by proper balancing of the parallel noise source contribution from M1 and the series noise contribution from the common

source transistor M2, whereas the input impedance still depends on the loop gain that can be tuned by proper choice of common source voltage gain.

Whatever its value, the transimpedance resistor also generates noise which ultimately depends on the parasitic capacitance on the output node; together they also set the second pole of the amplifier and their values need to be carefully selected. Furthermore, the interconnect parasitic inductance would eventually affect noise performance. In fact, it would isolate the large detector capacitance from input node N1, thus the parasitic capacitance on that node would now shunt to ground parallel noise at higher frequencies and the inductance would also deactivate the series noise at those frequencies. The discussion of low frequency noise is neglected at this stage in the analysis because it would eventually affect the baseline and a decision has been made to deploy a dedicated block for its stabilization that will be afforded in the next sections.

BASIC64 operates at 3.3 V supply voltage. It draws 3 mA DC current with a total power dissipation of 10 mW. Assuming this number as a benchmark for the new channel under development and considering that its differential structure is operated with a 1.2 V supply voltage, the biasing current for each of the two legs shall not exceed 4 mA. To minimize the parallel noise contribution, transistor M1 should be operated at relatively low bias current, whereas the transconductance (and the current) of the common source transistor should be as high as possible to reduce the series noise.

Just for one leg the first try is distributing a total amount of 3 mA of current as a five to one ratio between the current for the CS transistor and the current for the CG transistor.

It's a common practice to deal with input referred noise and Equivalent Noise Charge (ENC) for noise characterization of SiPM readout electronics; nonetheless, the choice of referring the contributions of all noise sources to the output for calculation is beneficial in simulations to extract jitter. In fact, the output rms σ_n noise is evaluated by means of AC noise simulations on the frequency band of

interest and is divided by signal slope to get jitter. The slope is obtained running transient simulations and picking off the maximum slope of the output transient response. Another method to extract jitter in simulation is to run a noise transient simulation in one go, and build up a histogram with the random delay samples registered between the time of SiPM firing and the time the output voltage signal crosses an optimum virtual threshold. The latter is set runtime by an iterative Ocean script. Both methods are viable and get to the same result.

The capacitive load of the voltage discriminator cascaded to the front-end is included in the model for jitter extraction; nonetheless, the discriminator stage undergoes a separate design flow and will be designed to feature extremely low jitter that doesn't impair the results coming from the preceding analog blocks. The RCG TIA configuration represents a viable solution for the addressed specifications, thus a differential architecture based on this kind of circuit will be further investigated.

3.6 ARCHITECTURE OF THE ANALOG CHANNEL

The architecture of the channel readout is shown in Fig. 3-6. The preamplifier is a differential RCG current buffer that receives the current signal from the SiPM and provides an output pulse that feeds two signal paths:

- Fast path for time measurement
- Slow path for Charge/Energy measurement.

The fast path uses a fast voltage discriminator with a fine programmable threshold to manage extremely fast and low magnitude signal, while the slow path uses an active R-C integrator followed by a peak detector and a slower comparator that can discard dark pulses depending on the amount of integrated charge. A dedicated control logic generates the digital signals that handle the TDC for timestamp and the ADC for the digital conversion of the output signal of the peak detector, which carries charge information.
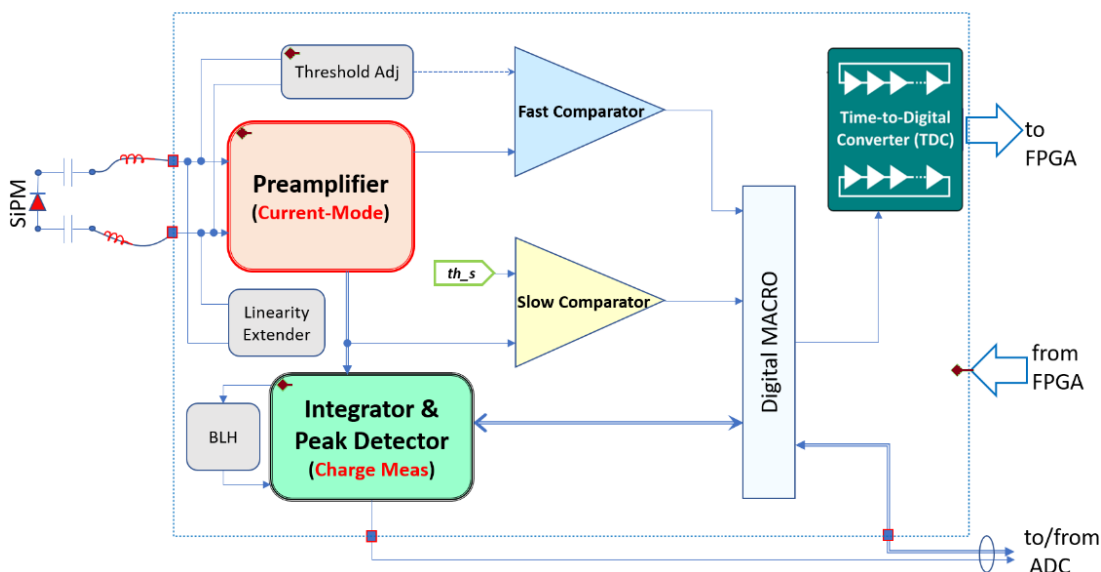
Fig. 3-6: System diagram.

A common biasing block has been designed and is used to provide biasing currents and to define voltage references throughout the channel blocks. The implementation of these blocks will be described in the following sections.

3.6.1 THE RCG TIA PREAMPLIFIER

By adding a booster amplifier to a CG current buffer, a new arrangement is obtained known as a regulated common gate (RCG) stage normally used as a TIA.

Fig. 3-7 shows the modified version of the circuit, with the ideal current source I_b and the resistor R_{CS} that loads the boosting transistor M2 in CS configuration.
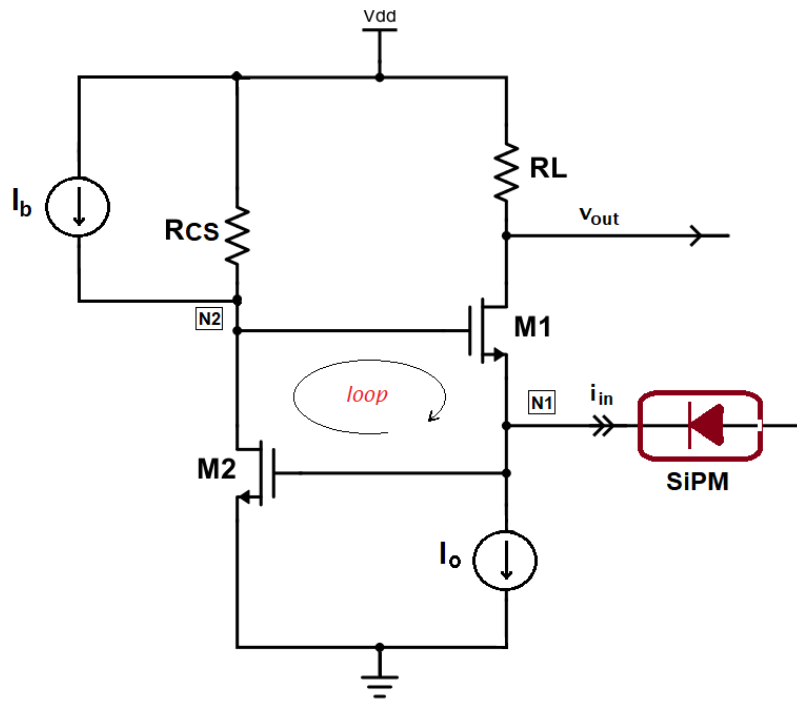

Fig. 3-7: RGC front-end for SiPM (half circuit of the differential structure): new version.

The CS amplifier (booster) senses any voltage change at the input terminal (the source of the CG transistor M1) and controls the gate voltage of the CG transistor to counteract it; as a result, the transconductance increases by the loop gain and the input resistance is scaled down resulting to be lower than the input resistance of a simple CG stage by the same amount of the loop gain itself.

The transistor-level implementation of I_b comprehends a circuital solution to perform a temperature and process compensation for the variations of the load resistor and prevents the DC voltage on the node N2 from shifting.

As discussed in the previous section, the first design choice is to fix I_o = 0.5 mA and calculate the transconductance of M1 with the square-law equation describing the MOSFET behaviour in strong inversion.

The drain current for the n-MOS transistor in saturation mode under strong inversion conditions is modelled approximately as:

$$I_D \cong \frac{\mu_n C_{ox}}{2} \frac{W_n}{L} (V_{GS} - V_{th})^2 \qquad (8)$$

where μ_n is the charge-carrier effective mobility, W_n is the gate width, L is the gate length and C_{ox} is the gate oxide capacitance per unit area. The channel length modulation parameter, λ, is neglected, preferring simplicity to accuracy.

While the channel length scales down with technology, Eq. (8) becomes inaccurate and this also applies to the TSMC 130 nm CMOS process we are using.

The transconductance is

$$g_m \cong \frac{2I_D}{V_{ov}}, \qquad (9)$$

where V_{ov} is the overdrive voltage. Choosing V_{ov} = 150 mV and $I_D=I_o$=0.5 mA, the transconductance of transistor M1 is equal to 6.7 mS and the transistor aspect ratio W_n/L can be calculated as:

$$\frac{W_n}{L} = \frac{g_m}{\mu_n C_{ox} \cdot (V_{GS} - V_{th})} = \frac{g_m{}^2}{2I_D \mu_n C_{ox}} \qquad (10)$$

Designing with non-minimum size and assuming L = 160 nm, $\mu_n C_{ox}$ = 280·10^{-6} A/V^2, I_D = 0.5 mA and g_m = 6.7 mS, from equation (10) the width of the CG transistor, W_n, is found to be equal to 26 μm (around 160 times the channel length) and rounded up to 30 μm for simulation to further extend the overdrive voltage for saturation and permit larger swing of the drain voltage of the transistor.

With a biasing current equal to 2.5 mA, the transconductance g_{m2} of the CS transistor M2 is five times the transconductance g_{m1} of CG transistor M1. Assuming L = 200 nm as channel length of the CS transistor, its width is found to be as large as 160 μm (around 800 times the channel length) with g_{m2} = 33.5 mS.

It was observed by simulation that replacing the ideal current source that sets the biasing current of CS transistor with a transistor active load may lead to poor control of the loop gain of the feedback; this is because of the channel length modulation that reduces the output resistance of active devices in deep sub-micron technology nodes. Furthermore, the transient response of the circuit tends to become unstable and ringing may occur. Therefore, it was decided to introduce the resistor RCS (Fig. 6) to set the voltage gain of the CS stage, relying on the good performance of resistors in polysilicon layer available in the design library.

The resistance needed to achieve Rin = 10 Ω can be calculated as:

$$R_{CS} \cong \frac{1}{g_{m2} g_{m1} R_{in}} = \frac{10^6}{10 \cdot 27 \cdot 5.9} \cong 630 \ \Omega \qquad (11)$$

where the values of transconductance are retrieved from simulation and are 10% lower than the values obtained by hands calculation.

That is a too large resistor and it cannot be used for the intended purpose as it is. In fact, if a current equal to 2.5 mA flows through a resistance equal to 630 Ω it will induce a voltage drop across its terminal in excess of the supply voltage; on the other hand, lowering the resistor carries around the decrease of the loop gain and, as a consequence, the input resistance increases.

To overcome such an issue, it was decided to target a larger input resistance, in the range from 12 to 15 Ω, that also provides better matching over an extended frequency range with the constraint of almost no additional DC current burden; at the same time RCS should be reduced to one third of the value calculated in (11), by acting on the transconductance of both transistors. Some topology changes were necessary resulting in the final schematic Fig. 3-7. A DC current is injected in the drain of CS transistor to reduce the voltage drop and keep the transistor far

from operating in triode mode. Through several iterations made to try to achieve the required input resistance without success, the final chance comes from pushing the transistors to operate in sub-threshold by resizing them at expense of silicon area. Finally, the output load resistor R_L equal to 1 kΩ should allow for quite large signal swing at the drain of the CG transistor while the output DC voltage is half of the supply voltage thus leaving the biasing current mirror operate in strong inversion with enough headroom.

3.6.1.1 FINAL TRANSISTOR SIZING

Noise, bandwidth and DC gain are conflicting requirements in the realization of amplifiers. As already pointed out, use of the square-law design equations permits to find a rough estimate of parameters though proving unsuitable to fully validate deep sub-micron CMOS design, whose physical parameters (μC_{ox}, V_{t0}, V_{dsat}, etc.) are badly defined for short channel transistors. This means that achieving optimal performances by properly sizing the active devices of the circuit is extremely difficult unless one resorts to specific design methodologies, e. g. g_m/I_D, that fall outside the scope of this work.

Assuming the increase of transconductance is the goal to pursue under the tight conditions exposed in the previous section, the size of the transistors has been progressively increased pushing them to operate at the edge of weak inversion region, where a good compromise between speed, noise and power can be found. Starting from the established circuit model with the transistor in strong inversion already devised, the width of transistor has been enlarged moving towards the moderate inversion, where the transconductance for a given current amount is improved. The jitter of the circuit has been calculated at each iteration and the resizing process has been stopped upon reaching the low jitter limit of 20 ps, where the transconductance hits a maximum and is no longer influenced by the area of the devices but the parasitic capacitance starts to plague bandwidth.

The conservative value for jitter has been addressed to circumvent the impairment of jitter performance caused by layout parasitic components, which results in

degradation of signal speed and increase of noise. Finally, overall noise minimization has been preferred to speed increase because a further increase in speed would have required a progressive incursion into strong inversion thus resulting in progressive degradations of the gain and transconductance. Noise has been minimized by unbalancing the total biasing current in favour of CS transistor whose transconductance has undergone a sensible further increase, whereas the shot noise of the CG transistor has slightly decreased. The final compromise between speed, gain and noise has been reached in moderate inversion. In fact, high g_m efficiency of moderate inversion is the best choice for circuits that has not been demand the highest speed but cannot afford neither the low speed of weak inversion nor the power inefficiency of strong inversion.

The schematic core structure of the differential input RCG TIA depicted in Fig. 3-8 has been captured using the models of the devices from the PDK libraries. Several simulation tests have been carried out in different operating conditions of process, supply voltage and temperature to validate the performance in terms of jitter.
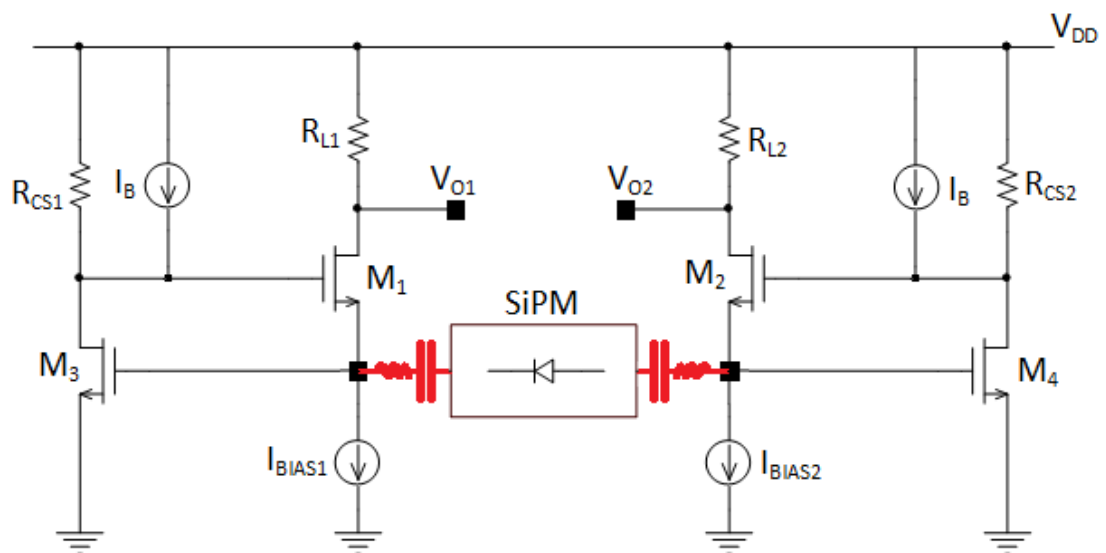
Fig. 3-8: Differential structure of the input stage.

The ideal current sources have been replaced with current mirrors that are not shown in the Fig. 3-8. The transistors of the mirrors have been sized to give low transconductances, thus minimizing their contribution to the integrated noise.

The final sizes of the transistors are summarized in the Table 3-3.

Table 3-3: Drawn dimensions of the n-type MOSFETs of the circuit in Fig. 3-8.

Parameter	M1, M2 1 out of 20 in parallel	M3, M4 1 out of 20 in parallel
W/L	30	276
W	4.8 μm	55.2 μm
L	0.16 μm	0.2 μm

Each of the transistor M1, M2, M3, M4 is composed of as many as 20 elementary transistors with the sizes reported in Table 3-3.

The following Table 3-4 reports the values of the load resistances and of the biasing currents of the transistors along with their transconductances.

Table 3-4: Nominal DC current and transconductances for the n-type MOSFETs of the differential circuit in Fig. 3-8, simulated @ room temperature of 25 °C.

Parameter	M1, M2	M3, M4
I_{DC}	0.4 mA	2.8 mA
g_m	7.8 mS	54.8 mS

The resistors $R_{CS1,2}$ are equal to 167.5 Ω each and the resistors $R_{L1,2}$ are equal to 735 Ω each (see Fig. 3-8).

The loop gain of the circuit about 10.

The value of the differential input impedance is shown in Fig. 3-9. It remains flat and equal to about 30 Ω up to 270 MHz, when it starts to rise till reaching a maximum peak at 1.1 GHz.
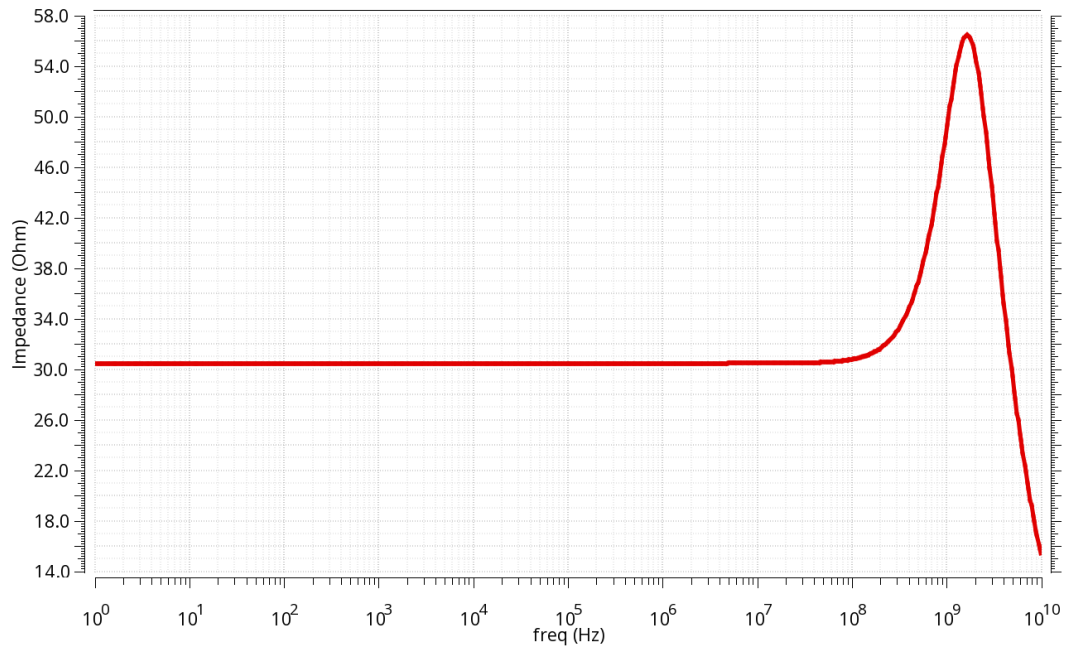
Fig. 3-9: Magnitude of the differential input impedance of the RCG front-end: new version.

3.6.1.2 SIMULATION RESULTS: FROM SiPM TO TIMING

Simulations of the comparator coupled to the front-end have been carried out, to assess the timing performance which can be achieved with the designed processing chain. As input stimulus, the familiar model of the 3600 micro-cell SiPM already employed in the previous simulations has been used, along with a total parasitic inductance of 10nH due to the interconnections between the detector and the front-end. The comparator output is loaded by an equivalent capacitance of 300 fF, which is a conservative estimate of metal tracks and an equivalent capacitance accounting for the TDC and digital circuitry. Aside from the input capacitance of the comparator, the output of the preamplifier is also burdened with the charge reconstruction circuit whose equivalent capacitive load is also included for simulation purpose.

A summary of the results of the transient noise analysis carried out on the overall circuit will be reported hereafter.

To fully characterize the response of the preamplifier to the single photon in the time domain, the circuit has been AC coupled to the SiPM.

A parasitic inductance of 10 nH in series with the detector has been introduced in the simulation model to mimic the effect of interconnections. For the purposes of simulation accuracy and reliability the SiPM electrical model which appears to be appropriate has been reported in Fig. 3-10 (see Chapter 2 of the thesis).
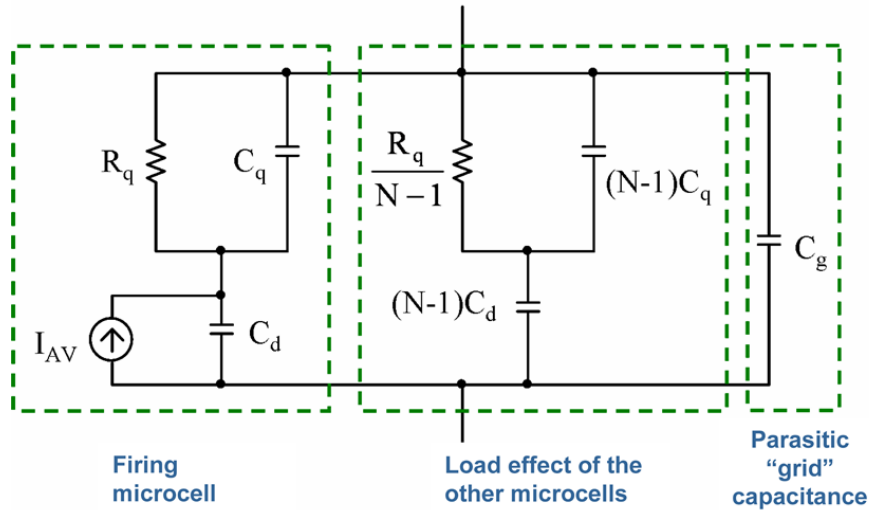
Fig. 3-10: Electrical model of a SiPM.

The electrical parameters for the SiPM Hamamatsu S10931-050P are reported in Table 3-5.

Table 3-5: **Results of the parameter extraction procedure applied to the SiPM**

Model parameter	SiPM Hamamatsu, N=3600
R_q	182.75 kW
C_q	17.7 fF
C_d	75.2 fF
C_g	36.85 pF

It is a 3 x 3 mm^2 large area device with 3600 micro-cells and is suitable to meet the specification of the SiPM types demanded in the preliminary project requirements.

Fig. 3-11 shows the output pulse of the readout chain composed of the SiPM coupled with the front-end whose output terminals are loaded with two

capacitances of 200 fF each that represent the equivalent capacitances of the differential input terminals of the comparator (not connected).
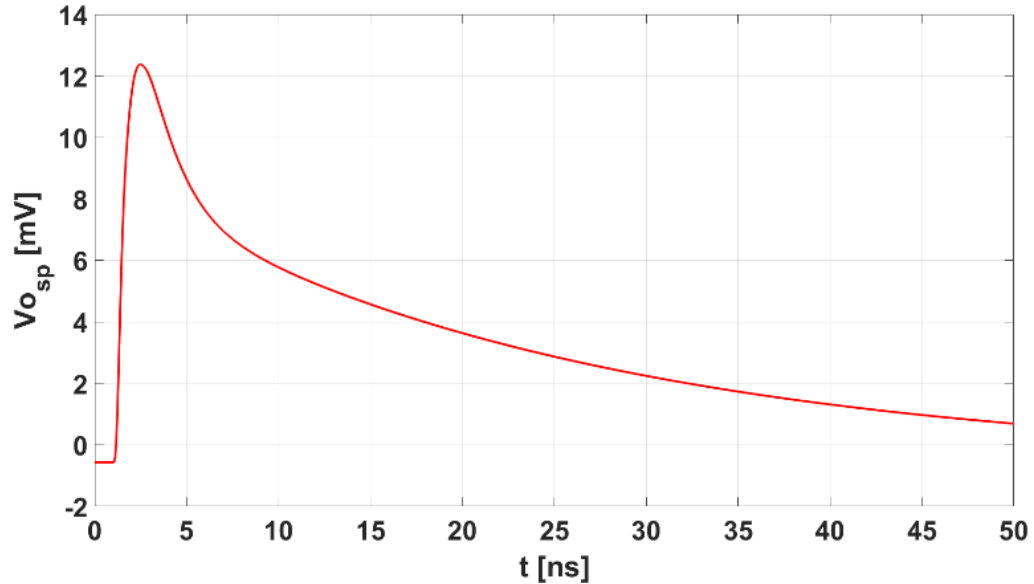
Fig. 3-11: Simulated response of the whole front-end to a single photon.

The peak amplitude of the output voltage signal is equal to 13 mV that corresponds to a transimpedance gain of 64 dBΩ, considering a peak input current of 10 μA. The rise time is 700 ps.

As apparent from the plot in Fig. 3-11, the baseline at the onset of the voltage pulse is shifted with respect to zero-volt level. This is an intentional offset introduced by setting the comparator threshold as low as 500 mV. A detailed description of the comparator and of the triggering mechanism adopted will be exposed later in a dedicated section.

Fig. 3-12 shows the output pulse generated by the front-end as a response to the single SiPM micro-cell hit, i.e. the input of the fast comparator with no threshold applied other than mismatch compensation.
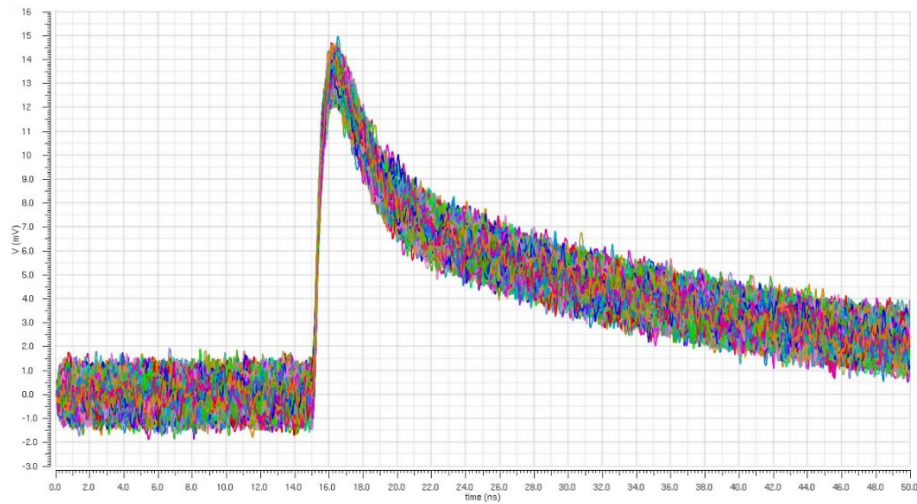
Fig. 3-12: Output of the RCG TIA with the full load (single photon, transient noise analysis).

The previous plot represents the result of a transient noise analysis for 100 trials in a typical case for process, supply voltage and temperature. The average rise time of the noisy pulse response is about 0.8 ns, and its amplitude is slightly less than 13 mV.

Fig. 3-13 shows the response of the comparator with a threshold of 3.2 mV and the voltage pulse generated in response to the photon that hits the detector. For simulation purpose, the hit time (reference for delay calculation) coincides with the onset of the current stimulus (not shown) in the circuit model of the SiPM.
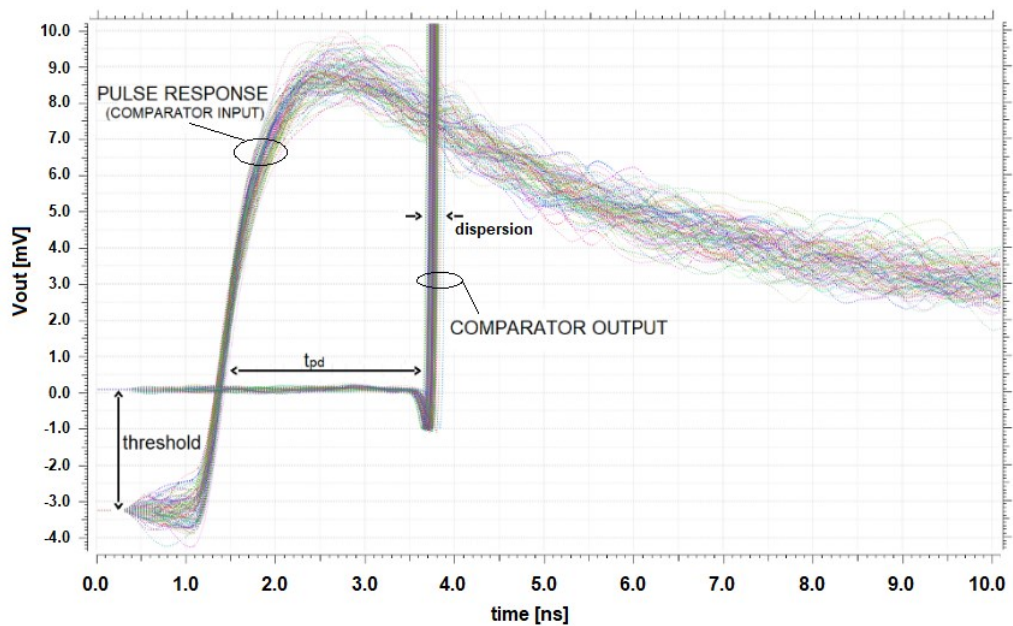
Fig. 3-13: Output of the comparator (single photon, transient noise analysis).

The comparator has a propagation delay, t_{pd}, of 2.3 ns but a tight dispersion: the former depends on voltage overdrive and determines time-walk, instead the latter accounts for jitter.
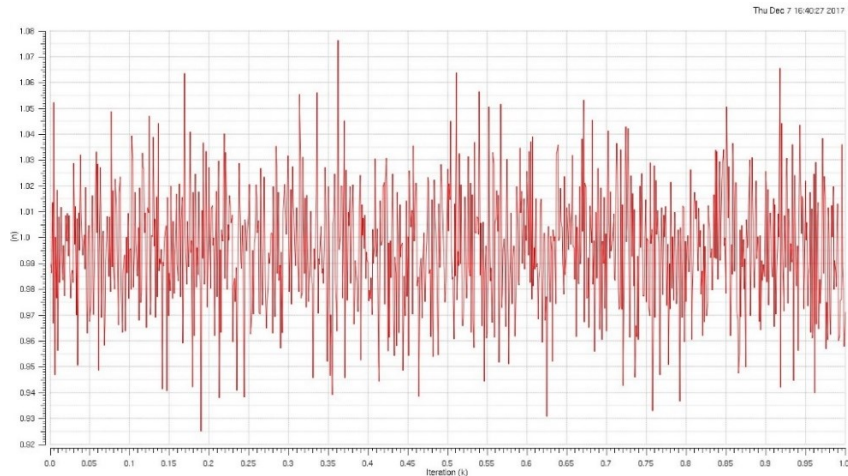
Fig. 3-14: Time jitter of the comparator response as a function of the transient noise iterations.

In Fig. 3-14 the delay between the avalanche trigger time (hit time) and the response of the comparator as a function of the iteration of the transient noise analysis is reported. The figure shows the time jitter of the leading edge of the comparator response, which is the time jitter of the system, for a total 1,000 simulation trials.

The resulting time jitter is about 22 ps *rms*.

Time jitter can also be calculated as the ratio between the *rms* output noise and the slope of the response measured at the threshold crossing point. The *rms* noise at the output of the differential front-end is the intrinsic noise of the preamplifier resulting from AC simulation while the slope of its response to the single-photon and is taken as the time derivative of the signal waveform that comes from a noiseless transient simulation.
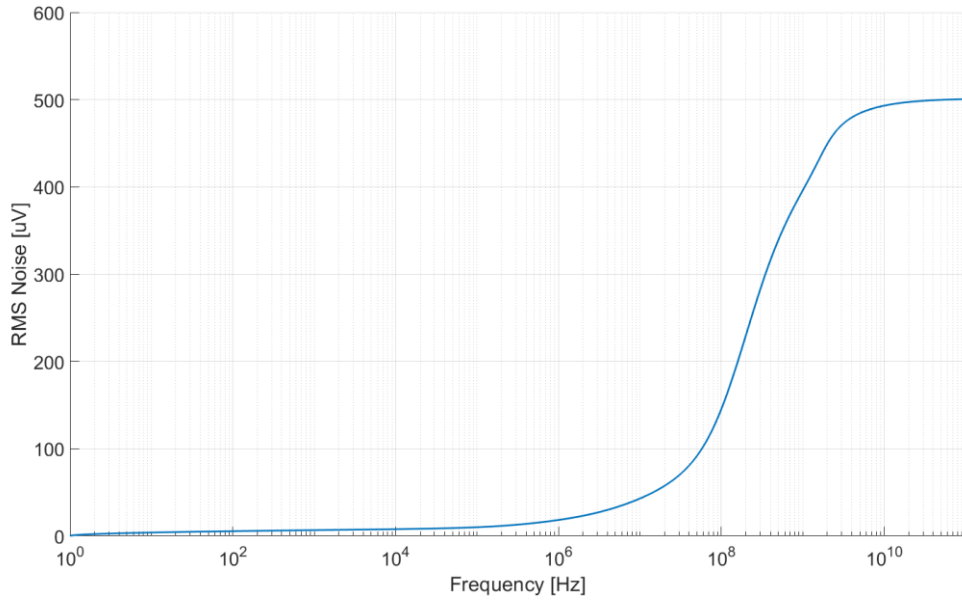
Fig. 3-15: Simulation of the rms output noise of the differential front-end.

In Fig. 3-15 the integrated noise over frequency is shown. The curve saturates at around 2 GHz where the noise level is around 510 μV_{rms}.

Fig. 3-16 reports two curves representing the voltage waveform at the comparator input, e.g. preamplifier output, and its derivative as a function of time.
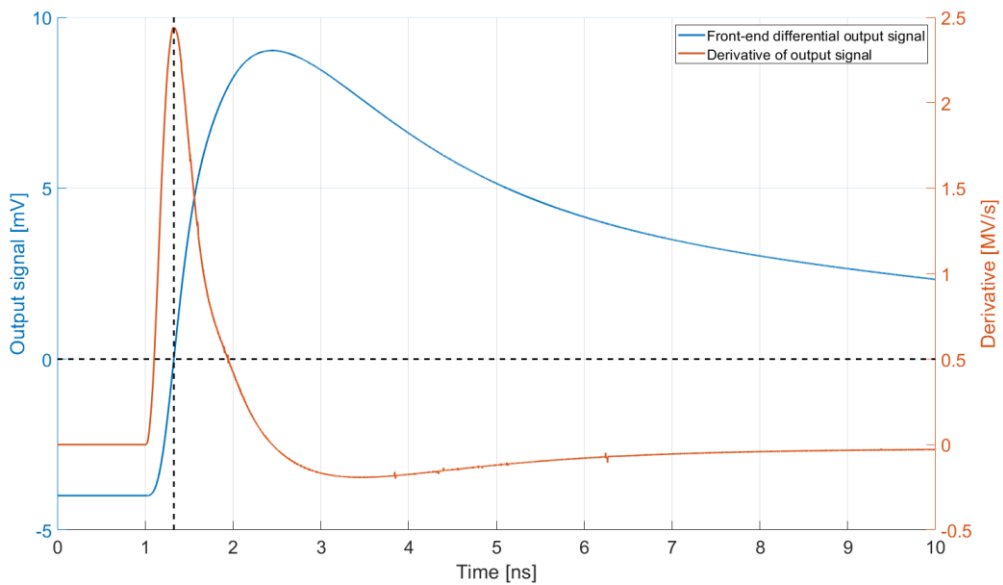

Fig. 3-16: Response of the differential front-end to the single photon.

The maximum slope is around 24.5 V/µs, calculated at that time when the signal crosses the threshold. The resulting time jitter is 20.8 ps. As expected, this value fits quite well to the previously reported value, when the jitter was calculated as delay time dispersion using noise transient simulation, though slightly better.

3.6.2 THE VOLTAGE DISCRIMINATOR

The comparator has been made be fast enough to preserve the timing accuracy performance of the front-end. The block structure of the comparator that has been designed is depicted in Fig. 3-17.
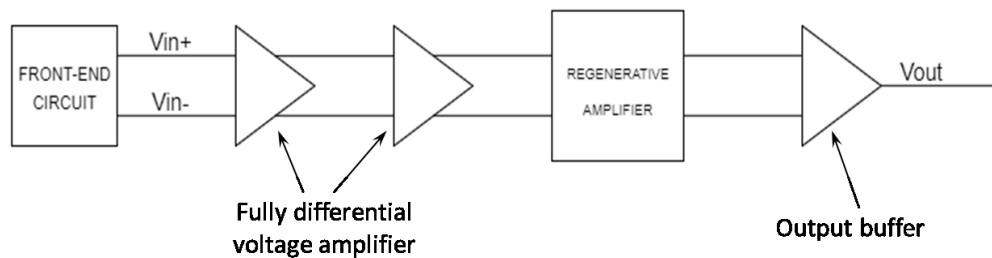
Fig. 3-17: Block diagram of the comparator.

It was planned to cascade a proper number of large-bandwidth fully differential amplifiers with passive loads to increase the amplitude of the signal, along with a decision-making circuit with some amount of positive feedback, i.e. a regenerative amplifier; finally, an output buffer is there to convert the differential signal to single-ended, and generate the correct logic levels and drive the capacitive load of the circuit. The cascaded gain stages are simple large bandwidth differential amplifiers with passive loads, as depicted in Fig. 3-18.
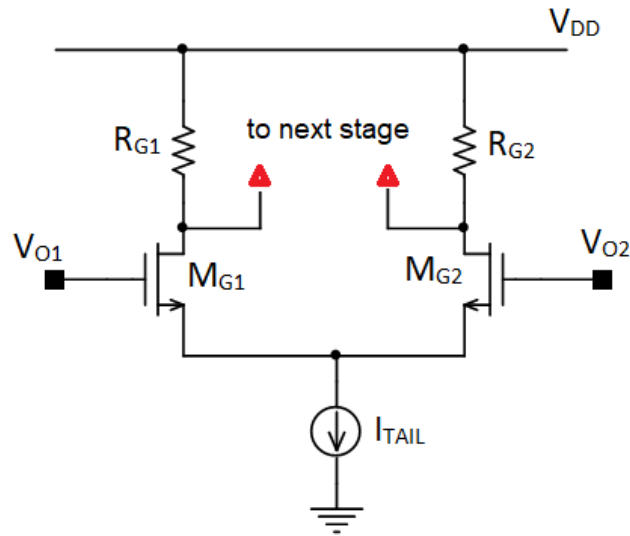
Fig. 3-18: One of the differential gain stages of the comparator.

A single gain stage is biased with a tail current of 0.25 mA, with a differential gain of 4 V/V, obtained with transistors operated in strong inversion (g_m = 1.25 ms) and a half-side load resistance R_G of 1.8 kΩ.

The peak amplitude of the differential voltage signal produced by a single micro-cell of the SiPM undergoing avalanche breakdown is about 210mV at the output of the two cascaded gain stages.

The specialized next stage of the chain, i.e. the decision circuit, has the structure reported in Fig. 3-19.
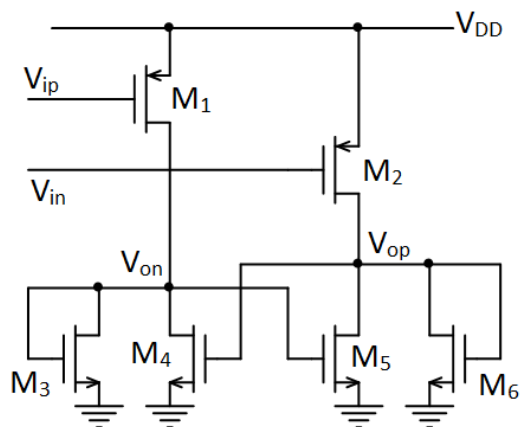
Fig. 3-19: Regenerative stage of the comparator.

It is a regenerative stage with some amount of positive feedback, used to produce the fast switching of the output differential voltage as soon as the signal corresponding to a single photon is detected.

The structure of the comparator is completed by an output stage, which performs the differential to single-ended conversion and drives the load of the comparator, represented by the input capacitance of the TDC and of digital circuitry. The schematic of the output driver of the comparator is illustrated in Fig. 3-20.
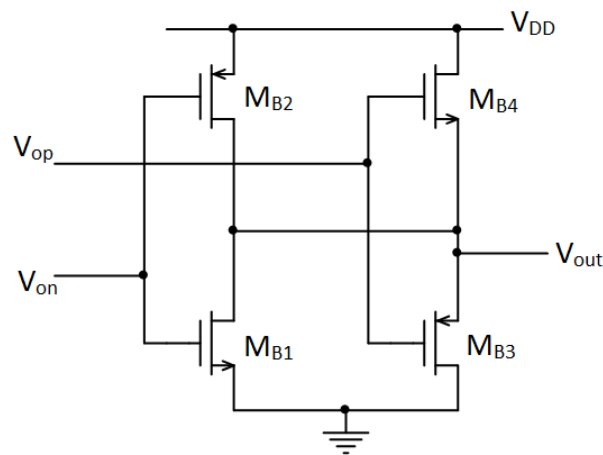

Fig. 3-20 : Output buffer of the comparator.

All the blocks which compose the structure of the comparator have been properly sized and the total power consumption of the circuit is about 0.8 mW. The comparator introduces a jitter of 4 ps on time measurements thus negligibly affecting the jitter performance of the overall system.

Implementation of the threshold

The threshold used for time pickoff on the fast path can be set by introducing an imbalance between the two output sides of the differential RCG-TIA that also offsets the differential input of the comparator by the desired amount. The schematic of the Threshold Adjustment System (TAS), is shown in Fig. 3-21.
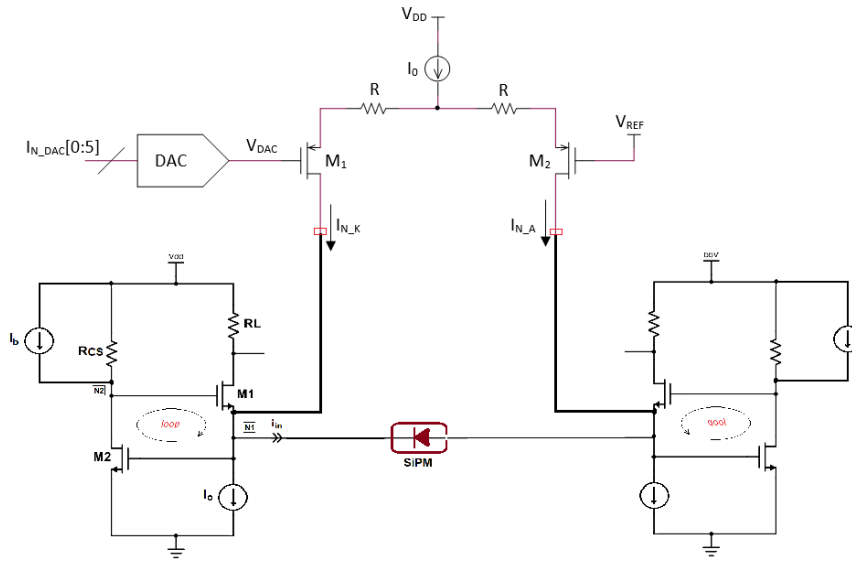
Fig. 3-21 : Differential pair for the adjustment of the threshold of the fast comparator.

A differential pair is connected to the input terminals of the preamplifier as in the Fig. 3-21. With the differential pair at the balance point (no differential voltage applied) its DC biasing current is ideally split in two equal DC currents that flow into the input nodes of the preamplifier. As soon as a differential voltage is applied across the linearized PMOS transistors of the differential pair, the current is steered and the differential output voltage of the preamplifier gets unbalanced. This introduces an offset at the comparator input that can be easily controlled and represents the timing threshold. The offset can be controlled using a programmable embedded Digital-to-Analog Converter (DAC) that allows for very fine step size.

The threshold of the fast discriminator can be adjusted by changing the configuration settings of the embedded 6-bit DAC according to the following description:

- 3 bits for the configuration of the DAC gain set three modes: ultra-fine, fine, normal
  - Ultra-fine from 0 to 3 photoelectrons (pe$^-$) with a step size of 1/15 pe$^-$
  - Fine from 0 to 7 pe$^-$ with a step size of 1/8 pe$^-$
  - Normal from 0 to 10 pe$^-$ with a step size of 1/5 of pe$^-$
- 6 bits for DAC code set the number of steps of each gain set

Fig. 3-22 illustrates the three major settings and the dependence of the threshold on the 6-bit DAC code.
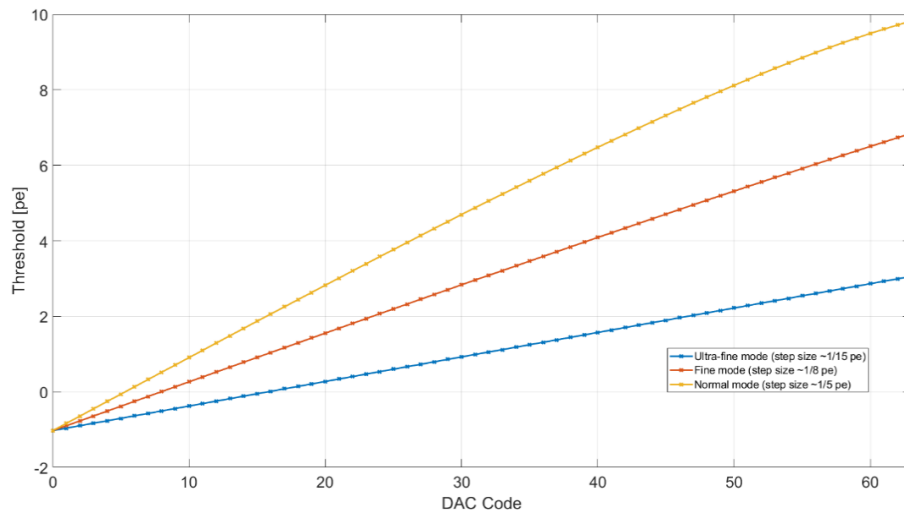

Fig. 3-22: Threshold of the fast comparator as a function of the DAC code.

Note that it is possible to realize also slightly negative values of the imbalance between the two branches of the input differential front-end, i.e. negative values for the threshold. This feature makes possible the compensation of the input stage offset caused by mismatch.

3.6.3 THE CIRCUIT FOR CHARGE MEASUREMENT

An accurate charge measurement can be done if the slow path of the front-end is able to reconstruct with adequate precision the same waveform that represents the total current generated by the SiPM. For this purpose, an *ad hoc* current reconstruction circuit is used. Fig. 3-23 shows schematically the circuit solution that has been adopted: the output voltage V_{O1} on the left side of the preamplifier is 'copied' by means of a feedback loop and is applied on a replica of the parallel R_{L1}-M_{L1}, thus allowing the replica of the current generated by the SiPM to flow through the dummy load.
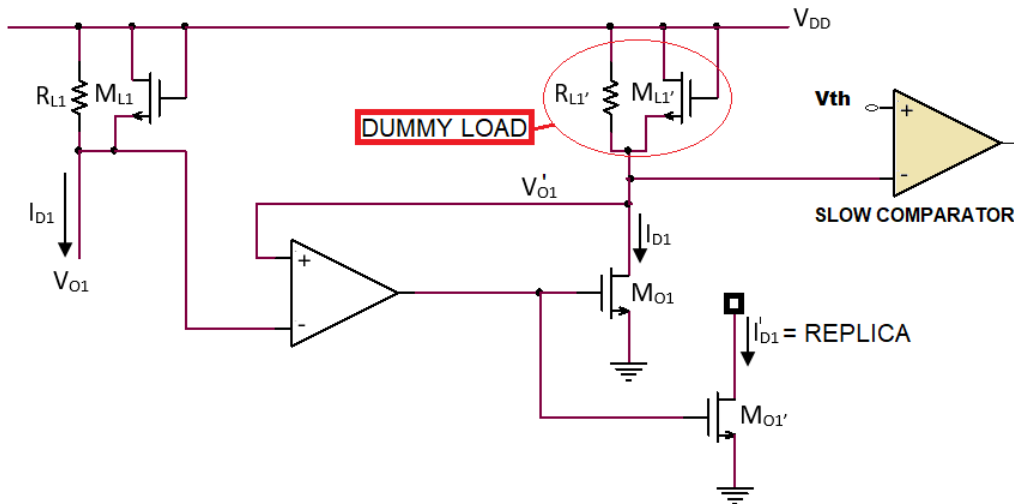
Fig. 3-23: Current reconstruction circuit and the replica current of the SiPM.

A "slow" comparator is placed on the replica of the output node of the RCG-TIA amplifier and implements a filtering strategy against dark pulses. In fact, to allow for single photon detection, the threshold of the fast comparator must go down to a very low level. Thus, a mechanism for an early discarding of the triggers caused by dark pulses has also been considered, useful in applications where single or very few photon events are not interesting, but the threshold should be very low for timing accuracy. In practice, if the signal that has caused a trigger does not overcome a further, higher threshold (the "high threshold", whereas the "low threshold" is the timing comparator threshold) within a short time window, the TDC will stop to process the trigger and is reset. The slow comparator is a standard differential pair with regeneration. It is less demanding in terms of performance than the fast comparator, and has been designed to feature a higher threshold, and to produce a trigger validation signal that will be used by the internal logic of the analog channel, according to the automatic procedure that is explained later.

The replica current signal is now ready to be further processed to extract the charge information it carries. This is accomplished using the charge measurement circuit whose complete structure, with all its basic blocks, is schematically represented in Fig. 3-24.
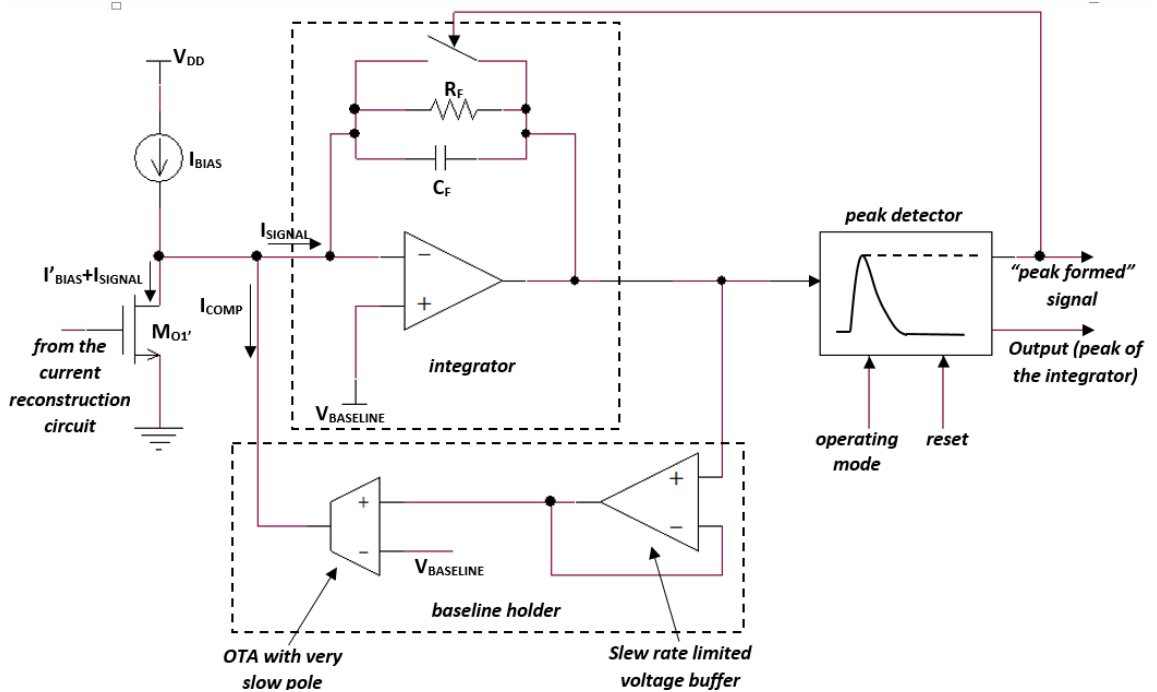
Fig. 3-24: Schematic structure of the charge measurement circuit.

The reconstructed current flows through the feedback network represented by the capacitance C_F in parallel to the resistor R_F thanks to the core amplifier; in fact, it sources the signal current, maintaining a virtual short circuit at its input terminals. Almost all the current is integrated onto C_F, since the time constant $R_F C_F$ is slow enough compared to the duration of the current pulse. Consequently, the output voltage of the integrator increases on top of the integrator output DC voltage, which, with no signal applied, is kept equal to the DC voltage $V_{BASELINE}$ for the joint effect of both the feedback resistance R_F and the baseline holder circuit. The nominal value of $V_{BASELINE}$ has been set to 300mV. The integration capacitance can be also quickly reset and the circuit can be rapidly brought back to its operating point by means of a switch, that is closed whenever a valid event has been detected as soon as the cascaded peak detector has stored the peak of the signal.

To accommodate for the large input dynamic range required by the target specifications (see section 3.4), different scaling factors of the replica current at the input of the integrator can be introduced current, by choosing different mirroring factors in the output branch of the current reconstruction circuit.

The RC-active current integrator has been made configurable, to allow for four possible gain values as defined in the following Table 3-6.

Table 3-6: Gain configurations of the front-end.

Configuration	Gain [mV/pe⁻]	Gain [μV/fC]	Dynamic range [expressed in no. of pe⁻]	Maximum charge [pC]
1	4	25	$1 \div 150$	24
2	0.31	1.9	$15 \div 2000$	320
3	0.15	0.95	$30 \div 4000$	640
4	0.08	0.49	$60 \div 8000$	1280

The gain specifications which have been summarized in Table 3-6 above call for some explanations.

The dynamic range of the circuit is limited by two different causes: a) the maximum current signal height which can be managed by the differential front-end; b) the maximum charge signal which can be managed by the integrator.

The values contained in Table 3-6 are referred to the assumption that one photon (i.e. one photo-electron pe⁻, or one firing microcell of the SiPM) corresponds to 160 fC (i.e. the gain of the SiPM is about $1 \cdot 10^6$).

In case all the photons hit the detector at the same time, the peak current can be very high, since the currents of the firing microcells (about 10μA for one microcell) add up and the limitation a) is dominant and can lead to undesired behavior of the front-end, despite the existence of the linear extender circuit. Otherwise, when the light comes from a scintillator, the photons do not arrive all at the same time, but according to the decay time constant of the scintillator, thus the dominant limitation is b). The first configuration enlisted in Table 3-6 has been introduced to make possible the detection of single photons: up to 150 concurrent photons can be handled. The dynamic range for configurations 2, 3 and 4 have been evaluated for signals obtained with the use of a scintillator characterized by a time constant of about 40ns (for instance LYSO).

Corresponding to the lower limit of each range, the signal to noise ratio of the differential front-end is about 20dB.

Two different values of the integration time constant R_FC_F have been also implemented to comply with the approach used to decide the possible values of the gain. In case of configuration 1, since it is oriented to photon counting applications, a smaller integration time constant, about 250ns, can be set, whereas for the other ranges this value has been increased to 350ns because the waveform of the current pulse produced by a scintillation event results from convolution between the photon emission law of the scintillators (with characteristic time constant used in PET applications (e.g. LYSO) around 40 ns) and the pulse waveform corresponding to a single microcell of the SiPM (at least 20ns), which produces a relatively long tail.

To fine-tune the time constant against process fluctuations both resistance and capacitance of the feedback network can be trimmed, by means of MOS switches driven by configuration bits that are not shown in Fig. 3-24. A good linearity of the charge-to-voltage characteristics of the whole analog chain is obtained, as the following two plots in Fig. 3-25 and Fig. 3-26 show. In all the figures, the peak detector output and the maximum deviation from the linear fits are reported as a function of the injected charge for the two extreme values of the gain.
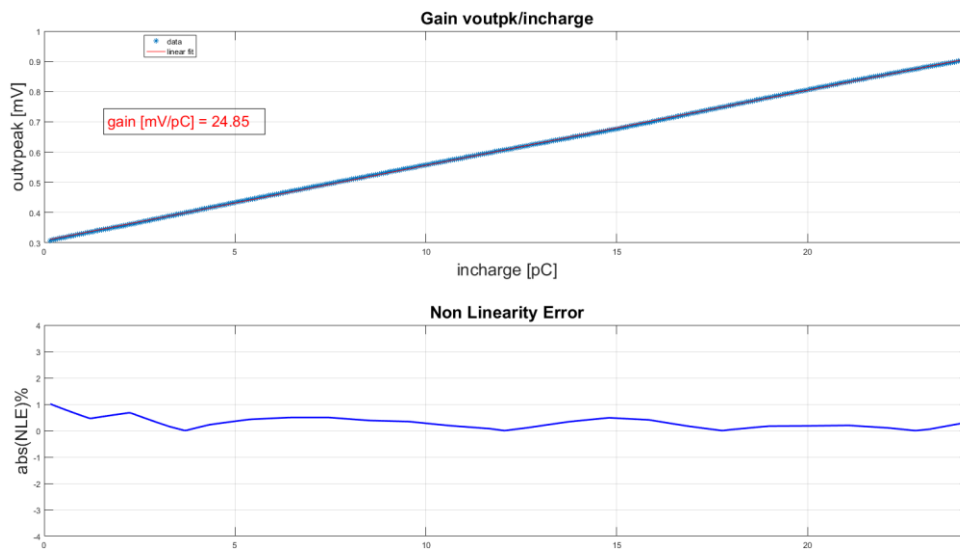

Fig. 3-25: First gain configuration: from 1phe to 150phe @T=27°.
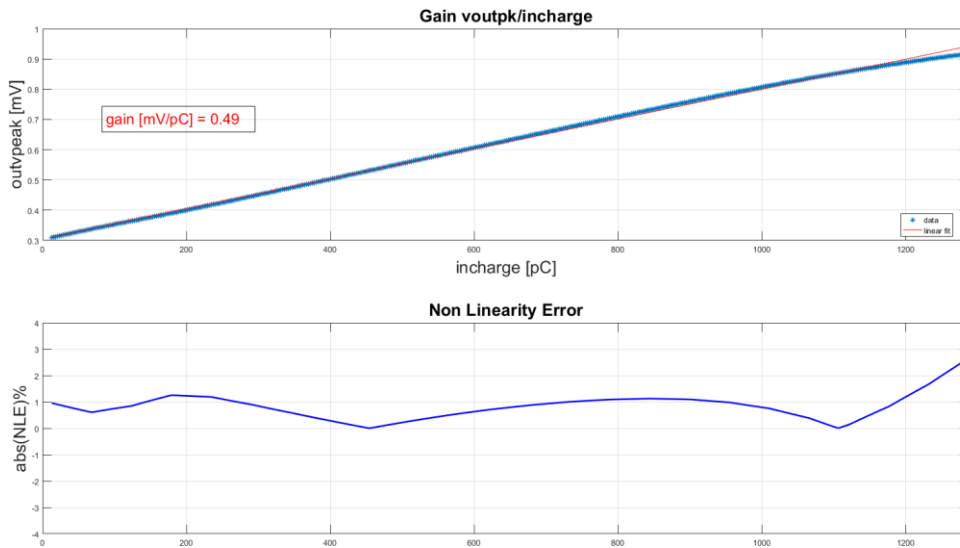
Fig. 3-26: Fourth gain configuration: from 60phe to 8000phe @T=27°.

For all the configurations, the simulated fluctuations of the gain as a function of temperature in the range from -40 °C to +85 °C are not relevant.

3.6.3.1 THE BASELINE STABILIZER

During a charge measurement operation, the RC-active integrator responds to the input signal with an output voltage signal resulting from the integration of the current, that is a scaled replica of the SiPM current according to the scaling factor of the selected gain. With an adequate choice of the RC time constant, that depends on the individual characteristic recovery time of the SiPM employed, a complete charge collection can be assumed, and the peak of the output voltage signal is proportional to that charge. The current-to-voltage gains have been properly calculated (see Table 3-6) to exploit the full dynamic range of the cascaded peak detector circuit with voltage ranging from a baseline of 300 mV, with no signal applied, up to approximately 900 mV, when saturation occurs.

The front-end, as mentioned above, operates according to the following description, in case of the detection of a valid event. After the peak is formed in the peak detector, the reset switch of the integrator is closed to rapidly discharge the integration capacitance, thus shortening the recovery time and minimizing the negative effect of pile-up on the resolution of the energy spectrum. This would be

137

fine enough in an ideal situation, with no additional sources disturbing the return of the output voltage to the baseline. Unfortunately, due to the current sensitivity of the integrator any unintentional current signal with a longer time drift than the shaping time of the integrator itself might contribute to displace the output voltage from baseline, which in turn leads to non-linear behaviour of the peak detector. A robust low frequency current removal block, the baseline stabilizer, has been designed to stabilize the output voltage of the integrator (see Fig. 3-24). The original idea [7], [8] has also been implemented and tested [9], [10]. Some adjustments have been introduced to the original design to comply with the technology node adopted, and to allow for a high event rate with negligible dead time. It keeps the output voltage out of the saturation as it can both source and sink excess currents, thus removing low frequency disturbing signals at the input of the integrator. A major disturbance is represented by the uncompensated DC leakage current of the current mirrors that convey the scaled replica current to the integrator. The leakage current can be as large as 30 μA and it is turned into approximately 150 μV of baseline voltage shift at the output, thanks to the effect of the feedback; in fact, the current-to-voltage gain around DC is less than 15 dBΩ, as shown in Fig. 3-27, meaning that the corresponding offset of the output voltage is the current value times a transresistance factor of about 5Ω.
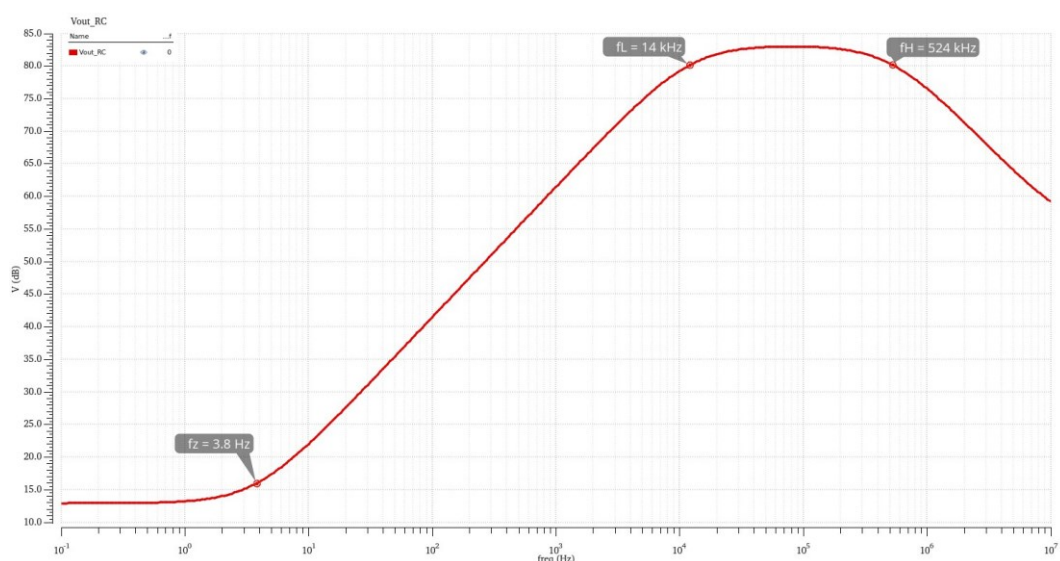

Fig. 3-27: Magnitude of the transimpedance gain of active RC filter with baseline stabilizer.

The maximum integrator gain has been chosen so that peak output voltage corresponding to a single photon is about to 6 mV, compatible with the sensitivity of the cascaded peak detection circuit. This means that a 1 kHz current disturbance with 0.5 µA peak amplitude at the input will be superimposed, at the output, to the signal generated by the single photon with a peak amplitude equal to 6 mV, like a noise voltage source of 0.5 mV (gain is 60 dBΩ at 1 kHz), that negligibly contributes to the degradation of the signal-to-noise ratio.

The loop gain G(jω) has a frequency dependent transfer function with a very slow pole p_{fb} = 25 rad/s introduced by the transconductor in the feedback network and a DC gain G_{loop} (0) = 3500. The transfer function of the integrator is responsible for the fast pole p_H that, according to the selected time constant, can be either equal to 350 ns or to 250 ns. Closing the loop causes the slow pole to become the zero shown in Fig. 3-27, whereas a slow pole p_L is created in the closed loop transfer function, at the unity gain frequency of the loop gain, as shown in Fig. 3-28:
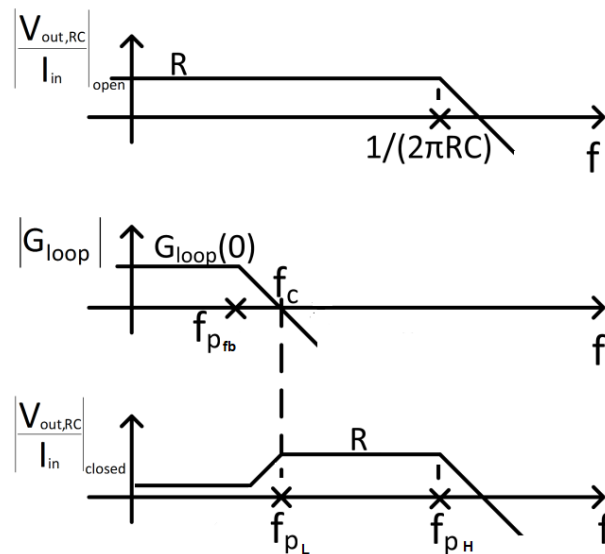

Fig. 3-28: Sketch of the Bode diagrams composition of the stabilized R_FC_F filter (from [10]).

The pole can be calculated as $p_L \cong p_{fb} \cdot G_{loop}(0) \cong 88 \ krad/s$, whereas the pole frequency due to R_FC_F integrator is equal to $p_H \cong 1/\tau_{RC} \cong 2.9 \ Mrad/s$. In case τ_{RFCF} = 350 ns is considered.

To avoid an excessive area occupancy, the slewing capacitor of the slew limited buffer in Fig. 3-24 is equal to 1.5 pF. This capacitance is introduced to smooth abrupt changes at the output of the integrator while the pole capacitance C_{pole} that determines the low frequency pole in the feedback network. The pole capacitance is equal to 2 pF, thus, the transconductance of the transconductor block is $G_m \cong p_{fb} C_{pole} \cong 50\ pA/V$ which is an extremely low value that can be obtained only with transistor operating in deep subthreshold

3.6.4   THE LINEARITY EXTENDER CIRCUIT

The reduced dynamic range is a relevant issue of the RCG TIA preamplifier. Overload currents generated by intense flashes of light on the surface of the SiPM don't comply with the preamplifier's ability to detect extremely low-level signals. In fact, keeping bias currents small to meet power consumption constraint is unfit to handling large signals for two reasons (referring to Fig. 3-8):

- Possible passage of M1 into the linear operation region in the left leg of the differential structure, due to the decrease of its drain voltage
- Possible cutoff of M2, due to the increase of its source voltage, with V_{O2} stuck at VDD in the right leg.

Apart from DC current shortcoming, these different limitations are due to the differential structure of the front-end and to the connection type of the SiPM. In more detail, looking back again to Fig. 3-8, the SiPM current signal comes out of the left-side input terminal of the circuit and flows into the right-side terminal in such a manner that, when the signal magnitude increases, the voltage V_{O1} (drain) decreases and M1 enters the linear region of operation. On the contrary, in the same conditions, the current in the RCG of the right side of the circuit decreases and M2 enters cut-off. Consequently, the total impedance seen by the SiPM dramatically increases, the recovery time of the detector becomes very slow and a non-linear behavior can possibly arise, because of the bias voltage variations experienced by the detector, even though the DC decoupling of the series input capacitance may help to minimize the latter effect.

To circumvent drawbacks due to impedance changing and nonlinearity, a linearity extender circuit is designed that extends the dynamic range of the front-end, without interfering with its small signal behavior.

Concerning the MOSFET M1, the increase of the linearity range can be obtained by inserting a diode-connected MOSFET, ML1, in parallel to the load resistor, as reported in Fig. 3-29.
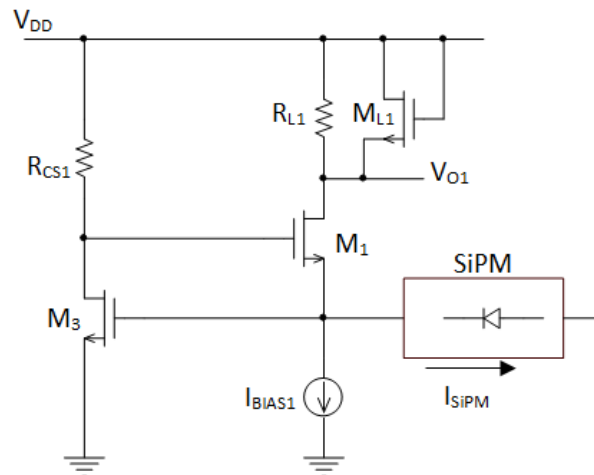

Fig. 3-29: Solution to avoid the transition of M1 in triode region.

M_{L1} is normally off, so during the formation of the trigger, it does not give any contribution to the behavior of the circuit (except a small capacitive load to the output node). When the SiPM current I_{SiPM} becomes larger and larger, it turns on, limiting the excursion of the voltage V_{O1} and preventing the transition of M_1 into the triode region. When this happens, the SiPM signal splits in two contributions, one flowing through the load resistor, the other through M_{L1}.

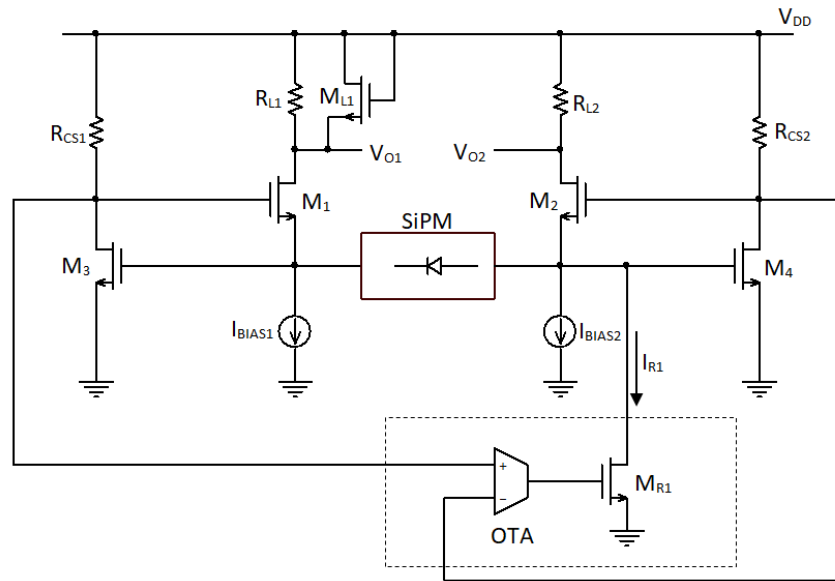The solution found for the right side of the preamplifier is schematically shown in Fig. 3-30.

Fig. 3-30: Schematic representation of the solution proposed to extend the dynamic range of the front-end.

The structure of the large signal OTA used in the circuit has been designed and is very simple: it is based on an p-type MOSFET differential pair with an active load, unbalanced by means of a current source in one of the two branches of the circuit. It can sense the voltage differences across the drains of the common source MOSFETs of the differential regulated common gate front-end and is able to turn on the MOSFET M_{R1} only when the circuit offset, intentionally introduced, has been recovered, thanks to the current of the SiPM pulse. In fact, at the bias point, M_{R1} is completely off, because of the offset introduced in the OTA (easily tunable by means of a given current), that is in negative saturation. In presence of a valid event, after the trigger of the comparator has occurred and before M2 turns off, the differential input voltage of the OTA is able to recover the offset and M_{R1} turns on. This closes the feedback loop formed by the OTA, M_{R1} and M4 and M_{R1} starts to take a substantial fraction of the large current of the SiPM.

Extensive simulations have been performed to evaluate the effectiveness of the proposed solution and make it reliable under different process and temperature conditions. This block should not affect the behavior of the front-end until the threshold is overcome and should also guarantee linearity in the reconstruction of

the charge. In fact, as it is clear from Fig. 3-31 that follows, the MOSFET prevents saturation by absorbing all the SiPM current in excess to the DC current mirror that sets the bias current for the CG transistor of the right side. This happens as soon as the current exceeds 400 µA which is the limit current before saturation occurs.
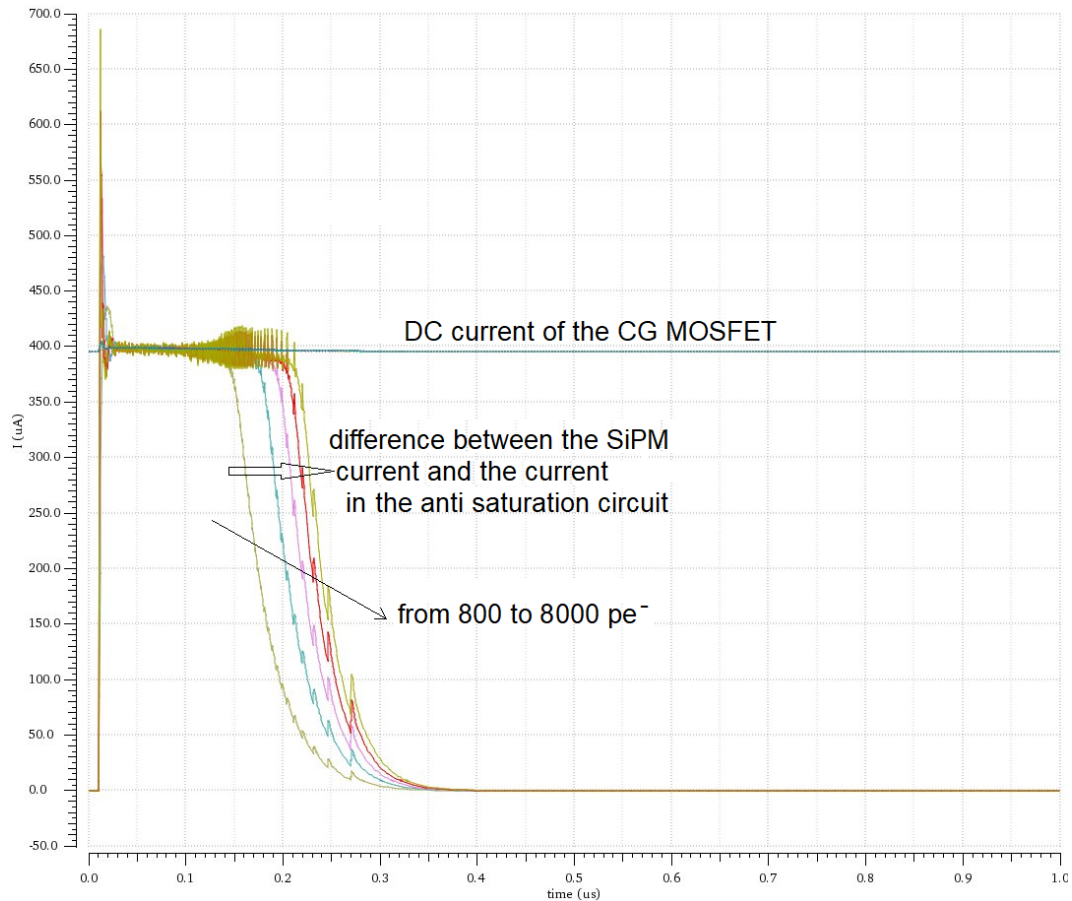

Fig. 3-31: Time diagram of the current difference between the SiPM and the anti-sat circuit.

The large signal in the SiPM is usually generated by a scintillator that produces intense light flashes with thousands of photons arriving in accordance to the statistic of a Poisson process. By design, the linearity extender circuit is able to prevent saturation for signals up to 8000 pe$^-$ emitted by the source during the time defined by the time constant of a scintillator (around 40 ns for a Lutetium based scintillation crystal like LYSO: Ce).

.

3.6.5 THE PEAK DETECTOR

As already mentioned, the Peak Detector (PD) is used to extract the peak of the output voltage of the integrator, which is proportional to the total amount of charge contained in the SiPM pulse; it also notifies the ADC that a stabilized voltage is available and can be converted into a digital format. The architecture of the designed PD is shown, in principle, in Fig. 3-32. It is based on a feedback loop containing a PMOS current mirror M_1-M_2 which works as a rectifying element [11]. When a valid event is recognized, the PD is configured by the internal logic to track the voltage signal from the integrator, to detect its peak and to store it in the analog memory, ready to be digitized by the ADC. In fact, with both switches S1 and S2 closed and S3 opened, if the input voltage from the integrator increases, the "hold" capacitor C is charged by the mirror and the output voltage follows the input. When the input pulse reaches its peak value, the current in the PMOS mirror reverts to zero, the loop is broken, S3 is closed while S1 and S2 opened, thus the charge stored on the capacitor is frozen.
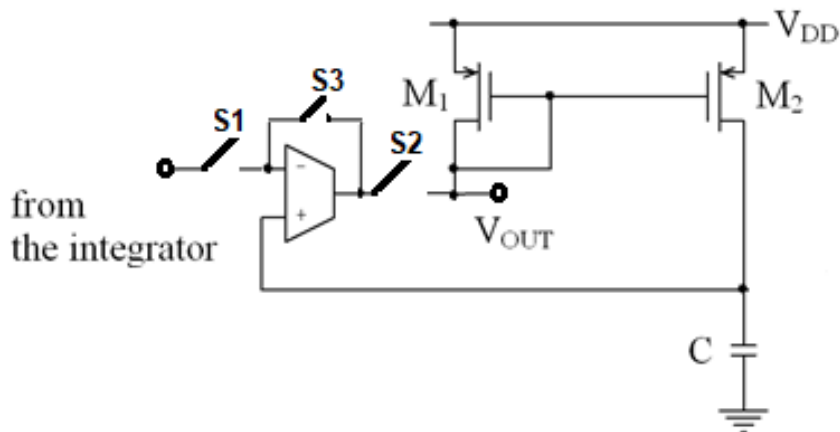

Fig. 3-32: A schematic picture of the Peak Detector.

The actual operating mode sequence of the PD is decided by means of control signals generated by the same front-end and will be discussed in the next section. In Fig. 3-33, the charge-to-voltage gain curves are plotted. Each of them can be conveniently set according to the required input range from 1 pe$^-$ to roughly 8000 pe$^-$.
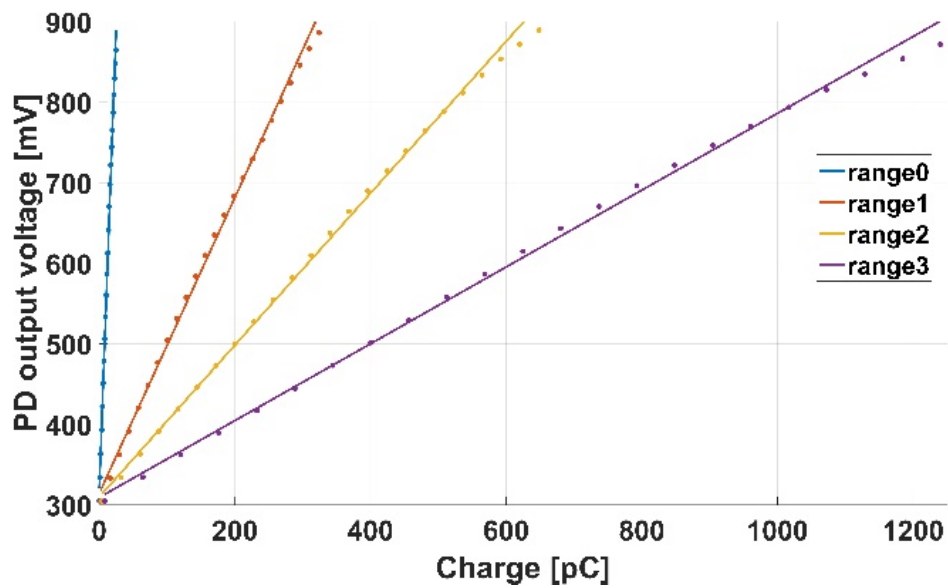
Fig. 3-33: Gain curves reporting PD output voltage as a function of the charge for all the selectable ranges.

3.6.6 AN OVERVIEW OF THE CHANNEL OPERATION MODES

Upon a trigger issue of the fast timing discriminator, the channel undergoes a sequence of steps to ensure the proper handling of the information carried by the incoming signal. In fact, as already pointed out, the voltage threshold of the fast timing discriminator is usually set at very low levels, to allow for single photon early detection and improve timing accuracy; this may provoke false triggers due to noise and dark pulses that would disturb the correct processing of concurring valid events, unless some strategy is deployed that recognizes and stops them. To do this, the slow comparator is used. It employs a high threshold to issue a trigger validation signal.

The connection of the "slow" comparator to the output node of the current replica circuit is schematically proposed in Fig. 3-34 and some control signals involved in the trigger validation process are reported.
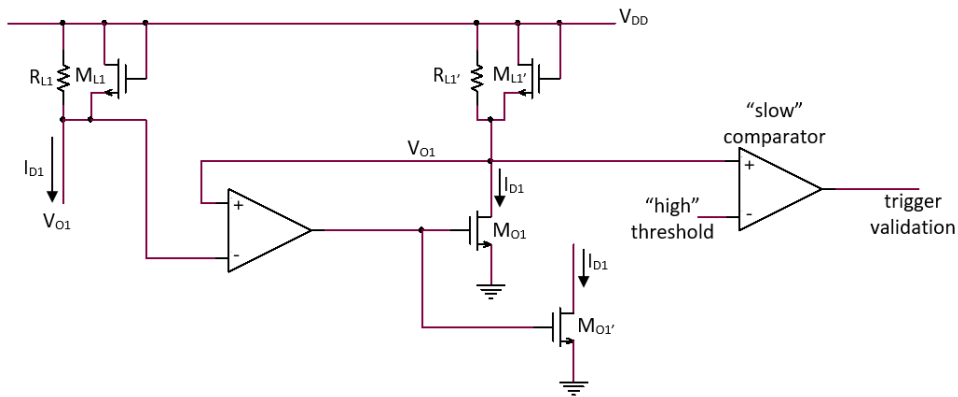
Fig. 3-34: Some of the signals involved in forming the validation trigger.

Fig. 3-35 shows a simplified representation of the architecture of the channel, with the signals and the blocks involved in either the validation or exclusion of an event.
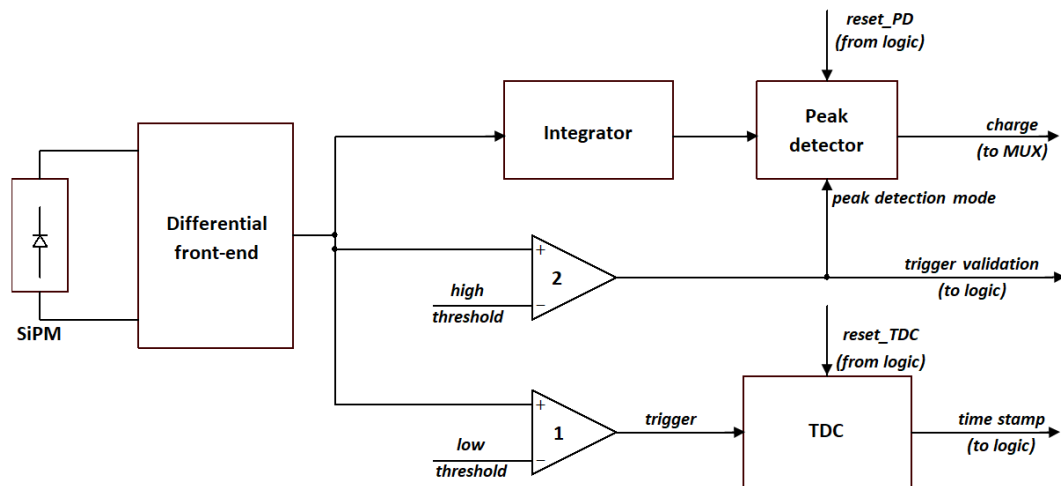

Fig. 3-35: Simplified functional block diagram of the channel.

Starting from the structure of the analog channel, schematically illustrated in the following Fig. 3-36, the sequence of the states that the circuit must go through and the signals used to control the behavior of the different blocks are detailed.
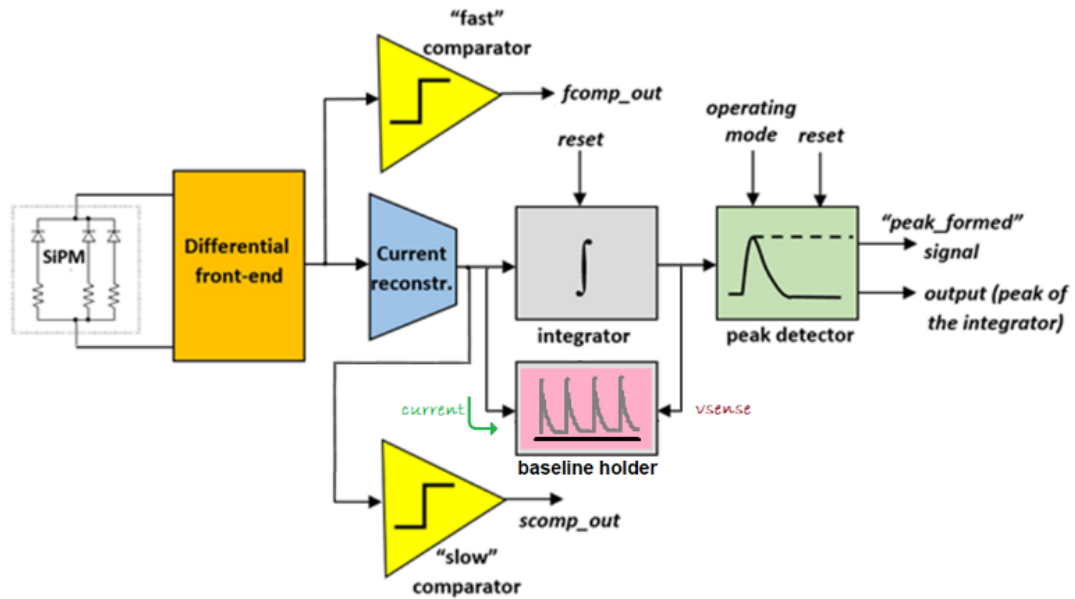
146

Fig. 3-36: Structure of the analog channel with digital signal highlight.

In the following the operations of the channel are described.

Operations for valid signals.

In the initial state, the circuit is waiting for a trigger. The reset of the integrator is inactive and the peak detector (PD) is working in "voltage follower" mode. The "fast" comparator fires when the output signal of the differential front-end overcomes the low threshold, whereas the "slow" comparator has a higher threshold and is used only to discriminate signals that are small against a programmable high threshold.

As soon as the fast comparator fires, the operating mode of the PD is switched to "peak detection" and the circuit starts to track the peak of the integrator output. At the same time, a short time window of duration TW is opened using a delayed replica of the comparator output, *fcomp_out*. The rising edge of *fcomp_out* is also used to activate the TDC that will measure the activation time and mark it with a timestamp. If, during the time window TW, the "slow" comparator fires as well, the integrator, the PD and the TDC simply go on with their operations.

Once the peak of the input signal is reached, the PD activates the *peak_formed* signal and goes into "analog memory" operation mode. In this operation mode,

147

the PD stores the analog level corresponding to the peak of the input signal and the operation results are not influenced by the output of the integrator.

As soon as the *peak_formed* signal becomes active and the PD is switched in analog memory mode it turns out that:

a) a *service_request* output signal is activated reaching out the external logic. After receiving the *service_request* signal, the external logic, as soon as the ADC is available, directs the PD output towards the ADC and the A/D conversion starts.

b) the integrator is reset: the *reset* signal of the integrator is kept active to avoid integration of possible further SiPM pulses which can be produced during the read-out phase. When the A/D conversion has been completed, the external logic brings back the analog channel to the initial state. The PD is reset and comes back into the "voltage follower" operation mode. The TDC is reset.

Then, there are two cases:

1. if the fast comparator output *fcomp_out* is inactive (which means that no other SiPM pulses are injecting current in the front-end), the integrator reset is removed, and the circuit comes back into the initial state.

2. If the fast comparator output *fcomp_out* is active (which means, for instance, that other SiPM pulses, uncorrelated with the event just read-out, are injecting current in the front-end), the reset of the integrator is kept active until the *fcomp_out* signal becomes inactive. As soon as this happens, the reset of the integrator is removed. In this way, the integrator output is at its baseline and it is not influenced by spurious charge contributions from previous events, when the circuit starts again waiting for a new trigger.

Fig. 3-37 shows a schematic representation of the described operation sequence.
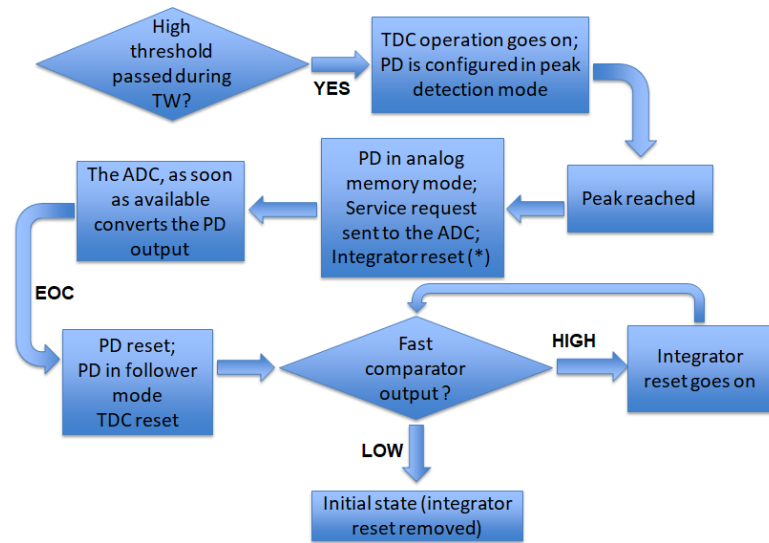
Fig. 3-37: Sequence of the read-out operation: valid signals (high threshold overcome during time window TW).

Fig. 3-38 shows a simplified time diagram of the signals involved in the read-out operation.
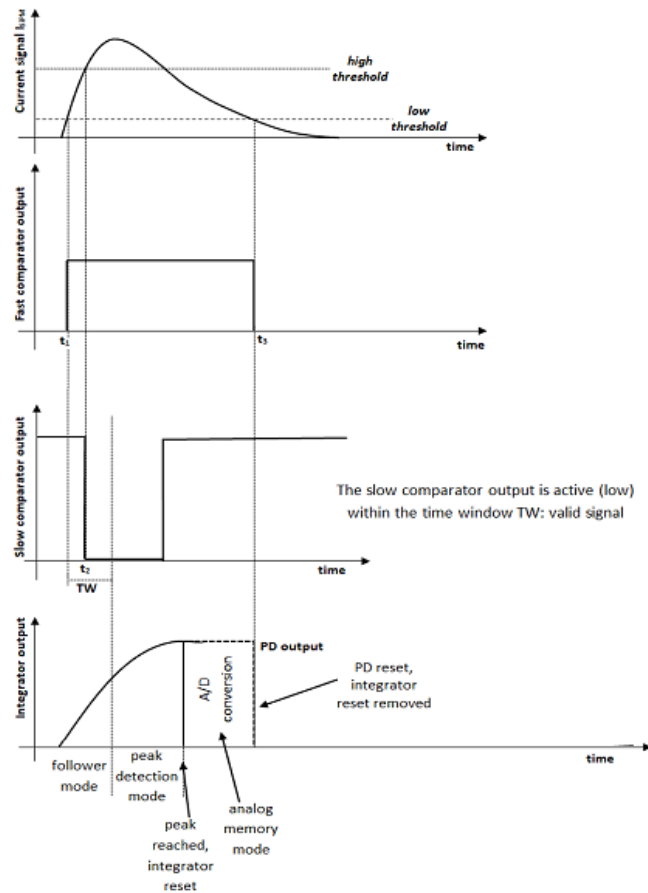

Fig. 3-38: Operation of the front-end: qualitative time diagram for a valid signal.

Operations for invalid signals.

In case the high threshold is not overcome within the time window TW, i.e. the signal that has caused the trigger is not considered a valid signal since it is too low, as soon as TW expires, a reset for the TDC is generated. At the same time, the integrator, which has started to integrate the invalid signal, is reset and its reset signal is kept active until the fast comparator comes back to its initial state, i.e. until the current pulse of the SiPM comes back to the baseline. In this way, the integrator can recover its own baseline and be ready for a new signal and a new trigger. The PD does not switch to peak detection mode thus remaining in voltage follower mode. Fig. 3-39 illustrates the operations for an invalid signal.
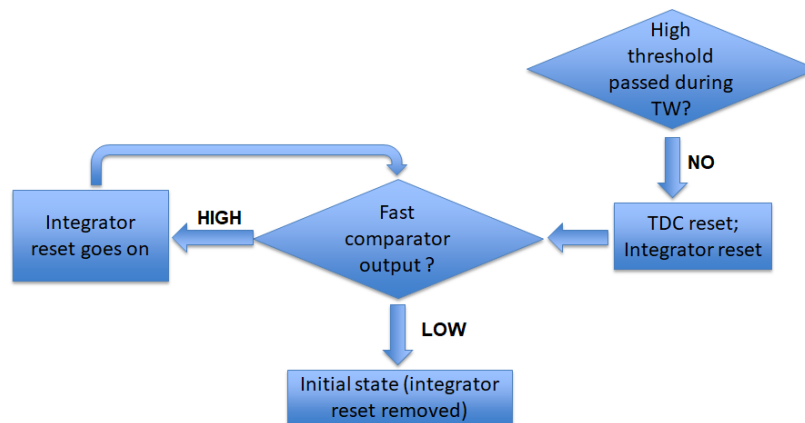
Fig. 3-39: Sequence of the read-out operation: invalid signals (high threshold is not overcome during time window TW)

The following Fig. 3-40 shows the qualitative time diagram of the signals involved.
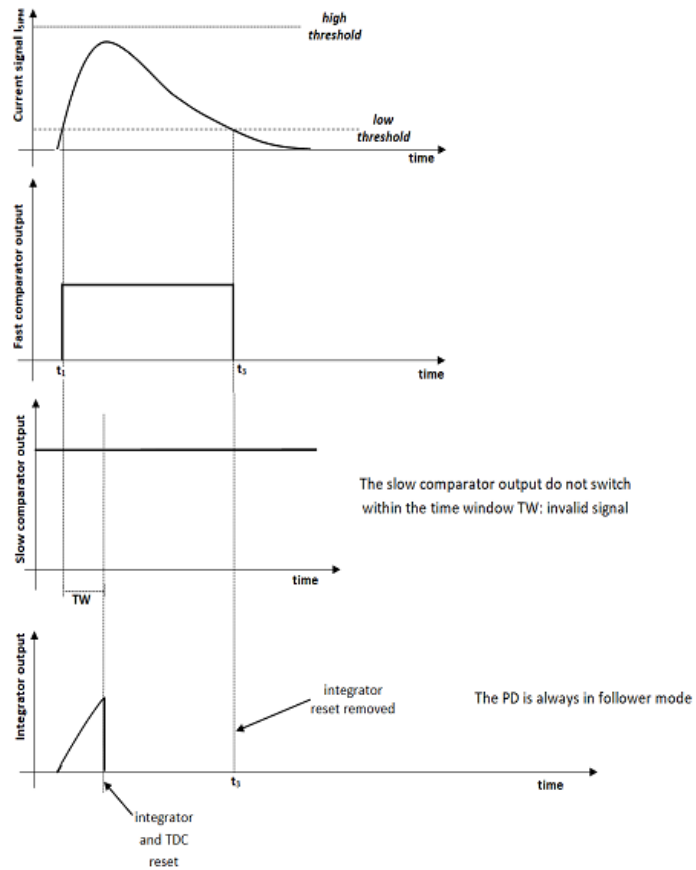
Fig. 3-40:. Operation of the front-end: qualitative time diagram for an invalid signal.

A VHDL description of the state machine which must manage all the involved signals has been developed and validated using a suitable testbench.

3.6.6.1 THE FINITE STATE MACHINE (FSM)

The digital machine is a control digital macro conceived as an asynchronous state machine with 7 states. Fig. 3-41 shows a simplified state transition diagram, which illustrates the sequence of the states that the digital machine passes through.
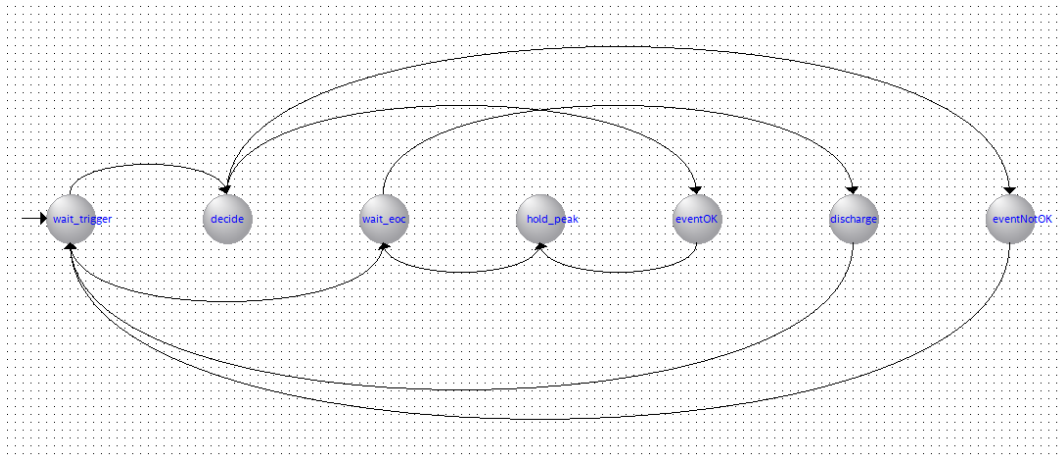

Fig. 3-41: State transition diagram of the digital machine.

Table 3-7 is the flow table of the machine, which is of the Moore type, where the transitions between the states and the signals which trigger and/or control them are defined.

Table 3-7: State transition table of the digital machine.

CURRENT STATE	/Output vector	Input_1 scomp /single	Input_2 fcomp	NEXT STATE	STATE TRANSITION triggering signals
wait_trigger	/Out_0	'X'	'X'	decide	⌐fcomp
decide	/Out_1	'=LOW'	'X'	eventNotOK	⌐fcomp_d
decide	/Out_1	'=HIGH'	'X'	eventOK	⌐fcomp_d
eventOK	/Out_2	'X'	'X'	hold_peak	⌐peak_formed
hold_peak	/Out_3	'X'	'X'	wait_eoc	⌐peak_hold
wait_eoc	/Out_4	'X'	'=LOW'	wait_trigger	⌐eoc_ADC
wait_eoc	/Out_4	'X'	'=HIGH'	discharge	⌐eoc_ADC
eventNotOK	/Out_5	'X'	'X'	wait_trigger	⌐fcomp_2d
discharge	/Out_6	'X'	'X'	wait_trigger	⌐fcomp_n
Not_used	/Out_7	'X'	'=LOW'	wait_trigger	⌐autoreset_d
Not_used	/Out_7	'X'	'=HIGH'	discharge	⌐autoreset_d

The inputs which control the next state where the machine moves starting from the current state are the following:

- *scomp_or_single*, which represents the logic OR between the output of the slow comparator and a configuration flag which, if set to HIGH value, disables the discrimination of valid events based on the amplitude of the signals and makes valid all the signals which trigger the fast comparator;

- *fcomp* is the output of the fast comparator used for the timing.

The inputs that trigger the transition from the current state to the next one are the following:

- ⎍ *fcomp* is the leading edge of the output of the fast comparator;

- ⎍ *fcomp_d* is the leading edge of a delayed replica of the output of the fast discriminator;

- ⎍ *fcomp_2d* is the leading edge of a delayed replica of the signal *fcomp_d*, thus *fcomp_2d* is a replica of the output of the fast comparator further delayed;

- ⎍ *peak_formed* is the leading edge of the signal generated by the peak detector when the peak of the signal has been reached;

- ⎍ *peak_hold* is the leading edge of the signal generated by the peak detector when the peak of the signal is stable and available at the output of the peak detector;

- ⎍ *eoc_ADC* is the leading edge of the signal generated by the ADC when the conversion has been done;

- *autoreset_d* is a signal generated when the machine accidentally goes into the Not_used state (for noise or whatsoever). Its leading edge brings the machine in the wait_trigger state.

The states of the machine have been defined according to the read-out behavior already described in previous section. The following

Table 3-8 defines the outputs of the digital machine contained in Table 3-7.

Table 3-8: Outputs of the digital machine.

/Output_ vector	STATEOUT [2:0]	reset_TDC	start_TDC	reset_SHAPER	trackmode_PD	request_ADC
/Out_0	"000"	'1'	'0'	'0'	'1'	'0'
/Out_1	"001"	'1'	'1'	'0'	'1'	'0'
/Out_2	"010"	'1'	'0'	'0'	'0'	'0'
/Out_3	"011"	'0'	'0'	'1'	'0'	'0'
/Out_4	"100"	'1'	'0'	'1'	'1'	'1'
/Out_5	"101"	'0'	'0'	'1'	'1'	'0'
/Out_6	"110"	'1'	'0'	'1'	'1'	'0'
/Out_7	"111"	'0'	'0'	'1'	'1'	'0'

The variable STATEOUT in

Table 3-8 is the code associated to each state of the machine.

The description of the outputs reported in

Table 3-8 follows.

1) **reset_TDC**
   It's an active LOW and normally HIGH reset signal for the TDC. It resets the TDC each time an event is discarded by the machine (whenever the slow comparator threshold has not been overcome within the time window defined by the delayed replica of the *fcomp* signal, i.e. *fcomp_d*). When the leading edge of the *fcomp* signal is generated (i.e. the fast comparator fires), the machine goes into the decide state and in any case the TDC is started (see the following description of the start_TDC signal).

   The reset_TDC signal is generated when the digital machine enters the eventNotOK state. Furthermore, the TDC is kept in the eventNotOK state for at least 10ns (6ns for reset and further 3ns to be ready to convert again) before the machine goes back to the wait_trigger state, where the reset_TDC signal is immediately set to HIGH by the machine and the system can accept and handle the next event.

   The reset_TDC signal is also activated to reset the TDC after a complete conversion of a valid signal, which occurs after the machine goes into the eventOK state.

2) **start_TDC**
   It's a normally LOW signal to the TDC, which starts a TDC conversion on the edge of its LOW to HIGH transition. The machine assigns a HIGH value to this signal, kicking-off a TDC conversion whenever the decide state is entered, due to LOW to HIGH transition of the fast comparator output.

The start_TDC signal is brought back by the digital logic to the LOW value on the transition from the decide state to whatever different state (eventNotOK state or eventOK state).

3) reset_SHAPER

It's an active HIGH and normally LOW signal to the integrator-shaper of the analog channel. It resets the output level of this circuit to the baseline by discharging the capacitor of the RC feedback network of the integrator.

Once the peak is formed and the peak detector evolves from peak detection mode to hold mode, the integrator is reset to the baseline. This happens when the machine enters the hold_peak state, i.e. when the peak detector generates the leading edge of the *peak_formed* signal.

The reset_SHAPER signal is released only when the machine comes back to the wait_trigger state.

4) trackmode_PD

It's an active LOW and normally HIGH signal to the peak detector. It is asserted when the eventOK state is entered and is deasserted only after the *peak_hold* signal is issued by the peak detector and the machine goes into the wait_eoc state. When asserted, this signal forces the peak detector to go out of the "follower" mode and enter the "track" mode, aimed at peak finding.

5) request_ADC

It's an active HIGH and normally LOW signal to the ADC. Upon entering the wait_eoc state, triggered by the *peak_hold* signal from the peak detector, a request of service is issued by the digital machine to the ADC for the conversion of the analog value held on the peak detector output. The request_ADC signal is kept active as long as the ADC is busy in the conversion process and is deasserted on the rising edge of the *eoc_ADC* signal (issued from the ADC), once the conversion has been completed.

Concerning the physical implementation of the asynchronous digital machine, the following approach has been adopted. The machine state undergoes a state transition on the leading edges of the previously defined signals. The signals are multiplexed, and the resulting output works as a sort of clock signal for the machine, though not free-running. Starting from the current state, the multiplexer brings out the signal which is responsible of the transition to the next state, one at a time, according to the state transition table. An asynchronous reset signal, active LOW, has been also added. The following Fig. 3-42 depicts the structure of the FSM.
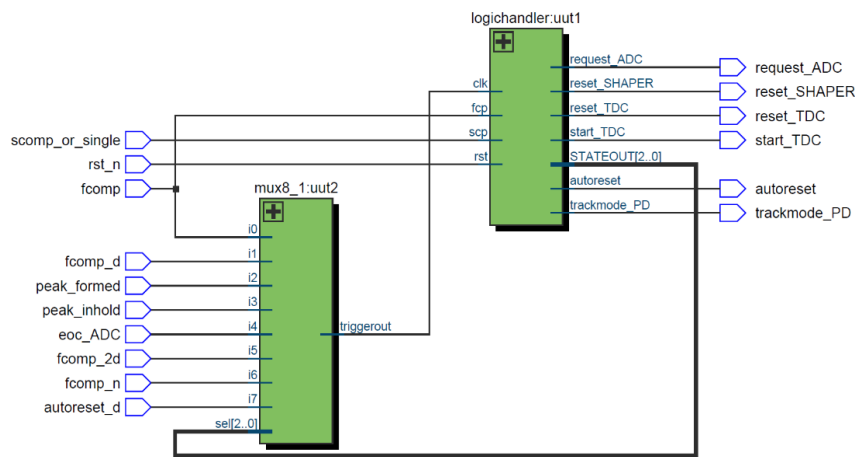

Fig. 3-42: Structure of the digital machine, with its inputs and outputs.

The "triggerout" signal in Fig. 3-42 is the output signal of the multiplexer used to cause the transition between the states of the machine and the "rst_n" signal is the external reset.

The digital machine has been synthesized and described using a VHDL code. The following Fig. 3-43 depicts the schematic diagram of the machine corresponding to its VHDL implementation.

The behavior of the digital macro has been simulated and verified using suitable test-benches, providing the expected results.
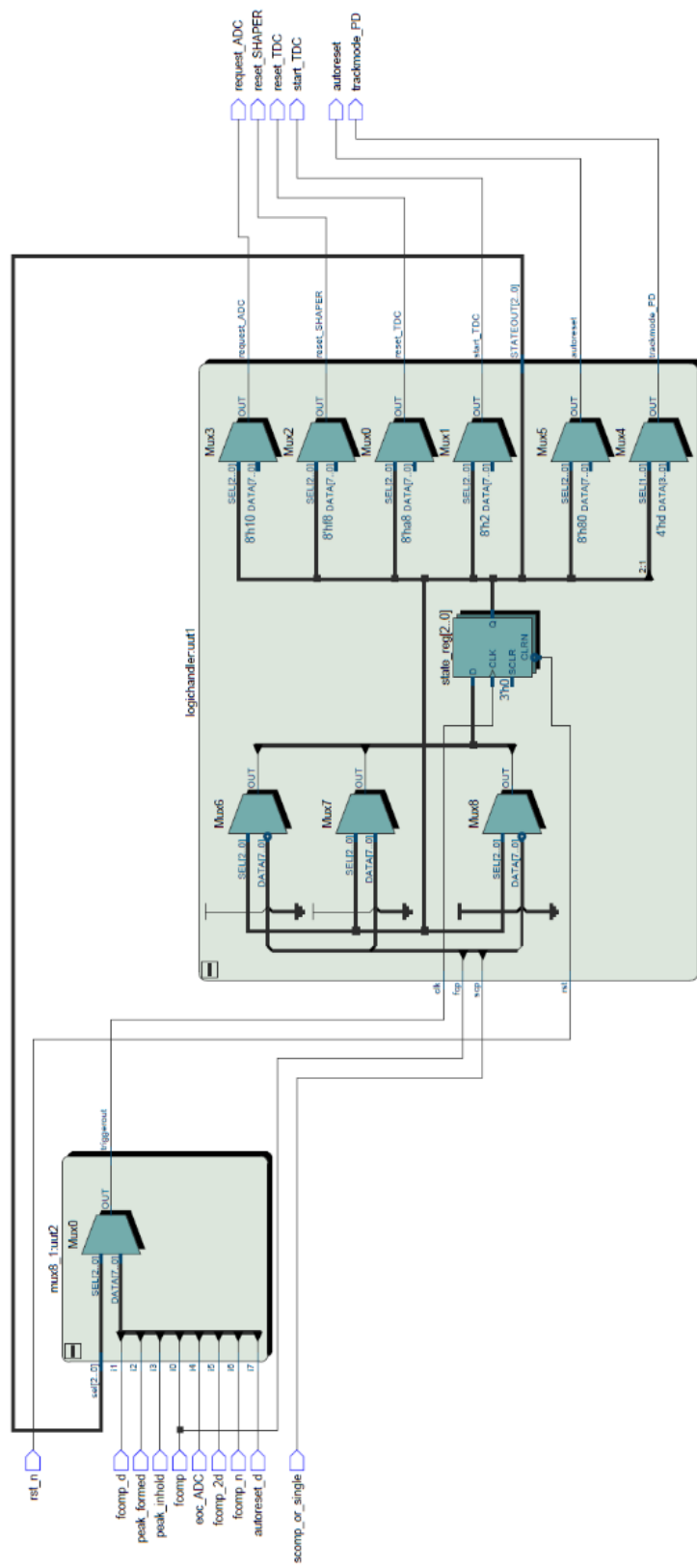
Fig. 3-43: Schematic diagram of the digital machine resulting from VHDL code.

3.7 LAYOUT

The channel layout measures approximately 1.2 x 0.5 mm^2 die area. Fig. 3-44 shows the layout of the channel ASIC with the TDC and DAC taking large part of the area and leaving room for future optimizations.
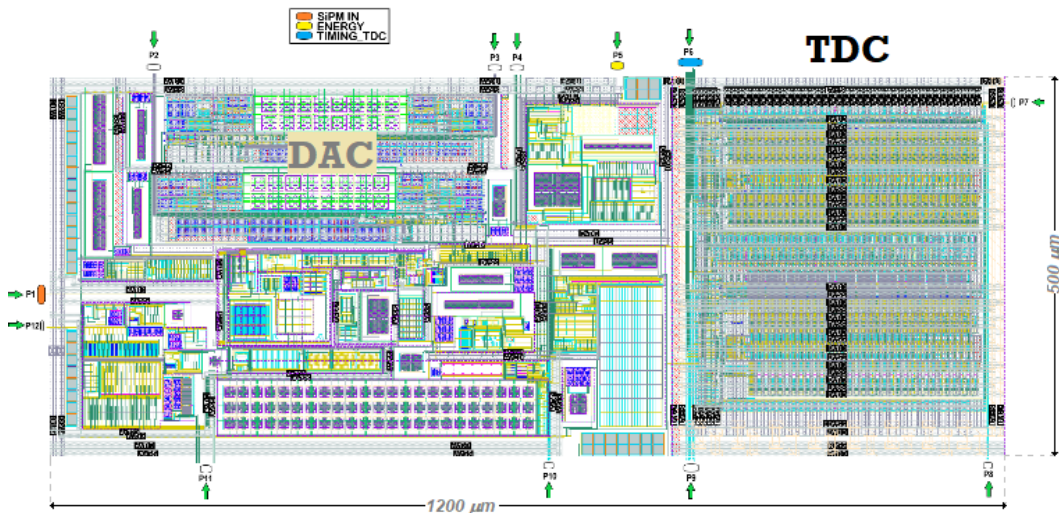

Fig. 3-44: Layout of the channel.

Pre-layout analysis has been useful to the investigation of certain design configurations to find an optimum solution early. Pre-layout analysis has also allowed for the setup of a bundle of design rules for subsequent design stages (for example, minimum distance of traces to keep crosstalk low).

The criteria adopted for the placement of the blocks in the top channel layout are described in the following. The placement of the differential front-end is the result of a compromise. The two input lines from the SiPM are on the extreme left-hand side of the channel (in the space between the block for the extension of linearity and the BIAS blocks), but, since the front-end is a low impedance current-buffer, it can afford to be moved towards the centre of the channel, so that the overall path of the signals interesting for the timing performance is shortened. The output of the front-end is very close to the fast comparator, which, in turns is very close to the digital part. With this arrangement, the distance between the digital part and the TDC, which must be necessarily placed on the extreme right-end side, due to its size and shape, is minimized.

The placement of the two DACs has been decided according to their size and aspect ratio and the location of their outputs. The upper DAC is oriented with the output on its left side, so that it is easy to reach the Threshold Adjustment block to be placed on the left side of the fast comparator, in the small area between the linearity extender, the lower DAC and the differential input stage.

3.8 POST-LAYOUT PERFORMANCES

Extensive post-layout simulations have been performed to validate the circuit performances. It is important to evaluate the contribution of the parasitic components which are non-negligible given the small feature size of this technology. The reliability of the circuit has been tested through parametric simulations with a sweep on the process corners and temperature.

Extensive simulations have been carried out to explore the behavior of the circuit in terms of linearity in all the operating conditions (e.g. temperature variations) and gain configurations.

To make more effective the verification process, an Ocean script has been set up, which has been used to measure the simulated time jitter of the single-photon response of the front-end. For all the corner PVT simulations, this script does the following:

1.   Configuration of the trimming resistor to obtain the correct value of the bias currents for the circuit;

2.   Configuration of the DAC of the Threshold Adjustment System to compensate the imbalance of the voltages on the output nodes of the preamplifier;

3.   Identification of the point with maximum slope of the preamplifier response, by means of the evaluation of the first derivative;

4.   Configuration of the DAC of the Threshold Adjustment System to set the threshold of the fast comparator corresponding to the point of maximum slope;

5.   Evaluation of jitter, average rise time and delay time, by means of Noise Transient simulations;

6.   Storage of the output data in a text file.

A plot of the noisy output pulse of the whole channel corresponding to a single photon impinging on the SiPM is shown in Fig. 3-45.
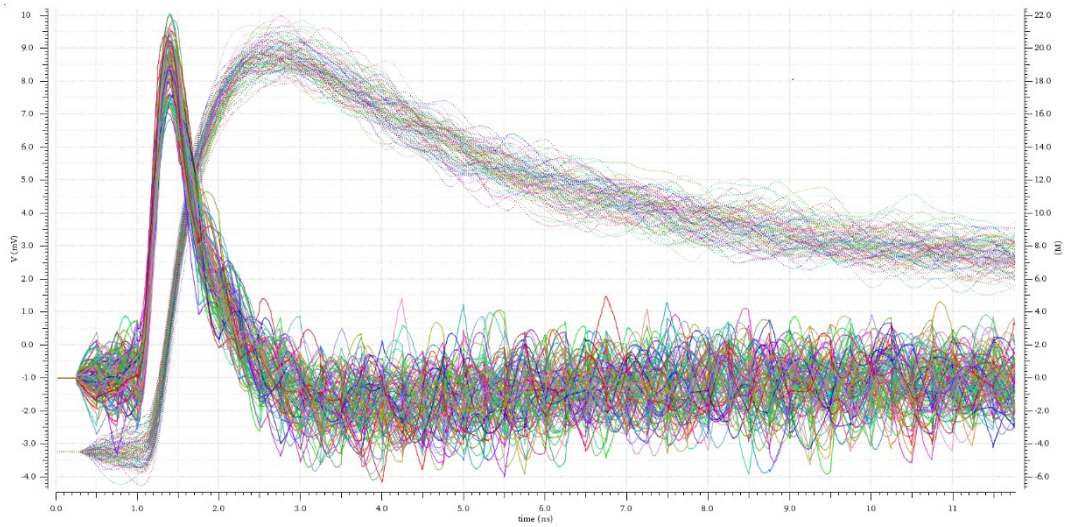
Fig. 3-45: The pulse and its derivative with the threshold set at max slope.

The main performance associated to the waveform shown in Fig. 3-45 are summarized in the following Table 3-9.

Table 3-9: Summary of the main performance and specifications of the front-end.

Parameter	TARGET SPECIFICATIONS	POST-LAYOUT SIMULATIONS
Rise time	1 ns	0.6 ns
Max slope	41.6 MV/s	21 MV/s
Peak value	25 mV	12.5 mV
Output noise (rms)	1 mV	0.65 mV
SNR (peak to rms)	28 dB	19 dB
Time accuracy (rms)	23.6 ps	33 ps
Dynamic range	1250 pC	1280 pC
Power consumption	10 mW	11 mW

3.9 REVIEW AND OUTLOOK

In this chapter of the thesis the design of a channel front-end for the readout of SiPM has been presented. Input impedance has been among the earliest metrics to drive the choice of the input stage configuration whenever the large equivalent input capacitance of the detector and the parasitic series inductance may result in speed limitation. To fit to the strict requirements imposed by fast timing applications, the current mode approach has been proved to be a viable solution. It offers the flexibility of a current buffer that provides input impedance tunability whilst fully preserving the current waveform of the SiPM. Focusing on jitter, speed is not the only characteristic to consider for design. In fact, minimization of the electronic noise is much relevant for timing and it can become a critical point to be addressed. Low level noise floor improves signal-to-noise ratio with the consequence that the leading-edge discriminator threshold can be moved down to the first initial onset of the waveform raise in search of the maximum signal slope that would further reduce time jitter.

A revised version of the channel, modified based on the simulation results, is at an advanced design stage so far, and a fabrication run will be launched at the beginning of 2020.

4 CHAPTER 4

4.1 INTRODUCTION

To demonstrate the main differences between current readout mode and voltage readout mode of the SiPM in terms of timing performance, it was decided to employ a ready-to-use SiPM interface board that had already undergone a pre-characterization test [1]. The preliminary evaluation tests were promising at that time, indeed not completely satisfying in terms of signal integrity; actions would have been taken to increase noise immunity and further simulations scheduled for final validation of circuit performance. At cost of minor adjustments this has been successfully accomplished so far, and the PCB is now being used to 'see' the single photon under the form of a dark pulse.

The interface board is assembled around the core amplifier composed of two discrete Bipolar Junction Transistors (BJT), Q_1, Q_2, and six passive components, R_1, R_2, R_3 R_E, R_C, C_B, as shown in Fig. 4-1.
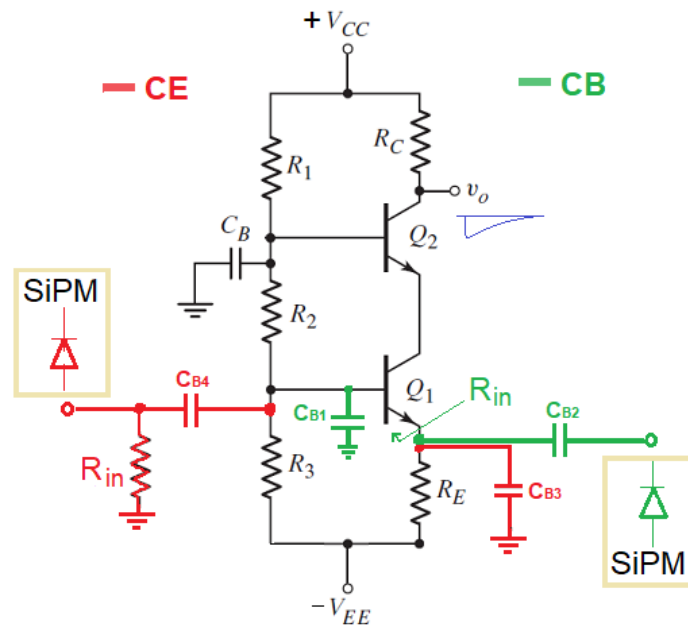

Fig. 4-1: Cascode amplifier with reconfigurable input section CE, CB.

Using jumpers on the Printed Circuit Board (PCB) to select one out of three values for the resistors R_E, the operating point of the amplifier can be changed, thus

setting the required DC current through Q_1, Q_2 (see Table 4-1 below). The input section can be manually configured to be used either as a Common Base (CB) stage or as a Common Emitter (CE) stage, acting in the former case as a current buffer (current-mode readout approach), in the latter as a voltage amplifier (voltage-mode readout approach). The DC current determines the transconductance of the current buffer and therefore the input differential resistance for the CB option, whereas, for the CE option, the same input resistor can be reproduced by soldering an equal value physical resistor R_{in} on the board. The evaluation tests aim at demonstrating how diversely the input impedance can affect the time jitter for the two basic configurations of SiPM amplifier. To make a significant comparison, input resistance and power consumption have been changed in the same way for both circuits and while in the CB, changing the current intrinsically changes the incremental input resistance, for the CE the current has been changed so far and a physical resistor of the same value of the incremental resistance of CB has been soldered on the PCB; meanwhile the load resistance and bandwidth have been left unchanged for both configurations, thus the output pulses are supposed to have also the equal heights for both configurations.

Table 4-1: Operating point settings and input resistances.

Biasing Resistance R_E	Biasing Current	CB Input resistance R_{IN} (typical)	CE Input resistance R_{IN} (nominal)
5.6 kΩ	2.41 mA	10.63 Ω	10 Ω
10 kΩ	1.36 mA	18.65 Ω	18 Ω
15 kΩ	0.91 mA	27.84 Ω	27 Ω

To preserve the signal speed and relax the gain-bandwidth trade-off, two BFR92P NPN Silicon RF transistors from NXP were used in an open-loop cascode structure, with the circuit operating with a dual ± 15 V supply voltage.

The cascode configuration consists of a CE stage driving a CB stage. It combines the advantages of the common-emitter and common-base circuits. Specifically, it

provides isolation between the input and the output of the amplifier by reducing the effect of the Miller capacitance that would otherwise limit the high frequency pole of a simple CE configuration. When used as a current buffer, the input CB of the cascode is inherently not affected by the Miller effect and already has a wide bandwidth. To limit the high frequency noise of the amplifier on the PCB and provide further amplification, two inverting stage realized with two non-inverting Op-Amps THS3001ID from Texas Instruments are cascaded and an additional voltage gain of 25 dB is obtained up to 250 MHz of frequency bandwidth.

The series inductance introduced by the interconnection wires has been estimated to be roughly 10 nH and it has been included in the comprehensive electrical model that has been used to investigate the front-end behavior, to simulate the signal shape and amplitude, and to retrieve the signal slope and the output voltage noise that contribute to jitter. The well-known electrical model of SiPM Hamamatsu MPPC S10931-050P has been used as a single photon signal source.

4.2 EVALUATION BOARD

The layout of the circuit that has been transferred onto the PCB and the final prototype with components after assembly is shown in Fig. 4-2. Two 3×3 mm^2 SiPMs Hamamatsu MPPC S10931-050P with 3600 micro-cells have been mounted, either of which can be used according to in-line configuration option.
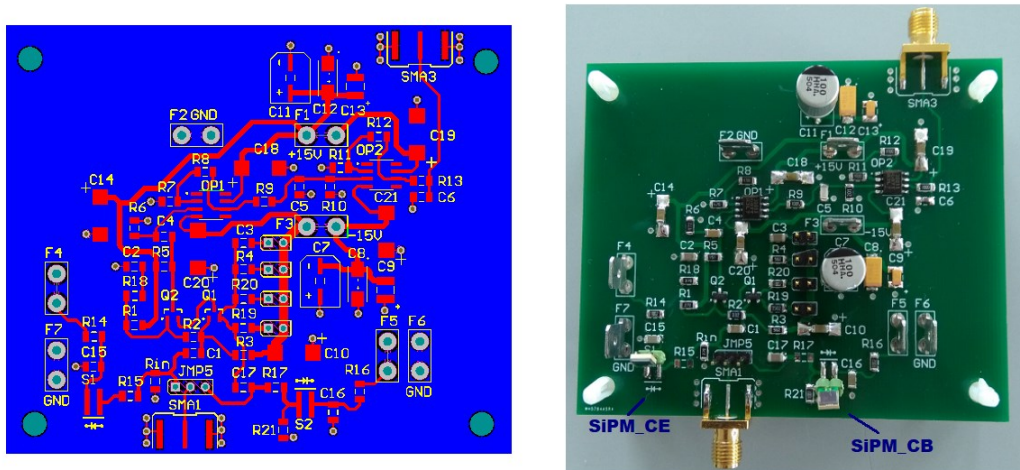

Fig. 4-2: Overall view of the PCB layout and components placement

The board provides the electrical connections to the bias supplies for both the SiPMs and the electronic interfaces that also share the same ground plane; it also features an analog output for direct connection to an oscilloscope for signal capture and analysis.

The evaluation tests have been run in an Electro-Magnetic Interference (EMI) sensitive environment with some major impact on the functionality of the system. Some simple EMI counter-measures, such as cutting ground current loops and shielding, have been deployed to dampen or even eliminate the electromagnetic disturbances.

4.3 PRELIMINARY TEST

To evaluate the performance of the proof-of-principle prototype, a PC based acquisition system has been arranged as reported in Fig. 4-3. A MATLAB® script is run to manage and control data communication and storage.
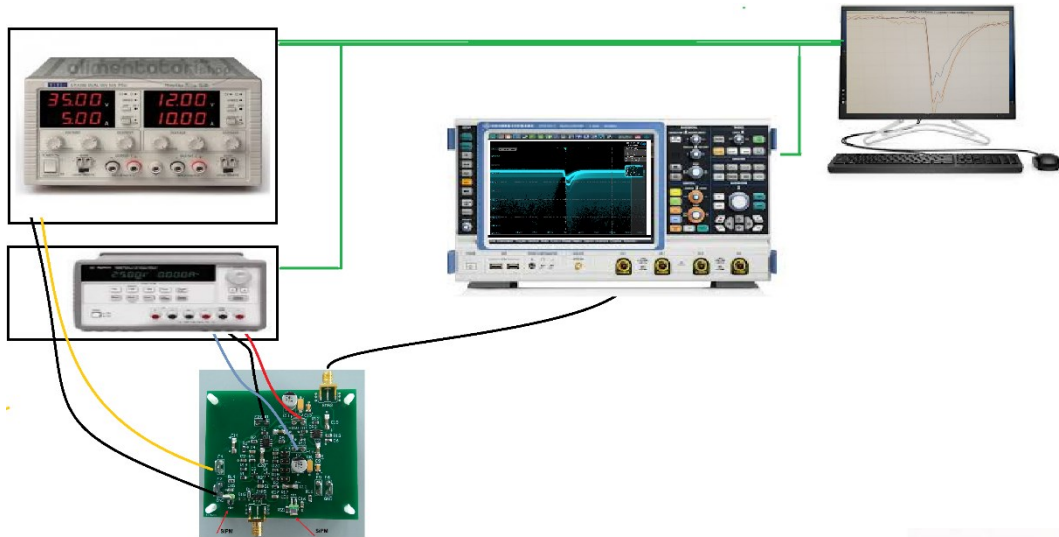

Fig. 4-3: Measurement setup using a digital oscilloscope for noise and transient recording.

To prevent the SiPM from saturation, it was sealed in a light-proof enclosure and to shield electromagnetic noise, an aluminium sheet was used to wrap up both the board and the power cables, realizing a Faraday cage. Two external power supplies from the equipment available in the laboratory were used to generate supply voltage for biasing the SiPM (overvoltage of 3 V on top of the 70.5 V breakdown voltage) and the PCB (±15 V). The output signals were captured using a Tektronix oscilloscope RTO-1044 with 4 GHz bandwidth and a sampling rate up to 20 GS/s on a single channel. The output measurement signal was AC coupled to the oscilloscope, because removing the DC component of the signal the measurement resolution can be increased.

The dark pulses thermally generated in the SiPM produce output voltage pulses that are captured by the oscilloscope. They are undistinguishable from the pulse waveform generated when a light photon impinges on the surface of the SiPM,

and consequently they can be used to assess the timing performance of the circuit, evaluating the electronic noise and the signal slope. The electronic noise was measured as the standard deviation of the Gaussian fit function obtained by the recorded histogram of the values of the pedestal baseline, with the SiPM switched off.

An oscilloscope screenshot of the measured output pulse is presented in Fig. 4-4.
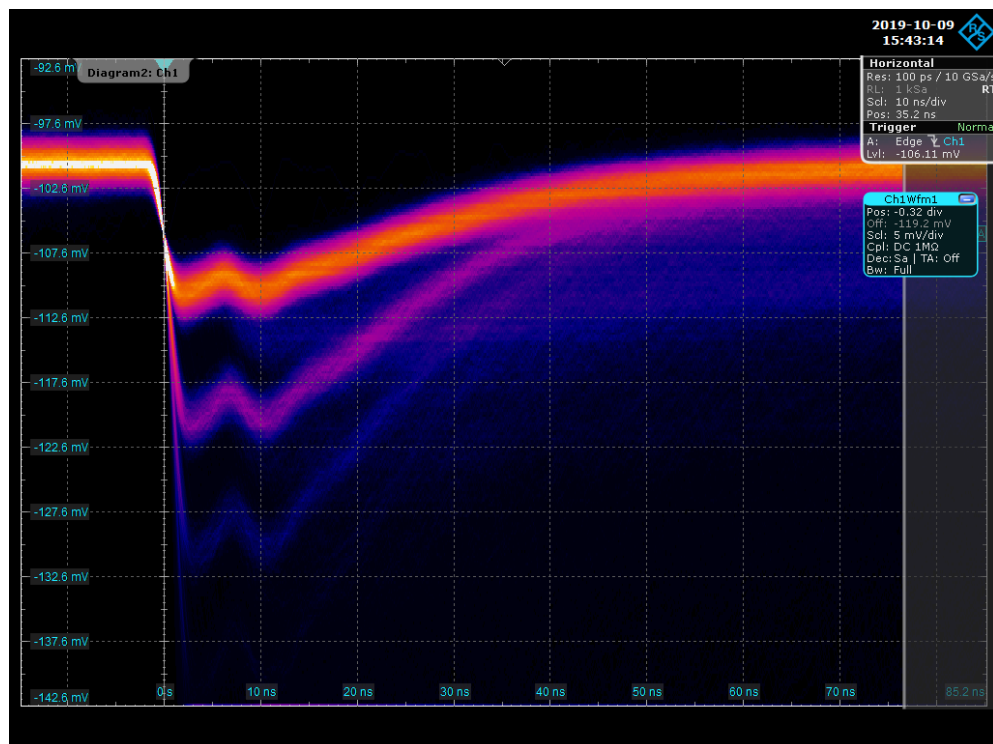

Fig. 4-4: Oscilloscope screenshot showing the pedestal, single and multiple photoelectrons.

The aim of the measurements is to experimentally demonstrate the effects of input resistance on time jitter, again calculated as the ratio between electronic noise and signal slope, for both a current mode and a voltage mode approach. To complete the test, the effect of interconnection parasitic components on jitter should have been investigate using different test load combinations but, as will be clearly demonstrated in the proceeding discussion, the present setup and equipment proved inadequate to achieve this objective.

4.3.1 THE DARK PULSE METHOD

A typical temporal trace captured from the oscilloscope in single shot mode and plotted with MATLAB® is shown in Fig. 4-5.
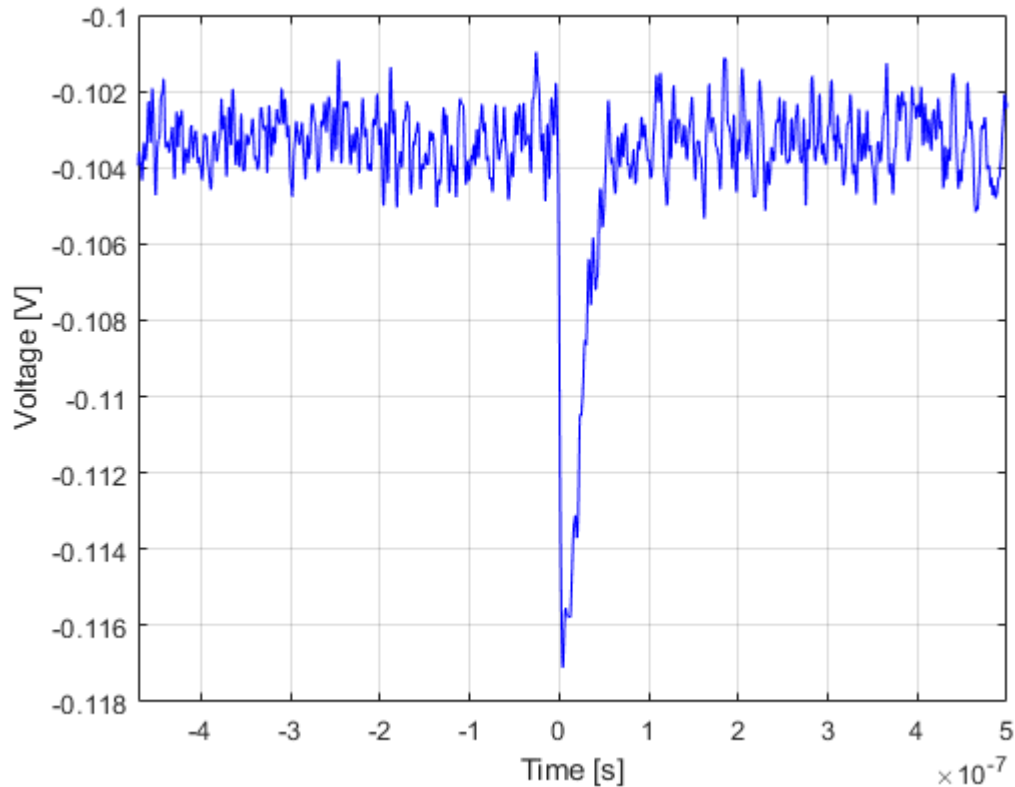

Fig. 4-5: Measured response of the circuit in CB configuration to a dark pulse.

It represents the output pulse of a dark event occurred in the SiPM, for the case CB with a medium biasing current of the cascode.

The falling edge is very steep, and the amplitude is 12 mV. The peak-to-peak noise is around 2 mV which, with a crest factor of 5, makes less than 400 µV rms.

With simulation data in mind and an eye-check to the acquisitions on the oscilloscope in run mode the trigger level was first regulated to a level suitable to acquire the single dark pulses; thanks to a customized acquisition procedure, the measurement was almost completely automatized and as many as 10000 waveforms were sent to the PC over a fast Ethernet link with 40 ps resolution over an observation window of 200 ns and stored for post-processing. This was done

for CB and CE configuration with three values of input resistance, i.e. 10 Ω, 18 Ω, 27 Ω.

Using a MATLAB® script developed on purpose, the waveforms were parsed in search for 'golden' ones and the following procedure has been applied:

1. A loose low-pass filter with 800 MHz roll-off frequency has been applied to the measured waveform to get rid of the white noise superimposed, without significantly changing the slope of the signal.

2. Waveforms with either one or more other peaks preceding the pulse that has caused triggering were discarded to prevent from signal degradation caused by baseline fluctuation and pile-up

3. The derivative has been obtained and the time when the derivative reaches the peak magnitude has been stored for each sample waveform; furthermore, an average occurrence time has been calculated. Subsequently, taking that average time as reference, all the waveforms have been time shifted of an amount equal to the residual difference between the reference time and the individual peak time of maximum slope. Finally, all the waveforms have been realigned to virtual "trigger" reference.

The results are now illustrated.

4.3.2   DATA ANALYSIS AND DISCUSSION

Fig. 1-6 and 1-7 show the plots of output pulses from the simulations and from the experimental tests, for both CE and CB configurations.
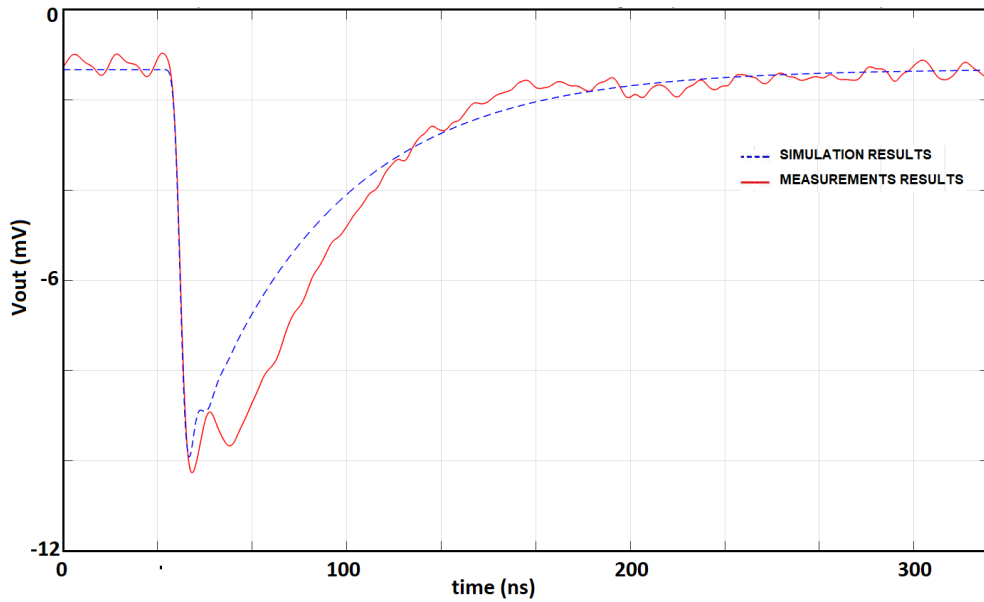

Fig. 4-6: Comparison between simulated and measured waveforms in CE configuration (R_{IN} = 27 Ω, I_{bias} = 0.91 mA).
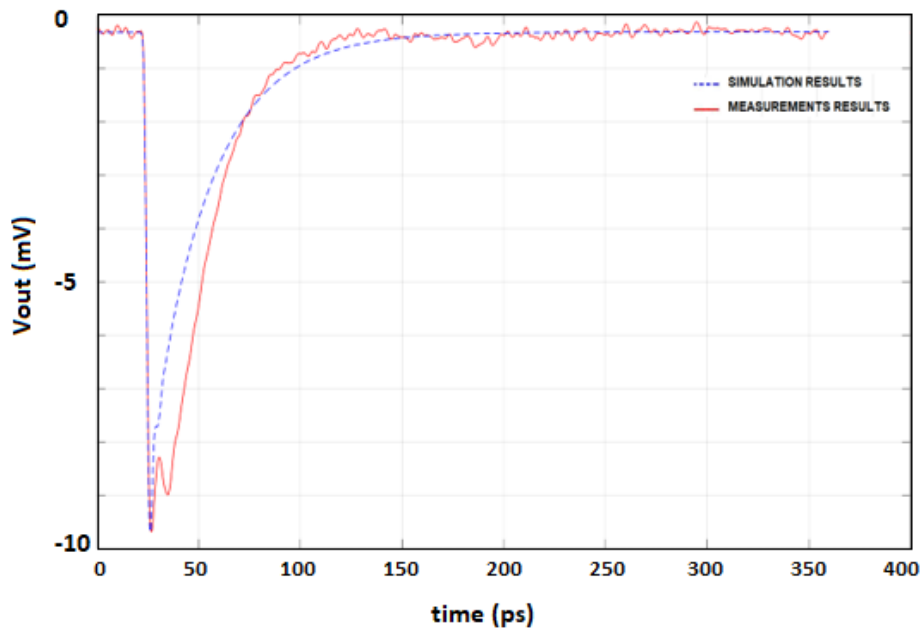

Fig. 4-7: Comparison between simulated and measured waveforms in CB configuration (R_{IN} = 27 Ω, I_{bias} = 0.91 mA).

By comparison, it is apparent that the matching between simulations and measurements is excellent for both current and voltage mode, especially as far as

173

the leading edge of the acquired pulses is concerned, which is the most interesting part of the waveforms, for the intended purpose of the test. Considering that the simulations have been carried out with a series inductance of 10 nH it can be inferred, from the good matching between simulated and measured curves in Fig. 4-6 and, that the initial estimate for L_{par} was correct.

The jitter has been plotted as a function of input resistance for the CB configuration, and the curves are shown in Fig. 4-8.
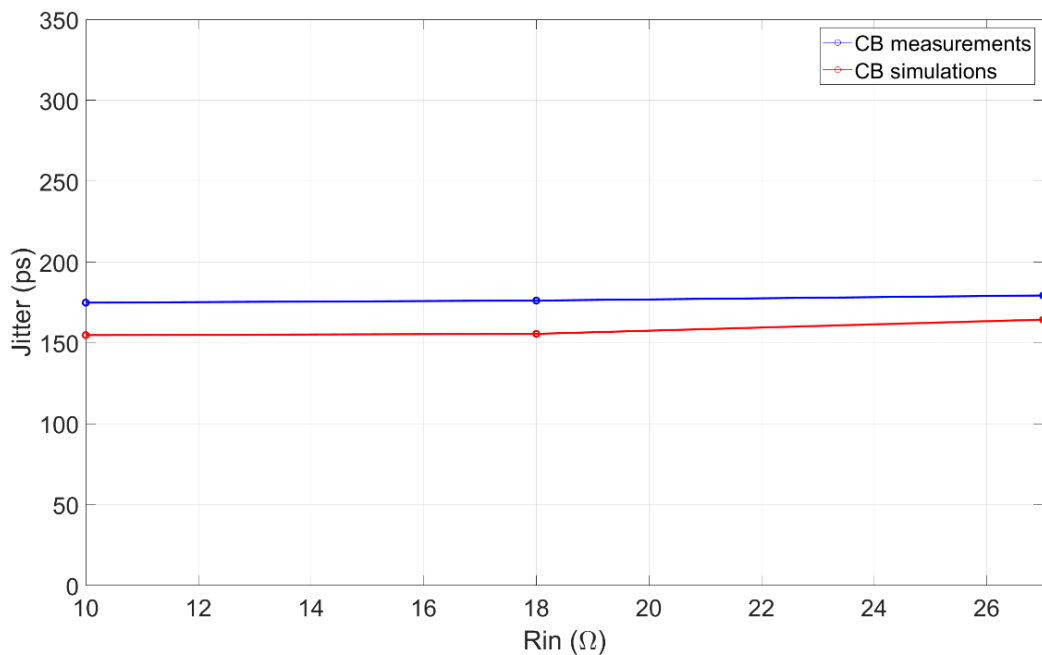

Fig. 4-8: Jitter as a function of input resistor in CB configuration.

The numerical data of the simulations and the results of measurements for CB configuration are reported in Table 4-2.

Table 4-2: Measured and simulated data for the CB configuration.

R_{IN} [Ω]	Noise rms [µV]		Slope [V/µs]		Jitter [ps]	
	Simulated	Measured	Simulated	Measured	Simulated	Measured
10	911	1034	5,9	5,9	150	**175**
18	750	818	4,8	4,6	156	**176**
27	663	700	4,0	3,9	164	**179**

The jitter has been plotted as a function of the input resistance also for the CE configuration, and the curves are shown in Fig. 4-9.
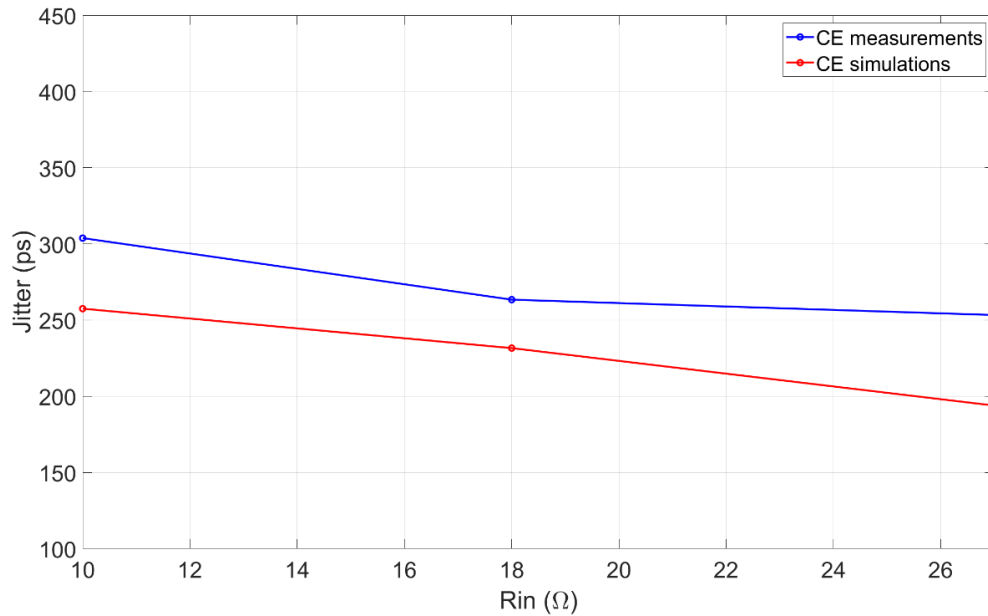

Fig. 4-9: Jitter as a function of input resistor in CE configuration.

The numerical data of the simulations and the results of measurements for CE configuration are reported in Table 4-3: Measured and simulated data for the CE configuration.

Table 4-3: Measured and simulated data for the CE configuration.

R$_{IN}$ [Ω]	Noise rms [μV]		Slope [V/μs]		Jitter [ps]	
	Simulated	Measured	Simulated	Measured	Simulated	Measured
10	1455	1647	5,6	5,4	257	**304**
18	1028	1148	4,4	4,4	232	**263**
27	688	889	3,5	3,5	194	**252**

The main difference that has been observed is the level of the output noise. It is higher for the CE than for the CB, and this makes the jitter to be higher for the Common Emitter configuration no matter what input resistance is chosen.

4.4 EXPERIMENTAL TESTS WITH A LASER LIGHT SOURCE

Many experimental tests have been carried out on the PCB module using a laser source available at Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Bari. A photograph of the apparatus is shown in Fig. 4-10.
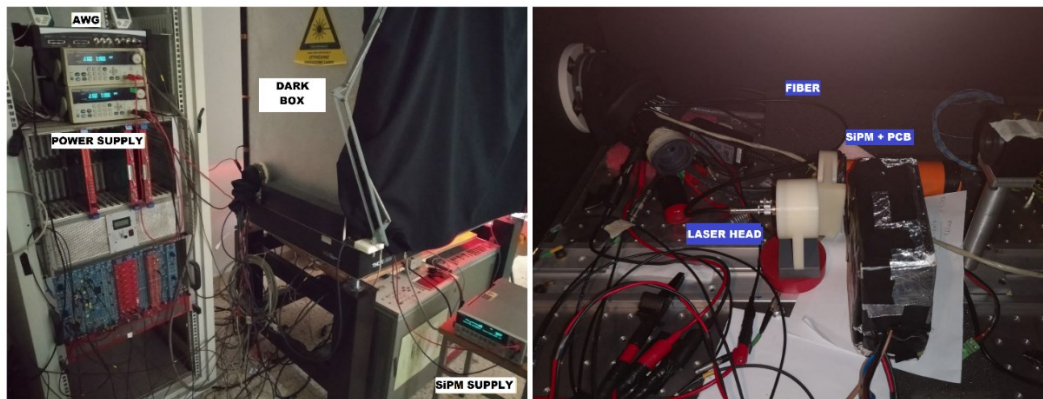

Fig. 4-10: Photos of the apparatus used for the test (on the left) and a typical arrangement inside the dark box (on the right).

The aim of the experimental tests using the laser facility is to demonstrate on a statistical basis that the model equations that have been developed throughout the present study are a good approach to predict the behavior of a real readout system as it would allow for fast simulations with acceptable accuracy.

This proves particularly useful to choose the most suitable front-end architecture for SiPM detectors, since the performance of the whole detection system, especially in terms of dynamic behavior and timing resolution, is approximately ruled by simple equations that account for the detector parameters, the main characteristics of the coupled electronics and the parasitic interconnections.

Both the light source employed for the test and the PCB module that hosts the SiPM detector and the front-end electronics have been deployed inside a 1 m^3 volume dark box that replicates an optically isolated environment thus minimizing the impact of external light sources on the detectors.

The dark box allows signal and power cables to run into it to power the electronic stuff and read the signals out to the oscilloscope. The PCB with the detector mounted on it has also been encapsulated in a smaller dark box covered with a

black blanket and aluminium sheets with just a tiny aperture left to enlighten the SiPM. It also acts as a Faraday cage to shield electronics against the effects of electromagnetic noise.

4.4.1 MEASUREMENT SETUP AND METHODS

Fig. 4-11 shows a schematic representation of the acquisition system used during the test of the PCB module.
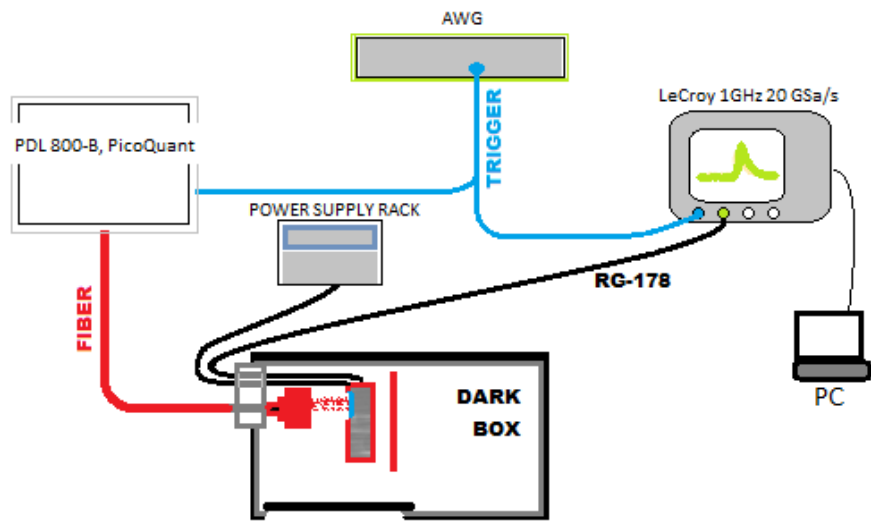

Fig. 4-11: Schematic diagram of the acquisition system.

It consists of the PCB module itself with mounted SiPM, a *Keithley 2400 SourceMeter* to bias the SiPM, a Picosecond laser emitting light with a wavelength of λ ≈ 380 nm operating in pulse mode and controlled by a Pulse Diode Laser Driver *PDL 800-B, PicoQuant*. The output responses are acquired and digitized with a *Teledyne-LeCroy WaveRunner* oscilloscope featuring 1 GHz analog bandwidth and 20 GSa/s; therefore, they are stored on the PC in *.txt* format files for subsequent processing. The light pulses have a width less than 50 ps, which is much shorter than the long recovery time of the SiPM of ≈ 40 ns. To reduce considerably the amount of light that reaches the SiPM, making negligible the probability that more than one photon hits the same microcell and increasing the probability of single photon detection, a diffuser is installed in front of the laser head. To avoid heavy contributions from dark pulses, produced by the SiPM with relatively high rate at

room temperature, which can corrupt the signals associated to the laser pulses, the measurements have been carried out exploiting coincidence. This has been simply achieved using a waveform generator that triggers both the oscilloscope and the laser driver. This guarantees that only the SiPM pulses in time coincidence with the laser light are acquired by the oscilloscope, thus rejecting the undesirable effect of dark pulses. The *LeCroy ArbStudio 1104* arbitrary waveform generator is used to generate a 1 kHz square wave that synchronizes both the PDL and the oscilloscope, with a calibrated delay useful for compensating the different signal paths.

The layout of the PCB has been reviewed and optimized to increase the number of different electrical configurations under test. To scale up and down the input resistance of both CB and CE amplifier, a set of as many as four different input resistances has been considered, while keeping one of the following series inductances that have been soldered in turn on the PCB:

- L = 0 nH (R = 0 Ω)
- L = 51 nH
- L = 100 nH

The input resistances have been varied, either by changing, on the field, the component on the PCB (Common Emitter) or by changing the DC emitter current (Common Base); this has been accomplished using the appropriate biasing emitter resistor R_E to get the required incremental input resistance (Common Base), as shown in Table 4-4.

Table 4-4: Input resistances for different electric configurations of the amplifier board.

Biasing Resistance R_E	I_E (V_T = 26 mV)	CB Input resistance R_{IN} (typical)	CE Input resistance R_{IN} (nominal)
3 kΩ	2.59 mA	10.0 Ω	10 Ω
5.6 kΩ	1.45 mA	17.9 Ω	18 Ω
10 kΩ	0.79 mA	32.9 Ω	33 Ω
15 kΩ	0.51 mA	51.0	51 Ω

Compared to the first prototype, the hardware has undergone few changes to tweak the electric coupling of the module output with the 50 Ω input impedance of the oscilloscope. The amplifier output of the first prototype was indeed unbuffered; it carried an output RC low pass filter, being terminated with the series resistor R = 100 Ω and a capacitor C = 2.2 pF to ground. Once directly connected to the oscilloscope, the amplitude and shape of high frequency output pulses appeared degraded, due to attenuation and signal reflections. To improve the matching, R has been lifted off and replaced with a 0 Ω resistor in series with a decoupling capacitor of 1 μF. Doing like that, the final op-amp amplification stage can be schematized as a voltage signal source with a low output impedance that is better suited to drive the 50 Ω input impedance of the oscilloscope.

A 3 m long, 50 Ω coaxial cable covers the distance from the dark box containing the electronic module which hosts the SiPM to the oscilloscope (see Fig. 4-10). It would rarely represent a real issue for a typical impedance-matched system; on the contrary, the performance of the system under test has been affected by non-ideal cable dispersive effects. An accurate six poles and five zeroes SPICE model of the cable from *Maxime Integrated* Application Note 5141 has been used to avoid that pitfall by accounting it in the comprehensive simulation model of the system.

AC simulation shows that a system bandwidth limitation down to 75 MHz has occurred along with few decibel units of gain loss. It has certainly come about in the real system so far as measured data are consistent with simulation model.

To summarize the final performance of the circuit, its frequency response is adequate to guarantee that any output signal in response to a single photon with risetime down to 4.5 ns is getting neither attenuated nor distorted.

Following the technical datasheet of Hamamatsu, all the measurements have been carried out with two SiPMs reverse-biased at 71.8 V, that work separately for either the CB and the CE configuration with 1.5 V overvoltage, thus providing an average delivered charge Q equal to 120 fC for the single photon detection.

4.4.1.1 WAVEFORM ACQUISITION PROCEDURE

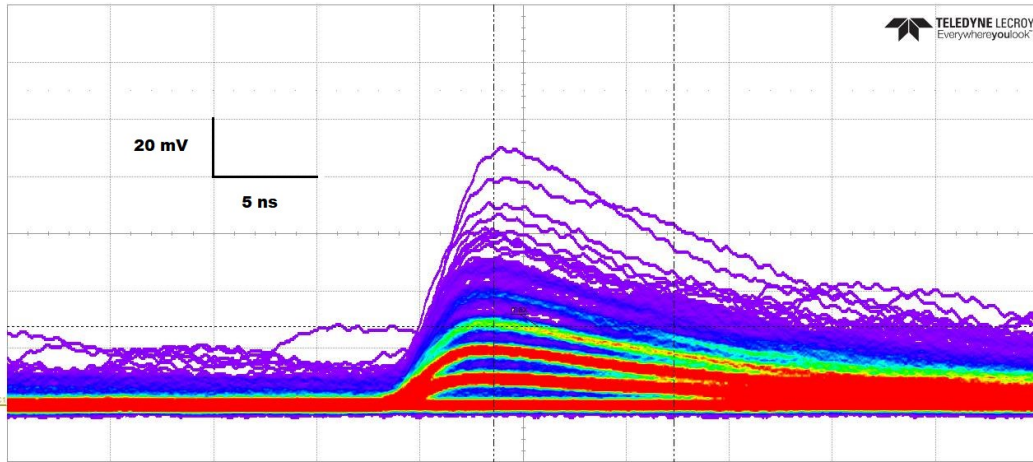The discrete nature of the acquired signals is apparent in the snapshot of Fig. 4-12.

Fig. 4-12: Snapshot of the oscilloscope screen for photon illumination of SiPM with persistence ON.

As a first step in data analytics, visualization is employed, this being a simple yet powerful way to evaluate correlations and retrieve data with a small number of descriptors (height and position). The histogram in Fig. 4-13 represents the typical amplitude distribution of the acquired signals. The peak around the origin is the electronic noise of the data acquisition system (pedestal).
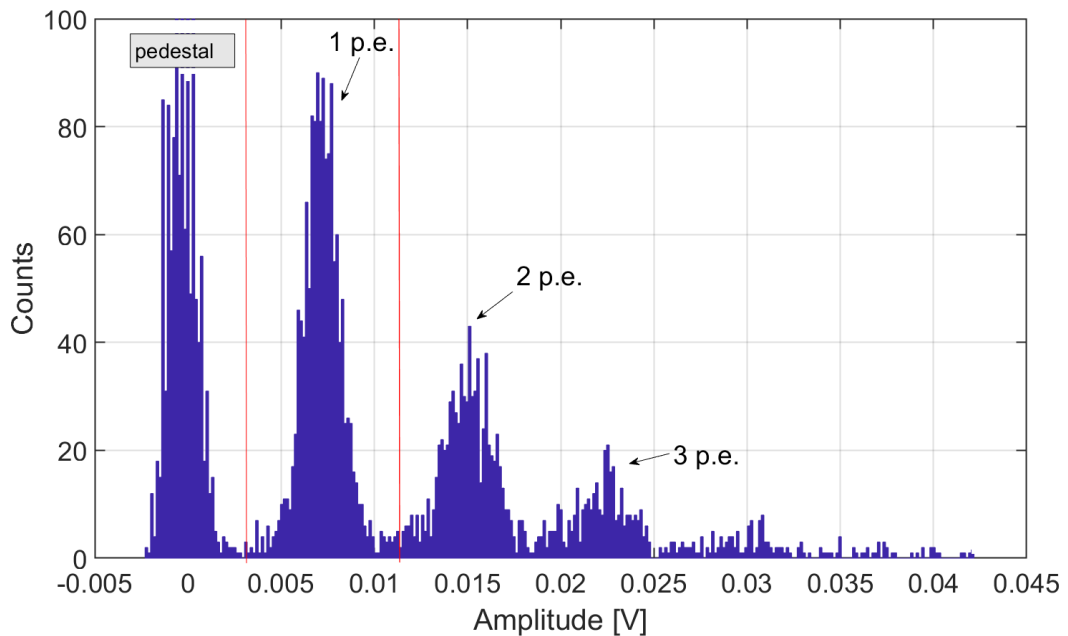

Fig. 4-13: A histogram of the amplitude of the front-end response when the SiPM is illuminated using low levels of light, as in the experimental test.

The other peaks represent the height of the front-end response when *n* photoelectrons (in Fig. 4-13 the events corresponding to *n*=1,2,3 can be easily distinguished) are simultaneously detected by the SiPM. Pulses generated by a single photon impinging on the SiPM are typically identified to lay between a specific range of amplitudes whose endpoints are used as thresholds to instruct the software that will select the corresponding traces. Everything that doesn't match the requirements is marked as outlier and discarded.

OUTLIER SCREENING STEPS:

1. Remove higher amplitudes waveforms (e.g. 2 p.e., 3 p.e.) and noisy signals (no photon detected) after setting the optimal thresholds based on the histogram;

2. Convert the primary data set to a filtered primary data subset; in practice, this is usually achieved by discarding the samples that present at least one more pulse in the time interval that precedes the triggering point, and thus stabilizing the baseline before the onset of the signal.

Signal acquisition has been carried out for each of the 24 circuit arrangements, 12 for the Common Base and 12 for the Common Emitter, stemming from all the possibilities to combine 4 input resistances with 3 inductances.

Each primary dataset consists of at least 4,000 waveforms. After the primary dataset is filtered, around 800 waveforms are left, that represent the single photon pulses, and can be used for processing.

From a statistical point of view, those numbers cannot be referred to as quantitative data; indeed, the common sense suggests that taken datasets are far from having the right sample size and that, just like the mean, the standard deviation can be sometime deceptive if taken alone. For example, if the data have many outliers, then the standard deviation doesn't provide the accuracy you need for a nuanced decision. This problem may probably have affected the accuracy of time jitter calculation, although the size of the taken datasets and the statistical tools used to process them in the present work have answered to the questions

of the research topic giving meaningful information that will be discussed before long.

The resulting waveforms of a primary dataset after the first two screening steps are presented in Fig. 4-14.
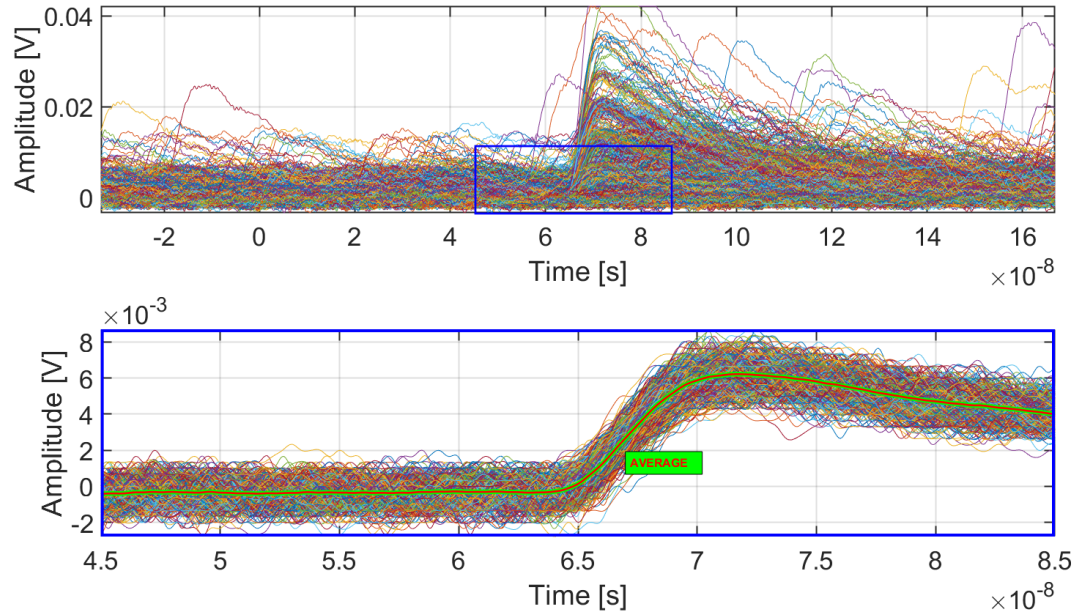

Fig. 4-14: A bunch of four thousand waveforms (top graph) digitized and the selected eight hundred single pulse samples with the superimposed average golden pulse in green.

The averaged green-coloured pulse shown in the picture above represents the golden reference of a single photon response of ideally noiseless readout system. It has been obtained by averaging all the waveforms together and its profile has been kept for further off-line analysis. The baseline is very stable around 0 V thanks to the DC blocking of the output series capacitor. The standard deviation of the electronic noise, σ_n, is calculated as the standard deviation of a Gaussian fit to the values of all the selected waveforms at a given time ahead of the pulse onsets.

4.4.2 DATA ANALYSIS AND DISCUSSION

Before starting any discussion about the PCB readout performance involving either configuration of the preamplifier (Common Base and Common Emitter), it is worthy to notice that the measurement conditions may have been changing over the day time required to accomplish the task; for example, neither the

temperature has been accurately and actively controlled nor the biasing currents and SiPM parameters with a direct dependence on it have been strictly monitored. However, the risk has been carefully assessed and the expected deviation of results from ideal conditions has been deemed not large enough to compromise their validity; besides, it has been paid off with a simple measurement setup.

Furthermore, intrinsic physical differences between ideally identical components along with a poor modelling of the active devices (BJTs) under peculiar biasing conditions may have been responsible for spot model unfitting, especially in the Common Emitter configuration, where the complex parameters involved in the determination of the voltage gain and bandwidth can be easily affected by non-idealities.

4.4.2.1 COMMON BASE CONFIGURATION

A typical set of 1 p.e. signals amplified and shaped by the PCB front-end measured with the described experimental setup and off-line processed is shown in Fig. 4-15 for the Common Base configuration with L = 0 nH (no intentional series inductance soldered but a 0 Ω resistor).
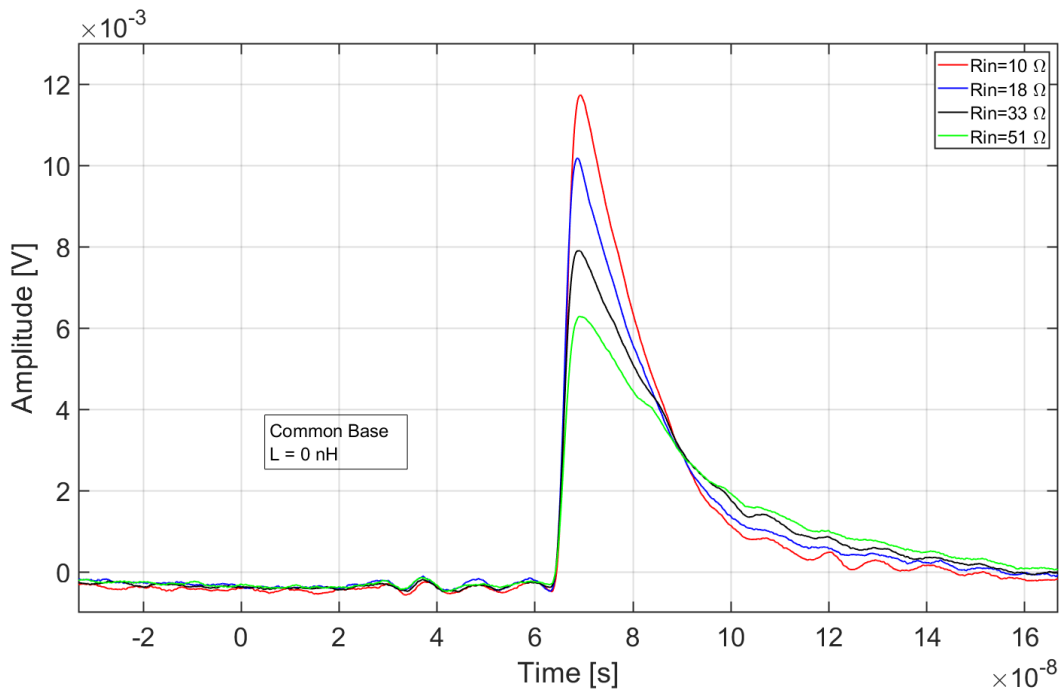

Fig. 4-15: Comparison between the average responses to pulsed laser light for each of the four Common Base configurations obtained with L = 0 nH (0 Ω resistor) and different values of R_{in}.

After soldering the inductor (in this case the 0 Ω resistor) and keeping it on, the input resistance (incremental emitter resistance) has been varied in discrete non-uniform steps, by changing the DC emitter current of the preamplifier with appropriate choice of the physical emitter resistor, as in Table 4-4. The real signals, that have been inverted for the analysis of their timing characteristics, are very close to the transient simulation carried out in CADENCE.

The comparison between the output pulses for all the combinations of R_{in} and L for the Common Base configuration is shown in Fig. 4-16.
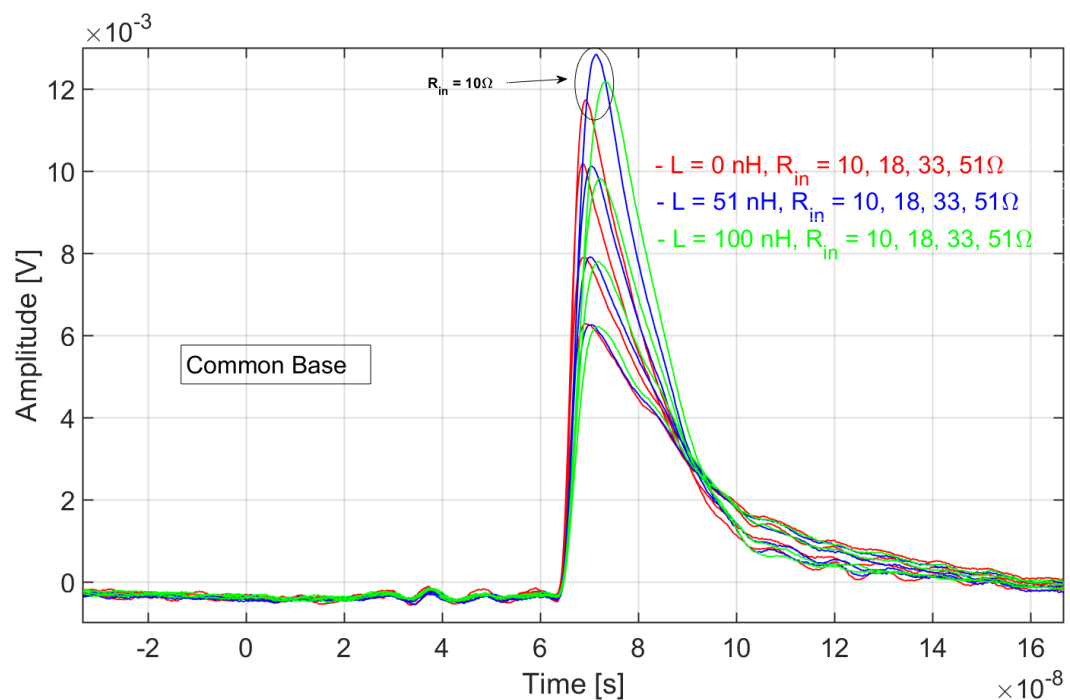

Fig. 4-16: Comparison between the average output pulses for all the combinations of R_{in} and L.

All the golden waveforms of the CB configuration can be sorted out with an eye-check: with a fixed inductance, the lower the resistance, the higher the amplitude of the baseline-to-peak amplitude; moreover, the lower the inductance the higher the peak and the steeper the rising edge, except for L = 0 nH and R_{in} = 10 Ω. In the latter and unique case (for CB configuration) the peak is lower than expected while the slope is the steepest, as expected. This behavior hasn't been further investigated. The derivative has been calculated in MATLAB® for all the golden pulses and the maximum of the resulting functions has been calculated to show

184

the trendline of the maximum slope as a function of the input resistance for different inductances. Fig. 4-17 shows an example derivative of the 1 p.e. output.
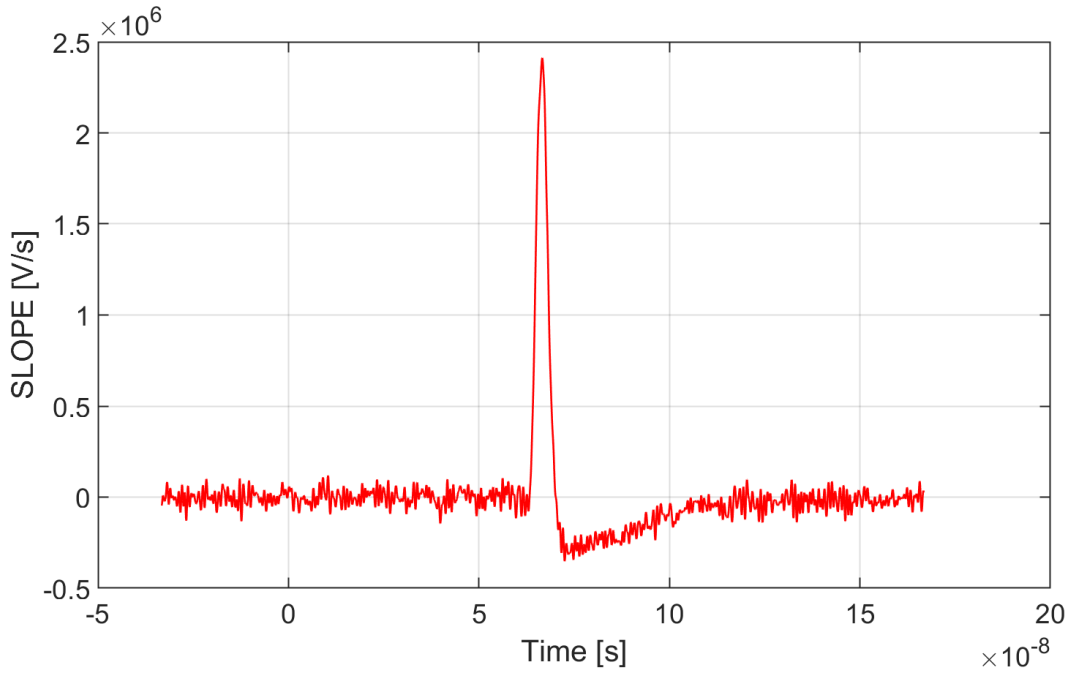

Fig. 4-17: An example of the first order derivative of the output 'golden' pulse.

In accordance with the Leading-Edge Discrimination (LED) technique for time pickoff, the peaking time in Fig. 4-17 is normally brought back onto the temporal trace of the output pulse to spot the point where the threshold should be set thus minimizing the time jitter. The standard deviation σ_n of the electronic noise distribution has also been calculated from measurements. The σ_n values are lower than 1 mV rms, as shown in Table 4-5.

Table 4-5: Electronic noise as a function of R_{in} and L for CB configuration.

COMMON BASE	Electronic Noise (mV$_{rms}$)		
R_{IN} (Ω)	L = 0 nH	L = 51 nH	L = 100 nH
10.0	0.98	0.82	0.76
17.9	0.84	0.74	0.67
32.9	0.76	0.66	0.68
51.0	0.69	0.64	0.68

Fig. 4-18 shows the maximum values of the slopes for all the pulses that have indirectly been retrieved from the measurements, as a function of R_{in} and L. Their fitting curves are in red colour. Both the simulation results and the approximate model curves have been reported on the same graph.
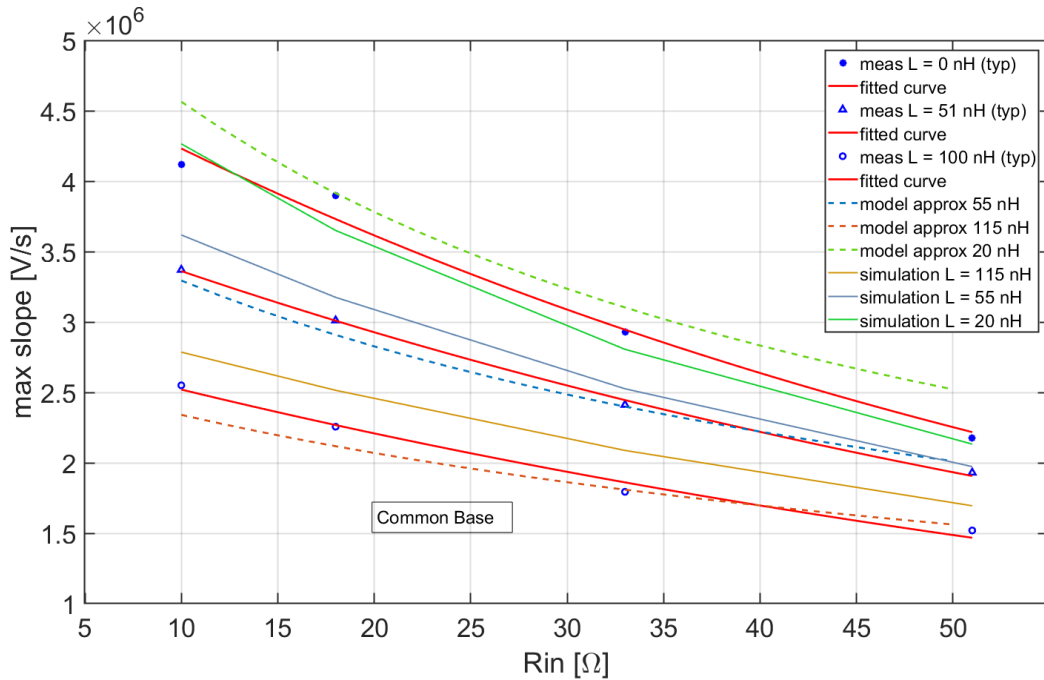

Fig. 4-18: The maximum slopes of the CB SiPM readout output, predicted by the model and simulated, compared with the experimental data results, represented by symbols and red lines.

Either model and simulation seem to be quite close to measurements, but, as previously pointed out, the case with L = 0 nH and Rin = 10 Ω represents an exception. In fact, the signal slope doesn't increase at the rate predicted by the model, probably due to a complex interaction of some parasitic components with a performance drop of the transistor at higher rate of biasing current. Additionally, to fit the measurements when L = 0 nH, both the approximate model equation and the SPICE model require L = 20 nH; this can be explained with the presence of a parasitic series inductance on the board associated with the 0 Ω series resistor.

It is apparent from the curves that, whatever the series inductance, while the input resistance R_{in} of the CB amplifier increases, the maximum of the slope decreases. This proceeds against intuition because the time constant $\tau_A = L_{par}/(R_{in}+R_{par})$ in the approximate model presented in Chapter 2, becomes smaller when R_{in} increases

and, consequently, the system should be faster as its bandwidth is supposed to get wider. This is likely to happen, so far: the pole is truly pushed away to higher frequencies because the higher R_{in} the faster the pulse attains its peak (see Fig. 2-19 and Fig. 4-15); however, due to the fact that the amplitude of the current signal that determines the shape of the rising edge is concurrently getting smaller and considering that this occurs even more rapidly than the system grows faster, its slope diminishes on the whole. The same effect can be observed when it comes to changing the series inductance. In this case, the peak time increases, as apparent, for example, in Fig. 4-16 because the system gets slower and the baseline-to-peak signal magnitude decreases; therefore, the slope decreases even further.

4.4.2.2   COMMON EMITTER CONFIGURATION

A typical set of 1 p.e. signals amplified and shaped by the PCB front-end measured exploiting the previously described experimental setup and off-line processed is shown in Fig. 4-19 for the Common Emitter configuration with L = 100 nH.

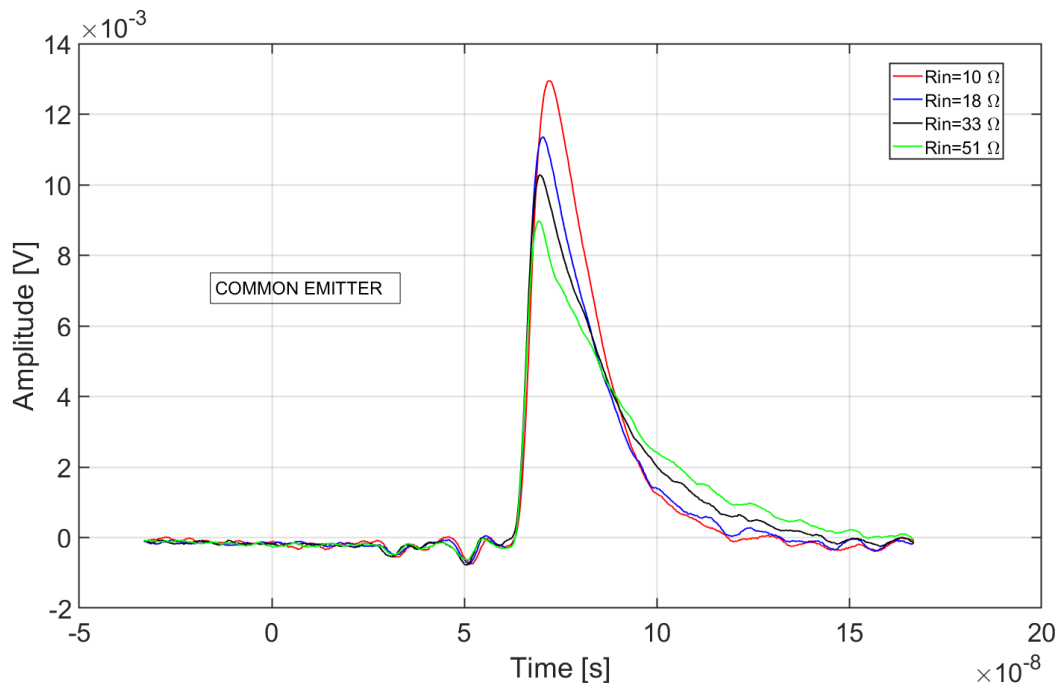*Fig. 4-19: Comparison between the average responses to pulsed laser light for each of the four Common Emitter configurations obtained with L = 100 nH and different values of R_{in}.*

It is even more apparent, compared to the situation illustrated for the CB, that in this case the peaking time of the curves is moving left while R_{in} increases. This represents a further proof that time constant $\tau_A = L_{par}/(R_{in}+R_{par})$ is suited to describe the dynamic behavior of the system in the approximate model presented in Chapter 2, despite the configuration (either CB or CE). Nonetheless, the trend of the signal heights as a function of R_{in} requires further explanations. When dealing with a typical voltage-mode readout system, the current signal of the SiPM is converted into a voltage signal across the input resistance that is amplified with a fixed gain voltage amplifier. Once the voltage gain is given, it is expected that the baseline-to-peak magnitude of the output pulse increases when the input resistance is increased: this behavior has already been confirmed in Chapter 2 by the model (see Fig. 2-19). Eventually, the Common Emitter configuration under test doesn't behave the same way, so far. In fact, to obtain the same BJT operating point and power consumption of the corresponding CB versions of the circuit, when R_{in} is increased, at the same time the transconductance g_m is intentionally scaled down of the same amount, by changing the DC emitter current, thus the voltage gain decreases by the same amount and the baseline-to-peak magnitude of the output pulses decreases as well. Therefore, in these conditions, the Common Emitter behaves as a Common Base is expected to do and this likelihood is also well supported by the simulation results of the slopes shown in Fig. 4-20.
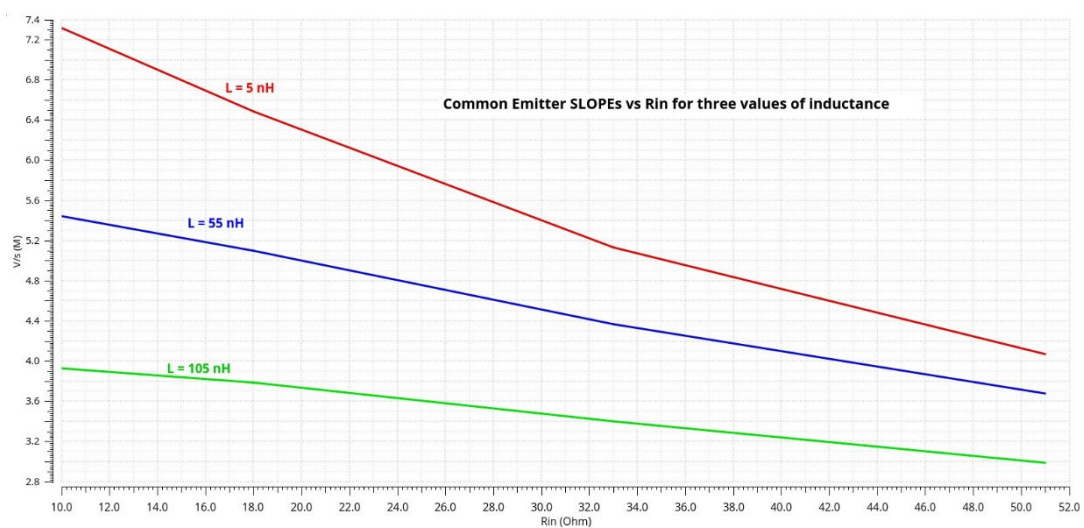

Fig. 4-20: Comparison of slopes in simulation for the Common Emitter configuration.

Fig. 4-21 shows the maximum values of the slopes of all the pulses that have indirectly been retrieved from the measurements, as a function of R_{in} and L. Data fitting curves are in a light blue colour while the approximate model curves in red.
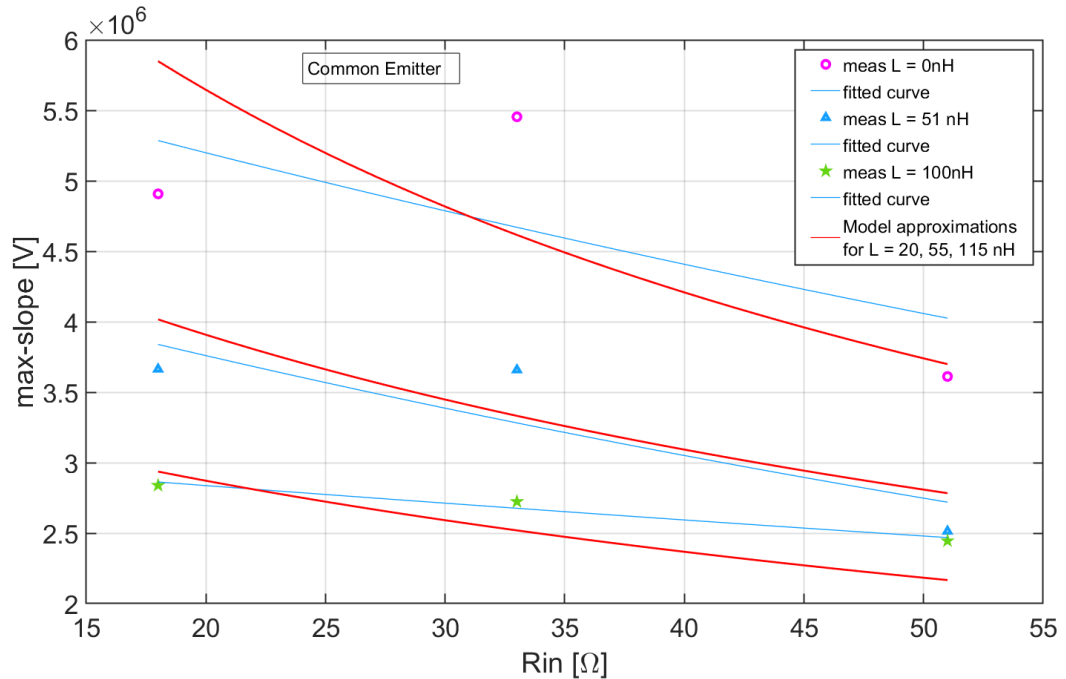

Fig. 4-21: The slopes of the CE SiPM readout output, predicted by the model, compared with the experimental data results represented by symbols and light blue lines.

After looking at the results of both model and simulations, larger values of slope were expected from measurement data points corresponding to low R_{in}. Therefore, the data points corresponding to $R_{in} = 10\ \Omega$ have been discarded.

Despite their exclusion, data dispersion is quite large, though the data trend is still close to the trend predicted by the approximate model (and simulations too). An explanation can be found in a puzzling effect due to combination of poor SPICE modelling of the discrete bipolar transistor at low input resistance and large DC current level and parasitic components on the board.

The standard deviation σ_n of the electronic noise distribution has also been calculated from measurements. The σ_n values are larger than the corresponding values found for the Common Base configuration and quite larger than 1 mV *rms*, as shown in Table 4-6.

Table 4-6: Electronic noise as a function of R_{in} and L for CE configuration

COMMON EMITTER	Electronic Noise (mV$_{rms}$)		
R$_{IN}$ (Ω)	L = 0 nH	L = 51 nH	L = 100 nH
18	1.39	1.18	1.16
33	1.32	1.02	1.03
51	1.04	0.99	0.90

4.4.3 TIMING CAPABILITIES

Time jitter is one of the key parameters that describe the timing capabilities of any detection system composed by a SiPM coupled to the front-end electronics. Timing jitter is also defined as the deviation of a signal's timing delay from the ideal timing reference. To make a comparison between CB and CE configuration and explain the differences, the jitter has been evaluated by means of delay measurements, with the assumption that each golden pulse, derived for each circuit configuration, can be used as the ideal timing reference for the corresponding dataset. For this purpose, the peaking times of the derivatives of all the golden pulses have been calculated and the associated amplitude levels and their time of occurrence have been evaluated.

Eventually, each amplitude value has been used as a threshold for the corresponding set of pulses and applied on the associated single photon filtered dataset to calculate the arrival delays of all the acquired waveforms.

Moreover, for each dataset, the arrival delays have been worked out as in Fig. 4-22, sorted in a histogram and a Gaussian fit has been considered. The standard deviation of the Gaussian fit represents the time jitter.

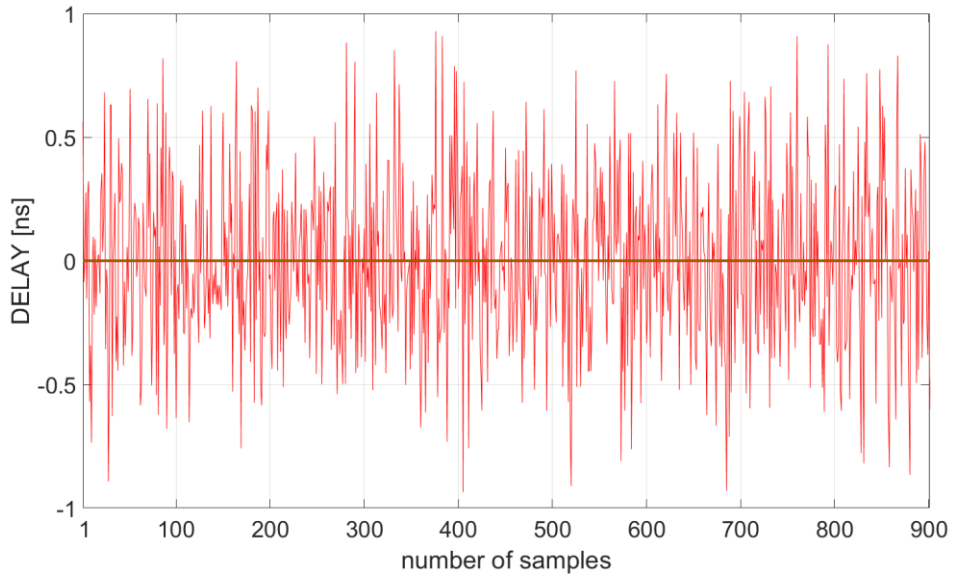*Fig. 4-22: Diagram of the arrival delays. The delay of the golden pulse is equal to zero.*

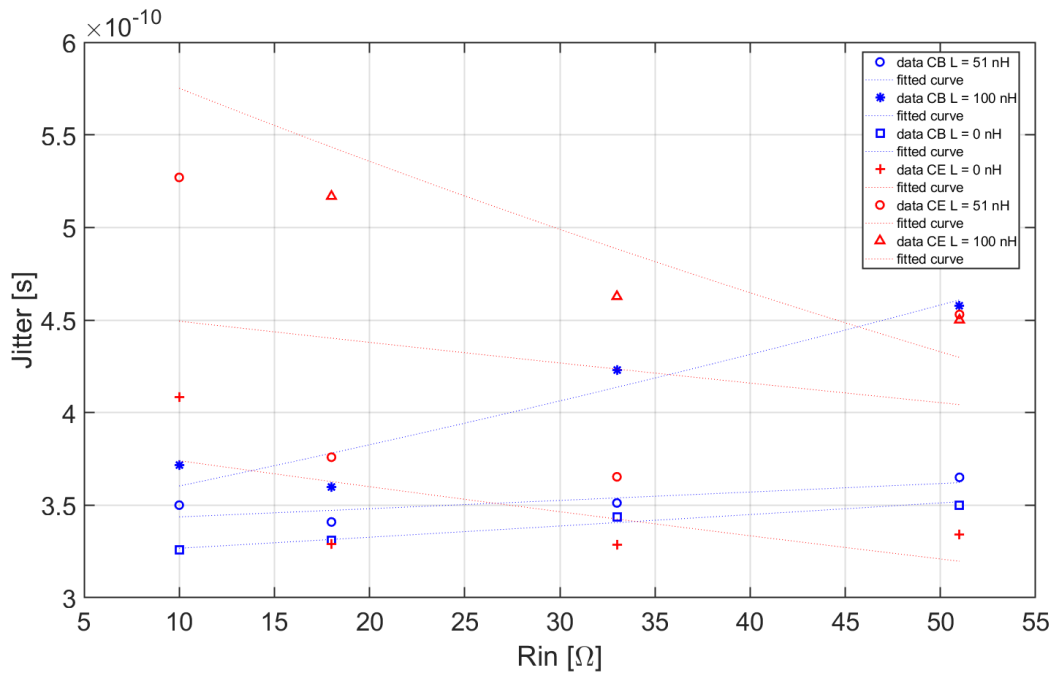Fig. 4-23 shows the jitter variation with input resistance and series inductance for both CB and CE configuration.

Fig. 4-23: Jitter calculated with delay method as a function of Rin and L for CB (in red) and CE.

The light dotted curves (barely visible) are exponential fittings of the first order.

Despite the accuracy of the statistical results depend on the size of dataset and the presence of the outliers may have contributed to increase the uncertainty of the measurements (in step with the decrease of Signal-to Noise Ratio), it is apparent that the jitter is globally lower for the CB configuration than for the CE. As far as low input resistances are selected, all the CB jitter curves, irrespective of the value of the inductance, will lay below the best performing CE jitter curve (obtained with L = 0 nH).

Both configurations have been compared on equity terms of power dissipation and input resistances. It has been demonstrated that the current-mode approach is the preferred solution as it can guarantee the best timing performance in the low and medium range of input resistance, whatever the parasitic series inductance.

4.4.4 COMMENTS

Aligned with the thesis that has been supported throughout the present work, it is confirmed that the jitter, as a function of the input resistance, changes according to the circuit configuration.

The results of the evaluation test have demonstrated that the measured jitter is higher in CE configuration than in CB configuration.

With respect to the CE configuration, the noise dominates the slope in the determination of the jitter trendline as a function of the input resistance. Moreover, the slope decreases when the input resistance increases (Table 4-3). This may seem to disagree with Eq. (32) of Chapter 2, although it is just another confirmation of that theory. Indeed, when the input resistance was changed during test, the DC biasing current of the CE was varied accordingly; to operate the amplifiers under the same power conditions, the current was trimmed to the value that obtained the same input resistance in the CB counterpart. Consequently, the higher the input resistance the lower the polarization current and the voltage gain. Therefore, when R_{in} is increased the system is faster but the voltage gain decreases to such an extent that the signal slope diminishes.

On the other hand, the slope of the CB slightly decreases as the input resistance increases and this is coherent with what has been theoretically argued; furthermore, the noise is lower and almost constant in CB configuration rather than in CE, thus giving a slowly varying jitter over the input resistance in the former case.

Eventually, even including the effects of parasitic inductance, the current mode solution is preferred because the jitter is always lower with a lower input resistance; moreover, higher event rates are sustained without suffering excessive degradations of the timing performance caused by the large time constant normally associated to the tail of the SiPM signal.

5 CONCLUSIONS

The present work demonstrates the validity of a systematic approach to choose and design high performance analog front-ends for fast timing using the SiPMs.

SiPM is best candidate to read out fast and weak light signals, down to the single-photon level thanks to their high photodetector gain, excellent timing resolution and single photon-counting capability. Nonetheless, when coupled with readout electronics, their characteristics can be impaired because of noise, bandwidth limitation and parasitic components that can have a detrimental effect on timing performance and energy resolution of the detection system on a whole.

The development of an analog front-end for the readout of SiPMs in a standard CMOS 130 nm process from TSMC has been presented. An outstanding timing resolution of 33 ps *rms* for the detection of single-photon events and a maximum dynamic range from 1 to 8,000 photoelectrons have been achieved and validated with post-layout simulations, thanks to the deployment of innovative circuit techniques and design guidelines that have even been explained in detail.

A front-end architecture based on a current buffer as input stage is suited to applications demanding accurate timing, as proven by design and simulation.

Pending the starting date for the fabrication of the chip prototype on Silicon, extensive measurements have been carried out on a test preamplifier coupled to a commercial SiPM from Hamamatsu. It has been implemented and mounted on a printed circuit board using discrete components with different configurations to validate the results of the theoretical analysis.

Lastly, the experimental tests on the PCB module with the SiPM detector excited by a pulsed laser have confirmed its functionality in accordance with the analytical expressions that has been devised throughout the present dissertation; its validity is expected to be fully acknowledged by the results of the evaluation tests that will eventually be carried out on the forthcoming chip prototype on Silicon.

Bibliography

References CHAPTER 1

[1] Einstein and our world. (1995). *Choice Reviews Online*, 33(01), pp.33-0277-33-0277.
doi:10.5860/choice.33-0277

[2] G.F. Knoll, *Radiation Detection and Measurements.* John Wiley and Sons, New York, 2012.

[3] G. Lutz, *Semiconductor Radiation Detectors.* Springer, Berlin, 2007.

[4] Chunnilall, C. J., Degiovanni I. P., Kück S., Müller I., Sinclair A. G. (2014). Metrology of single-photon sources and detectors: a review. *Optical Engineering*, 53(8), p.081910.
doi: 10.1117/1.OE.53.8.081910

[5] Derenzo, S., Choong, W. and Moses, W. (2014). Fundamental limits of scintillation detector timing precision. *Physics in Medicine and Biology*, 59(13), pp.3261-3286.
doi: 10.1088/0031-9155/59/13/3261

[6] Seifert, S., van Dam, H., Vinke, R., Dendooven, P., Lohner, H., Beekman, F. and Schaart, D. (2012). A Comprehensive Model to Predict the Timing Resolution of SiPM-Based Scintillation Detectors: Theory and Experimental Validation. *IEEE Transactions on Nuclear Science*, 59(1), pp.190-204.
doi: 10.1109/tns.2011.2179314

[7] Piemonte, C., Acerbi, F., Ferri, A., Gola, A., Paternoster, G., Regazzoni, V., Zappala, G. and Zorzi, N. (2016). Performance of NUV-HD Silicon Photomultiplier Technology. *IEEE Transactions on Electron Devices*, 63(3), pp.1111-1116.
doi: 10.1109/ted.2016.2516641

[8] Acerbi, F., Paternoster, G., Gola, A., Regazzoni, V., Zorzi, N. and Piemonte, C. (2018). High-Density Silicon Photomultipliers: Performance and Linearity Evaluation for High Efficiency and Dynamic-Range Applications. *IEEE Journal of Quantum Electronics*, 54(2), pp.1-7.
doi: 10.1109/jqe.2018.2802542

[9] Surti, S. and Karp, J. (2016). Advances in time-of-flight PET. *Physica Medica*, 32(1), pp.12-22.
doi: 10.1016/j.ejmp.2015.12.007

[10] Lutz, B. (2012). Upgrade of the CMS Hadron Outer Calorimeter with SiPM sensors. *Journal of Physics: Conference Series*, 404, p.012018.
doi: 10.1088/1742-6596/404/1/012018

[11] Adamo G., Busacca A., Proc. of AEIT Int. Ann. Conf. (AEIT) (2016).
doi: 10.23919/EETA.2019.8804612

[12] E. Martinenghi E., *et al.*, Rev. Sci. Instrum. 87 (2016) 073101.
doi: 10.1063/1.4954968

[13] Dolgoshein, B., *et al.*, I. (2006). Status report on silicon photomultiplier development and its applications. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 563(2), pp.368-376.
doi: 10.1016/j.nima.2006.02.193

[14] Piemonte, C. and Gola, A. (2019). Overview on the main parameters and technology of modern Silicon Photomultipliers. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 926, pp.2-15.
doi: 10.1016/j.nima.2018.11.119

[15] Acerbi, F. and Gundacker, S. (2019). Understanding and simulating SiPMs. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 926, pp.16-35.
doi: 10.1016/j.nima.2018.11.118

[16] Wilmshurst TH. Signal Recovery from Noise in Electronic Instrumentation. Bristol: Hilger; 1990.

[17] Corsi, F., Dragone, A., Marzocca, C., Del Guerra, A., Delizia, P., Dinu, N., Piemonte, C., Boscardin, M. and Dalla Betta, G. (2007). Modelling a silicon photomultiplier (SiPM) as a signal source for optimum front-end design. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 572(1), pp.416-418.

[18] Calò, P., Ciciriello, F., Petrignani, S. and Marzocca, C. (2019). SiPM readout electronics. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 926, pp.57-68.
doi: 10.1016/j.nima.2018.09.030

[19] Ciciriello, F., Corsi, F., Licciulli, F., Marzocca, C., Matarrese, G., Del Guerra, A. and Bisogni, M. (2013). Accurate modelling of SiPM detectors coupled to FE electronics for timing performance analysis. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 718, pp.331-333.
doi: 10.1016/j.nima.2012.10.103.

[20] Huizenga, J., Seifert, S., Schreuder, F., van Dam, H., Dendooven, P., Löhner, H., Vinke, R. and Schaart, D. (2012). A fast preamplifier concept for SiPM-based time-of-flight PET detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 695, pp.379-384.
doi: 10.1016/j.nima.2011.11.012

[21] Bousselham, A. and Bohm, C. (2007). Sampling Pulses for Optimal Timing. *IEEE Transactions on Nuclear Science*, 54(2), pp.320-326.
doi: 10.1109/tns.2007.892692

[22] Powolny, F., Auffray, E., Brunner, S., Garutti, E., Goettlich, M., Hillemanns, H., Jarron, P., Lecoq, P., Meyer, T., Schultz-Coulon, H., Shen, W. and Williams, M. (2011). Time-Based Readout of a Silicon Photomultiplier (SiPM) for Time of Flight Positron Emission Tomography (TOF-PET). *IEEE Transactions on Nuclear Science*, 58(3), pp.597-604.
doi: 10.1109/TNS.2011.2119493

[23] Träff, H. (1992). Novel approach to high speed CMOS current comparators. *Electronics Letters*, 28(3), p.310.
doi: 10.1049/el:19920192

[24] Ritt, S. (2004). The DRS chip: cheap waveform digitizing in the GHz range. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 518(1-2), pp.470-471.
doi: 10.1016/j.nima.2003.11.059

[25] Sarasola, I., Nemallapudi, M., Gundacker, S., Sánchez, D., Gascón, D., Rato, P., Marín, J. and Auffray, E. (2017). A comparative study of the time performance between NINO and FlexToT ASICs. *Journal of Instrumentation*, 12(04), pp.P04016-P04016.
doi: 10.1088/1748-0221/12/04/p04016

[26] Stankova, V., Shen, W., Briggl, K., Chen, H., Fischer, P., Gil, A., Harion, T., Kiworra, V., Munwes, Y., Ritzert, M. and Schultz-Coulon, H. (2015). STIC3 – Silicon Photomultiplier Timing

Chip with picosecond resolution. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 787, pp.284-287.
doi: 10.1016/j.nima.2014.12.073

[27] Munwes, Y., Briggl, K., Chen, H., Harion, T., Schultz-Coulon, H., Shen, W. and Stankova, V. (2015). Single photon time resolution with silicon photomultipliers using the STiC readout chip. *2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*.
doi: 10.1109/NSSMIC.2015.7581738

[28] Francesco, A., Bugalho, R., Oliveira, L., Pacher, L., Rivetti, A., Rolo, M., Silva, J., Silva, R. and Varela, J. (2016). TOFPET2: a high-performance ASIC for time and amplitude measurements of SiPM signals in time-of-flight applications. *Journal of Instrumentation*, 11(03), pp.C03042-C03042.
doi: 10.1088/1748-0221/11/03/C03042

[29] Fischer, P., Peric, I., Ritzert, M. and Koniczek, M. (2009). Fast Self Triggered Multi Channel Readout ASIC for Time- and Energy Measurement. *IEEE Transactions on Nuclear Science*, 56(3), pp.1153-1158.
doi: 10.1109/TNS.2008.2008807

[30] P. Calò, F. Ciciriello, F. Corsi, C. Marzocca, G. Matarrese and M. G. Bisogni (2016). BASIC64: A new mixed-signal front-end ASIC for SiPM detectors. *IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*, pp. 1-5.
doi: 10.1109/NSSMIC.2016.8069651

[31] Bisogni, M., Calò, P., Ciciriello, F., Corsi, F., Marzocca, C., Matarrese, G. and Petrignani, S. (2019). Experimental test and characterization of BASIC64, a new mixed-signal front-end ASIC for SiPM detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* 936, pp.313-315.
doi: 10.1016/j.nima.2018.10.133

References CHAPTER 2

[1] E. Lamprou et al., "Progress Report for an Accurate PET Detector Based on SiPMs and the TOFPET ASIC," *2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, Atlanta, GA, 2017, pp. 1-3, doi: 10.1109/NSSMIC.2017.8533123.

[2] G. Cozzi et al., "High-Resolution Gamma-Ray Spectroscopy With a SiPM-Based Detection Module for 1" and 2" LaBr$_3$:Ce Readout,", *IEEE Transactions on Nuclear Science*, vol. 65, no. 1, pp. 645-655, Jan. 2018, doi: 10.1109/TNS.2017.2784238

[3] W. Becker, "Overview of photon counting techniques," in *Advanced Time-Correlated Single Photon Counting Techniques*, 1st ed. Berlin, Germany, Springer-Verlag, 2005

[4] G. Adamo and A. Busacca, "Time Of Flight measurements via two LiDAR systems with SiPM and APD," 2016 *AEIT International Annual Conference (AEIT)*, Capri, 2016, pp. 1-5,

[5] P. A. P. Calò, F. Ciciriello, S. Petrignani, C. Marzocca, "SiPM readout electronics", *Nucl. Instr. and Meth. in Phys. Res. A,* vol. 926, pp. 57–68, ISSN: 0168-9002, doi: 10.1016/j.nima.2018.09.030, 2019.

[6] J. Huizenga, S. Seifert, F. Schreuder, H.T. van Dam, P. Dendooven, H. Löhner, R. Vinke, D.R. Schaart, "A fast preamplifier concept for SiPM-based time-of-flight PET detectors", *Nuclear Instruments and Methods in Physics Research Section A,* Volume 695, 2012, Pages 379-384, doi: 10.1016/j.nima.2011.11.012.

[7] F. Corsi, A. Dragone, C. Marzocca, A. Del Guerra, P. Delizia, N. Dinu, C. Piemonte, M. Boscardin, G.F. Dalla Betta, "Modelling a Silicon Photomultiplier (SiPM) as a Signal Source for Optimum Front-End Design", *Nucl. Instr. and Meth. in Phys. Res. A,* vol. A572, pp. 416-418, 2007.

[8] F. Ciciriello, F. Corsi, F. Licciulli, C. Marzocca, G. Matarrese, A. Del Guerra, M. G. Bisogni, "Designing the front-end electronics of a SiPM based gamma-ray detection system for optimal time resolution", 2012 *IEEE Nucl. Science. Symp. and Med. Imag.Conf. (NSS-MIC'12)* Conference Record, pp. 827-833, 2012, doi: 10.1109/NSSMIC.2012.6551220.

[9] F. Ciciriello, F. Corsi, F. Licciulli, C. Marzocca, G. Matarrese, A. Del Guerra, M.G. Bisogni, "Accurate Modeling of SiPM Detectors Coupled to FE Electronics for Timing Performance Analysis", Nucl. Instr. and Meth. in Phys. Res. A, vol. 718, pp. 331-333, ISSN: 0168-9002, doi: 10.1016/j.nima.2012.10.103, 2013.

[10] F. Licciulli and C. Marzocca, "Parameter Extraction Method for the Electrical Model of a Silicon Photomultiplier," *IEEE Transactions on Nuclear Science,* vol. 63, no. 5, pp. 2517-2526, Oct. 2016, doi: 10.1109/TNS.2016.259458.

[11] H. Spieler. 2005. Semiconductor Detector Systems. Oxford University Press.

[12] F. Ciciriello, F. Corsi, F. Licciulli, C. Marzocca and G. Matarrese, "Comparing front-end alternatives for SiPM's in single-photon time resolution applications," 2015 *IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, San Diego, CA, 2015, pp. 1-8. doi: 10.1109/NSSMIC.2015.7581821

[13] Seifert van Dam Huizenga Vinke Dendooven lohener Schaaart *IEEE transaction on nuclear science* NS-56 (2009) 3726

[14] Thomas F. Schubert, Jr. and Ernest M. Kim, A Short Study on the Validity of Miller's Theorem Applied to Transistor Amplifier High-Frequency Performance*" IEEE Transactions on Education,* vol. 52, no. 1 pp. 92-98, February 2009.

References CHAPTER 3

[1] A. Otte, D. Garcia, T. Nguyen and D. Purushotham, "Characterization of three high efficiency and blue sensitive silicon photomultipliers", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 846, pp. 106-125, 2017.
doi: 10.1016/j.nima.2016.09.053

[2] F. Powolny, E. Auffray, S. Brunner, E. Garutti, M. Goettlich, H. Hillemanns, P. Jarron, P. Lecoq, T. Meyer, H. Schultz-Coulon, W. Shen and M. Williams, "Time-Based Readout of a Silicon Photomultiplier (SiPM) for Time of Flight Positron Emission Tomography (TOF-PET)", *IEEE Transactions on Nuclear Science*, vol. 58, no. 3, pp. 597-604, 2011.
doi:10.1109/tns.2011.2119493

[3] F. Corsi *et al*. (2007). Preliminary results from a current mode CMOS front-end circuit for silicon photomultiplier detectors. *IEEE Nuclear Science Symposium Conference Record*, Honolulu, HI, 2007, pp. 360-365.
doi: 10.1109/NSSMIC.2007.4436349

[4] Huizenga, J., Seifert, S., Schreuder, F., van Dam, H., Dendooven, P., Löhner, H., Vinke, R. and Schaart, D. (2012). A fast preamplifier concept for SiPM-based time-of-flight PET detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 695, pp.379-384.
doi: 10.1016/j.nima.2011.11.012

[5] F. Ciciriello, F. Corsi, F. Licciulli, C. Marzocca and G. Matarrese, "Interfacing a SiPM to a current-mode front-end: Effects of the coupling inductance," *2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, Seattle, WA, 2014, pp. 1-6.
doi: 10.1109/NSSMIC.2014.7431046

[6] D. Comer and D. Comer, "Using the Weak Inversion Region to Optimize Input Stage Design of CMOS Op Amps", *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 51, no. 1, pp. 8-14, 2004.

[7] A. Rivetti, P. Delaurenti, "A fast large dynamic range shaping amplifier for particle detector front-end", *Nuc. Instr. and Meth. A,* Vol. 572, Issue 1, pp. 392-393, March 2006

[8] G. De Geronimo, P. O'Connor and J. Grosholz, "A CMOS baseline holder (BLH) for readout ASICs, *IEEE Transactions on Nuclear Science*, vol. 47, no. 3, pp. 818-822, 2000.
doi: 10.1109/23.856523

[9] M. Rolo, "Amplificador CMOS de baixo ruído para imagiologia médica", MSc Universidade de Aveiro 2010

[10] P. Trigilio, "Development of an ASIC for SiPM readout in SPECT applications", PhD thesis, Politecnico di Milano, 2015

[11] G. De Geronimo, P. O'Connor, A. Kandasamy, "Analog CMOS peak detect and hold circuits. Part 1. Analysis of the classical configuration", *Nuclear Instruments and Methods in Physics Research*, vol. A 484, 2002.

[12] H. Spieler, "Semiconductor detector systems," Oxford University Press, 2005.

[1] M. Di Gioia, "Preamplificatori a larga banda per la lettura di rivelatori SiPM", *Master Thesis at Politecnico di Bari,* 2018.

[2] B. Hyland, "An improved and simple cable simulation model", *Application Note 5141,* 2012. Retrieved from Maxime Integrated website
https://pdfserv.maximintegrated.com/en/an/AN5141.pdf