High Timing Resolution Front-end Circuit for Silicon Photomultiplier Detectors

Politecnico di Bari

LIBERATORIA PER L'ARCHIVIAZIONE DELLA TESI DI DOTTORATO

Al Magnifico Rettore
del Politecnico di Bari

Il sottoscritto **SAVINO PETRIGNANI** nato a **TRANI (BT)** il **4/12/1993,** residente a **TRANI (BT)** in via **SANDRO PERTINI 78/C**, e-mail **savinopetrignani@outlook.com,** iscritto al 3° anno di Corso di Dottorato di Ricerca in **INGEGNERIA ELETTRICA E DELL'INFORMAZIONE** ciclo **XXXIII**

ed essendo stato ammesso a sostenere l'esame finale con la prevista discussione della tesi dal titolo:

HIGH TIMING RESOLUTION FRONT-END ELECTRONICS FOR SILICON PHOTOMULTIPLIER DETECTORS

DICHIARA

1) di essere consapevole che, ai sensi del D.P.R. n. 445 del 28.12.2000, le dichiarazioni mendaci, la falsità negli atti e l'uso di atti falsi sono puniti ai sensi del codice penale e delle Leggi speciali in materia, e che nel caso ricorressero dette ipotesi, decade fin dall'inizio e senza necessità di nessuna formalità dai benefici conseguenti al provvedimento emanato sulla base di tali dichiarazioni;
2) di essere iscritto al Corso di Dottorato di ricerca **INGEGNERIA ELETTRICA E DELL'INFORMAZIONE** ciclo **XXXIII**, corso attivato ai sensi del "*Regolamento dei Corsi di Dottorato di ricerca del Politecnico di Bari"*, emanato con D.R. n.286 del 01.07.2013;
3) di essere pienamente a conoscenza delle disposizioni contenute nel predetto Regolamento in merito alla procedura di deposito, pubblicazione e autoarchiviazione della tesi di dottorato nell'Archivio Istituzionale ad accesso aperto alla letteratura scientifica;
4) di essere consapevole che attraverso l'autoarchiviazione delle tesi nell'Archivio Istituzionale ad accesso aperto alla letteratura scientifica del Politecnico di Bari (IRIS-POLIBA), l'Ateneo archivierà e renderà consultabile in rete (nel rispetto della Policy di Ateneo di cui al D.R. 642 del 13.11.2015) il testo completo della tesi di dottorato, fatta salva la possibilità di sottoscrizione di apposite licenze per le relative condizioni di utilizzo (di cui al sito http://www.creativecommons.it/Licenze), e fatte salve, altresì, le eventuali esigenze di "embargo", legate a strette considerazioni sulla tutelabilità e sfruttamento industriale/commerciale dei contenuti della tesi, da rappresentarsi mediante compilazione e sottoscrizione del modulo in calce (Richiesta di embargo);
5) che la tesi da depositare in IRIS-POLIBA, in formato digitale (PDF/A) sarà del tutto identica a quelle **consegnate**/inviate/da inviarsi ai componenti della commissione per l'esame finale e a qualsiasi altra copia depositata presso gli Uffici del Politecnico di Bari in forma cartacea o digitale, ovvero a quella da discutere in sede di esame finale, a quella da depositare, a cura dell'Ateneo, presso le Biblioteche Nazionali Centrali di Roma e Firenze e presso tutti gli Uffici competenti per legge al momento del deposito stesso, e che di conseguenza va esclusa qualsiasi responsabilità del Politecnico di Bari per quanto riguarda eventuali errori, imprecisioni o omissioni nei contenuti della tesi;
6) che il contenuto e l'organizzazione della tesi è opera originale realizzata dal sottoscritto e non compromette in alcun modo i diritti di terzi, ivi compresi quelli relativi alla sicurezza dei dati personali; che pertanto il Politecnico di Bari ed i suoi funzionari sono in ogni caso esenti da responsabilità di qualsivoglia natura: civile, amministrativa e penale e saranno dal sottoscritto tenuti indenni da qualsiasi richiesta o rivendicazione da parte di terzi;
7) che il contenuto della tesi non infrange in alcun modo il diritto d'Autore né gli obblighi connessi alla salvaguardia di diritti morali od economici di altri autori o di altri aventi diritto, sia per testi, immagini, foto, tabelle, o altre parti di cui la tesi è composta.

Luogo e data **BARI, 25/05/2021** Firma Savino Petrignani

Il/La sottoscritto, con l'autoarchiviazione della propria tesi di dottorato nell'Archivio Istituzionale ad accesso aperto del Politecnico di Bari (POLIBA-IRIS), pur mantenendo su di essa tutti i diritti d'autore, morali ed economici, ai sensi della normativa vigente (Legge 633/1941 e ss.mm.ii.),

CONCEDE

- al Politecnico di Bari il permesso di trasferire l'opera su qualsiasi supporto e di convertirla in qualsiasi formato al fine di una corretta conservazione nel tempo. Il Politecnico di Bari garantisce che non verrà effettuata alcuna modifica al contenuto e alla struttura dell'opera.
- al Politecnico di Bari la possibilità di riprodurre l'opera in più di una copia per fini di sicurezza, back-up e conservazione.

Luogo e data **BARI, 25/05/2021** Firma Savino Petrignani

Department of Electrical and Information Engineering
ELECTRICAL AND INFORMATION ENGINEERING
PH.D. PROGRAM
SSD: ING-INF/01 - ELECTRONICS

High timing resolution front-end electronics for Silicon Photomultiplier detectors

by

Savino Petrignani

Supervisor:
Prof. Cristoforo Marzocca

Coordinator of Ph.D. Program:
Prof. Mario Carpentieri

Course XXXIII, 01/11/2017 - 31/03/2021

Politecnico di Bari

Department of Electrical and Information Engineering
ELECTRICAL AND INFORMATION ENGINEERING
PH.D. PROGRAM
SSD: ING-INF/01 - ELECTRONICS

High timing resolution front-end electronics for Silicon Photomultiplier detectors

by

Savino Petrignani

Referees:
Prof. Stefano D'Amico
Dott. Michele Caselle

Supervisor:
Prof. Cristoforo Marzocca

Coordinator of Ph.D. Program:
Prof. Mario Carpentieri

CARPENTIERI MARIO
2021.05.27 17:19:22
CN=CARPENTIERI MARIO
C=IT
2.5.4.4=CARPENTIERI
2.5.4.42=MARIO
RSA/2048 bits

Course XXXIII, 01/11/2017 - 31/03/2021

A mia sorella Marika

Contents

Acknowledgment

First of all, I wish to show my deep appreciation to Professor Cristoforo Marzocca, my supervisor, for guiding me during these three years (and half) with his invaluable advice and support. I feel very grateful and honored to have worked together with him and to have had the chance to learn from one of the most passionate and experienced professionals I have ever known.

I would also like to sincerely acknowledge my supervisor at SLAC National Accelerator Laboratory, Dr. Angelo Dragone, for allowing me to work for six months in his IC design group. It has been a unique opportunity for my professional development in a challenging and inspiring environment.

Moreover, I want to thank all the people of the Applied Electronics and Microelectronics group of Politecnico di Bari and of the IC group of SLAC for their constant help. I feel very fortunate to have met colleagues like them, always keen to share their expertise and to cooperate with each other.

Last, but definitely not least, I would like to express my gratitude and love for my parents, my sister, my grandparents and those closest to me. Without their support and guidance nothing would have been possible.

Grazie!

Summary

Over the last few decades, the technological advancement in planar processes for the production of CMOS integrated circuits has also enabled the development of new high performance solid-state detectors, among which it is worth mentioning silicon photomultipliers. The distinctive feature of such sensors, also known with the acronym SiPMs, is the intrinsic amplification that they exhibit when operating in Geiger mode; in this condition these devices are able to generate a fast current signal with an adequate amplitude even in response to the detection of a single impinging photon. Moreover, given their low cost, robustness and insensitivity to magnetic fields, SiPMs represent a valid alternative to the most established Photomultiplier Tubes (PMTs). Therefore, the application of these detectors is taking hold in a number of fields, especially where low light levels and fine time resolutions are concerned. This is the case of the *Positron Emission Tomography* (PET), a medical imaging technique aimed at diagnosing specific diseases, where photomultipliers are employed to detect the gamma-ray photons emitted by the radiotracer injected into the patient's body.

During this doctoral program, a new front-end circuit for silicon photomultipliers has been designed in a standard 130 nm CMOS technology. This project has been carried out in collaboration with the IC group of SLAC National Accelerator Laboratory based in Menlo Park, California, with the aim of developing an analog channel for PET system with groundbreaking performances in terms of temporal resolution. Indeed, this electronic circuit is able to provide not only the energy of the detected event, but also the occurrence time of the photon absorption with a time resolution of just few tens of picoseconds, compliant with the design specifications. Subsequently, a multichannel *Application Specific Integrated Circuit* (ASIC) has been developed with the purpose of testing the analog front-end, by implementing all the circuit blocks useful for the conversion, parsing and transmission of the digital data.

In the first chapter, after a brief introduction on optical detectors, the working principle and the electrical model of SiPMs will be discussed. Furthermore, the most widespread techniques for reading out the current signal of these devices will be presented, along with an example of front-end already designed and tested. In the second

chapter, instead, the development of the new readout circuit for SiPMs will be described, also providing some details about the performance parameters obtained by means of post-layout simulations. Moreover, the third chapter includes an overview of two systems that have been implemented in the ASIC: the 12-bit ADC, which handles the conversion of the analog output of the channel in digital data, and the TDC, for the digitization of the time information of the event detected by the SiPM. As regards to the ADC, the optimization process of its array of capacitors will be also presented; this procedure has allowed the converter to improve its performances in terms of ENOB. In addition, the design of a capacitor-less LDO regulator will also be introduced; this device will be implemented in the second version of the chip in order to internally regulate the power supply voltage. Last, in the fourth chapter the development of a multichannel ASIC is presented; it comprises nine aforementioned front-end circuits and all the electronic systems required for their test. Additionally, the description of the full-chip simulations, performed to verify the functioning of the entire chip, is also provided.

Chapter 1

Readout electronics for SiPMs

1.1 Introduction

Since the beginning of the 17th century, the study and observation of phenomena related to the propagation of light led some of the most important physicists and scientists to develop theories about its nature. Among all the proposed models, the two that are worth mentioning are *the corpuscular theory* and the *wave theory*. The former, pioneered by René Descartes and Isaac Newton, states that the light consists of a flux of discrete particles, which propagate in straight lines according to the laws of geometric optics. However, as a result of the research work carried out by Huygens and Fresnel, the corpuscular theory was not able to explain other lighting effects, such as interference and diffraction, which can only be described considering the light as an electromagnetic wave, to which the Maxwell's equations may be applied.

In 1901 the German theoretical physicist Max Planck, in one of his most famous articles about the radiation emitted by a black body, assumed that the energy radiated by a generic system was quantized and its value was multiple of a fundamental unit, called *quantum* E, which depends on the radiation itself. It represents the lowest amount of energy that can be emitted and it is proportional to its frequency ν:

$$E = h\nu \tag{1.1}$$

The term h represents the Planck constant, whose value is around $6.626 \times 10^{-34}\,\mathrm{J \cdot s}$.

After few years, Albert Einstein published a breakthrough article in the *Annalen der Physik* journal, wherein he described for the first time the *photoelectric effect* [1]. During his experiments, Einstein noticed that, exposing the surface of certain kind of metals to a radiation, with a wavelength lower than a critical value, there was a non-zero probability that some electrons might be ejected from the material. In order to explain

this phenomenon, he did not consider the light beam energy uniformly distributed, but quantized, as proposed by Planck earlier, and made up of moving particles, called *quantum of light*. The energy of these particles, now commonly known as *photons*, corresponds to a quantum of energy and thus it is given by equation (1.1).

In the last century, thanks to the several scientific achievements of Einstein and other scientists on the behavior of matter interacting with electromagnetic waves, different techniques concerning the detection of small quantities of photons have emerged, which led to the development of new high performance photosensitive devices. Among all these sensors, it is worth remembering the *Single Photon Avalanche Photodiodes (SPAD)* and the *Silicon Photomultipliers (SiPM)* that, due to their design and operating mode, exhibit an intrinsic amplification. This feature guarantees an appropriate level of the output signal even when a single photon is detected, allowing them to be deployed in *photon counting* system, where low levels of light are concerned [2].

In this first chapter, after a brief introduction about the physics behind photodetectors, the operating mode and the electrical model of silicon photomultipliers are presented. Moreover, before outlining the project on which this doctoral program is based, an overview on the readout techniques for SiPM is reported, along with an example of readout circuit already developed.

1.2 Interaction between matter and radiation

In order to describe the physical processes that may happen when a radiation interacts with matter, let us consider an atomic system with two energy levels only, denoted by E_1 and E_2, where $E_2 > E_1$ and the difference between the energy levels is $\Delta E = E_2 - E_1$. In the literature, the lower level is commonly called *ground* and it corresponds to a minimum of potential energy associated to a generic particle; on the other hand, E_2 is called excited or unstable level. Moreover, let us suppose that the difference of the energy levels is much greater than the thermal energy, given by the product of the Boltzmann constant, $K_B = 1.38 \times 10^{-23}\,\mathrm{J \cdot K}$, and the absolute temperature T:

$$\Delta E \gg K_B T \tag{1.2}$$

Under this assumption, the phenomena described in this section are not encouraged by the atomic vibrations, due to the fact that the temperature of the material is different from absolute zero.
The three mechanisms of interaction between matter and radiation are:

- Absorption;

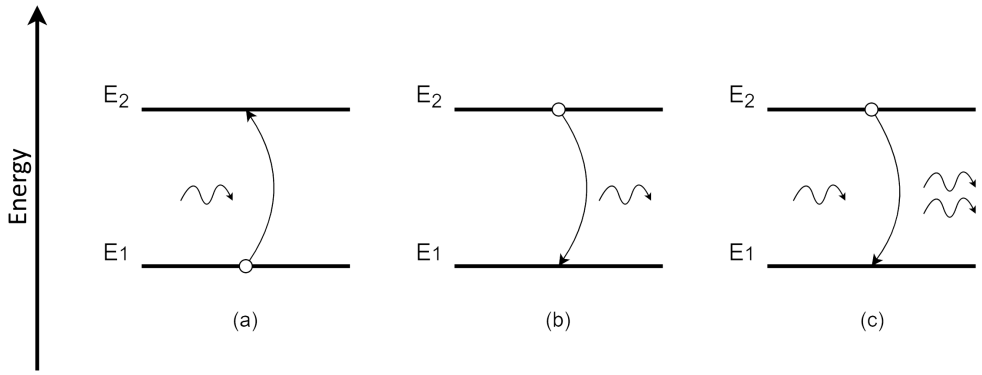

Figure 1.1: Matter-radiation interactions: (a) Absorption (b) Spontaneous emission (c) Stimulated emission.

- Spontaneous emission;
- Stimulated emission.

The absorption can occur when a photon with a frequency $\nu = \Delta E/h$ hits the system. In these conditions, with a probability that depends on the material characteristics, an electron on the ground level may be excited, by absorbing the photon to the unstable level E_2. In contrast, the spontaneous emission is a random process that concerns the electrons on the excited level, which, dropping from E_2 to E_1 may emit photons, whose frequency ν is still related to the energy gap ΔE. Lastly, the stimulated emission takes place when a photon induces an electron on the excited energy level to drop to E_1. As result, a new photon is released with the same frequency, phase, polarization and direction of the one that triggered it. The energy level diagrams in figure 1.1 briefly sum up the three processes just described [3].

The physical mechanism underlying the operation of SiPMs, the main topic of this chapter, is the absorption. Nevertheless, since these devices are made of countless atoms of silicon, whose energy diagram does not consist of just two levels, it is necessary to point out what happens when two or more atoms interact with each other.

Let us consider a crystalline solid made of multiple, identical atoms in a highly ordered and periodic microscopic structure. If the interatomic distance a is such that the wave functions of each particles do not overlap, the levels of each atom have exactly the same energy. Decreasing a, the higher energetic states start interacting with the adjacent elements and they split into as many different levels as are the atoms. This effect is a consequence of the Pauli exclusion principle, according to which, in an atom or a molecule, no more than two electrons can occupy the same energy state at the same time.

The difference among the energy levels is so low that it is more practical to con-

sider a quasi-continuous states distribution, which is called *band*. Keep decreasing the interatomic distance, the lower levels start interacting as well, creating other bands. However, for each crystal, there is an equilibrium value of a, often denoted by a_0, where attractive and repulsive forces among particles are balanced.

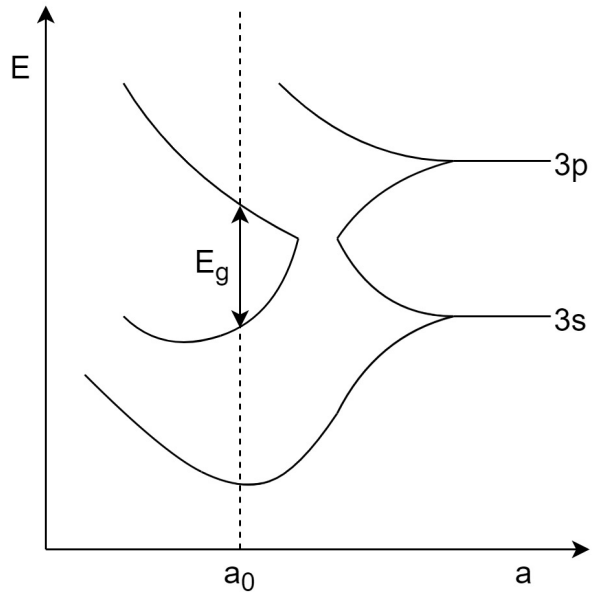

Figure 1.2: Simplified energy diagram of silicon.

The most used material in electronics for the design of integrated circuits and some kind of detectors, including SiPMs, is silicon, whose atomic number is 14. In each atom, ten electrons reside in the three innermost orbitals $1s$, $2s$ and $2p$, while the other four electrons are involved in chemical reactions. Since the first three orbitals are fully occupied and their radius are also much smaller than the others, in the energy diagram it is possible to consider just the $3s$ and $3p$ orbitals. The simplified energy band diagram of a silicon crystal, with many atoms, appears to be similar to the one in figure 1.2. At the equilibrium condition a_0, there are two distinct bands, the *valence band* below and the *conduction band* above, separated by the *energy gap* E_g, also known as *band gap* or *forbidden band*. At room temperature (300 K), the energy gap in a silicon crystal is $E_g \approx 1.12\,\text{eV}$ [4]. All the materials are generally divided into three different categories, according to the value of their band gap; *insulators* exhibit a large value of E_g, mostly greater than 7 eV, *conductors* have a small band gap or none (valence and conduction bands overlap), while *semiconductors* present an intermediate value, as in the case of silicon.

Therefore, in order to have a relevant absorption rate when a radiation hits the surface of a semiconductor, the energy of the photons E has to be greater than E_g and thus their frequency ν needs to be greater than a critical value ν_c, given by:

$$E_g \leq E = h\nu \Longleftrightarrow \nu \geq \nu_c = \frac{E_g}{h} \tag{1.3}$$

1.3 Photomultipliers

A photomultiplier is an optic device able to convert a light signal into an electric one. There are different models of photomultipliers commercially available, which differ in the range of frequencies they are sensitive to. These sensors are mainly deployed to detect photons in the near ultraviolet (300 nm - 400 nm), visible wavelength range (400 nm - 760 nm) and the near infrared (760 nm - 1.4 µm), but they can also be useful in the detection of ionizing particles and high energy photons (X and gamma rays) if they are coupled with scintillators. The different kind of photomultipliers are compared taking into account the following parameters:

- Quantum efficiency QE;
- Bias voltage;
- Gain;
- Temperature dependence;
- Manufacturing costs.

The quantum efficiency is defined as the probability that an impinging photon is detected by the sensors and its value is a function of the radiation wavelength. There are two main categories of these devices: the *photomultiplier tubes* and the solid-state photomultipliers, made of semiconductor materials. They have different working conditions and performance, so according to the application it is necessary to pick the appropriate device.

1.3.1 Photomultiplier tubes

These detectors, known with the acronym PMT, exploit the photoelectric effect for the conversion of the light into a current signal and, as shown in the figure 1.3, they have a photocathode exposed to the radiation of interest.

Under specific working conditions, a light beam may be capable of ripping one or more electrons from the photocathode, which are focused, by means of an optical input system, into a multiplier. It consists of an array of cathodes, called *dynodes*, biased with a gradually increasing potentials, that trigger a cascade mechanism during which

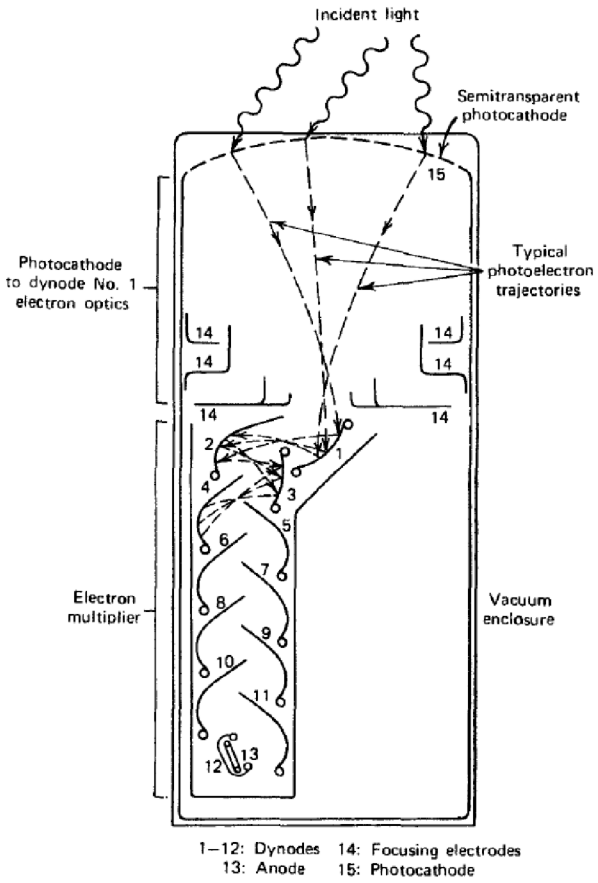

Figure 1.3: Basic structure of a photomultiplier tube [5].

the electron flux is amplified. The design of photomultiplier tubes is based on the selection of materials, number of dynodes and their voltages, on which the detectors gain and resolution depend [6].

The PMTs exhibit a quantum efficiency between 25 % and 40 %, they need high bias voltages (of the order of some kilovolts), but, due to their large gain ($10^4 - 10^9$), these sensors are able to generate a proper output signal even though a single photon hits the active area. Moreover, the performance of photomultiplier tubes does not vary much with temperature, in contrast to what happen in solid-state devices, but they are greatly sensitive to magnetic fields.

1.3.2 Solid-state photomultipliers

Solid-state photomultipliers are made of semiconductor materials, mostly silicon, and they are usually preferred over PTMs because provide many advantages, such as low production costs, low bias voltages, compactness, robustness and insensitivity to magnetic fields. These features make these detectors suitable for both space-based, astrophysics and medical applications, i.e. *Positron Emission Tomography* (PET), used for

early diagnosis of certain diseases.

In this paragraph, after introducing the photodiodes, the working principle of *avalanche photodiodes*, also known with the acronym APDs, is presented, which represent the first kind of photomultipliers realized with semiconductors. Then, silicon photomultipliers will be discussed in the following sections of this chapter, by analyzing their structure, noise sources, electric models, readout techniques and applications.

Photodiodes

A photodiode is an optical sensor able to generate a current in response to a lightning signal and it is made up of a reverse-biased *p-n* junction working in photoconductive mode.

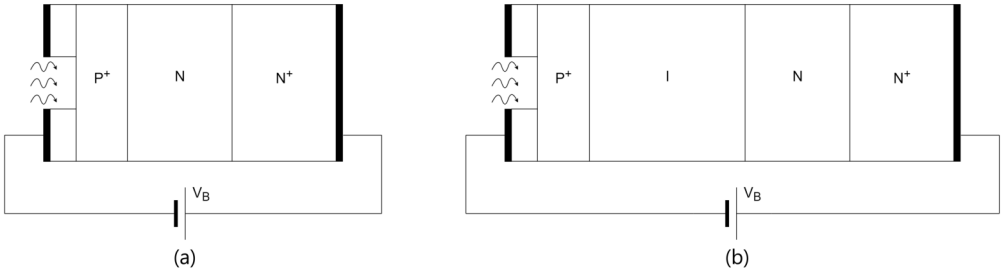

Figure 1.4: Sketches of a (a) *p-n* photodiode (b) *p-i-n* photodiode.

If a photon is absorbed into the depletion layer, it may excite an electron in the conductive band and create an electron-hole pair. These two particles are driven, in opposite direction, towards the detector edges by the electric field and those that do not recombine with other carriers are collected at the detector terminals and generate a photocurrent. To ensure the highest possible quantum efficiency, the absorption of photons should always happen in the space-charge region, depleted of mobile carriers, and in a material with moderate doping levels. Generally, in commercial detectors the *p*-type semiconductor, exposed to the radiation of interest as shown in figure 1.4.a, is heavily doped to provide a good external ohmic contact, while the impurities concentration into the *n* material is lower. Nonetheless, in case of high-energy photons (short wavelengths), the probability that absorption takes place nearby the exposure window is high. Therefore, in order to increase the quantum efficiency for short wavelength radiations, the *p+* region should be realized as narrow as possible [3].

As far as long wavelength radiations are concerned, it may happen that the space-charge region is not wide enough to guarantee the sensitivity required. This problem can be solved with a different kind of detector, called *p-i-n* photodiodes, in which the depletion region is properly expanded growing an intrinsic material (low concentration

of donor and acceptor atoms) between the two doped regions. It allows these sensors to achieve a high quantum efficiency, around 90 %.

In figure 1.4 the sketches of both regular and pin photodiodes are reported; they also have an $n+$ region which provide even in this case a good external ohmic contact with the other electrode. The main advantage of pin detectors is related to the bias voltage. Indeed, in order to deplete completely the space-charge region of any free carriers, the bias voltage V_B does not have to be very high, but just few volts are enough. Moreover, since the intrinsic layer width is much larger than the other two doped regions, the depleted area does not significantly vary with V_B, unlike what happens in p-n photodiodes.

However, such photodiodes do not have any intrinsic gain mechanism and they are not suitable for low-light level applications. When the power of incident radiation is not high, other optical detectors are preferred, like avalanche photodiodes, SPADs or silicon photomultipliers.

Avalanche photodiodes

The structure and the working principle of *Avalanche PhotoDiodes* (APD) are basically identical to regular photodiodes and the biggest difference concerns the reverse-biased power supply, whose value needs to be higher. The larger voltage applied to the p-n junction causes a strong distortion of the band diagram and thus an intense electric field $\vec{\varepsilon}$ in the space-charge region (SCR), as shown in 1.5.

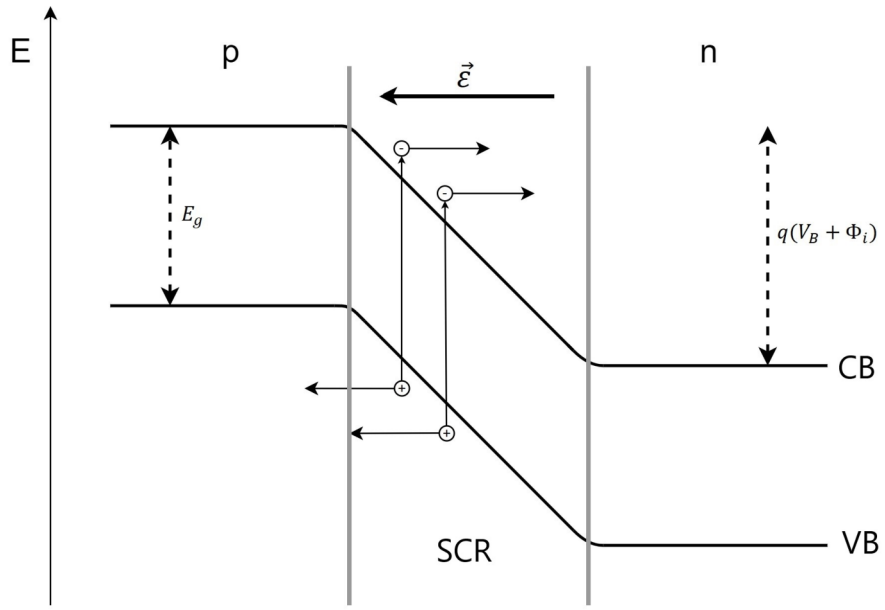

Figure 1.5: Simplified energy-band diagram of an avalanche photodiode.

As it happens for photodiodes, an impinging photon may be absorbed in the depletion layer by the APD and generate an electron-hole pair. Due to the high electric field, the electrons in conduction band inside the SCR are driven toward the cathode. The closer they are generated to the p-type material, the higher is the kinetic energy they acquire. Therefore, it is possible that excited electrons collide with others in the valence band, transferring a certain amount of their kinetic energy. If it exceeds the band-gap energy E_g, these particles may be excited in the conduction band as well. This phenomenon is known as *impact ionization*. The secondary electrons, in turn, are accelerated by the electric field in the space-charge region and may collide with further particles, triggering an *avalanche process*.

The avalanche multiplication in reverse-biased junction strongly depends on the electric field intensity in the SCR, hence on the voltage applied to the device and its temperature, and it does not make avalanche photodiodes suitable for linear applications. This happens mainly because of two reasons:

- A positive feedback due to the presence of two different kind of carriers, electrons and holes, that can ionize other particles;
- The ionization rate has an exponential dependence on the electric field intensity.

When an electron-hole pair is created in response to an absorption, both particles, driven in opposite directions by the electric field, can generate more pairs. For this reason, the multiplication factor M_n, related to a current of just electrons, depends on the ionization rate of both carriers:

$$M_n = \left[1 - \int_0^{w_d} \alpha_n \cdot exp\left(-\int_0^x (\alpha_n - \alpha_p)dx'\right) dx\right]^{-1} \tag{1.4}$$

In the previous expression w_d is the space-charge region width, while α_n and α_p are the ionization rates of, respectively, electrons and holes and they refer to the average numbers of ionizing collision per unity path length in the direction of $\vec{\varepsilon}$. It has been proven that the ionization rates are function of the electric field at an exponential rate and consequently the equation (1.4) shows a highly non-linear dependence between M_n and the field in the SCR. Nevertheless, as reported in [7], there is a way to reduce the effect of bias voltage and temperature on the multiplication factor. It consists in limiting the positive feedback process, by selecting a material in which the current is mainly due to the carriers with the highest α. In this manner M_n is poorly influenced by the positive feedback as long as it is lower than the ratio between ionization rate and the electric field.

In order to reduce the slope of $\vec{\varepsilon} - M_n$ characteristic, a $p+$ - i - p - $n+$ structure has been proposed. In this device the voltage drop is primarily located at the SCR

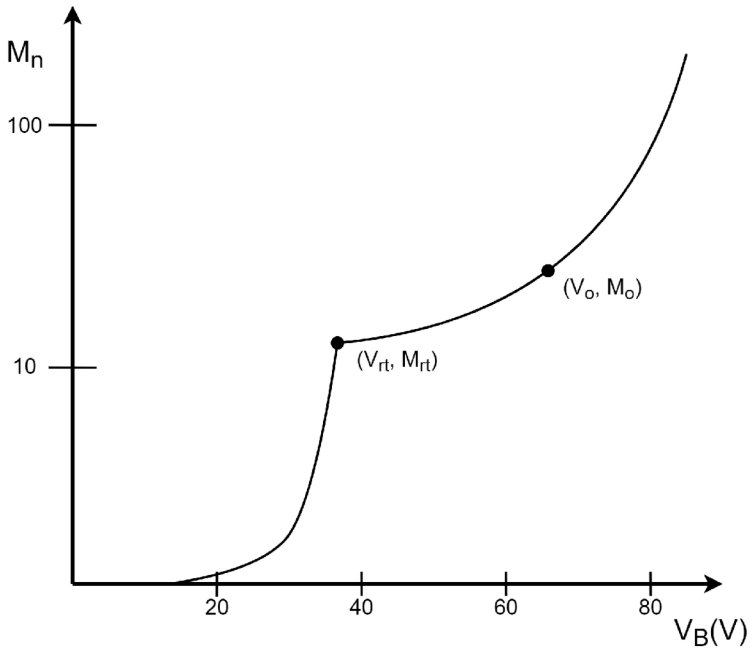

Figure 1.6: Multiplication factor - reverse-bias voltage characteristic.

edges between p and $n+$. When $V_B \geq V_{rt}$ (cf. figure 1.6) the depletion layer, which extends mainly in the p-type semiconductor, reaches the intrinsic material and any further excess of bias voltage do not affect the doped regions. If the intrinsic layer is much larger than p, the electric field, and thus also the multiplication factor, slowly vary with the power supply [7].

Single-photon avalanche photodiodes

The two optical devices taken into account previously are employed avoiding that the power supply reaches the breakdown voltage. If it is higher, the absorption of a single photon may trigger an avalanche multiplication that is impossible to stop, jeopardizing the sensing of further photons.

There is a particular class of APDs, called *SPADs (Single Photon Avalanche Photodiodes)*, for which the problem of the never-ending avalanche is solved and therefore they can be used in photon counting applications. A SPAD consists of an avalanche photodiode working in the so-called *Geiger mode*, where the power supply is 10% - 20% higher than the breakdown voltage, coupled with a series *quenching resistor*, acting as a current limiter as shown in figure 1.7. During the device operation, three different cyclical phases can be distinguished. In the idle phase, the SPAD just waits for a photon absorption in the active area. During this phase the current flowing through the device is really low, on the order of tens of picoamperes, and then negligible. The second phase starts when an impinging radiation generates at least one electron-hole pair

that triggers the avalanche. Due to the large quenching resistor, the voltage across the device decreases with the increase of the current and when it drops below the breakdown value, the avalanche is quenched and the third phase begins. During the last step, the recovery phase, the detector returns to the initial status to get ready for the following detection.

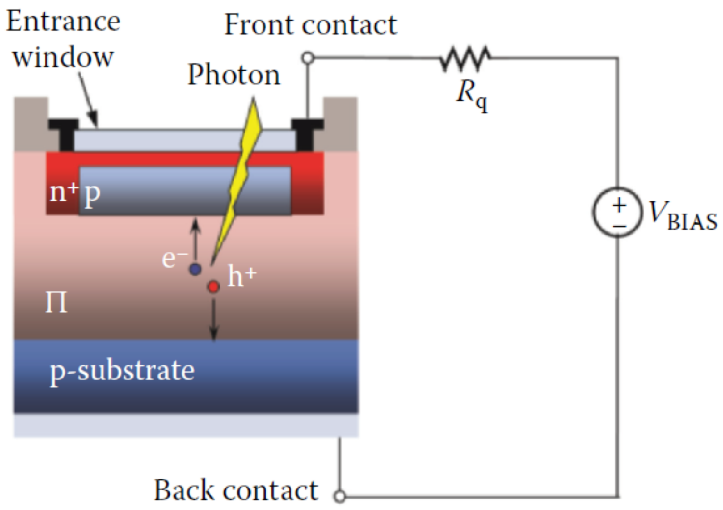

Figure 1.7: Simplified structure of a SPAD.

However, there are two different kind of current limiting circuits, called *quenching circuits*: active and passive. The latter, as already explain, consists of a resistor in series to the photodiode, whose value is between some tens of kΩ and few MΩ. The active circuits are made of transistors that can reset the detector right after the avalanche is triggered.

1.4 Breakdown voltage in SPADs

Expressed in mathematical terms, the breakdown voltage V_{BD} is defined as that value of power supply that applied to a generic diode make the multiplication factor M_n (or also M_p) infinite. Recalling equation (1.4), a p-n junction breaks when:

$$\int_0^{w_d} \alpha_n \cdot exp\left(-\int_0^x (\alpha_n - \alpha_p)dx'\right) dx = 1 \tag{1.5}$$

This entails that SPAD doping profiles define the electric field distribution in the structure and then its breakdown voltage. V_{BD} can be derived solving the Poisson equation of a diode and its value, in case of devices with a uniform impurity concentration in each region and an abrupt change in doping at the metallurgical junction (step junctions) is:

$$V_{BD} = \frac{\varepsilon E_m^2}{2eN} \tag{1.6}$$

In the previous equation, ε_s is the semiconductor relative dielectric constant, E_m is the maximum electric field in structure, e is the elementary charge and lastly N is the impurity concentration in the less doped region. Nevertheless, not all the junctions can be depicted as step junctions, but overall V_{BD} is always inversely proportional to N.

The breakdown voltage, as for all the semiconductor parameters, is also a function of the temperature T. Carrying out qualitative analysis, when the temperature is high, the probability that mobile carriers collide with fixed charges increases because of lattice vibrations. During impacts, a certain amount of their kinetic energy is converted in phonons by the atoms, and this energy loss does not allow them to ionize other particles. Therefore, they need more energy, and thus a more intense electric field, to generate secondary electron-hole pairs. Conversely, at low temperature the effect is the opposite and the breakdown voltage is lower.

It has been proven that, in some step junctions, V_{BD} varies linearly with the temperature T and, in a range between $-196\,°\mathrm{C}$ and $25\,°\mathrm{C}$, it is given by the following equation:

$$V_{BD}(T) = V_{BD}(T_0) \cdot [1 + \beta'(T - T_0)] \tag{1.7}$$

where $V_{BD}(T_0)$ is the breakdown voltage at T_0. In applications where high accuracy levels are concerned, a temperature compensation system is required. Different solutions have been suggested in literature, such as the use of a device with a feedback to set the temperature at the desired value or an instrument to change the bias voltage according to the variation of V_{BD} [8]. Besides these, other researchers even recommend the use a dummy SPAD as temperature sensor, in order to carefully sense the parametric variations [9].

1.5 Electrical models for SPADs

Denoting the total SPAD capacitance with C_{pixel} and the over-voltage with $\Delta V = V_{BIAS} - V_{BD}$, the total charge Q_{TOT} generated in response to the absorption of a photon is:

$$Q_{TOT} = C_{pixel}(V_{BIAS} - V_{BR}) = C_{pixel}\Delta V \tag{1.8}$$

Moreover, assuming that the avalanche multiplication process takes place much

quicker than the recovery phase, the output peak current of the SPAD, with a quenching resistor R_q, should be:

$$I_{peak} = \frac{V_{BIAS} - V_{BR}}{R_q} = \frac{\Delta V}{R_q} \tag{1.9}$$

However, empirical analyses have shown that equation (1.9) is not valid and that the SPAD peak current is actually more intense. Since the quenching resistance is usually at least of some hundreds of kΩ, the only possible explanation of this effect is the existence of an alternative path where the charge generated during the avalanche flows. It is due to a small parasitic capacitance in parallel to R_q through which most of the charge is collected almost immediately after the breakdown [10].

Among all the proposed circuit models for SPADs, it worth recalling the one from Cova et al. [11] and the one from Corsi et al. [12]. The former model, shown in figure 1.8, includes a switch with a series resistance R_d, to simulate the trigger of the avalanche multiplication (switch close) and its quenching (switch open), while the capacitor C_d represents the diode junction capacitance. When the switch is closed, the I_{AV} current abruptly changes from zero to $\Delta V/R_d$, while the voltage on node A drops from V_{BIAS} to the device breakdown value following an exponential decay. The expression of the current into the series resistor is given by:

$$I_{AV}(t) = \frac{\Delta V}{R_d} e^{-\frac{t}{\tau_d}} \tag{1.10}$$

where the time constant is $\tau_d = R_d\,(C_d + C_q)$. The avalanche process ends in correspondence of the switch opening that happens when the current is lower than a reference value I_q and the circuit starts going back to the initial condition. Afterwards, the voltage on the node A starts growing although with a slower time constant this time. Indeed, the recovery phase time constant does not depend on R_d, but it is function of R_q, which is usually three degrees of magnitude larger, and its value is $\tau_r = R_q\,(C_d + C_q)$. Photomultiplier performances are strongly related to t_r, especially in photon counting applications, since it is necessary to wait a not negligible recovery time before sensing another photon.

However, this model in real situation cannot be easily employed, because there are no established techniques for estimation of the switch series resistance and I_q. In addition, it requires a controlled source for any device and, in situation where thousands of SPADs are used, the setting of their parameters becomes time-consuming and the simulations more intricate. In order to solve the issues related to the circuit in figure 1.8, Corsi et al. have proposed a simplified SPAD model. The whole idea behind it is that it is not important to accurately reproduce the initial current transient, because it is too

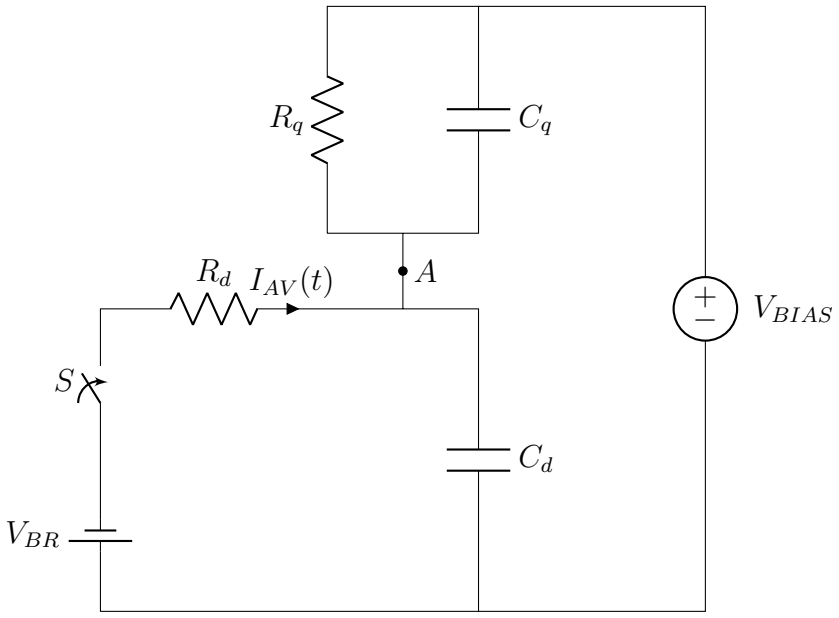

Figure 1.8: SPAD electric model proposed by Cova et al.

fast and, even considering the latest fabrication technologies, it cannot be read without distortion. In fact, the rising time of avalanche current is of just few picoseconds and a low-noise preamplifier with a bandwidth of several hundreds of gigahertz would be necessary.

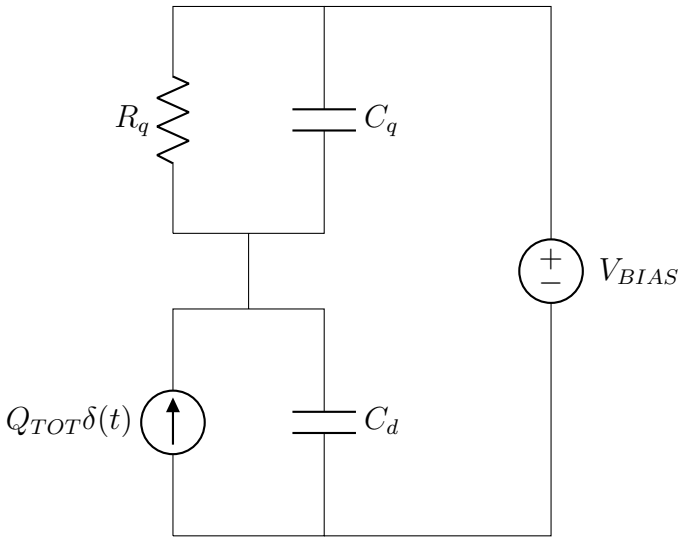

Figure 1.9: SPAD electric model proposed by Corsi et al.

The key element of the proposed circuit, shown in figure 1.9, is the current generator $Q_{TOT}\delta(t)$ that emulates the avalanche multiplication process with a Dirac pulse, so that the total charge in the readout circuit is the same generated by the SPAD. The Q_{TOT} value, as shown in equation (1.8), is given by:

$$Q_{TOT} = C_{pixel}\Delta V = C_q\Delta V + C_d\Delta V \tag{1.11}$$

where the first term is the amount of charge instantly located across the C_q capacitor and collected by the external electronics, while the second represents the current that flows during the recovery phase.

1.6 Silicon photomultipliers

Silicon photomultipliers, best-known with the acronym SiPMs, are optical devices consisting of a matrix of SPADs, which are all connected in parallel. In order to describe the shape of the output current signal of a single pixel, it is also necessary to consider the load effect due to the other cells. Moreover, the metal grid, used to distribute the power supply to every single SPAD, introduces a further parasitic capacitance C_g that must be taken in account because, especially in large SiPMs, its value can be of several tens of picofarads. The complete model of a silicon photomultiplier with N pixels is reported in figure 1.10.

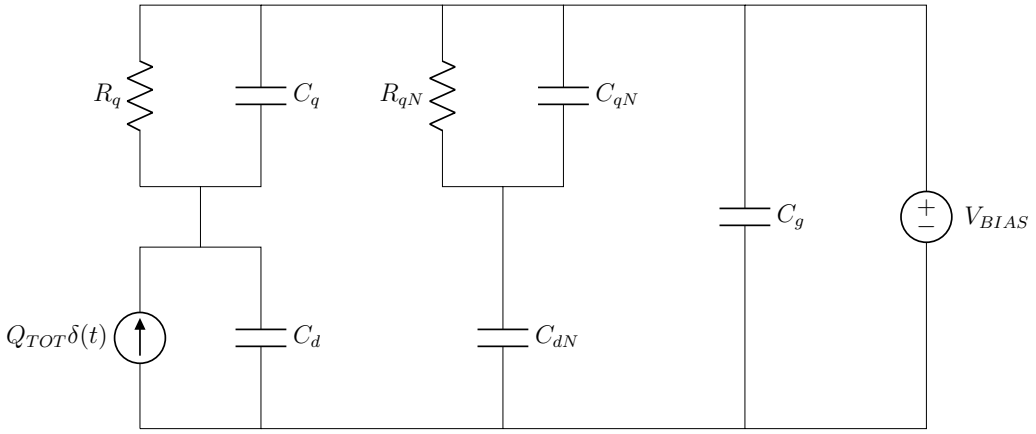

Figure 1.10: Silicon photomultiplier electric model proposed by Corsi et al.

where:

$$R_{qN} = \frac{R_q}{N-1} \qquad C_{qN} = C_q\,(N-1) \qquad C_{dN} = C_d\,(N-1) \tag{1.12}$$

One of the main advantages of the model proposed by Corsi et al. is the possibility of simulating the SiPM output current even if more than one pixel is fired, by just applying the superposition principle. Indeed, assuming that M photons are absorbed by the photomultiplier and that each cell generates during the avalanche a charge Q_{TOT}, the photocurrent value is given by the following equation:

$$I_{AV}(t) = Q_{TOT} \sum_{i=1}^{M} \delta(t - t_i) \tag{1.13}$$

The term t_i represents the time when the i-th photon is detected [12].

In all the simulations reported in this thesis work, the SiPM S10931-050P from Hamamatsu has been considered and its electrical parameters, reported in table 1.1, have been used.

Table 1.1: Electrical parameters of SiPM S10931-050P from Hamamatsu.

Parameter	Value
R_q	49.6 kΩ
C_q	15.49 fF
C_d	80.14 fF
C_g	18.24 pF
V_{BR}	70.35 V
Q_{TOT}	160 fC
N	3600

A circuit simulator has been employed to analyze the behavior of the SiPM model, using a pulse current source to stimulate the SiPM, which has to be able to deliver the same amount of charge generated by the detector when the avalanche is triggered in one of its pixels. Since the avalanche is one of the fastest physical processes and it takes place in just few picoseconds, a current source with a 10 ps pulse of 16 mA has been considered. In this way, the total injected charge, given by the area under the pulse, is specifically $Q_{TOT} = 160$ pC. Moreover, a load resistor $R_L = 20\,\Omega$ has been connected to the SiPM output terminal; as explained in the next paragraph, the input impedance is a key parameter of any SiPM readout, because it affects the output current shape and in particular its rise time and peak value.

Delaying the stimulus of 5 ns, the trend of the SiPM output current is the one displayed in figure 1.11. As already mentioned, the rise time is almost instantaneous, due to the time constant t_q, while the recovery phase is much slower.

1.7 SiPM performance parameters

In order to pick out the right silicon photomultipliers for each application, there are some figures of merit that allow SiPMs to be compared.

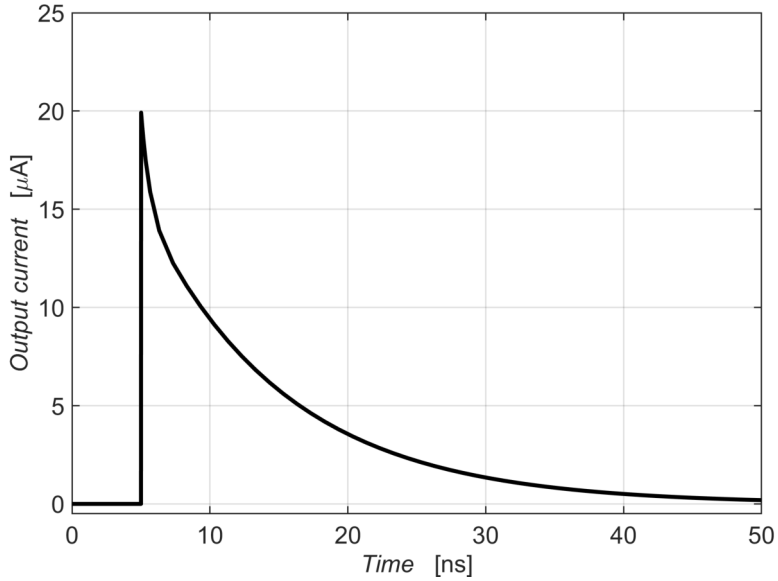

Figure 1.11: SiPM output current when a single pixel is fired ($R_L = 20\,\Omega$).

Gain

One of the main parameters is the *gain* and it is defined as the ratio between the total charge generated by a single pixel during the avalanche and the elementary charge e. Hence it denotes the number of photogenerated electrons in response to a photon absorption. Recalling the equation (1.11), the gain is given by the following expression:

$$G = \frac{Q_{TOT}}{e} = \frac{(C_q + C_d)}{e}\Delta V = \frac{C_q \Delta V}{e} + \frac{C_d \Delta V}{e} \tag{1.14}$$

Considering the SiPM parameters in table 1.1 and its bias voltage of about 80 V, the gain is around $G \approx 6 \cdot 10^6$. Thanks to their high-gain, silicon photomultipliers are suitable for photon counting applications, in which the lighting radiation power is low [13].

Photon detection efficiency

Another noteworthy SiPM figure of merit is the *Photon Detection Efficiency PDE*, which depends on the radiation frequency and bias over-voltage $\Delta V = V_{BIAS} - V_{BD}$. It can be calculated using this formula:

$$PDE = \varepsilon(\lambda, \Delta V) = \eta(\lambda) \cdot \varepsilon_G(\Delta V) \cdot \frac{A_{active}}{A_{total}} \tag{1.15}$$

Three different terms can be distinguished. The first is the quantum efficiency η, whose value is function of wavelength, while the other one is ε_G, which represents the probability that a carrier in the SPAD space-charge region may trigger an avalanche

multiplication and it is related to the over-voltage. It is thus quite clear that the product of these two quantities represents the single SPAD photon detection efficiency. Lastly, the coefficient A_{active}/A_{total} is called *fill factor* or *geometrical efficiency* and it is the ratio between the exposed and the whole matrix area. Its value is not exactly 1, because both the quenching resistor and the metal grid prevent the pixels to be fully illuminated. Recent research showed that ε_G is expressed by the formula:

$$\varepsilon_G = 1 - exp\left(-\Delta V \cdot \frac{e}{E_g}\right) \tag{1.16}$$

where E_g is the semiconductor band-gap energy and e is the elementary charge. The complete equation is consequently [14]:

$$\varepsilon\left(\lambda, \Delta V\right) = \eta\left(\lambda\right) \cdot \left[1 - exp\left(-\Delta V \cdot \frac{e}{E_g}\right)\right] \cdot \frac{A_{pixel}}{A_{total}} \tag{1.17}$$

Dynamic range

In all sensors, the measurements are limited by the *dynamic range* that is the interval of values in which the system response is linear. In silicon photomultipliers this range is determined by the finite number of cells N: when light intensity is high (the number of photogenerated electrons is comparable to the number of pixels), the SiPM may saturate, making the detection of further photons impracticable [15].

When N_{ph} photons hit the SiPM active area, the average number of triggered cells is:

$$N_f = N \cdot \left[1 - exp\left(-\frac{N_{ph} PDE}{N}\right)\right] \tag{1.18}$$

where N is the total number of pixels, while PDE is the photon detection efficiency.

In realty, because of noise sources intrinsically present in photomultipliers (cf. next paragraph), the equation (1.18) slightly changes:

$$N'_f = N \cdot \left[1 - exp\left(-\frac{N_{ph} PDE'}{N}\right)\right] + p_{ap} \cdot PDE' \cdot N_{ph} \tag{1.19}$$

The value of N'_f differs from the one in equation (1.18) because of the effects of two noise contributions: optical crosstalk and afterpulse. The former decreases the photon detection efficiency value according to the expression:

$$PDE' = PDE\left(1 - k\right) \tag{1.20}$$

in which k is the probability that crosstalk takes place. On the other hand, the afterpulse increases the number of triggered cells even if they are not exposed to the

radiation and this is taken into account with the expression $p_{ap} \cdot PDE' \cdot N_{ph}$, where p_{ap} is the probability that afterpulse occurs.

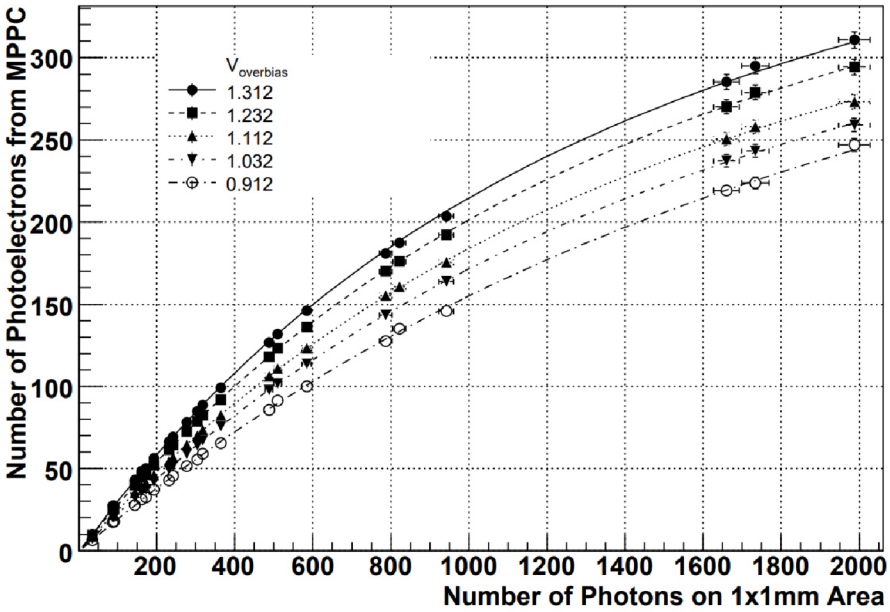

Figure 1.12: Linearity curves of the number of photoelectrons versus the total number of impinging photons on a SiPM with 400 pixels for different values of ΔV [16].

The plot in figure 1.12 shows how the number of photogenerated carriers varies with the number of impinging photons and bias voltage for a SiPM with 400 pixels. The higher the power supply, the larger the SiPM dynamic range, due to the increase of PDE. Moreover, this parameter can also be improved by increasing the device pixel density [16].

1.8 Noise sources in silicon photomultipliers

As it happens in all real systems, undesired random signal can be detected at the SiPM output node, not due to the detection of photons but to other physical processes. The three main noise sources in silicon photomultipliers are *dark current*, *optical crosstalk* and *afterpulse*.

Dark counts

In SiPMs, it may happen that some current pulses are generated even in absence of an incident radiation. They are indistinguishable from those produced in response to the absorption of photons and they are referred to as *dark-counts*. The primary cause of this phenomenon is the thermal generation of an electron-hole pair in the space-charge

region. Even in this case, these particles are driven by the electric field and may trigger the avalanche multiplication process, resulting in a current signal identical to the one generated by photoelectrons. It has been proven that the rate of such noise process, called *Dark Count Rate (DCR)*, does not change much with the over-voltage ΔV, but is highly dependent on temperature [17].

Optical crosstalk

When an avalanche process starts in any SPAD of a silicon photomultiplier, hot carriers in the depletion layer may emit other photons in the infrared range, because of their high kinetic energy. Hence these photons may be absorbed by surrounding pixels and trigger a further avalanche. The phenomenon under discussion, called *optical crosstalk*, happens almost immediately after the primary avalanche and its probability is proportional to the SiPM gain.

In order to decrease the impact of this unwanted effect, one might think to reduce the bias voltage; however, that would result in a lower photon detection efficiency. Typically, optical insulating grooves can be built around each pixel, reducing the probability that photons can travel in adjacent cells. Although, as reported in [18], the prevailing contribution is due to those photons that are reflected on the device substrate and these peripheral slots do not prevent infrared photons to reach other SPADs.

Afterpulse

The fabrication of SiPM is a challenging process, because the device performances are highly sensitive to imperfections of the material. Manufacturing defects are able to capture a certain amount of charge generated during the avalanches, particularly if they are located into the space-charge region. If this charge is released during the SiPM recovery phase, the carriers may ionize other particles and trigger another avalanche, causing the so-called *afterpulse*, associated to an amount of charge smaller than Q_{TOT}, since the device has not completed the recovery phase. It can be demonstrated that the probability of afterpulse is directly proportional to C_q+C_d and to the trap concentration N_t. This means that, to reduce the occurrence of afterpulses, it is not only necessary decrease the number of defects in the material, by improving the fabrication process, but also make the SPAD parasitic capacitance smaller [19].

1.9 Front-end electronics for SiPMs

The excellent timing properties and good linear response of silicon photomultipliers have enabled them to be an appealing alternative to APDs and PMTs in a variety of

applications where low-light levels are concerned. However, in order to achieve the best possible performance, each SiPM requires an appropriate readout circuit which does not affect the detector operating point and fully exploits its favorable features. The SiPM output signal allows the measurement of two quantities: the first one is related to the energy of the detected event, while the second is the photon occurrence time.

The energy of the incident light is proportional to the total charge collected by the external circuit, and thus to the number of triggered pixels in the SiPM, provided that saturation of the device does not occur. This kind of measurement is often influenced by the intrinsic noise sources (dark current, crosstalk and afterpulse), which can generate errors in the evaluation of the energy associated to the detected event.

For time measurements, instead, thanks to the sharp slope of the output current, with a rise time of just few nanoseconds, it is possible to identify the instant when one or multiple photons hit the SiPM active area. Besides Positron Emission Tomography, there are several other applications where the time difference between the detection of a photon and a reference signal is concerned, such as *LiDAR* (*Light Detection and Ranging*) or *Time Correlated Single Photon Counting* (*TCSPC*). In these systems, the time of arrival of the photon is determined by means of a voltage (or current) leading-edge discriminator so that, when the front-end output signal exceeds a threshold V_{TH} (or I_{TH}), its logic state changes from low to high (or vice versa). Nonetheless, time measurements are affected by the readout electronic noise, which causes statistical fluctuation of the discriminator output signal, whose uncertainty is estimated by the *time jitter* σ_t:

$$\sigma_t = \frac{\sigma_n}{\left.\frac{dV_{out}}{dt}\right|_{V_{out}=V_{TH}}} \tag{1.21}$$

where σ_n is the rms output noise of the readout electronics, while the denominator represents the slope of the front-end signal when it overcomes the chosen threshold V_{TH}. However, these parameters are related to each other through the circuit bandwidth, which determines both the noise contribution on the output node and the signal slope. In fact, a broader band not only entails a faster rise time, but also a higher integrated noise. In order to have the best time resolution possible, it is therefore necessary to select the circuit parameters so that the slope to noise ratio is maximized.

In most cases, when the application imposes stringent requirements, the SiPM circuits are designed in integrated technology. This solution allows the system to be more reliable and less bulky, especially if large arrays of SiPMs are employed. Indeed, multiple front-end circuits can be realized in the same ASIC (Application Specific

Integrated Circuit), where analog signals are not only read, but also digitized, processed and transmitted to external computer systems.

In the following subsections, three different readout approaches for SiPMs are presented and, after introducing an example of front-end circuits reported in the literature, a new CMOS channel architecture is described.

1.9.1 Charge-Sensitive Amplifier

The *Charge-sensitive amplifier (CSA)* is the most basic approach to read the SiPM output signal and it is made of a low-noise inverting voltage gain stage and a feedback capacitor C_{FB}, as shown in figure 1.13. It allows the integration of the detector current I_{IN}, so that the output voltage is proportional to the total charge Q_{TOT} generated during the avalanche:

$$V_{out} = \frac{Q_{TOT}}{C_{FB}} \tag{1.22}$$

In this case, considering that the amplifier inverting input is a virtual ground, the SiPM reverse bias can be provided by connecting a supply voltage source with a negative output voltage $-V_{BIAS}$ to the sensor anode.

However, this circuit could not work in this simple configuration because it does not have a dc path to discharge the capacitor, thus any input with non-zero DC value or the input offset of the amplifier would cause the output voltage grow until the amplifier saturates. This problem can be fixed by adding a reset circuit to the integrator, which is commonly done by means of a feedback resistor R_{FB} that guarantees a stable quiescent point and a DC path for the DC current of the detector. In these conditions, the circuit time constant $\tau_{FB} = R_{FB}C_{FB}$ must be slower than $\tau_r + \tau_s$, which represent respectively the SiPM recovery time and the scintillator decay time, to ensure that most of the detector charge is collected and integrated by the CSA.

The CSA must be designed by carefully choosing the values of C_{FB} so that, according to the equation (1.22), it can collect all the SiPM charge. Nevertheless, the value of V_{out} cannot exceed the opamp output swing, which in turn is limited by its power supply that, in the latest IC technologies, it is not higher than a couple of volts. When a photomultiplier is coupled with a scintillator, the number of impinging photons that may trigger an avalanche can reach 2000 and, considering the SiPM S10931-050P, a total charge up to 320 pC may be generated by the detector. For instance, with an output voltage swing of 1 V, a 320 pF capacitor would be necessary, which requires too much silicon area, making the circuit not suitable for integrated applications, especially when a multi-channel ASIC needs to be designed.

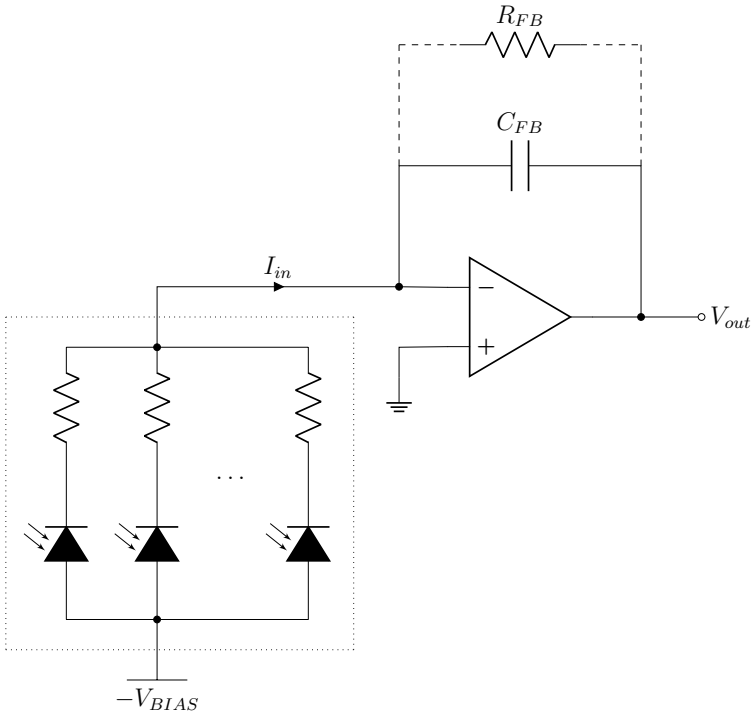

Figure 1.13: Charge-Sensitive Amplifier.

When a photon absorption triggers an avalanche in a SiPM pixel, most of the charge flows through the parasitic capacitor C_q and it is almost immediately collected at the input node of the readout circuit. Considering the equation (1.11), this amount of charge is given by:

$$Q_F = \Delta V C_q = Q_{TOT} \frac{C_q}{C_q + C_d} \tag{1.23}$$

Before the quenching, since the avalanche occurs in few picoseconds, the low-frequency contributions due to the resistors R_q are negligible, so they can be removed from the model. Therefore, for high-frequency analyses, the complete model in figure 1.10 can be replaced with an equivalent capacitance C_F and a current pulse source, where each pulse has an area equal to Q_F. The value of C_F is the sum of the metal grid parasitic capacitance and those of the remaining unfired pixels of the SiPM:

$$C_F = C_g + (N-1)\frac{C_q C_d}{C_q + C_d} \approx C_g + N\frac{C_q C_d}{C_q + C_d} \tag{1.24}$$

Thus, as far as time analyses are concerned, the SiPM in figure 1.13 can be replaced with its high-frequency model, as reported in figure 1.14.

One of the main parameters of this front-end circuit is the *Gain Bandwidth* product GBW and it should be considered to investigate the CSA performances. Considering the circuit in figure 1.14 without the feedback resistor R_{FB}, the closed-loop bandwidth

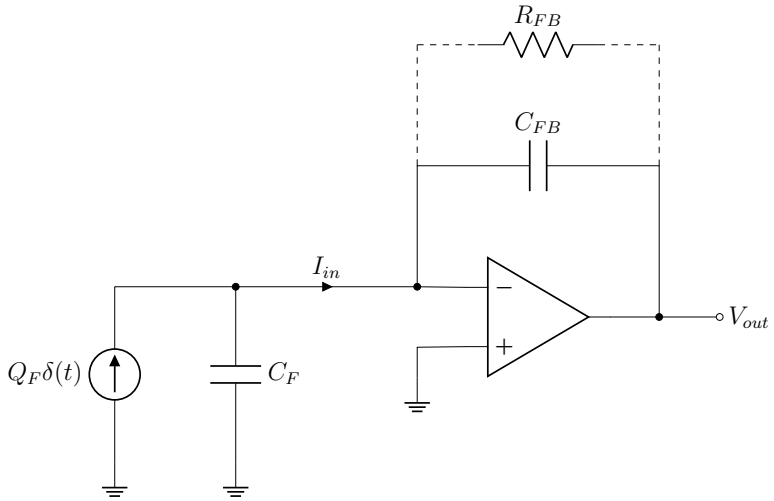

Figure 1.14: High-frequency SiPM model with a CSA front-end circuit.

is:

$$\omega_F = GBW \frac{C_{FB}}{C_{FB} + C_F} \tag{1.25}$$

The system bandwidth hence heavily depends on the equivalent capacitance C_F and this implies that, in SiPMs with a large number of pixels, the maximum slope of the output current is low, and so is the time resolution as pointed out in equation (1.21). Indeed, the maximum signal slope is given by:

$$\left.\frac{dV_{out}}{dt}\right|_{MAX} = \frac{Q_F}{C_{FB}}\omega_F = \frac{Q_F}{C_F + C_{FB}} GBW \tag{1.26}$$

As regards to the output noise, a voltage source with a spectral density e_n^2, measured in V^2/Hz, needs to be considered in series to the amplifier non-inverting input node. A good approximation of the total output noise σ_n is provided by the following formula:

$$\sigma_n \approx \sqrt{\int_0^\infty e_n^2 \left(1 + \frac{C_F}{C_{FB}}\right) \frac{1}{\left|1 + j\frac{\omega}{\omega_F}\right|^2} d\omega} = \frac{e_n}{2}\sqrt{\left(1 + \frac{C_F}{C_{FB}}\right) GBW} \tag{1.27}$$

where $1 + C_F/C_{FB}$ is the *noise gain* at high frequencies. Replacing equations (1.26) and (1.27) in (1.21), the estimation of the time jitter for a CSA front-end circuit is:

$$\sigma_t \approx \frac{e_n \left(C_F + C_{FB}\right)}{2Q_F} \sqrt{\left(1 + \frac{C_F}{C_{FB}}\right) \frac{1}{GBW}} \tag{1.28}$$

Therefore, time performances in CSA front-end circuit are strongly affected by the

equivalent capacitance C_F and thus also by the SiPM matrix size. Moreover, deriving equation (1.28) with respect to the feedback capacitance C_{FB}, σ_t has a minimum if $C_{FB} = C_F/2$, which again is too much wide to be implemented in an integrated circuit.

The CSA is not the best architecture to be employed in SiPM systems, because of its large value of feedback capacitance and its poor jitter. Nonetheless, if the application does not require high timing resolution and small area, it may be implemented on discrete printed circuit board which are faster to design and produce.

1.9.2 Voltage-mode readout

An alternative to the charge-sensitive amplifier is the *voltage-mode readout*, which requires the use of a resistor R_{IN} in series to the SiPM, as shown in figure 1.15, to convert the detector current into an equivalent voltage signal, which usually needs to be conveniently amplified by means of a further gain stage. If time measurements are concerned, it is possible to consider again the high-frequency SiPM model, as already done for the CSA. When an avalanche is triggered in one of the SiPM pixels, most of the generated charge, whose value Q_F is given by the equation (1.23), is instantly collected by the external circuit. Then, it begins to flow through the resistor R_{IN}, causing V_{IN_F} to decrease according to the following expression:

$$V_{in_F}(t) = \frac{Q_F}{C_F} \cdot e^{-\frac{t}{\tau_F}} \tag{1.29}$$

where, since the value of R_{IN} is generally low, the time constant τ_F can be approximated as follow:

$$\tau_F \approx R_{IN} C_F \tag{1.30}$$

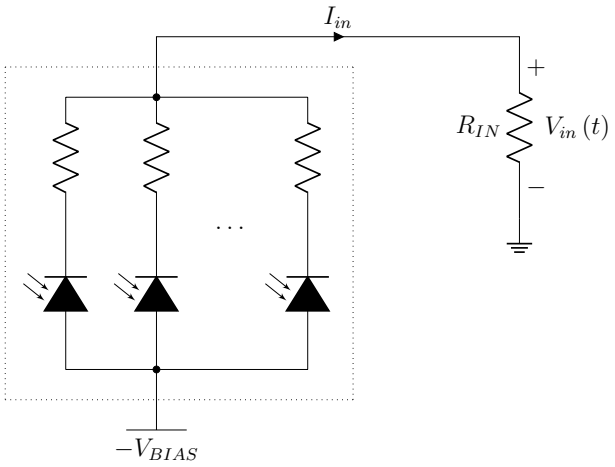

Figure 1.15: Voltage-mode readout.

As regards to the remaining amount of the total charge, $Q_S = Q_{TOT} - Q_F$, it is possible to demonstrate that its contribution on the front-end input voltage V_{IN_S} is approximately equal to:

$$V_{in_S}(t) = R_{IN}\frac{Q_S}{\tau_S - \tau_F} \cdot \left(e^{-\frac{t}{\tau_S}} + e^{-\frac{t}{\tau_F}}\right) \tag{1.31}$$

In the previous equation, the time constant τ_F defines the rise time of V_{IN_S}, while the tail is dominated by the slow time constant τ_S:

$$\tau_S \approx \tau_r + R_{IN}(C_g + NC_d) \tag{1.32}$$

The input voltages V_{in} for three different values of input resistances are reported in figure 1.16. The long tails of the signals depend on the time constant τ_S, which is not only function of $\tau_r = R_q(C_q + C_d)$, but also of R_{IN}. As can be noted, increasing the input resistance, the SiPM recovery time extends and the maximum rate that the detection system can handle is reduced. This happens because if an avalanche is triggered in another pixel during the recovery time, the second pulse overlaps the tail of the previous signal causing the pile-up effect, which involves a further degree of uncertainty in time and energy measurements.

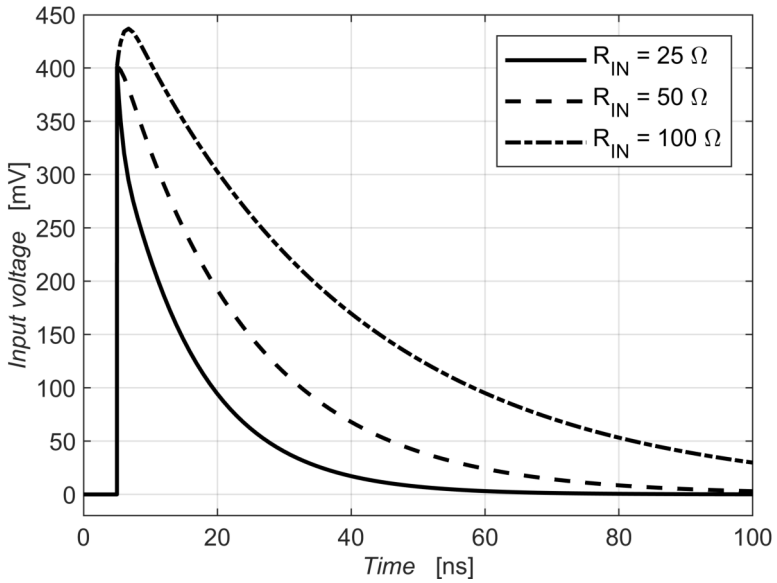

Figure 1.16: Input voltage for three different values of R_{IN}.

The bandwidth of the front-end circuit does not affect the energy measurement of the detected event. Indeed, even when several photons are absorbed in different SiPM cells at the same time, the peak of the output current is proportional to the charge generated in the photomultiplier, since the detection system is linear. Nevertheless, also in the case of the voltage-mode readout, the estimation of the energy must be

performed by integrating the output voltage of the preamplifier; such integrator needs be designed so that the time constant, related to its low-pass filter, is much slower than the scintillator decay time (if used) and τ_S.

As regards to the jitter estimation, as in the case of CSA circuits, it is possible to consider only the "fast" contribution Q_F of the total charge. The output voltage of the preamplifier, which reads the voltage across R_{IN}, with gain A_V and bandwidth $BW = 1/\tau_A$ is:

$$V_{out}(t) \approx A_V \frac{Q_F}{C_F} \frac{\tau_F}{\tau_F - \tau_A} \cdot \left(e^{-\frac{t}{\tau_F}} - e^{-\frac{t}{\tau_A}} \right) \tag{1.33}$$

Deriving the equation (1.33) with respect to time, the maximum signal slope, which is found at $t = 0$, is:

$$\left. \frac{dV_{out}}{dt} \right|_{MAX} \approx A_V \frac{Q_F}{C_F} \frac{1}{\tau_A} = GBW \frac{Q_F}{C_F} \tag{1.34}$$

First of all, as pointed out by equation (1.34), the maximum slope is independent from the input resistance R_{IN}. However, around the threshold the signal slope is lower than the one at time zero, but it slightly increases for increasing value of R_{IN}. In order to minimize the jitter, the input resistance can be risen as long as the maximum sustainable event rate is within the application requirements. Since the rms noise of the amplifier is

$$\sigma_n = \frac{e_n A_V}{2} \sqrt{BW} \tag{1.35}$$

the time jitter is given by the following expression:

$$\sigma_t = \frac{e_n A_V \sqrt{BW}}{2GBW \dfrac{Q_F}{C_F}} = \frac{e_n C_F}{2Q_F} \sqrt{\frac{1}{BW}} \tag{1.36}$$

This value of σ_t is much lower with respect to the equation (1.28), because it is not related to the feedback capacitance C_F, whose value could be large. Further improvements can be obtained with the following approach.

1.9.3 Current-mode readout

Since SiPMs generate current pulses as a result of the absorption of light, a further possible approach to read these signals involves the use of a current buffer.

Current buffers do not introduce much loading effect on the input node, because

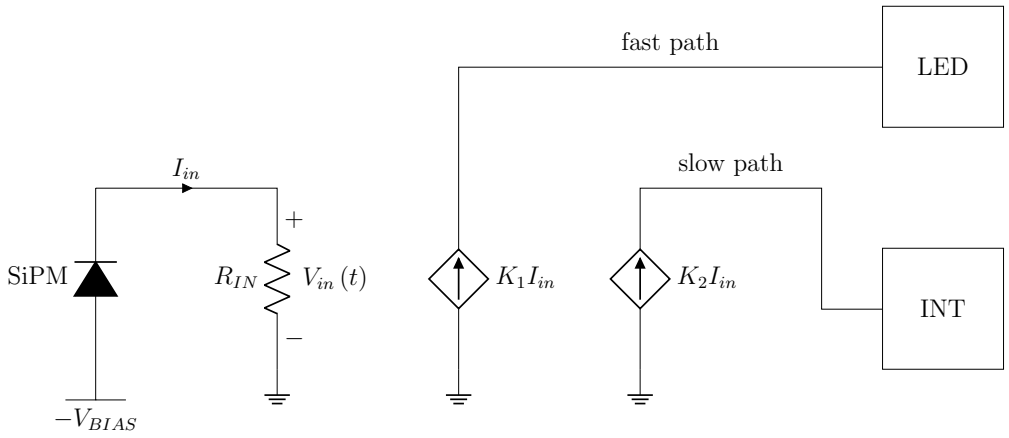

Figure 1.17: Current-mode readout.

they exhibit a very low input resistance R_{IN}; as reported in the previous section, this enables the system to handle a higher event rate than the voltage-mode readout, without affecting the signal slope. Indeed, owing to the absence of high impedance nodes, it is fairly straightforward to obtain large bandwidth in current amplifiers, preserving the fast-rising edge of the input signal and thus enhancing the system time resolution. Furthermore, as the block diagram in figure 1.17 suggests, the input signal can be reproduced in two different paths by means of current mirrors. Each replica is copied with a distinct scaling factor (K_1 and K_2), to fulfill the dynamic range requirements and thus optimize time and energy measurements. The "fast path" is coupled to a Leading-Edge Discriminator (LED) in order to determine the event occurrence time, while in the "slow path" the current is sent to an integrator (INT), whose output voltage is proportional to the number of triggered pixels.

As already done for the two previous architectures, in order to derive the time jitter, it is possible to take into account only the fast charge Q_F. The analytic expression of $I_{in}(t)$ is:

$$I_{in}(t) \approx \frac{Q_F}{R_{IN}C_F} \cdot e^{-\frac{t}{\tau_F}} \approx \frac{Q_F}{\tau_F} \cdot e^{-\frac{t}{\tau_F}} \tag{1.37}$$

where the approximation $\tau_F \approx R_{IN}C_F$ is valid only for small values of R_{IN}. Considering a unity gain current buffer with a bandwidth $BW = 1/\tau_A$, the output current is given by:

$$I_{out}(t) \approx \frac{Q_F}{\tau_F - \tau_A} \cdot \left(e^{-\frac{t}{\tau_F}} - e^{-\frac{t}{\tau_A}}\right) \tag{1.38}$$

Deriving the equation (1.38) with respect to t and evaluating it at time zero, the maximum slope of the output current is:

$$\left.\frac{dI_{out}}{dt}\right|_{MAX} \approx \frac{Q_F}{\tau_F}\frac{1}{\tau_A} = GBW\frac{Q_F}{\tau_F} \tag{1.39}$$

It has been proven that the peak current increases by reducing the input resistance, because τ_F if function of R_{IN}. Therefore, decreasing R_{IN} is not only convenient to have a steeper rising edge, but also a higher peak of the output signal and a faster recovery time.

In order to carry out the noise analysis, it is necessary to consider the most basic current-mode readout, shown in figure 1.18. The circuit consists of an MOS transistor in the *common-gate* configuration, where the SiPM current is injected into the source, while the output current is collected from the drain. The input resistance of this stage is:

$$R_{IN} = \frac{R_{OB}}{1 + g_m R_{OB}} \approx \frac{1}{g_m} \tag{1.40}$$

where g_m is the MOSFET transconductance, while R_{OB} is the output resistance of I_{BIAS}.

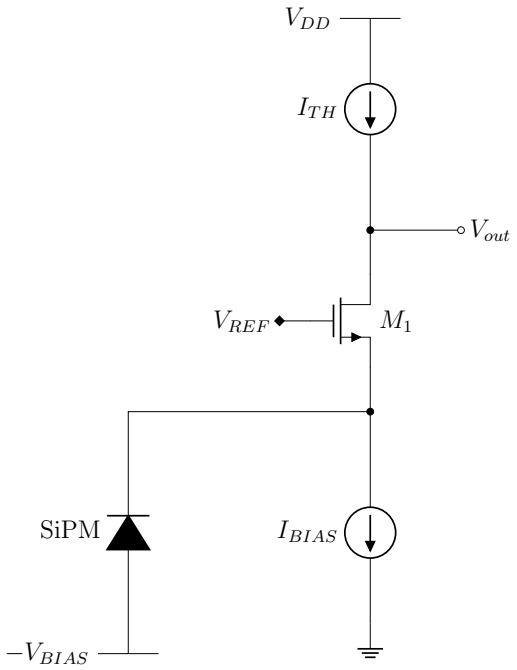

Figure 1.18: Example of a current-mode readout.

The noise on the output node is the sum of the contributions due to three devices: the MOSFET M_1 and the current sources I_{BIAS} and I_{TH}, which are used respectively to bias the transistor and establish a current threshold for the LED. The current spectral density of the transistor is:

$$i_{n1}^2 = 4K_B T \gamma g_m \tag{1.41}$$

In the previous equation, K_B is the Boltzmann constant, T is the absolute value and γ is a parameter which depends on the MOSFET size. Having to maximize g_m to reduce the input resistance, the noise contributions of current mirrors is negligible compared to that of M_1. Integrating the i_{n1}^2 in the whole frequency range, it is possible to obtain a conservative estimation of σ_n:

$$\sigma_n \approx \sqrt{\int_0^\infty i_{n1}^2 \cdot \frac{1}{\left|1 + j\frac{\omega}{BW}\right|^2} d\omega} = \frac{i_{n1}}{2}\sqrt{BW} \tag{1.42}$$

In fact, the contribution of i_{n1}^2 is significant only for high frequencies, where the SiPM equivalent impedance is really low and integrating over all the frequencies is just a worst-case scenario. The time jitter for this configuration is:

$$\sigma_t \approx \frac{i_{n1}\sqrt{BW}}{2\frac{Q_F}{\tau_F}BW} = \frac{i_{n1} R_{IN} C_F}{2Q_F}\sqrt{\frac{1}{BW}} = \frac{i_{n1} C_F}{2 g_m Q_F}\sqrt{\frac{1}{BW}} = \frac{e_{n1} C_F}{2Q_F}\sqrt{\frac{1}{BW}} \tag{1.43}$$

In the previous expression, $e_{n1} = \sqrt{i_{n1}^2/g_m^2}$ is the square root of the gate-referred equivalent voltage noise of M_1. The equations (1.43) and (1.36) look like the same; the only difference between the voltage-mode and the current mode read-out is the bandwidth. Generally, it is easier to design current amplifier than a voltage amplifier with the same bandwidths BW; consequently, with this architecture, it is also easier to have a lower jitter and better time performances.

In conclusion, this approach is the best among the three described in this paragraph for two reasons: first of all, decreasing the input resistance R_{IN} involves several advantages, such as faster rising edge and higher peak of the output current; moreover, the possibility of copying the current in multiple paths makes this architecture the most versatile. Indeed, this is the most commonly used architecture for SiPM front-ends [20].

1.9.4 Influence of parasitic components

The SiPM model by Corsi et al. outlined in the previous paragraphs well describes the output current pulses generated by the detector. However, in order to enhance

the accuracy, further components must be considered during the design phase of the front-end, especially when time resolution is a key parameter of the system.

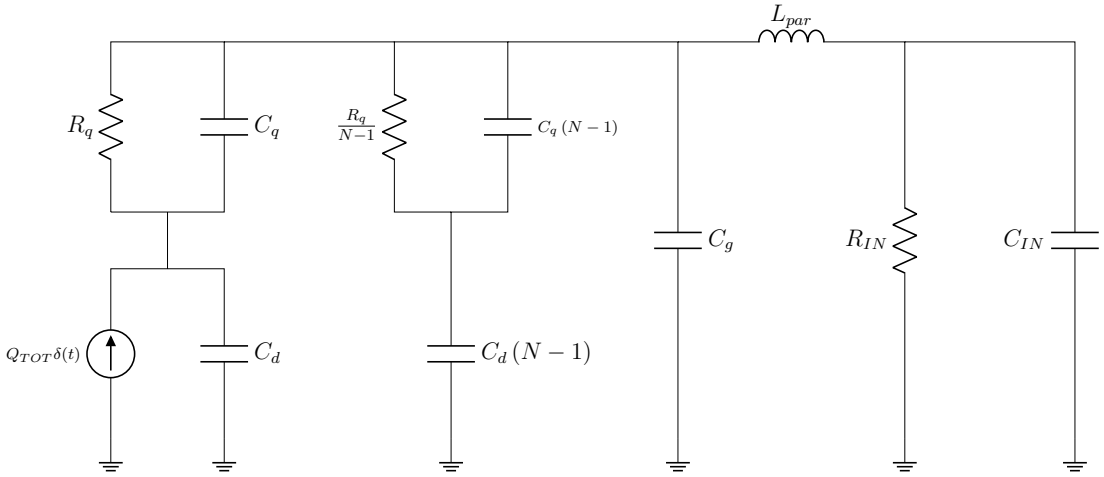

Figure 1.19: SiPM electrical model by Corsi et al. with the parasitic inductance.

Indeed, it has been observed that the simulations with the model as it is reported in figure 1.10 are not able to exactly reproduce the output signals obtained from measurements because the rising edge appears slower than one would expect. Nonetheless, this effect is not due to the SiPM itself, nor to the model, but to the interconnections between the sensor and the readout circuit, which affect the shape of the current by introducing a parasitic inductance L_{par}. The complete model of a silicon photomultiplier connected to a generic front-end circuit with an input impedance $R_{IN} - C_{IN}$ is shown in figure 1.19.

This parasitic inductance cannot always be neglected because its value can be between few nanohenries and several tens of nanohenries, depending on the trace lengths on the printed circuit board where the SiPM is mounted, on the wire bond structures and also on the traces inside the ASIC. L_{par} especially acts at high frequencies, and thus during the fast initial transient, decoupling the "fast" equivalent capacitance C_F of the SiPM from the electronics input impedance. This results in a decrease of the current signal slope, which negatively affects system performances in terms of time resolution according to the equation (1.21).

By analyzing the circuit in figure 1.19, it is possible to derive the transfer function of the front-end input current with respect to the total charge generated by the SiPM:

$$\frac{I_{in}\,(s)}{Q_{TOT}} = \frac{\tau_q s + 1}{a_4 s^4 + a_3 s^3 + a_2 s^2 + a_1 s + 1} \tag{1.44}$$

where and $\tau_q = R_q C_q$, while the coefficients at the denominator are given by:

$$\begin{aligned} a_1 &= \tau_{IN} + \tau_r + R_{IN}\left(C_g + NC_d\right) \\ a_2 &= \tau_{IN}\tau_r + C_g\left(R_{IN}\tau_r + L_{par}\right) + NC_d\left(R_{IN}\tau_q + L_{par}\right) \\ a_3 &= L_{par}\left[\left(C_g + NC_d\right)\left(\tau_{IN} + \tau_q\right) + R_q C_d C_g\right] \\ a_4 &= L_{par}\tau_{IN}\left[\left(C_g + NC_d\right)\tau_q + R_q C_d C_g\right] \end{aligned} \tag{1.45}$$

The time constant $\tau_{IN} = R_{IN}C_{IN}$ is associated to the front-end input impedance. The transfer function (1.44) is a fourth-order equation that is hard to handle by hand; therefore, in order to analyze or design a SiPM front-end circuit, it is convenient to use a circuit simulator or a numerical-analysis software. However, as far as jitter estimation is concerned, it is also possible to carry out an approximate analysis just taking into account the high-frequency equivalent Norton model of SiPMs (cf. figure 1.20), which consists of the "fast" capacitance C_F in parallel to a current generator $I_N(s)$:

$$I_N(s) = Q_F \frac{C_q}{C_q + C_d} \tag{1.46}$$

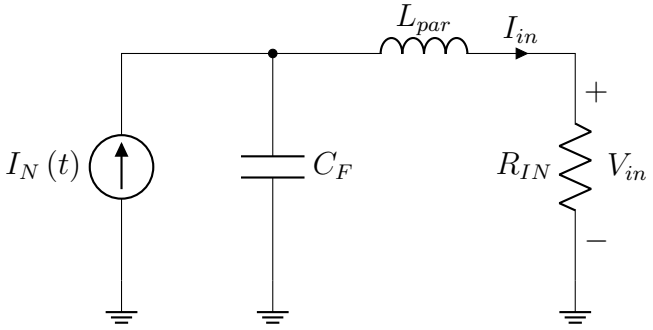

Figure 1.20: High-frequency equivalent Norton Model of a SiPM coupled with a generic input resistance R_{IN} with the parasitic inductance L_{par}.

Because the input resistance R_{IN} is commonly small, the front-end input capacitance can be neglected. The transfer function $I_{in}(s)/Q_{TOT}$ is a second-order low-pass filter:

$$\frac{I_{in}(s)}{Q_{TOT}} = \frac{1}{s^2 + \frac{\omega_0}{Q}s + \omega_0^2} \tag{1.47}$$

whose Q and ω_0 parameters are:

$$\omega_0 = \frac{1}{\sqrt{L_{par}C_F}} \qquad\qquad Q = \frac{1}{R_{IN}}\sqrt{\frac{1}{C_F}} \tag{1.48}$$

In this case, it is important to note that the quality factor Q is inversely proportional

to the R_{IN}, while ω_0 is not function of any front-end parameter. Therefore, decreasing the value of the input resistance entails a larger peaking time, which makes the system globally slower.

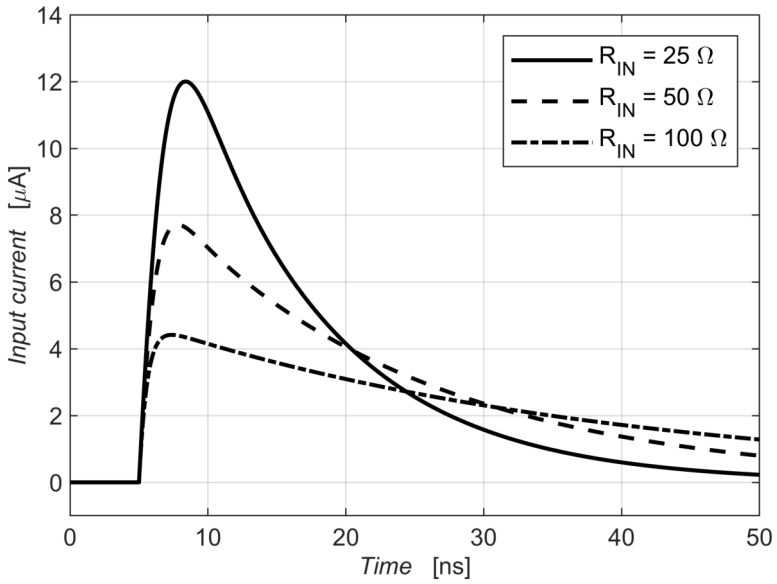

Figure 1.21: Input current for three different values of R_{IN} with a parasitic inductance of 50 nH.

This effect can be noted in the figure 1.21, where the three plots represent the front-end input currents for three different values of R_{IN} (25 Ω, 50 Ω, 100 Ω) with a 50 nH inductance between the SiPM and the electronics. Also here, as already mentioned for the current-mode readout, in correspondence of lower value of R_{IN}, the maximum current is higher, thus the reduction of R_{IN} can still be beneficial in terms of slope of the output pulse, due to its increased amplitude [21].

Another element that should not be overlooked is that the input current has a finite rise time of few nanoseconds, unlike what happened with the model without L_{par}. This implies that it is possible to identify a value of the front-end bandwidth, beyond which the input current slope does not increase and no improvements occur in terms of time performances; in fact, increasing BW the integrated noise on the output node rise and the time jitter worsen. Throughout the design of the readout electronics, it is thus necessary to find the values of input resistance and bandwidth that minimize the time jitter [22].

1.9.5 Analytic approximation

As already stated above, in order to analyze or design a SiPM readout circuit taking into account the parasitic components, it is possible to use the approximate model

presented in the previous section. However, such model does not usually guarantee the degree of accuracy required in most of the applications, especially when high time resolution is demanded. Moreover, since the frequency response of the electronics is a determining factor in the system performances, at least the dominant pole of the front-end needs to be considered in the equation (1.44), which makes the SiPM transfer function $I_{in}(s)/Q_{TOT}$ a fifth-order equation, even harder to manage by hand.

In this paragraph, a new and simpler analytic expression describing the rising edge of the SiPM current signal has been proposed. It allows a better understanding on how the front-end parameters and the parasitic components affect the photocurrent generated in response to the absorption of a single photon.

The analysis has been carried out not only considering the parasitic inductance L_{par} associated to the wires between the detector and the electronics, but also a further parasitic resistance R_{par} in series to the SiPM cathode, which represents the substrate ohmic resistance. It has been proven that adding this component to the model enables a better fitting between simulations and measurements. The complete model considered in this study is shown in figure 1.22.

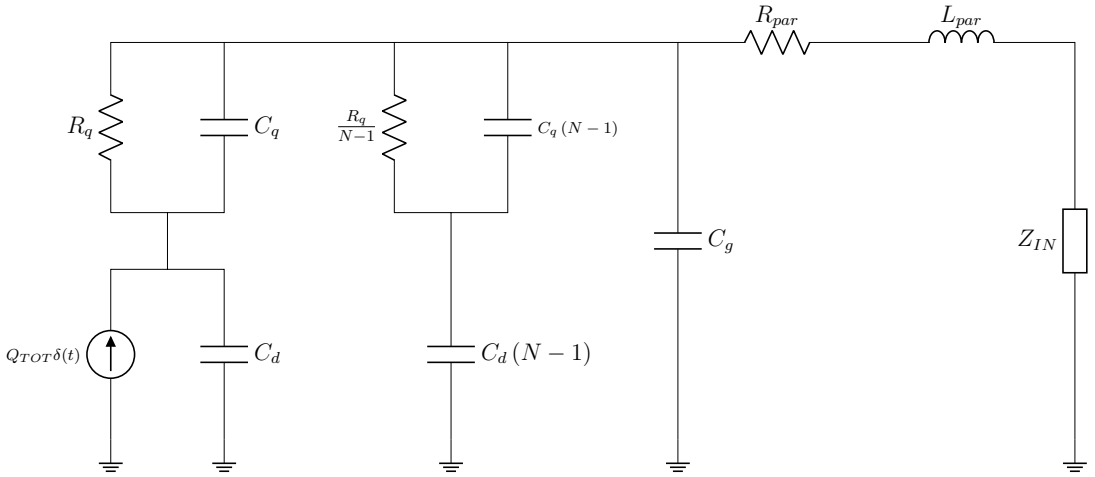

Figure 1.22: SiPM electrical model by Corsi et al. with the parasitic inductance and the substrate ohmic resistance.

As reported in [23], the analytic approximate expression of the front-end input current generated by the SiPM, neglecting the slow second order term, is given by:

$$I_{in}(s) \approx V_0 \cdot \frac{1+\tau_q s}{s(1+\tau_q s)} \cdot \frac{1}{R_s} \cdot \frac{\omega_n^2}{(s^2+2\zeta\omega_n s+\omega_n^2)} \tag{1.49}$$

In the previous equation $\tau_r = R_q(C_q + C_d)$ is the time constant associated to the SiPM recovery phase, $R_S = R_{IN} + R_{par}$ is the series between the input and the parasitic resistances, while the other parameters are:

$$V_0 = \frac{Q_{TOT}}{C_g + NC_d} \qquad \tau_p = \frac{C_g\tau_r + NC_d\tau_q}{C_g + NC_d}$$
$$\omega_n^2 = \frac{Rs}{L_{par}\tau_{IN}} \qquad 2\zeta\omega_n = \frac{1}{\tau_{IN}} + \frac{R_{par}}{L_{par}}$$

Let us now consider, following the current-mode approach, a generic transimpedance amplifier as front-end circuit, with a gain K_L and a bandwidth $BW = 1/2\pi\tau_L$. The resulting approximate expression of its output voltage in response to a single-photon current signal is:

$$V_{out}(s) \approx V_0 \cdot \frac{1+\tau_q s}{s(1+\tau_q s)} \cdot \frac{\omega_n^2}{R_s(s^2 + 2\zeta\omega_n s + \omega_n^2)} \cdot \frac{K_R}{1+\tau_L s} \tag{1.50}$$

The equation (1.50) consists of the product of three different functions, each of which is related to a specific section of the circuit under examination. The above-mentioned functions are:

$$G_1(s) = V_0 \cdot \frac{1+\tau_q s}{s(1+\tau_q s)} \quad G_2(s) = \frac{\omega_n^2}{R_s(s^2 + 2\zeta\omega_n s + \omega_n^2)} \quad G_3(s) = \frac{K_R}{1+\tau_L s} \tag{1.51}$$

The first one $G_1(s)$ is associated to the SiPM model, $G_2(s)$ takes into account the parasitic components due to the non-idealities of the detector and its interconnection with the front-end, and thirdly $G_3(s)$ is the transfer function of the transimpedance amplifier.

Taking as a starting point the expression (1.49), it is possible to make some practical considerations. First of all, the time constant of the input low-pass band filter $\tau_{IN} = R_{IN}C_{IN}$ is commonly much smaller than L_{par}/R_{par} and consequently it is possible to approximate $2\zeta\omega_n \approx 1/\tau_{IN}$. Then, after calling $\tau_A = L_{par}/R_s$, the expression of $G_2(s)$ becomes roughly equal to:

$$G_2(s) = \frac{\omega_n^2}{R_s(s^2 + 2\zeta\omega_n s + \omega_n^2)} \approx \frac{\dfrac{1}{\tau_{IN}\tau_A}}{R_s\left(s^2 + \dfrac{s}{\tau_{IN}} + \dfrac{1}{\tau_{IN}\tau_A}\right)} \tag{1.52}$$

Also in this case, it is possible to prove that in real situations $\tau_{IN} \ll \tau_A$ almost always, and that the pole associated to the preamplifier time constant in negligible. Therefore, the equation (1.52) is reducible to an approximate first-order system, whose expression is:

$$G_2(s) = \frac{\omega_n^2}{R_s\left(s^2 + 2\zeta\omega_n s + \omega_n^2\right)} \approx \frac{1}{R_s\left(1 + \tau_A s\right)} \tag{1.53}$$

Additional comments concerning $G_3(s)$ can be made by considering that at high frequency the following approximation is valid:

$$\frac{1 + \tau_q s}{1 + \tau_p s} \approx \frac{\tau_q}{\tau_p} = \alpha \tag{1.54}$$

This result implies that the equation (1.50) can be rewritten as:

$$V_{out}(s) \approx \frac{\alpha V_0}{R_s} \cdot \frac{1}{s\left(1 + \tau_A s\right)} \cdot \frac{K_R}{1 + \tau_L s} \tag{1.55}$$

Recalling the equation (1.21), by which the time resolution of a SiPM detection system is defined, it is necessary to derive the expression of the output voltage slope in the time domain starting from the equation (1.55). Since in the Laplace domain the derivative of a function is obtained by multiplying the same function by s, the output signal slope is therefore:

$$V_{out}^{(1)}(s) \approx \frac{\alpha V_0 K_R}{R_s} \cdot \frac{1}{\tau_A \tau_L} \cdot \left(\frac{a}{s + \dfrac{1}{\tau_A}} + \frac{b}{s + \dfrac{1}{\tau_L}} \right) \tag{1.56}$$

where

$$a = -\frac{\tau_A \tau_L}{\tau_L - \tau_A} \qquad\qquad b = \frac{\tau_A \tau_L}{\tau_L - \tau_A} \tag{1.57}$$

Finally, the slope expression of the output voltage is given by applying the inverse Laplace transform to (1.56).

$$\frac{dV_{out}(t)}{dt} = \mathcal{L}^{-1}\left\{V_{out}^{(1)}(s)\right\} = \frac{\alpha V_0 K_R}{R_s} \cdot \frac{1}{\tau_A \tau_L} \cdot \left(a e^{-\frac{t}{\tau_A}} + b e^{-\frac{t}{\tau_L}}\right) \tag{1.58}$$

As already stated before, in order to obtain the best time resolution, it is recommended to locate the threshold where the output signal has the maximum slope. This latter can simply be derived, by substituting in the previous equation the instant of time t_{MAX} where the function has a maximum. Its value can be evaluated by deriving the expression (1.58), equating it to zero and solving for t:

$$t_{MAX} = \frac{\tau_A \tau_L}{\tau_A - \tau_L} \ln \frac{\tau_A}{\tau_L} = \frac{\tau_A}{\Theta - 1} \ln \Theta \tag{1.59}$$

where $\Theta = \tau_A / \tau_L$. In conclusion, the maximum slope of the single-photon output signal

of a SiPM readout circuit is:

$$\left.\frac{dV_{out}(t)}{dt}\right|_{t=t_{MAX}} = \frac{\alpha V_0 K_R}{R_s} \cdot \frac{1}{\tau_A} e^{-\frac{t_{MAX}}{\tau_A}} \tag{1.60}$$

The previous equation can be further rearranged, by substituting the equation (1.59) in (1.60):

$$\left.\frac{dV_{out}(t)}{dt}\right|_{t=t_{MAX}} = \alpha V_0 \frac{A_I R_L}{L_{par}} \cdot \Theta^{\frac{1}{1-\Theta}} \tag{1.61}$$

where the preamplifier gain K_R has been replaced with the product between its current gain A_I and the load resistance R_L.

In order to validate the results just presented, a comparison between the maximum slope obtained by means of the complete and the proposed model has been carried out. As evidenced by the plot in figure 1.23, the equation (1.61), despite being of three order lower, approaches reasonably good the maximum slope obtained using the complete model; therefore, it might be used to estimate the maximum slope of the output signal of a analog front-end for SiPM, and thus its time jitter, during the design phase.

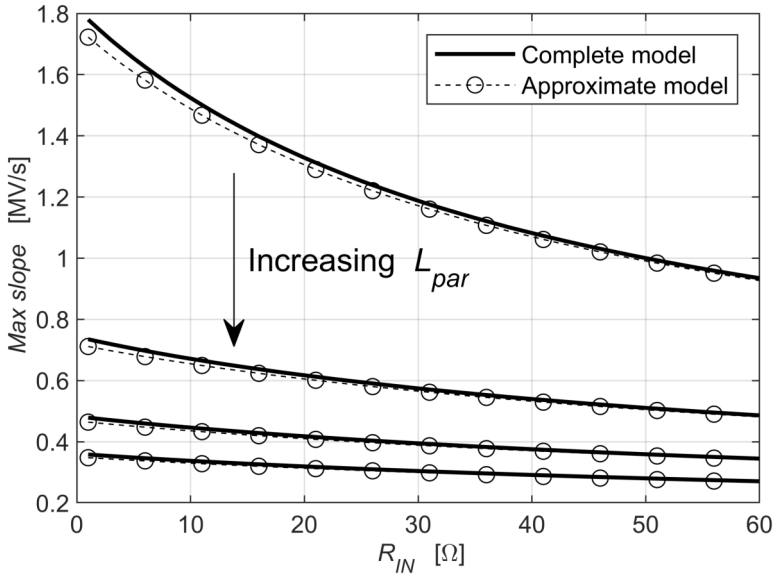

Figure 1.23: Comparison between the maximum slopes of the complete model and the approximation, using 4 different values of inductance L_{par} (10 nH, 40 nH, 70 nH and 100 nH) and keeping constant the input capacitance ($C_{in} = 0.5$ pF) and the bandwidth ($BW = 500$ MHz).

As a conclusion of this analytic work, a simple transimpedance amplifier has been designed and implemented on a printed circuit board, with the aim of empirically assessing the effectiveness of the proposed model. The schematic of the circuit is

reported in figure 1.24, where a cascoded bipolar transistor Q_1 has been employed in the common base configuration.

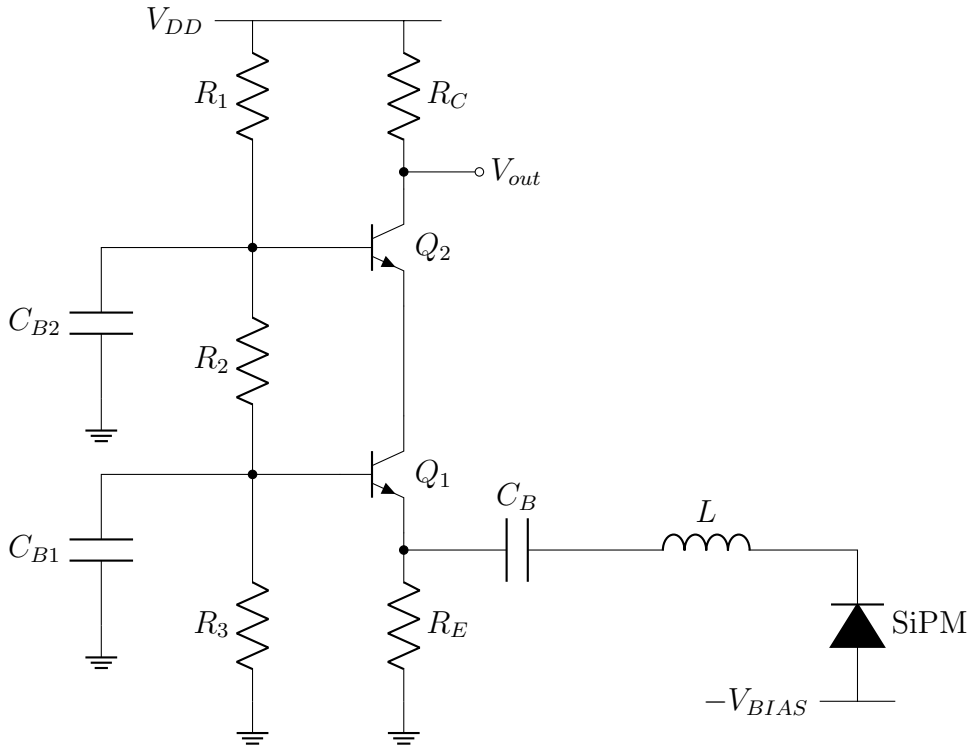

Figure 1.24: Schematic of the transimpedance amplifier implemented on the PCB.

This circuit has been designed in such a way that it was possible not only to pick one among four different values of the input resistance (10 Ω, 18 Ω, 33 Ω and 50 Ω), but also to select the desired inductance between the SiPM and the input terminal of the front-end. In particular, for this work it has been chosen to observe the behavior of the circuit considering three values of L_{par} commercially available: 0 nH, 51 nH and 100 nH. The measurement setup arranged for the test of this prototype mainly consists of a pulsed laser source, whose laser beam is conveyed to the SiPM (mounted on the PCB) by means of an optical fiber and a diffuser. For each of the 12 configurations (4 resistances and 3 inductors), the front-end output voltage has been amplified, acquired and then post-processed in order to filter out all the waveform not related to a single-photon event. At this stage, with the purpose of comparing the result of the proposed model with the measurement data, an estimation of the maximum slope at the output terminal of the circuit has been extracted and then plotted in figure 1.25 along with the ones obtained by means of equation (1.61). The graph highlights a good agreement between the results of measurements and the presented model.

It is appropriate to clarify that the values of L_{par} inserted in the analytic model do not matches with those of the discrete components mounted on the board, but they

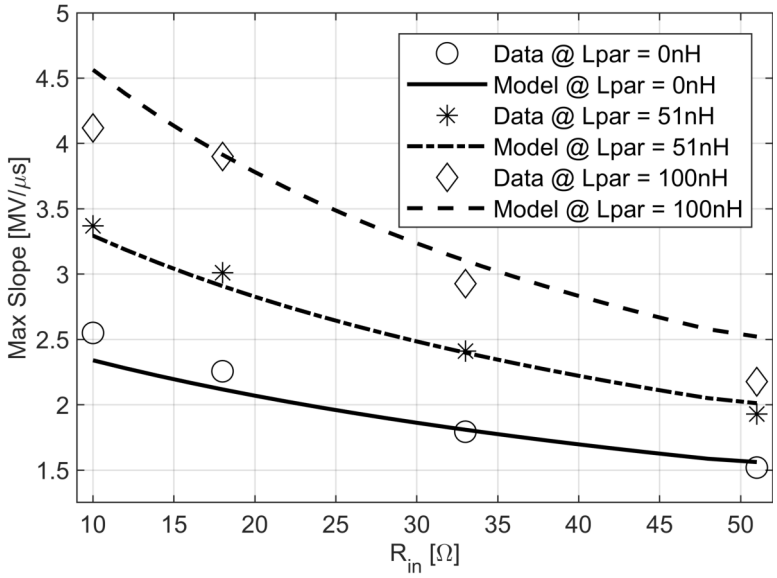

Figure 1.25: Comparison between the maximum slopes of the proposed model and the results of measurements, using 3 different values of inductance L_{par} (0 nH, 51 nH and 100 nH).

are slightly higher. This is firstly due to the tolerance of the devices employed, but also due to the parasitic inductance introduced by the tracks and the PADs along the way between the sensor and the emitter of the BJT. Moreover, after estimating the noise budget of the circuit, it has also been verified the compliance of the measurements with the model in terms of time jitter; even in this case, the empirical data reproduce the analytic ones with an error lower than 5 %.

A similar investigation might be carried out following the voltage-mode approach, using a voltage amplifier with an input resistance R_{IN} and a gain A_V in the place of a transimpedance one. The equation that approximates the value of the maximum slope of the output signal is:

$$\left.\frac{dV_{out}(t)}{dt}\right|_{t\,=\,t_{MAX}} = \alpha V_0 \frac{A_V R_{IN}}{L_{par}} \cdot \Theta^{\frac{1}{1-\Theta}} \tag{1.62}$$

Also in this case, the previous expression well approximates the maximum slope derived with the complete model. Nevertheless, although from equation (1.62) it appears that the maximum slope grows linearly with the input resistance, it is not feasible to increase indefinitely R_{in}, because, as emphasized in [23] and already stated in the previous sections, beyond a certain point it would limit the bandwidth of the amplifier, slowing down the output signal and thus lifting the jitter of the front-end. Moreover, it is worth recalling that, increasing the input impedance of a voltage-amplifier, the output signal would have not only a higher peak and slope, but also a longer tail, which

implies a longer recovery time, that reduces the maximum event rate that the system may handle. In summary, the current-mode approach offers a number of advantages in the design of a front-end for SiPM and this explains why it is the most employed technique for this type of circuits and it has been adopted for the design of the front-end described in this work.

However, before proceeding with the description of the novel preamplifier for silicon photomultipliers, it has been decided to present an example of readout circuit, which has been developed following the current-mode architecture and successfully tested.

1.10 An example of readout circuit: *BASIC64*

After discussing about the advantages of the current-mode approach over the other two techniques, an example of front-end preamplifier, where the SiPM signal is read by means of a current buffer, is presented. It is incorporated in a multichannel ASIC, called *BASIC64*, developed within the framework of the *INSIDE* (INnovative Solutions for Dosimetry in hadrontherapy) project. It not only allows the reading of the SiPM current pulses, but also provides the conversion and the transmission of the digitize measured data associated to a valid even detected by the readout chain.

The simplified preamplifier schematic of the analog channel is reported in figure 1.26. As it is possible to notice, the SiPM input current is split up into two different branches respectively for timing and energy measurements. Each branch consists of a common-gate transimpedance amplifier, whose gate voltage is controlled by an 8-bit DAC (Digital-to-Analog Converter), which enables the fine tuning of the input transistor and the SiPM operating point. The two MOSFET share the same bias current source I_{BIAS}, whose current is distributed proportionally to the device aspect ratios $S = W/L$, as well as the SiPM input current. In this circuit the ratio of S_T/S_C is equal to 20, in order to optimize time measurements, while guaranteeing that the specifications in terms of energy resolution and dynamic range are satisfied.

As far as the photon occurrence time is concerned, the current is converted into an equivalent voltage signal by means of a resistor R_T and a voltage leading-edge discriminator has been employed to identify the occurrence time of the photon absorption. Moreover, a clipping circuit, made of the p-type diode connected transistor M_D, has been included in parallel to the load resistor in such a way that in case of large input signals M_T does not end up working in triode region.

On the other and, regarding energy measurements, one-twentieth of the SiPM current flows through M_C and it integrated by means of the low-pass band filter $R_{INT}-C_{INT}$. Thereafter, V_C is sent to a peak detector that identifies and hold the peak

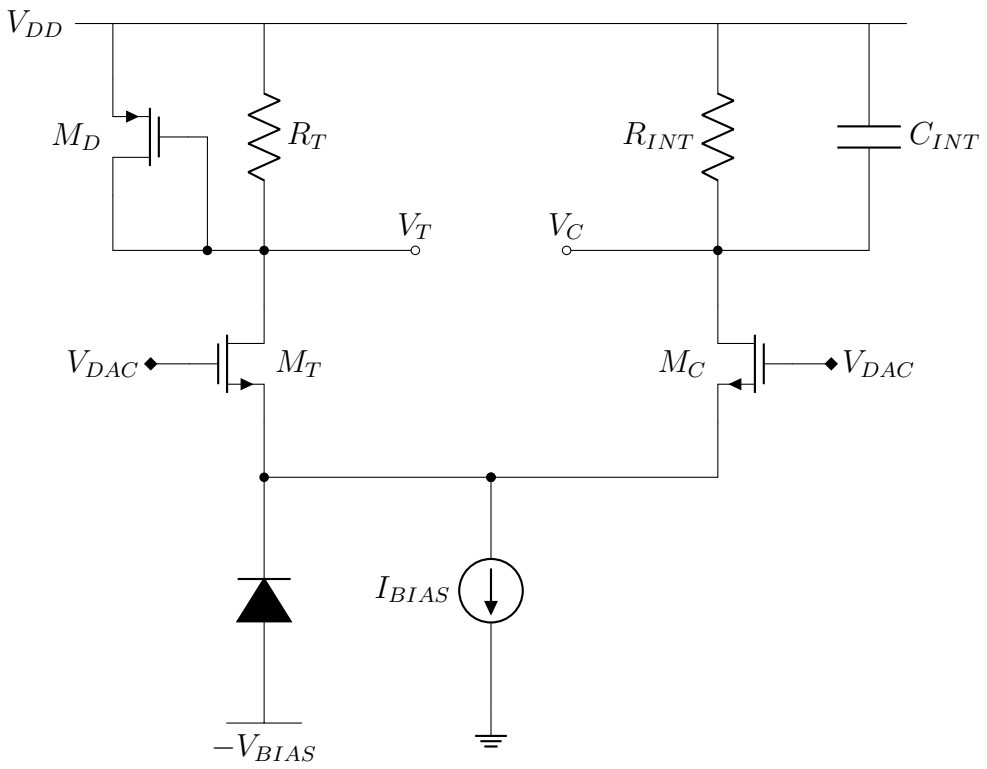

Figure 1.26: Preamplifier front-end circuit in BASIC64.

of the signal and thus converted thanks to the on-board 8-bit ADC (Analog-to-Digital Converter) [24].

In order to carry out time and energy measurements, an external pulser has been used to simulate the charge injection in the readout circuit through a 1 nF capacitor. In this circuit is possible to vary the input charge from 16 pC to 576 pC with a resolution less than 10 fC. Additionally, all the configuration signals and the output data have been sent and read by means of an FPGA development board.

The input signal has been injected in the electronics for multiple times and with different amplitudes, so that it has been possible to estimate the time jitter. Carefully selecting the discriminator voltage threshold, a σ_t of about 40 rms has been obtained, even with low values of charge. Moreover, the non-linearity error of the detection chain has been measured by changing the input signal from the lowest possible value to the highest. The characteristic peak detector output voltage - input charge exhibits a gain of 3 mV/pC with a non-linearity error lower than 2 % in the whole range, in good agreement with simulations. The measured performances of the BASIC64 are compliant with the requirements of the PET designed in the framework of the INSIDE project [25].

1.11 Positron Emission Tomography

Due to their extraordinary properties, SiPMs can be employed in a wide range of applications, such as 3D-Ranging, high-energy physics, biophotonics and medical imaging. The latter category includes the Positron Emission Tomography that represents the main application to whom this thesis project is addressed. Therefore, before wrapping this introductory chapter up, it seemed necessary to give a brief overview about its working principle.

1.11.1 Overview

Positron Emission Tomography (PET) is an imaging technique used in nuclear medicine and it is able to examine the activities of certain biological substances in the human bodies and detect their quantities. For each kind of analyses (neurological, cardiological, oncological, etc.), based on the molecules of interest, a different kind of radiotracer is employed. The most common isotopes are carbon-11 (^{11}C), oxygen-15 (^{15}O), nitrogen-13 (^{13}N) and fluorine-18 (^{18}F); they bind to specific particles and, without altering the physiological activities, allow the study of the behavior of the molecules and diagnosing certain kind of diseases.

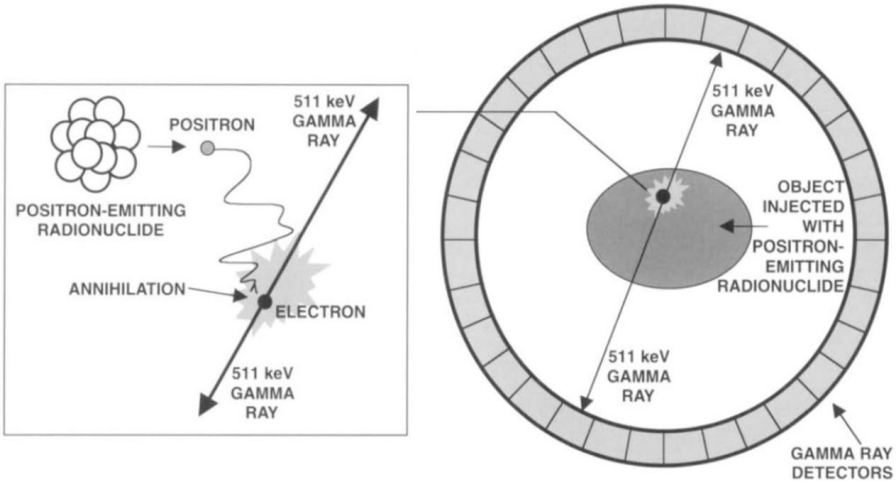

Figure 1.27: Overview of a PET scanner operation.

Positrons, naturally emitted by radioactive drugs, can travel thought body tissues, typically for few millimeters, until they meet an electron. These two particles can interact and create an exotic atom, called *positronium*. It is an extremely unstable system that may decay by annihilation and generate a pair of gamma photons, with an energy of 511 keV, which move in opposite directions. PET sensors are usually mounted on a ring structure, so that the two radiations are detected by collinear devices and the

reconstruction of their starting point is easier with *time of flight* measurements. The figure 1.27 gives an overview of the decay and annihilation phenomenon.

The system output data are collected and processed by a computer that, considering the photons directions and times of arrival, applies recursive and statistical algorithms and allows the production of images of the biological tissues under analysis.

Modern PET scanners are designed to operate even with light beams coming from very oblique directions and make the generation of 3D pictures possible [26]. Nevertheless, they require high-quality detectors with good PDE and with excellent energy and time resolutions. The first sensors to be employed in PET systems were photomultipliers tubes, but in recent years SiPMs have become a good alternative to PMTs, thanks to their very good timing performance. However, SiPMs do not have a good detection efficiency at gamma frequencies and, in order to solve this drawback, another device, called *scintillator*, must be inserted on the ring structure.

1.11.2 Scintillators

Scintillator crystals are used to perform the conversion of impinging gamma photons into visible light, at which SiPMs exhibit high photon detection efficiency. Therefore, in PET scanners, silicon photomultipliers are coupled to a scintillator to constitute what is called *scintillator detector*. The main parameters of these devices are:

- *Decay time*: it is the average time between the absorption of a gamma photon and the emission of light in the visible spectrum;
- *Energy resolution*: it is defined as the slightest energy change that can be distinguished at the output;
- *Stopping power*: it is the absorption efficiency of gamma photons;
- *Scintillation efficiency*: it is defined as the number of emitted photons per unit of absorbed energy.

In table 1.2 the parameters of few scintillators commercially available are reported.

The working principle of scintillators is based on the luminescence effect, proper to certain kind of inorganic crystals. The energy band diagram, reported in figure 1.28, may help to illustrate this phenomenon.

As it happens in photodiodes, an electron-hole pair may be generated in response to a photon absorption, but each carrier recombines after a certain interval that depends on their mean lifetime in the material. In ideal crystals, where there are no defects, the carrier recombination may create other photons with a frequency pretty close to

Table 1.2: Parameters of some scintillators commercially available.

Crystal	Density $[g/cm^3]$	Decay time $[ns]$	Total light output N_{ph}/MeV	Energy resolution @ 662 keV [%]
BGO	7.1	300	6000	10.2/20
$LYSO$	7.1	41	32000	–
LSO	7.4	40	32000	10.0
$LaBr_3$	5.1	16	63000	2.9
NaI	3.7	230	38000	6.6
BaF_2	4.9	0.8	12000	11.4

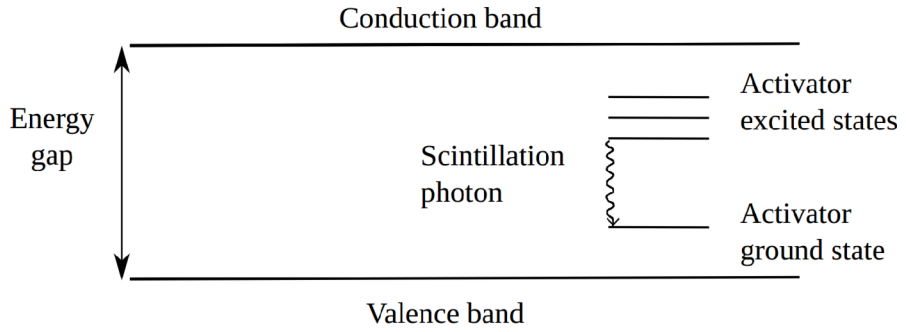

Figure 1.28: Energy band structure in a crystalline scintillator.

that of incident light. Nevertheless, scintillators are doped with impurity atoms, called activators, which create luminescence centers in the band gap. Therefore, excited electrons may decay in the valence band through these metastable states, emitting photons in the visible spectrum [27].

Chapter 2

A new readout channel for SiPMs

2.1 Introduction

The research project discussed in this doctoral thesis is the result of the ongoing collaboration between the *Integrated Circuit Department* of SLAC TID-AIR (Technology Innovation Directorate - Advanced Instrumentation for Research) based in Menlo Park, California, and the *Applied Electronics and Microelectronics group* of the Politecnico di Bari. The main target of this partnership has been the design and realization of a new state-of-art front-end integrated circuit for reading out the current signal generated by silicon photomultipliers used in PET scanners.

During the first meetings before the start of the project, the system requirements have been defined, also taking into account the PET specifications recommended by the division of Nuclear Medicine at Stanford University. This group indeed is currently working on a breakthrough instrumentation for tomography using SiPMs and, if the measurements will show good matching with the simulations, the proposed readout circuit may be a good candidate to replace the electronics now employed.

The front-end circuits for PET systems need to provide both time and energy information of the detected event with the best resolution possible and, in this case, a single-photon time jitter σ_t of 10 - 20 ps FWHM has been the target. Instead, as regards to energy measurements, the integration chain must guarantee a good linear response until at least 1300 pC, which corresponds to around 8000 photoelectrons if a SiPM with a gain of 10^6 is considered. Another important characteristic that this circuit must present is related to the analog interface between the detector and the preamplifier; the front-end needs to be fully differential, which means providing differential signals for both the input and the output. As far as the inputs are concerned, both SiPM terminals, cathode and anode, must thus be connected to the front-end with an AC coupling circuit that allows the rejection of the DC component of the SiPM bias.

The purpose of this structure is to reject the common-mode noises and make the circuit more robust against the interference due to the presence of digital circuit and clocks, especially in this application where the time resolution is a key parameter of the system. Moreover, it has been decided that the target CMOS technology was the well-established TSMC 130 nm, whose power supply for core circuits is 1.2 V. In this preliminary phase of the design there are no restriction about the power consumption, but usually it is a good practice to limit the preamplifier power to few tens of milliwatts in order to avoid the device overheating, thereby affecting the system performances, and to reduce the power consumption of the whole system.

As already stated in the previous chapter, the current-mode readout represents the most versatile and appropriate approach to achieve good time resolutions, because it is possible to increase both the SiPM current slope and its peak value by reducing the front-end input resistance. For this reason, it has been decided to design the preamplifier by adopting this architecture, which is the most widespread among all the other solutions available in the literature.

After briefly presenting the channel architecture chosen for this project, the description of each circuit block will be provided, by getting more into the details of those that are crucial for system performances. Besides the preamplifier, three main groups of circuits can be distinguished in the channel:

- The "fast path": it comprises the blocks that allow the identification of the occurrence time of the detected event;
- The "slow path": it includes the circuits involved in energy measurements;
- The digital state machine: it handles all the digital signal inside the channel and guarantees the proper functioning of some analog blocks.

The block diagram of the proposed SiPM channel is reported in figure 2.1 and, as observed, the differential preamplifier provides two signals for the two different paths. The time information is given by the "fast path", which consists of a Leading-Edge Discriminator (LED) and a Time-To-Digital Converter (TDC). On the other hand, the signal in the "slow path" goes through a voltage-to-current converter before reaching the integrator, whose output peak voltage is proportional to the charge contained in the input current pulse. A peak detector has been thus implemented to store the maximum value of the integrator output signal, until an external Analog-to-Digital Converter digitizes the energy information. Moreover, a digital state machine has been designed to filter out false positive event, to manage the communication with the external logic and to allow the channel to return to the initial state once the conversion has been carried out.

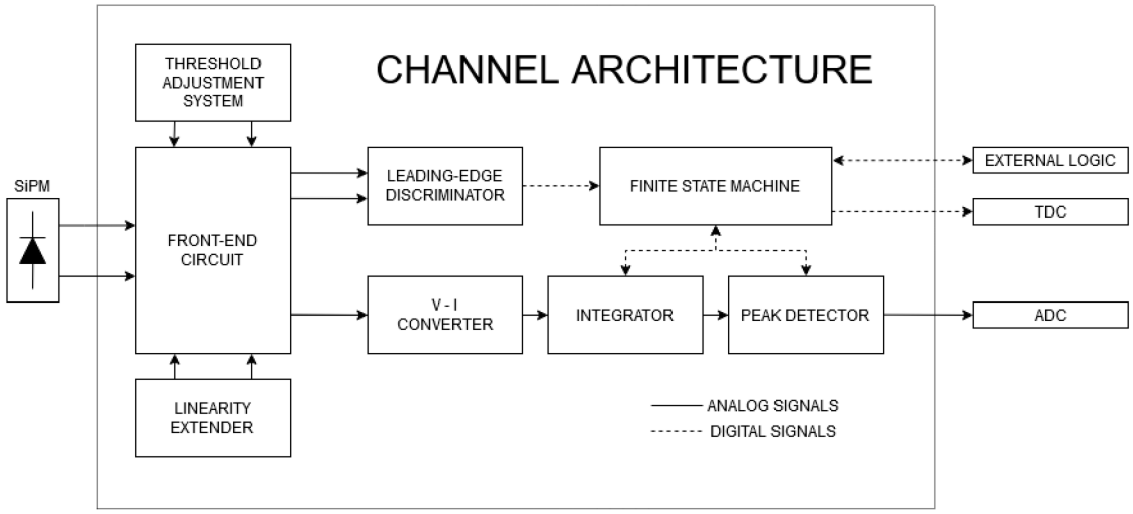

Figure 2.1: Block diagram of the channel.

2.2 Preamplifier

In order to choose the best topology for the preamplifier, several architectures have been studied and their performances have been analyzed. Eventually, the *Regulated Common-Gate Transimpedance amplifier* (RCG-TIA) has been deemed the most appropriate solution to be implemented in both the branches of the differential front-end circuit, due to its good features in terms of low input resistance R_{IN} and noise [28].

2.2.1 Regulated common-gate transimpedance amplifier

Before starting with the project of the differential structure, it is convenient to analyze the properties of the regulated common-gate transimpedance amplifier, and in particular its transfer function, its input resistance and its noise. As shown in the schematic reported in figure 2.2, the RCG-TIA is made up of a common-gate MOSFET M_1, biased by the current source I_{B1}, to which a feedback network, consisting of the common-source transistor M_2, has been added.

As far as the frequency response of the RCG-TIA is considered, it may be useful to make some considerations, since it is not easy to obtain a closed-form expression of $V_{out}(s)/I_{in}(s)$. At low frequencies, the circuit acts as perfect transimpedance amplifier, where the input current I_{in} is fully converted into a voltage signal $V_{out} = I_{in}R_L$ by means of the load resistor. In order to analyze the circuit behavior at higher frequencies, instead, it is important to point out that, when it is coupled with a SiPM, the equivalent capacitance on the input node C_{SiPM} can be of several tens of picofarad. In these conditions, the transimpedance function $V_{out}(s)/I_{in}(s)$ has three dominant poles: the first one related to the input node and hence to C_{SiPM}, another one associated to the

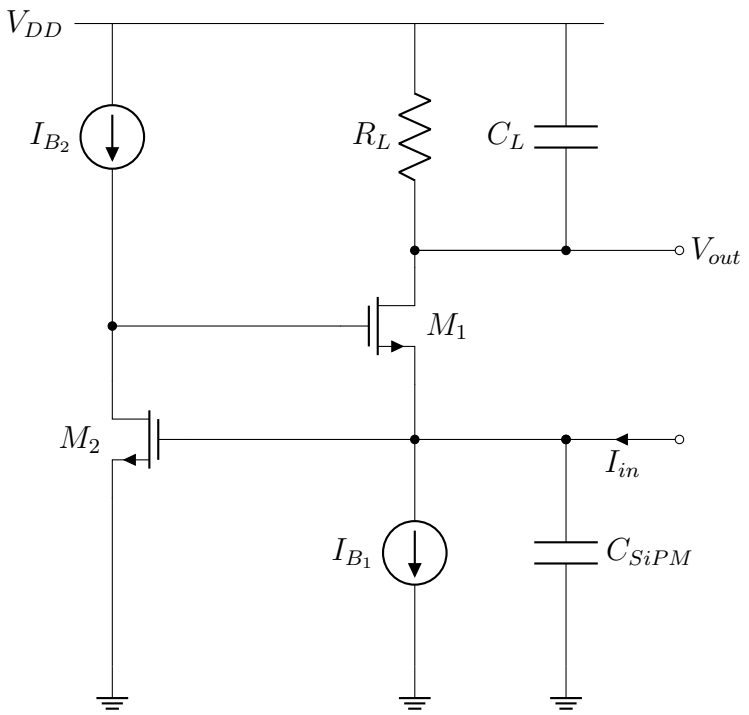

Figure 2.2: Regulated common-gate transimpedance amplifier.

feedback loop and the last one to the low-pass filter at the output node. Among these, the dominant pole is usually the one associated to the SiPM capacitance since the time constant $\tau_{IN} = R_{IN}C_{SiPM}$ is the slowest. Indeed, the value of the load capacitance C_L usually is not larger than some hundreds of femtofarads, because it is the gate capacitance of another transistor, and therefore the time constant $\tau_L = R_L C_L$ is much faster than τ_{IN}. The same comment can be done for the pole in the feedback, where the time constant τ_{FB}, due to the product between $C_{12} = C_{gs1} + C_{gd2}$ (parallel between the gate-source capacitance of M_1 and the gate-drain capacitance of M_2) and R_{O2} (parallel between the output resistances of M_2 and I_{B2}), is negligible compared to τ_{IN} even though, as it will be mentioned soon, the size of these MOSFETs can be considerable. Therefore, at least for frequencies lower than the first pole, the small-signal equivalent circuit can be drawn without considering C_L and C_{12}, as shown in figure 2.3.

Starting from the incremental circuit, it is possible to derive the approximate transfer function. Applying the Kirchhoff's current law (KCL) respectively to the nodes S_1, D_1 and G_1 in the circuit, the following three equations are obtained:

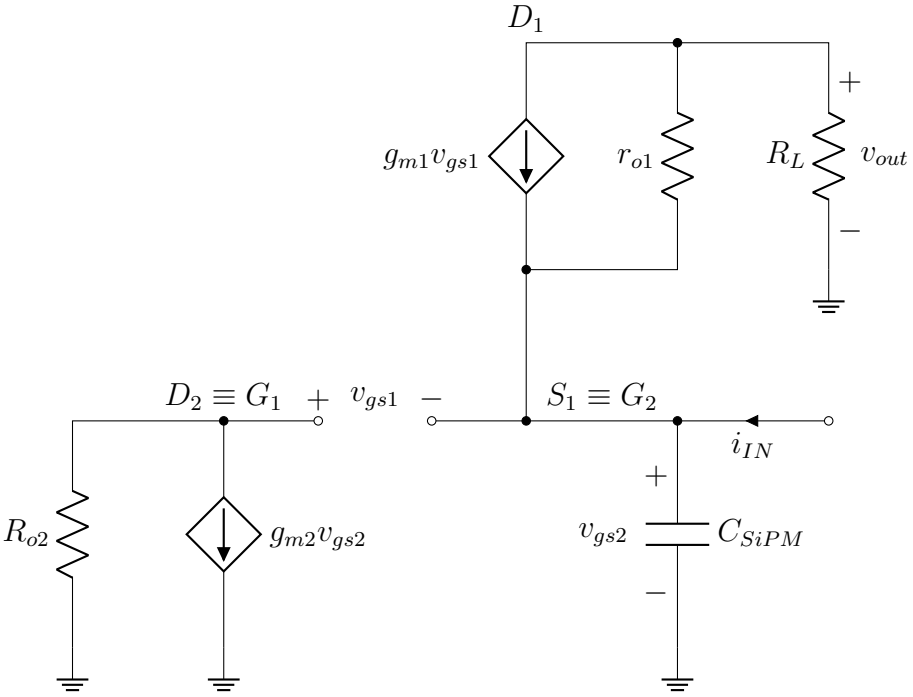

Figure 2.3: Approximate incremental circuit of the RCG-TIA in figure 2.2.

$$\begin{cases} g_{m2}v_{gs2} = -\dfrac{v_{gs1} + v_{gs2}}{R_{O2}} & \text{(2.1a)} \\ \dfrac{v_{out}}{R_L} + g_{m1}v_{gs1} + \dfrac{v_{OUT} - v_{gs2}}{r_{o1}} = 0 & \text{(2.1b)} \\ i_{in} = \dfrac{v_{out}}{R_L} + v_{gs2} \cdot sC_{SiPM} & \text{(2.1c)} \end{cases}$$

Considering the equation (2.1a), it may be written:

$$v_{gs2} = -\frac{v_{gs1}}{1 + g_{m2}R_{O2}} = -\frac{v_{gs1}}{1 + A_0} \tag{2.2}$$

where A_0 is nothing more than the absolute value of the common-source stage gain, which coincides with the low-frequency loop gain. From the equation (2.1b) it is possible to derive the value of v_{gs1} as function of v_{out}:

$$v_{gs1} = -v_{out}\left(\frac{1}{R_L} + \frac{1}{r_{o1}}\right)\left[g_{m1} + \frac{1}{r_{o1}\,(1 + A_0)}\right]^{-1} \approx -\frac{v_{out}}{g_{m1}R_L} \tag{2.3}$$

The last approximation can be done only assuming that $r_{o1} \gg R_L$ and $A_0 \gg 1$. Then, substituting (2.2) and (2.3) in (2.1c), the approximate transfer function is given by:

$$\frac{v_{out}(s)}{i_{in}(s)} = \frac{R_L}{1 + \dfrac{sC_{SiPM}}{g_{m1}(1 + A_0)}} = \frac{R_L}{1 + sC_{SiPM}R_{IN}} = \frac{R_L}{1 + s\tau_{IN}} \tag{2.4}$$

where the input resistance R_{IN} is:

$$R_{IN} = \frac{1}{g_{m1}} \frac{1}{1 + A_0} \approx \frac{1}{g_{m1}} \frac{1}{A_0} \tag{2.5}$$

It appears that R_{IN} is roughly A_0 times smaller than the input resistance of a simple common-gate stage and this allows the designer to decrease the input resistance without consuming an excessive amount of power. On the other hand, the loop gain is not constant in the whole range of frequencies, but its value decreases with increasing frequency, making the input resistance change. Nonetheless, this behavior of R_{IN} does not affect the current pulses, since the first pole of A_0 is usually located at a higher frequency than that of the preamplifier.

As regards to the noise analysis, there are three main noise contributions, due to the two transistors and the load resistor R_L. Actually, in this case, it is necessary to consider even the two bias sources I_{B1} and I_{B2}, because their current fluctuations can be relevant if the mirrors are not designed properly. For each noise source, its power spectral density $\overline{i_n^2}$ has been considered and its transfer function to the output of the circuit has been calculated. In this case, contrary to what has been done before, also the second pole has been taken into account, in order to obtain closed-form equations of the rms noise contributions, otherwise the result would have been an infinite value. To avoid stability and noise issues, typically this preamplifier is designed so that the second pole is located on its output node; in the following analysis the load capacitance C_L is therefore considered again.

In order to derive the rms voltage noise at the output node, the following well-known formula has been applied for each noise source:

$$v_{on,rms} = \sqrt{\int_0^\infty \overline{i_n^2} \cdot |H(f)|^2 \, df} \tag{2.6}$$

where $H(f)$ is the transfer function of the source to the output node of the circuit. However, the goal of this noise analysis is not to derive the precise value of the rms output noise, but it is to understand how the parameters of the RCG-TIA affect the noise and find some guidelines to follow during the design.

In figure 2.3 the approximate small-signal equivalent circuit has been redrawn adding the five current noise sources. Since the r_{o1} output resistance of M_1 is negligible, it has been excluded to perform the noise analysis of the circuit.

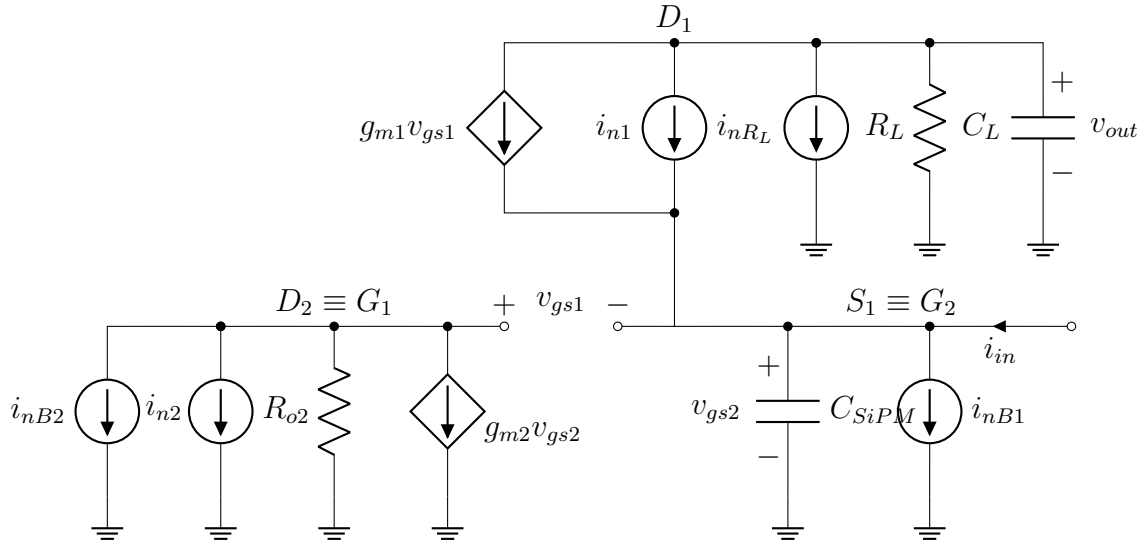

Figure 2.4: Approximate incremental circuit of the RCG-TIA in figure 2.1, including the noise sources.

As shown in figure 2.4, the noise produced by a MOSFET can be considered by adding between the drain and source terminals a current noise source with a power spectral density of:

$$\overline{i_n^2} = 4K_BT\gamma g_m \tag{2.7}$$

where γ is a coefficient that depends on the device size and on its working region; for example, for long channel MOSFET in saturation it is around 1.5, but it can be even higher than 2 for sub-micron transistors.

As regards to the noise contribution of M_1, the transfer function $H_{M_1}(s)$ has been calculated, by applying again the KCL to the three nodes of the incremental circuit:

$$H_{M_1}(s) = R_L \frac{s\tau_{IN}}{(1+s\tau_{IN})(1+s\tau_L)} \tag{2.8}$$

where $\tau_L = R_L C_L$. By substituting the (2.7) and (2.8) in (2.6), the rms output noise contribution due to M_1 is obtained:

$$v_{on,rms,M_1} = \sqrt{K_BT\gamma g_{m1}R_L^2 \frac{\tau_{IN}}{\tau_L(\tau_{IN}+\tau_L)}} \tag{2.9}$$

In the same way, the transfer function $H_{M_2}(s)$ has been calculated and it has been substituted in the equation (2.6). The rms output noise due to M_2 is:

$$v_{on,rms,M_2} = \sqrt{\frac{2K_BT\gamma R_L^2}{g_{m2}} \frac{1}{\tau_{IN}\tau_L(\tau_{IN}+\tau_L)}} \tag{2.10}$$

On the basis of equations (2.9) and (2.10), it appears that the output noise directly depends on M_1 transconductance, while it decreases with increasing g_{m2}. Therefore, it is a good practice in this case to not have a large aspect ratio for M_1 and avoid large values of its bias current. In order to decrease the input resistance, it is preferable to have a higher A_0, by using a bigger M_2.

As far as the current bias I_{B2}, its transfer function corresponds to $H_{M_2}(s)$ and the rms output noise is:

$$v_{on,rms,I_{B2}} = \sqrt{\frac{2K_BT\gamma g_{mB2}R_L^2}{g_{m2}^2}\frac{1}{\tau_{IN}\tau_L\left(\tau_{IN}+\tau_L\right)}} \tag{2.11}$$

In this case, it is necessary to reduce the ratio g_{mB2}/g_{m2}^2 in order to limit the noise contribution of I_{B2}. The current source I_{B1}, instead, is connected to the input node S_1; hence g_{mB1} must be decreased by properly choose the aspect ratio of M_{B1}, being careful to keep its output resistance negligible with respect to R_{IN}.

Last, the power spectral density of the noise associated to the load R_L, which is located at the output node, is given by:

$$\overline{i_n^2} = \frac{4K_BT}{R_L} \tag{2.12}$$

The respective transfer function is only affected by the low-pass filter R_LC_L. In order to reduce its noise contribution, it turns out that the load capacitance must be increased. However, this cannot be done without any cost, because it would affect the preamplifier bandwidth causing not only stability issues, but also limiting the maximum slope of the output signal.

2.2.2 Differential RCG-TIA design

The simplified schematic of the differential regulated common-gate transimpedance amplifier is reported in figure 2.5, where the SiPM cathode is connected to the left side, while the anode to the right branch.

After defining the design guidelines, described in the previous section, the schematic has been implemented in the circuit simulator and its performances has been studied for different device parameters. It has been decided at the beginning to consider, for the input transistors $M_{1/3}$, an over-voltage $V_{gs1} - V_{th1}$ around 150 mV to make them work in saturation, with a bias current no higher than 500 µA, in order to keep their noise contributions low. The known equation of the drain current in a n-type MOSFET in saturation is:

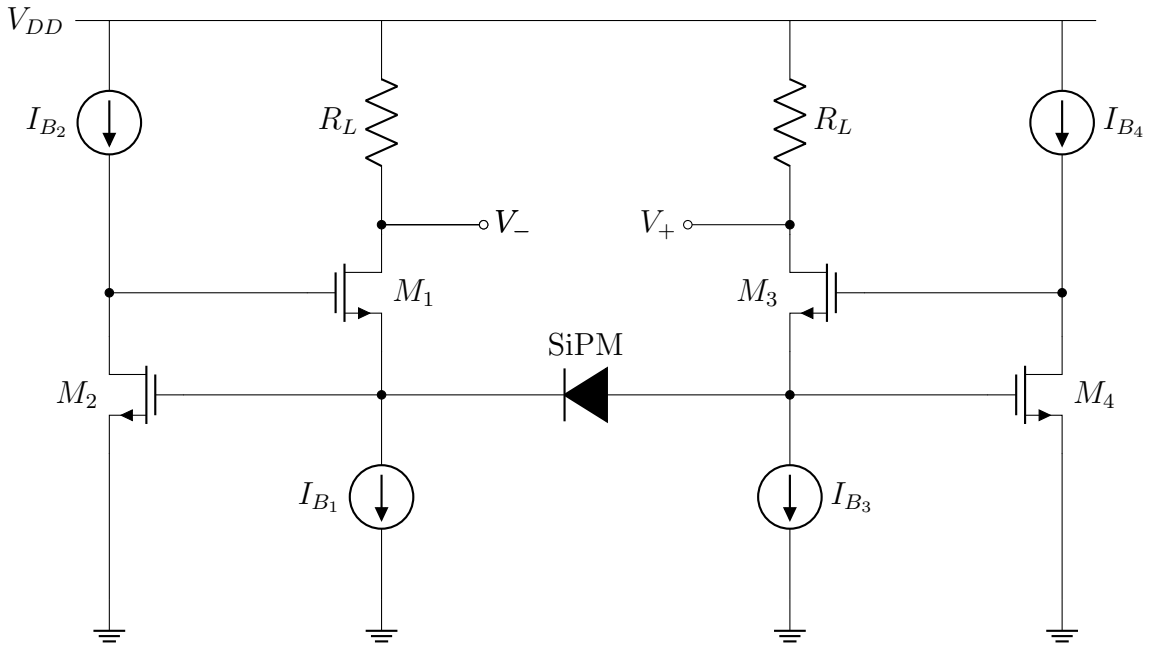

Figure 2.5: Differential regulated common-gate transimpedance amplifier (The bias of the SiPM is not reported in the schematic).

$$I_D = \frac{k'_n}{2}\frac{W}{L}\left(V_{GS} - V_{th}\right)^2 \tag{2.13}$$

where V_{th} is the threshold voltage, W and L are respectively the width and the length of the MOSFET channel, while $k' = \mu_n C_{ox}$ is the product between the average electron mobility and the gate oxide capacitance per unit area. For the TSMC 130 nm, $k'_n \approx 280\,\mu\text{A/V}^2$. From the equation (2.13), the M_1 aspect ratio W/L can be derived:

$$\left(\frac{W}{L}\right)_1 = \frac{2I_{D1}}{k'_n}\frac{1}{\left(V_{GS1} - V_{th}\right)^2} \approx 160 \tag{2.14}$$

As regards to the transconductance g_m, it is defined as:

$$g_{m1} = \frac{\partial I_D}{\partial V_{GS}} = \frac{2I_D}{V_{GS} - V_{th}} \tag{2.15}$$

Then, in this case the value of g_{m1} and the open-loop input resistance are:

$$g_m = \frac{2I_{D1}}{V_{GS1} - V_{th}} \approx 6.7mS \rightarrow R_{IN,ol} = \frac{1}{g_{m1}} \approx 150\Omega \tag{2.16}$$

The most of the current in the preamplifier has been used to bias the transistors M_2 and M_4, so that A_0 is high and its noise contribution is low. Choosing an initial amount of current of 2.5 mA and keeping the same over-voltage of the input MOSFETs, the transconductance of the feedback devices are $g_{m2} = g_{m4} \approx 33.5\,\text{mS}$, which is five times greater than g_{m1}, while the aspect ratio is around 800. It has been chosen to not

use the minimum device length permitted by the technology (130 nm), but to set the lengths to 160 nm for M_1 and 200 nm for M_2, which result in $W_1 \approx 26\,\mu m$ and $W_2 \approx 160\,\mu m$.

In this phase, it has been observed that the transient response of the RCG-TIA had some stability issues due to current source I_{B2}. In fact, the current pulses caused a variation of the drain voltage of M_2, and in turn of the output resistance R_{O2} because of the channel length modulation effect. This problem has been solved introducing in parallel to I_{B2} a further resistor R_{B2} to adjust the value of the loop gain. The presence of the current source allows a suitable setting of the gate voltage of M_1, which otherwise would have been too low to make this circuit work. A first attempt to design the value of R_{B2} has been done using the equation (2.5), trying to achieve a loop-gain of at least 10. The values of g_{m1} and g_{m2} considered in the formula have been obtained performing a DC simulation of the circuit, which led to value of R_{B2} equal to 630 Ω. As far as the load resistor is concerned, an initial resistance of 600 Ω has been used, so that the output DC level was 900 mV, a reasonable value to bias the n-type input MOSFETs of the following stages.

After implementing the SiPM model in the simulator, several transient, AC and noise simulations have been carried out to minimize the time jitter, by changing in turn all the parameters of the circuit. Based on the considerations done in the previous chapter, a coupling parasitic inductance of 5 nH has been included in the schematic, one for each SiPM terminal, so that the undesired effect due to the interconnection wires is taken into account.

For each iteration, the jitter has been estimated in two different ways:

1. by evaluating the maximum signal slope and the rms noise on the output node and then using the equation (1.21);

2. by running a transient noise simulation and measuring the root mean square of the time delay between the SiPM stimulus and the output voltage of an ideal discriminator cascaded to the RCG-TIA.

The first analyses revealed that both the methods led to the same result, but, for practical reasons, the second technique has been often preferred.

The final schematic of the preamplifier is reported in the figure 2.6. It includes not only the resistors $R_{B2/4}$ and the parasitic inductances (in grey because they do not represent actual devices), but also two diode-connected transistors on the output node. M_5 prevents M_1 to operate in the triode region when a current pulse with a considerable amplitude is generated by the SiPM. On the other hand, M_6 has been placed just to match both parts of the differential stage, but it does not intervene in

practice, as well we will see in the following. To extend the dynamic input range of the right branch, a linear extender has been designed, which will be discussed further on in the paragraph about the auxiliary circuits.

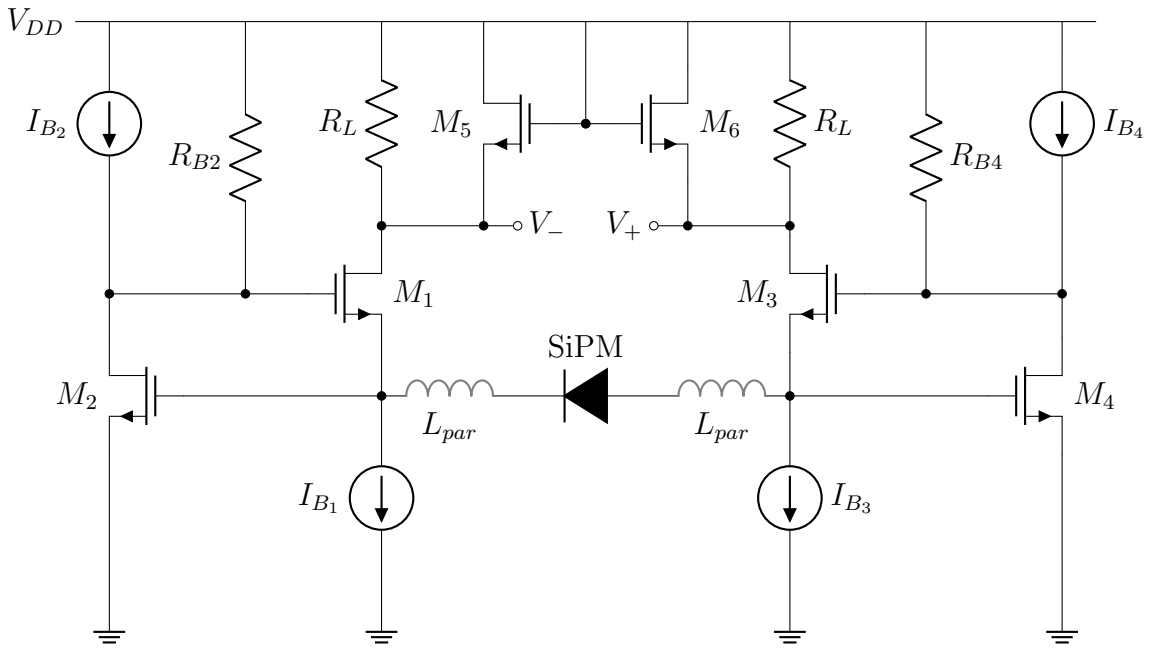

Figure 2.6: Final schematic of the differential RCG-TIA (The bias of the SiPM is not reported in the schematic).

Table 2.1: Size of the devices in the preamplifier.

Device	Value	Width	Length	Multiplicity
$M_1 \equiv M_3$	-	55.2 µm	200 nm	20
$M_2 \equiv M_4$	-	4.8 µm	160 nm	20
$M_5 \equiv M_6$	-	60 µm	130 nm	1
$R_{L1} \equiv R_{L3}$	735 Ω	2 µm	8.25 µm	2 in parallel
$R_{B2} \equiv R_{B4}$	168 Ω	2.34 µm	8.91 µm	8 in parallel
$I_{B1} \equiv I_{B3}$	400 µA	-	-	-
$I_{B2} \equiv I_{B4}$	2.4 mA	-	-	-

In table 2.1 the final sizes of each device and resistor are reported, which in some cases are far from the starting values. All the transistors in the RCG-TIA have been designed at the end to work in weak inversion, where the gate-source voltage is lower than the threshold and equation (2.13) is no longer valid. Making them work in this region, the circuit exhibit the best performances possible, because the transconductance approaches a value independent of the MOSFET size, comparable to the one that would be obtained with a bipolar device. As already said, this has been done to increase the time resolution of the circuit.

Before concluding this paragraph, it is appropriate to provide few plots about the preamplifier performances considering the electrical parameters corresponding to the Hamamatsu S10931-050P in the SiPM model. In particular, let us get started with the frequency response. As shown in figure 2.7, the preamplifier does not exhibit a large bandwidth, owing to the large SiPM input capacitance. The transimpedance gain is approximately equal to 1.47 kΩ (63.3 dB Ω), while its bandwidth is around 14 MHz.

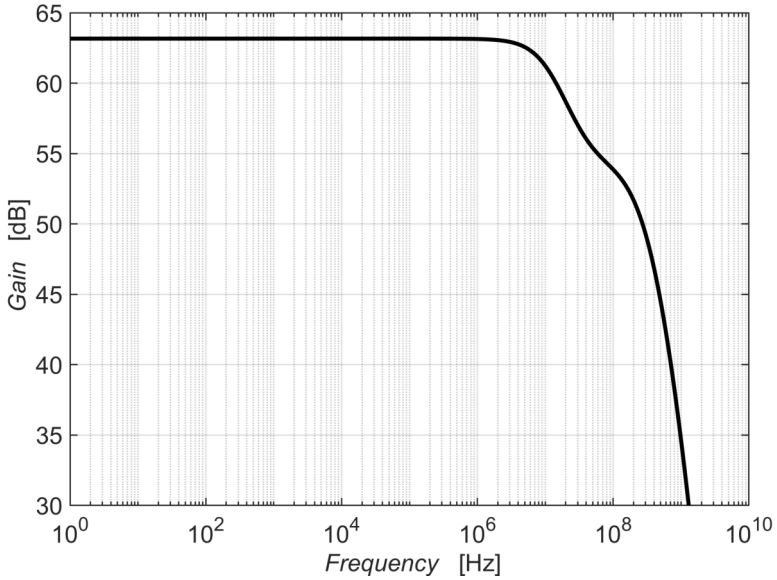

Figure 2.7: Frequency response of the RCG-TIA.

Another important parameter is of course the differential input resistance R_{IN}. In contrast to what is usually done in the design of current buffers, the input resistance has not been pushed down as much as possible, because it would have led to a lower time resolution due to the unavoidable presence of the parasitic inductance. In this way, all the parameters have been optimized to achieve the lowest time jitter possible. The behavior of the input resistance as a function of frequency is plotted in figure 2.8, where a value slightly higher than 30 Ω can be noted at low frequency, which then decreases for higher frequencies.

As far as the rms noise is concerned, an AC noise simulation has been run in order to obtain the power spectral density $\overline{v_{no}^2}$ at the output node. From the latter, the rms output noise has been derived integrating and squaring $\overline{v_{no}^2}$. Both quantities are shown in figure 2.9, where rms value of the output noise of about 600 µV can be inferred.

The last graph that needs to be drawn is the transient differential output signal for a single-photon event, from which it is possible to estimate the maximum slope, and thus the jitter σ_t. By observing the figure 2.10, a maximum slope of roughly 30 MV/s has been found in simulation; considering the equation 1.21 and assuming to set the threshold of an ideal leading-edge discriminator precisely in correspondence of

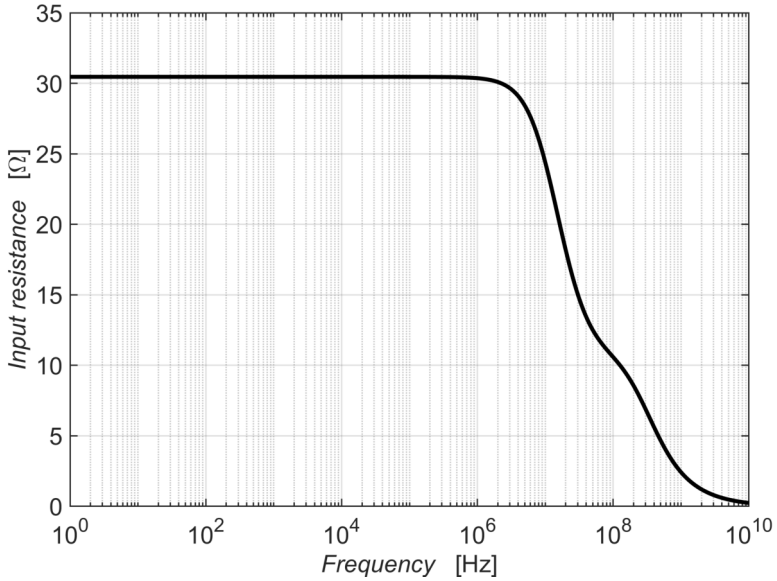

Figure 2.8: Differential input resistance R_{IN}.

the maximum slope, the time jitter is:

$$\sigma_t = \frac{\sigma_n}{\left.\frac{dV_{out}}{dt}\right|_{V_{out}=V_{TH}}} \approx \frac{600\,\mu\text{V}}{30\,\text{MV/s}} \approx 20\,\text{ps} \tag{2.17}$$

In the section dedicated to the leading-edge discriminator, more results will be reported, including those obtained with transient noise simulations [29].

2.3 Fast path

As already pointed out, the occurrence time of the detected event can be identified by exploiting the fast path of the front-end. In the proposed channel, it consists of just two blocks: the *leading-edge discriminator* (LED) and the *threshold adjustment system* (TAS). The former takes the output nodes of the preamplifier, $V+$ and $V-$, as inputs and generates a high logic signal as soon as the $V+$ terminal overcomes $V-$. In order to define the threshold of the leading-edge discriminator, the TAS is employed, which changes the DC output levels, by slightly varying the currents of the two branches of the differential structure.

2.3.1 Leading-edge discriminator

The leading-edge discriminator (LED), i.e. the comparator, is a non-linear circuit whose output is a logic one if the voltage at the non-inverting input terminal is greater

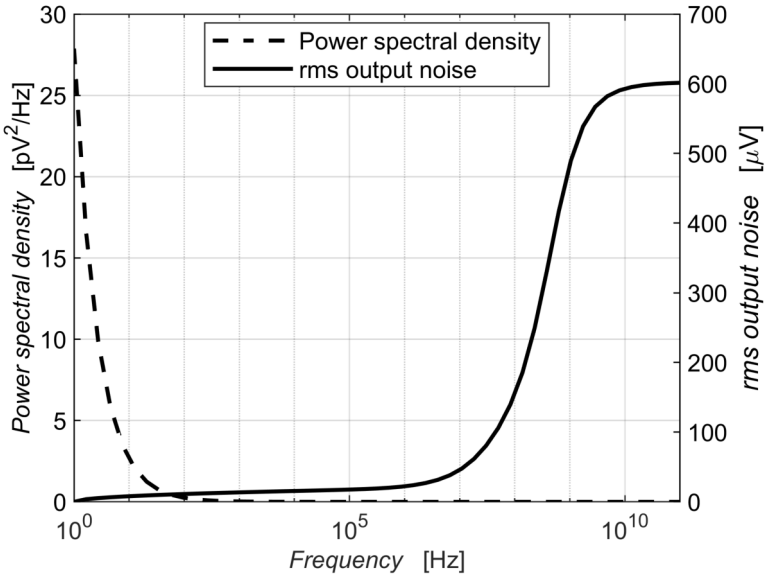

Figure 2.9: Power spectral density and rms noise at the output node.

than that on the inverting one, and vice versa. In order to preserve the time jitter σ_t of the preamplifier, it is required that the time uncertainty in the generation of the LED output signal is much lower than the one of the input stages, and thus negligible.

For this purpose, the high-performance architecture presented in [30] has been adopted with some changes, as reported in figure 2.11, where the voltage comparison is performed by means of the cascade of four different parts: the preamplification circuit, a transconductance stage, the decision block and the output buffer.

As regards to the preamplification circuit, it has been designed by paying attention to different factors. First of all, the LED must not throw in an excessive capacitive loading, otherwise the optimization work that has been done for the preamplifier would have been spoiled. Moreover, it must have an adequate gain, so that even the single-photon signal reaches the decision circuit with a reasonable voltage level, without introducing a excessive delay. Last but not least, its noise contribution must be low, to not add a further uncertainty to time measurements. For all these reasons, it has been decided to not use a single circuit with all the required gain, but to split it up into three different identical differential amplifiers, with a lower gain and size.

After several simulations, including the SiPM model and the preamplifier, a gain of 5 has been decided for each stage, in order to have a whole gain of 125. The simplified schematic of each stage is reported in figure 2.12. The differential gain of this kind of amplifier is given by:

$$A_d = -g_m R_L \tag{2.18}$$

and it has been chosen to start considering a not very high transconductance (for

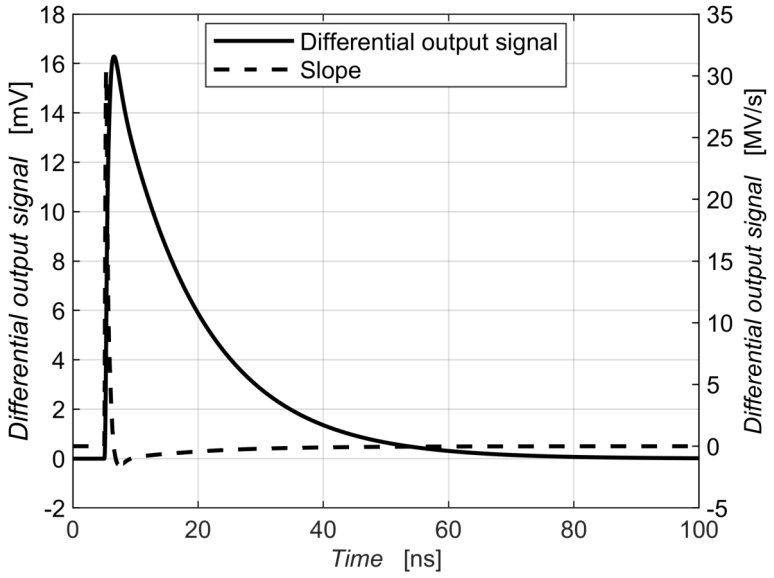

Figure 2.10: Differential output voltage of the RCG-TIA and its slope for a single-photon event.

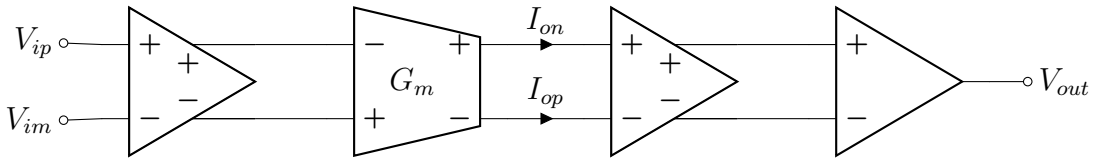

Figure 2.11: Architecture of the leading-edge discriminator with the preamplification circuit, a transconductance stage, the decision block and the output buffer.

limiting the noise) of 1 mS. In these conditions, in order to obtain the desired gain, a 5 kΩ load resistor must be considered. The value of the bias current I_D has been decided considering that a DC level of 900 mV is preferred at the output nodes, so that the following stages can be properly biased. Therefore, for each branch a current of 60 µA is needed, so that $I_D = 120\,\mu A$. Also in this case, the square-law equation (2.13) has been used to have an idea of the MOSFET aspect ratio, always considering an over-voltage of 150 mV:

$$\left(\frac{W}{L}\right)_1 = \left(\frac{W}{L}\right)_2 = \frac{I_D}{k'_n}\frac{1}{(V_{GS}-V_{th})^2} \approx 19 \tag{2.19}$$

rounded to 20. With a non-minimum channel length of 200 nm, the width of the transistor channel in the differential pair is 4 µm. Again, starting from these values, the size of all the transistors have been tuned to improve the performance. The final values are reported in table 2.2, corresponding to a final gain around 5.4.

The three following stages will be briefly described, without giving the details of the design. All the component values and dimensions may be seen in table 2.2.

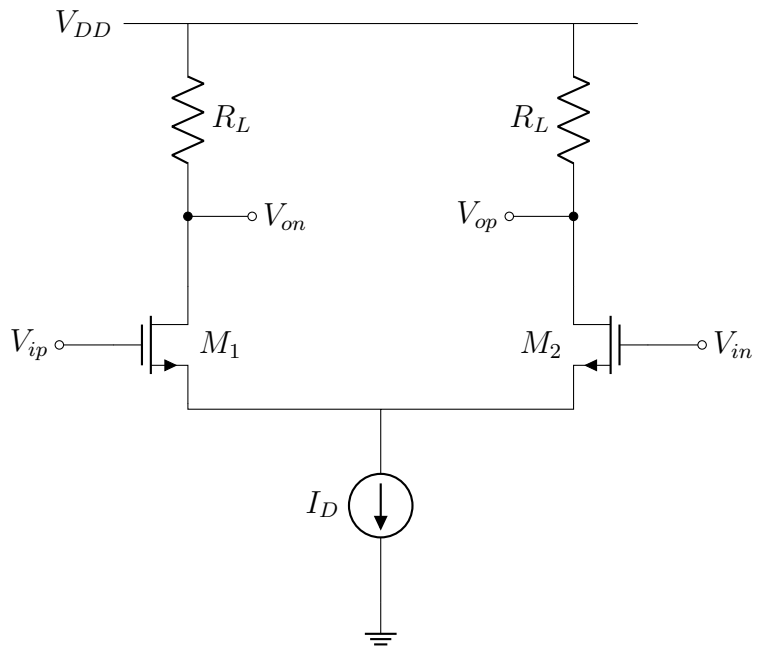

Figure 2.12: Preamplifier circuit.

The schematic of the transconductance stage is shown in figure 2.13. The resistors in the circuit 2.12 have been replaced with two active loads, which copy the drain currents of M_3 and M_4 and deliver them to the decision circuit. Since the gain of the previous stage is high enough, the signal level at the input of the transconductance stage is large; therefore, the G_m has been kept low in such a way that the noise contributions of M_3 and M_4 are irrelevant, accordingly to the equation (2.7).

The decision circuit, whose schematic is reported in figure 2.14, represents the core of the leading-edge discriminator. Its architecture exhibits a positive feedback, due to the cross-connections of M_{10} and M_{11} transistors and it is able to discriminate low voltage signals, generating thereby a fast output. The diode-connected MOSFET M_{13} has been included to shift the DC level of the outputs V_{op} and V_{on}.

Last, the output buffer performs two different tasks: the conversion of the signal from differential to a single-ended and the connection between the analog and the digital domains. This has been done by using the circuit in figure 2.15 with a digital inverter in cascade, which increase the LED capability of driving further capacitive loads and whose transistor widths are four times bigger (x4) than the ones of the smallest inverter available in the design kit. After analyzing the circuit functionality, it has been decided to supply the digital buffer with the digital supply, different from the one used for all the other analog blocks, which allows the reduction of the ground bouncing that happens whenever the inverter changes its logic state.

Once the leading-edge discriminator has been extensively simulated and its performances have been checked, the intrinsic jitter has been evaluated, by applying a

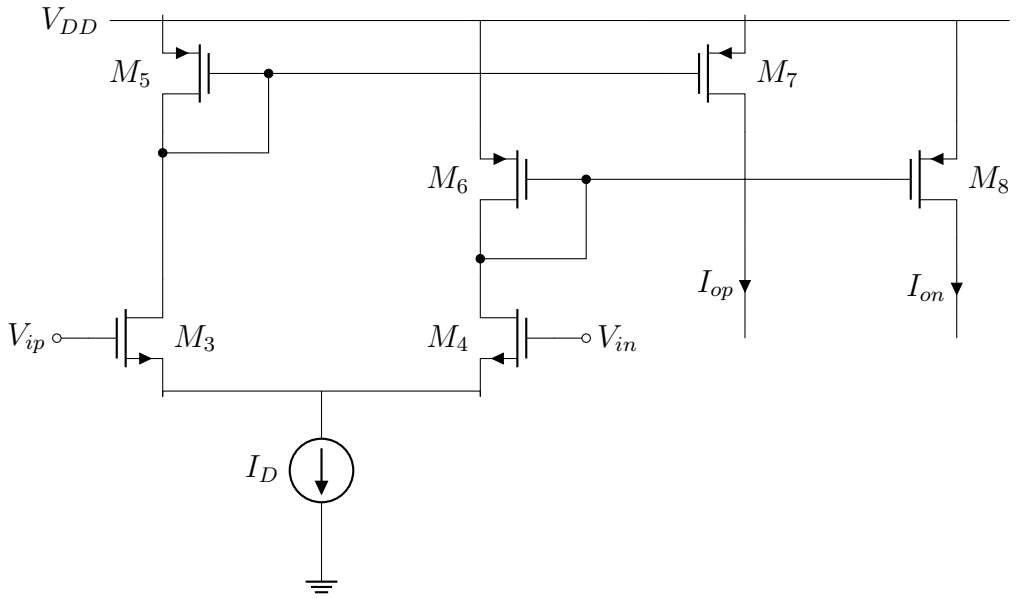

Figure 2.13: Transconductance stage.

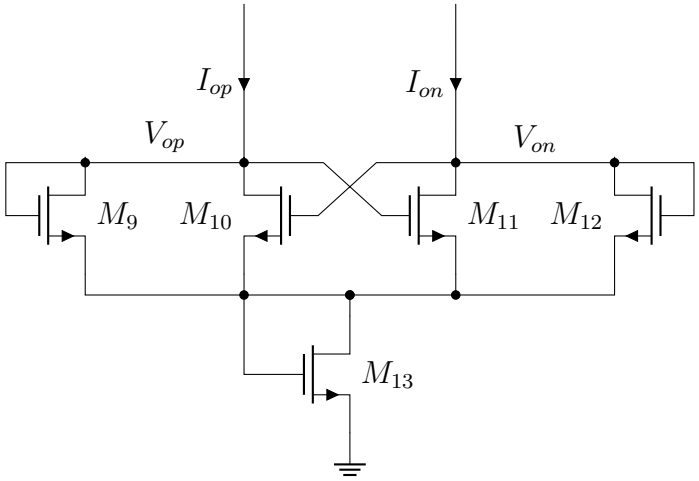

Figure 2.14: Decision circuit.

differential input signal similar to that on the output terminals of the preamplifier when a single-photon event is detected. As it is possible to check directly from the plot in figure 2.10, this signal has a peak voltage close to 16 mV, while the simulated rise time (10 % - 90 %) is 730 ps. Under these conditions, a transient noise simulation with 1000 iterations has been run and the delay between the ideal input signal (without noise) and the output logic state has been measured (cf. figure 2.16) which exhibits an average value of about 2 ns. Then, by calculating the standard deviation of the delays, the time resolution of the LED has been derived obtaining hence an intrinsic jitter $\sigma_{t,LED} \approx 7.4\,\text{ps}$. Its value is negligible with respect to the one of the preamplifier, because it must be added quadratically to the RCG-TIA jitter estimated in the previous paragraph ($\sigma_{t,RCG-TIA} \approx 20\,\text{ps}$):

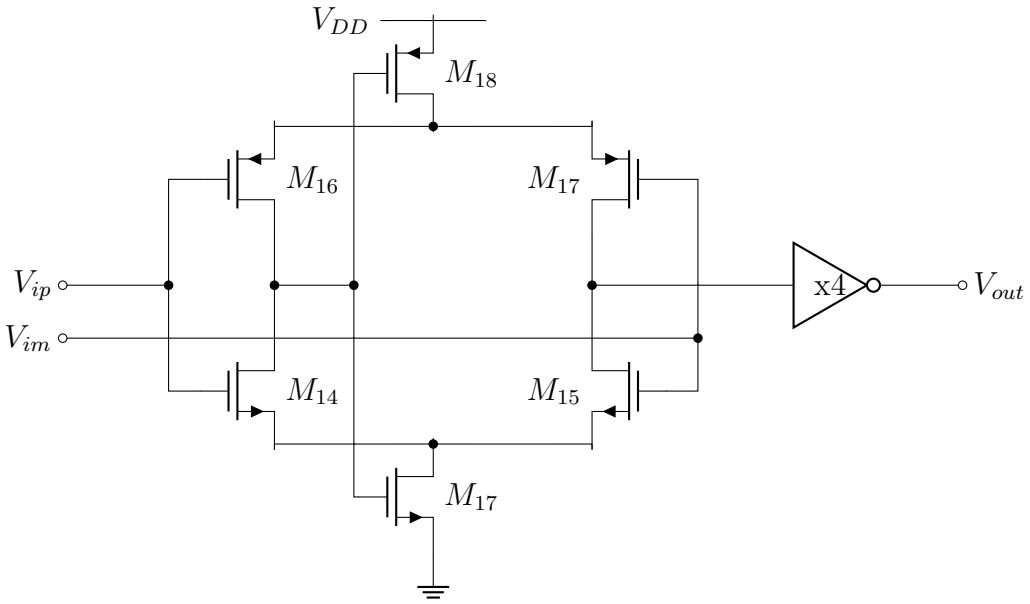

Figure 2.15: Output buffer.

Table 2.2: Size of the devices in the leading-edge discriminator.

Device	Value	Width	Length	Multiplicity
$M_1 \equiv M_2$	-	4 µm	200 nm	2
R_L	7.5 kΩ	-	-	-
I_{D1}	80 µA	-	-	-
$M_3 \equiv M_4$	-	1 µm	800 nm	2
$M_5 \equiv M_6$	-	2 µm	400 nm	2
I_{D2}	60 µA	-	-	-

$$\sigma_t = \sqrt{\sigma_{t,RCG-TIA}^2 + \sigma_{t,LED}^2} \approx 21.3\,\text{ps} \tag{2.20}$$

It actually means that the comparator contribution to the jitter is just over 1 ps.

In this phase, the whole system single-photon time resolution has been also evaluated by running 1000 transient noise simulations with both the preamplifier and the LED. The delays are now measured with respect to the current pulse contained in the model of the SiPM, which deliver the precise amount of charge generated by the detector. Moreover, as it will be described in the next section, the threshold has been placed in correspondence of the maximum slope of the output signal by means of the threshold adjustment system. It should also be mentioned that other circuits, belonging to the slow path and not yet analyzed, are connected to the RCG-TIA; therefore, this simulation takes into account all these capacitive loads, which can introduce some undesired effects on the jitter. In the plot of figure 2.17, the current stimulus $Q_{TOT}\delta(t)$ has been reported along with 50 among the 1000 iterations of the LED output digital

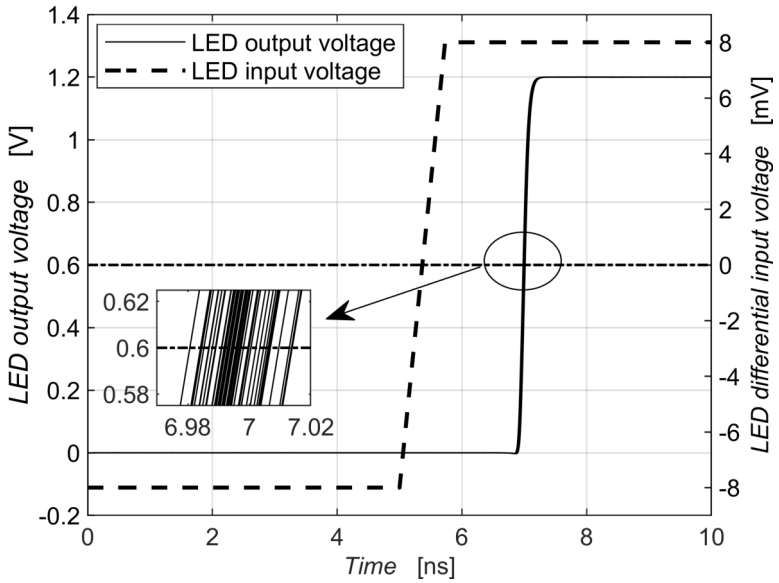

Figure 2.16: Leading-edge discriminator output signal and its differential input signal.

signal. The estimated time jitter is in this case $\sigma_t \approx 21.5\,\text{ps}$.

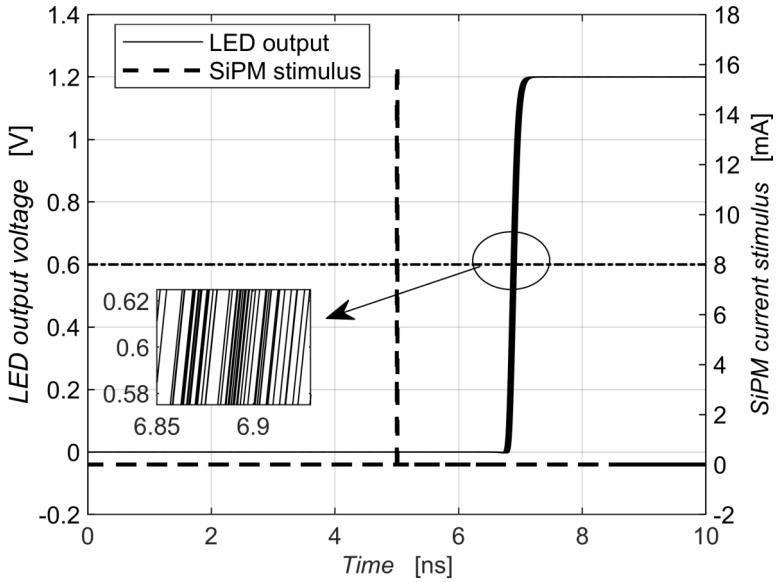

Figure 2.17: Leading-edge discriminator output signals and the SiPM current stimulus $Q_{TOT}\Delta\,(t)$.

2.3.2 Threshold adjustment system

Since both the preamplifier output terminals are connected to the leading-edge discriminator, it was not possible to set the threshold directly on the LED with a voltage reference signal, but another solution had to be found. The only viable solution was to come up with a new circuit to slightly change the current through the load resistors of

the RCG-TIA, so as to vary the DC levels on the output nodes in opposite directions. Not wanting to increase the capacitive load on the preamplifier outputs, which may have caused issues on the performances, a simple *Threshold Adjustment System* (TAS) has been designed to be attached to the RCG-TIA inputs, where the capacitive load is mainly dominated by the SiPM. The maximum threshold required for this system is eight photoelectrons, which correspond to a voltage eight times larger than the peak of the differential signal at the output of the preamplifier when a single-photon event occurs ($\approx 16\,\mathrm{mV}$, cf. figure 2.10).

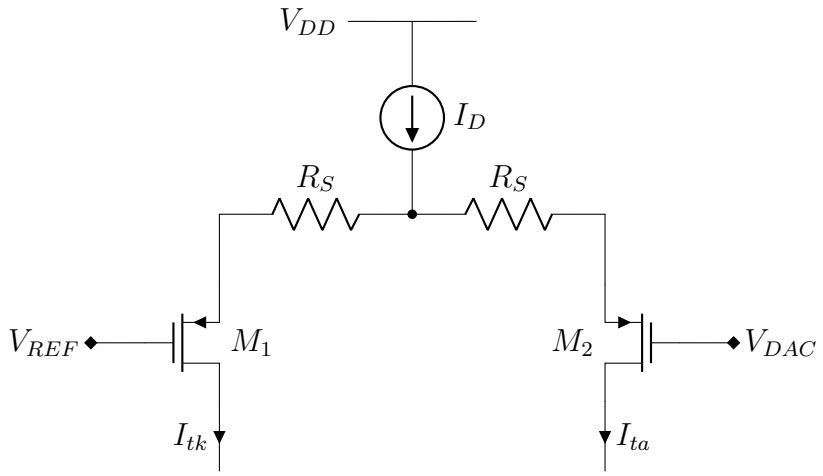

Figure 2.18: Threshold adjustment system.

The current can be varied through a differential pair, whose output currents I_{tk} and I_{ta} are respectively connected where the SiPM cathode and anode are coupled, as the schematic in figure 2.18 shows. It is biased by a current source I_D, so that in steady state conditions (when the gate voltages are exactly the same) the current that flows in each transistor is of course $I_D/2$. Moreover, the gate voltage of M_1, V_{REF}, is constant while the one of M_2, V_{DAC}, can be changed by means of a 6-bit DAC (provided by the IC group of SLAC), whose characteristic code-output voltage can be changed according to four different gains:

- Gain 0: from 100 mV to 300 mV, with a step of 3.125 mV;
- Gain 1: from 100 mV to 500 mV, with a step of 6.25 mV;
- Gain 2: from 100 mV to 700 mV, with a step of 9.375 mV;
- Gain 3: from 100 mV to 900 mV, with a step of 12.5 mV;

When an increasing V_{DAC} is applied to the gate of M_2, its source-gate voltage decreases and thus the current I_{ta} reduces as well; moreover, since $I_{tk} + I_{ta}$ must

always be equal to I_D, I_{ta} grows. This implies that the current that flows through the input MOSFET of the right part of the RCG-TIA, and hence also through its load resistor, is higher, making the potential of the corresponding output node drop. The opposite situation happens in the left part of the preamplifier, where an increasing V_{DAC} results in higher potential of the output node. The differential input signal seen by the leading-edge discriminator exhibits a negative DC level, so that its output logic state is normally low. When the SiPM generates a current pulse that produces a differential voltage greater than the threshold, the LED changes its state.

First of all, the value of the current source I_D needs to be chosen. Considering that the peak of the output signal for a single-photon event is approximately 16 mV, the differential DC level must vary until at least 128 mV, and hence 64 mV for each side. Furthermore, because the load resistor is $R_L = 735\,\Omega$, the bias current has to change at least up to 87 µA. It has been decided to round this value to 100 µA, so that the current that the source I_D must provide is 200 µA. Nonetheless, it has been necessary to increase the bias current of I_{B1} and I_{B3} (cf. figure 2.5) from 400 µA to 500 µA in order to accommodate also the TAS current and not change the operating points of M_1 and M_2.

The current-voltage characteristic of the differential pair has been conveniently linearized, by introducing a couple of degeneration resistors R_S on the sources of the devices. Given the possibility to change the gain of the DAC, four different modes of operation have been defined and their characteristics are reported in figure 2.19. After normalizing the threshold with respect to the single-photon signal, the four different ranges, expressed in terms of photoelectron [*pe*] are:

1. *Ultra-fine*: from 0 pe to 2.3 pe with a step of 1/20 pe;
2. *Fine*: from 0 pe to 5.3 pe with a step of 1/10 pe;
3. *Normal*: from 0 pe to 7.8 pe with a step of 1/6 pe;
4. *Coarse*: from 0 pe to 8.5 pe with a step of 1/5 pe;

For the first codes, the plots in the figure exhibit a "negative threshold", which can just be used to balance the DC levels during the calibration of the channel. After that, it is possible to exactly adjust the threshold to obtain the best possible performances.

2.4 Slow path

Beyond the identification of the instant of arrival of a light beam on a SiPM, the analog front-end is also able to provide the measurement of the energy of the detected

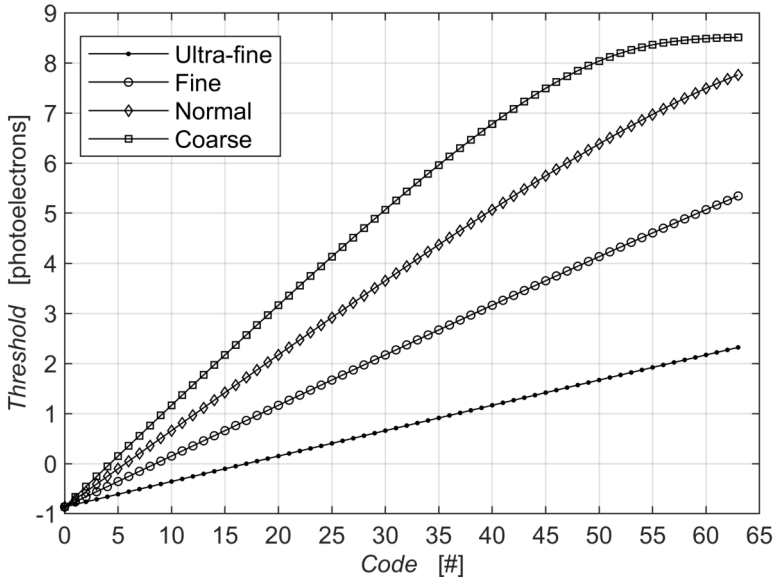

Figure 2.19: Characteristics of the threshold adjustment system.

event. Specifically, the channel generates an analog signal, whose value is proportional to the input charge, and hence also to the number of pixels that have been triggered. Therefore, in this section, all the circuits of the slow path, the measurement chain aimed at the estimation of the energy of the event, will be described. It is directly connected to one of the outputs of the differential transimpedance amplifier and it is essentially made up of a voltage-to-current converter, an integrator and thirdly a peak detector. Nonetheless, the V-to-I converter has also been equipped with a current divider, which guarantees a wider dynamic range when high amplitude current pulses are concerned. Moreover, between the input and the output terminals of the integrator, a baseline holder has been implemented; in case of two or more pulses one right after the other, it ensures the mitigation of the pile-up effect, which can introduce a further uncertainty during the measurement, and allows the system to sustain a higher event rate.

2.4.1 V-to-I converter

Before proceeding with the integration of the signal, it became necessary to provide an auxiliary circuit capable of reducing, whenever prompted, the value of the current so that to accommodate also the pulses produced as a result of the absorption of multiple photons by several cells of a single SiPM. In this case the amplitude of the input current may turn out to be remarkable and it can induce the saturation of the integrator. One of the most effective way to accomplish this result consists in converting the output voltage signal of the differential front-end into a current one and scale it down by means

of various current mirrors. The simplified schematic of the V-to-I converter together with the current scaler is reported in figure 2.20.

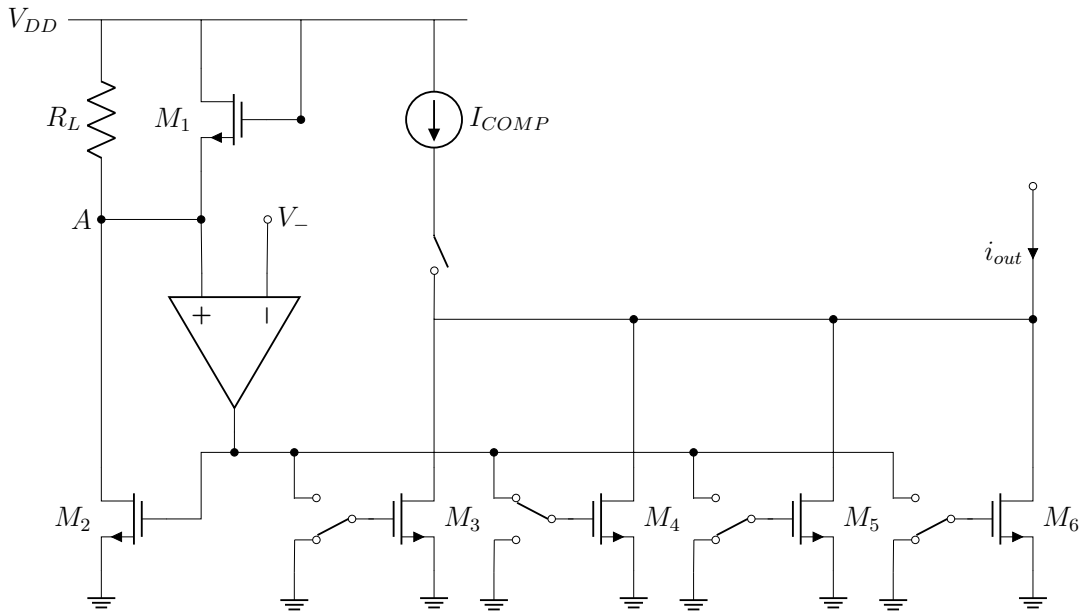

Figure 2.20: Schematic of the V-to-I converter.

In particular, the inverting input of the operational amplifier has been connected to the V_- terminal of the differential RCG-TIA, while the other input is attached to a replica of the front-end load (resistor R_L and the diode-connected MOSFET M_5). Under these circumstances, thanks to the virtual short circuit between the input terminals of the op amp, the voltages across the load on the left branch of the transimpedance amplifier and the replica feature the same value, as well as thus the currents that flow thought them.

In order to obtain an estimation of the input charge related to the detected event, the integration of this current has to be accomplished. However, in case of signals generated by multiple photons hitting the SiPM at the same time, the peak of the current can be very high and, as already said in the first chapter regarding the CSA, a massive value of integration capacitance in the order of nanofarads would have been needed, with an excessive waste of silicon area. Thereby, the drain current of the transistor M_2, which follows the trend of the input current, is than scaled down by the MOSFETs M_3, M_4, M_5 and M_6 connected in parallel, so that the user can choose the most suitable range for the application. Nevertheless, their gate terminals have been connected to the output of different switches which link them either to the output of the operational amplifier or to ground. Since the signals that drive the switches are the outputs of a decoder $1-4$, these transistors can only be enabled one at a time, so as to set up the desired range. The scale factors chosen for this application have been

reported in table 2.3, together with the number of fingers of each transistor.

Table 2.3: Scale factors of the current mirrors.

Device	Fingers	Scale Factor
M_2	72	–
M_3	28	$7/18 \approx 0.389$
M_4	4	$1/18 \approx 0.056$
M_5	2	$1/36 \approx 0.028$
M_6	1	$1/72 \approx 0.014$

As far as the operational amplifier is concerned, it has been designed following the two-stage Miller architecture with a RC compensation scheme.

2.4.2 Integrator

In order to produce a voltage signal whose peak value is proportional to the input charge, the current coming from the mirrors described in the previous paragraph must be conveniently integrated. It has been chosen to employ an active integrator, consisting of an operational amplifier and a R-C feedback network between the output and the inverting terminal; the schematic is shown in figure 2.21.

Based on the range selected, this circuit can have different integration time. In particular, if the detection of few photons is concerned, the time constant of the feedback network has been designed to be around 250 ns, meanwhile for the other three ranges, its default value is 350 ns. The latter value has been chosen considering that, when the system is implemented in a PET scanner, the current pulses are generated by a photomultiplier coupled with a scintillator, whose intrinsic time constant can reach even 100 ns, and thus exhibit a very long tail that cannot be disregarded during the integration process. Nevertheless, in both situations, the capacitor and the resistor values can be adjusted by means of 3 and 4 configuration bits respectively, so as to define the set of parameters suited for the acquisition in progress. Specifically, accordingly to the needs, the time constant R-C can be configured between 120 ns and 300 ns for single-photon applications, while it may be set up from 324 ns to 610 ns for all the others.

During the design of both the integrator and the V-to-I converter introduced in the previous paragraph, four different ranges have been defined. In particular, taking into account a SiPM with a gain of 10^6, the measurement intervals that can be selected are:

- $1 \div 150$ photoelectrons with a maximum input charge of 24 pC;
- $15 \div 2000$ photoelectrons with a maximum input charge of 320 pC;

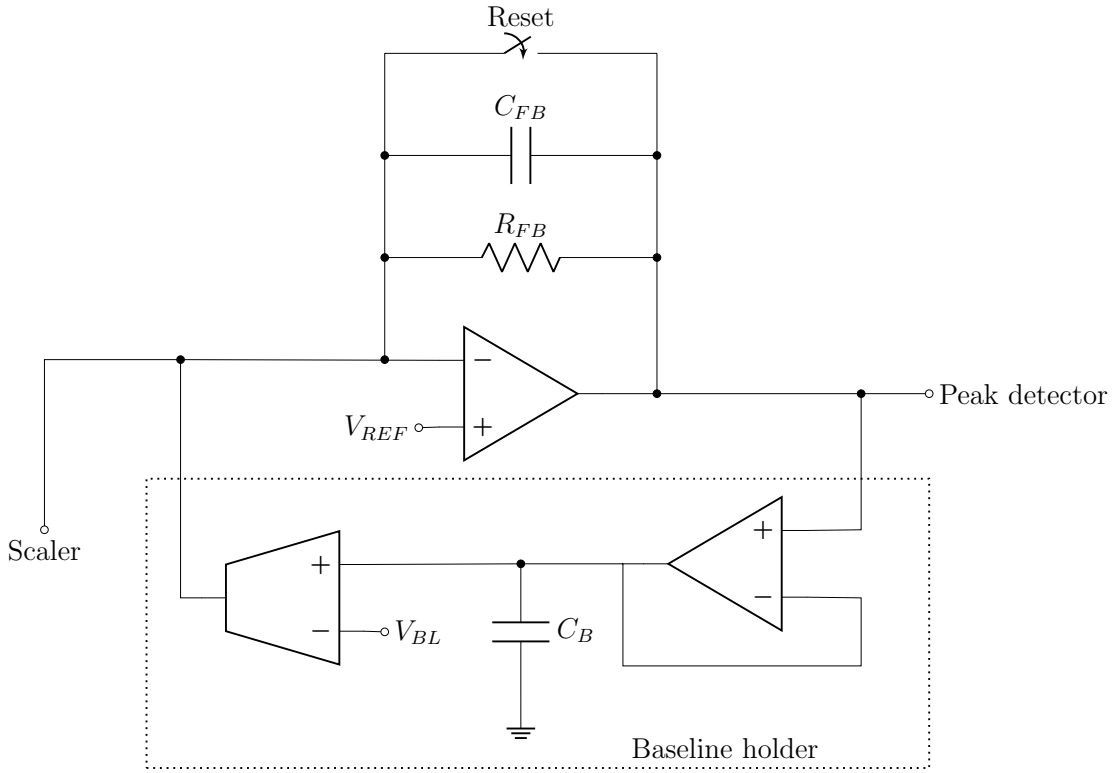

Figure 2.21: Simplified schematic of the integrator. In the dotted box the sketch of the baseline holder, described below, is reported.

- $30 \div 4000$ photoelectrons with a maximum input charge of 640 pC;
- $60 \div 8000$ photoelectrons with a maximum input charge of 1280 pC;

Subsequently, for each range multiple transient simulations have been performed to assess the performance of the slow path and its gain. After selecting the range, the input charge has been swept from the lower to the upper bound with a total number 16 steps, and the maximum value of the integrator output terminal has been plotted with respect to the input charge. As reference, the output voltage - input charge characteristic of the first range has been reported in figure 2.22, along with the non-linearity error that is always lower than 10 %. Actually, not including the last points of the characteristics, it would be even lower, less than 5 %, because the output begins saturating approaching its maximum value. Indeed, if a number of photons close to the upper bound of the first range are expected, it would be preferable to switch to the second configuration, where the characteristic is more linear.

A summary of all the features derived for each of the four configurations of the slow path is reported in table 2.4. For all the ranges, a non-linearity error lower than 10 % has been estimated. For the single-photon range, the stimulus has been generated as if all the pixels were fired at the same time, while, for the configurations 2,3 and 4,

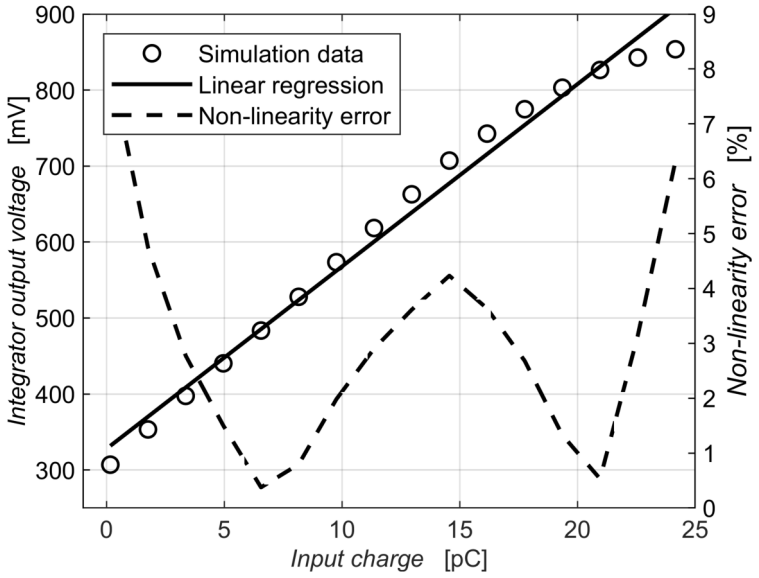

Figure 2.22: Linearity of the slow path for the first range.

the reference signal mimics the behavior of a scintillator with a time constant of 40 ns, whose process for the generation of photons follows a Poisson distribution.

Table 2.4: Ranges of the slow path.

Range	Gain	Gain	Dynamic range [photons]	Maximum charge
1	4.00 mV/phe	26.50 mV/fC	1 ÷ 150	24 pC
2	0.32 mV/phe	20.50 mV/fC	15 ÷ 2000	320 pC
3	0.16 mV/phe	1.05 mV/fC	20 ÷ 4000	640 pC
4	0.08 mV/phe	0.53 mV/fC	30 ÷ 6000	1280 pC

2.4.3 Baseline holder

When the rate of the detected events becomes sustained, it may happen that the output baseline of the integrator moves up from its nominal value (300 mV) affecting the result of the operation. Nevertheless, the shift of the baseline is not only due to a close succession of incoming pulses, but also to the leakage currents that the scaler (to which the integrator is attached) may introduce. Indeed, because of the feedback resistor, even a small DC current can produce a non-negligible variation of the DC level. Since the specifications of this project require the ASIC to handle a event rate up to 100 kHz, a *baseline holder* has been included between the output and the input terminals of the integrator. As showed in figure 2.21, the first part of this circuit consists of a voltage buffer with a 1.5 pF capacitance on its output terminal; it allows this block to sense the output voltage of the integrator, without influencing

its performances. Moreover, the capacitance guarantees the filtering of the high-speed pulses, so that the transconductance amplifier place in cascade to the buffer sees only the low frequency components of the output voltage. This signal is hence compared with a reference, set to the desired voltage of the baseline, by means of an operational transimpedance amplifier, which has the ability of sinking or sourcing current into the input node of the integrator in order to adjust its DC output voltage [31].

2.4.4 Peak detector

The last part of the slow path comprises a *peak detector*, whose purpose is to identify the peak of the input signal and maintain its value as long as needed. This circuit has been designed using the De Geronimo architecture presented in [32] and [33], whose diagram is shown in figure 2.23.

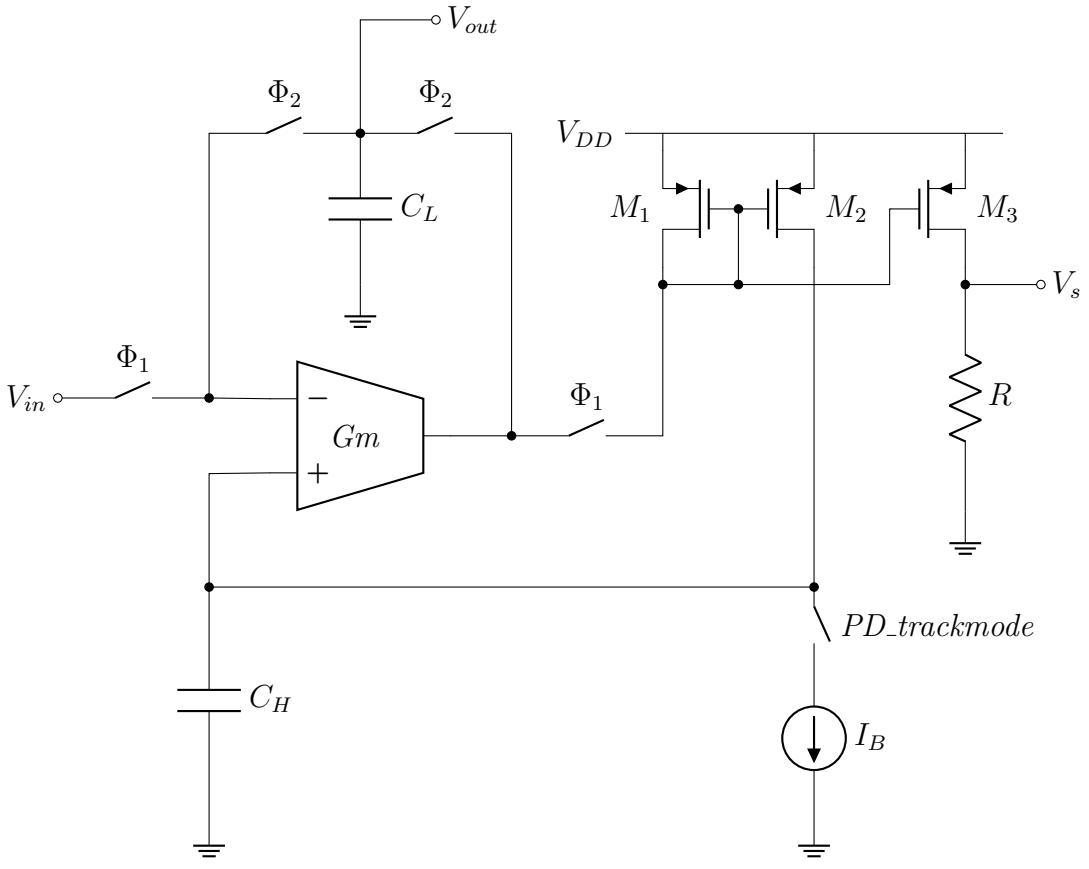

Figure 2.23: Simplified schematic of the peak detector.

The operation of the peak detector is mainly handled by an external active-low digital signal coming from the digital state machine of the channel (cf. next section). This signal, called *PD_trackmode*, is ordinarily high, but it is reset by the FSM only when an event has been detected and the identification of the peak of the integrator

output voltage is demanded to complete the measurement process. Within this block, in addition to the component drawn in the schematic, a simple logic has been implemented to identify the instant when the peak has been reached and regulate the opening and the closing of the switches, by means of two control signals Φ_1 and Φ_2. Specifically, this system monitors the node V_s, whose voltage grows on the rising-edge of the input signals and drops approximately to zero as soon as V_{in} begins decreasing. Moreover, this system provides two output signals for warning the state machine of the channel when the peak has been found (PF) and when the output terminal is ready to be read by the ADC (PH). It is possible to divide the functioning of the peak detector into three different phases, described below:

- *Voltage follower*: Initially, when *PD_trackmode* is high, the internal logic set Φ_1, while the switches driven by Φ_2 remain open. In this situation, the voltage across the capacitor C_H matches the one of the input node V_{in}.

- *Peak detector*: When the tracking is required, in response to the detection of a signal from the SiPM, *PD_trackmode* is reset and the circuit follows the input as long as the peak of the pulse is detected by the internal logic.

- *Analog memory*: As soon as the logic realizes that the peak has been attained, both Φ_1 and Φ_2 change their logic state respectively to low and high, so that the circuit behaves as a unity-gain voltage buffer, charging the output capacitor C_L to the value of the peak.

The output capacitor works as an analog memory which maintain the value of the peak for as long as necessary. The only factor that can alter the voltage across C_L is related to the leakage currents due to the presence of the switches Φ_2 on the same node. Nevertheless, a thorough design of the switches can restrict the value of this current down to some hundreds of femtoamperes [34].

As an example, a typical peak detection process is given, using as V_{in} the signal that would be produced in response to the simultaneous triggering of 20 pixels of a SiPM. In figure 2.24, the two inputs of the peak detector are reported; in particular the analog signal coming from the integrator is depicted with a solid line, while *PD_trackmode* driven by the digital state machine is dashed.

In figure 2.25, instead, all the output signals of the block are drawn; the analog output of the peak detector is shown in the upper subplot (along with V_{in} to point out the difference between the peak and the voltage to which the output terminal settles). As regards to the digital outputs, PF and PH are presented in the second subplot. Specifically, the internal logic of the peak detector sets PF right after the peak has

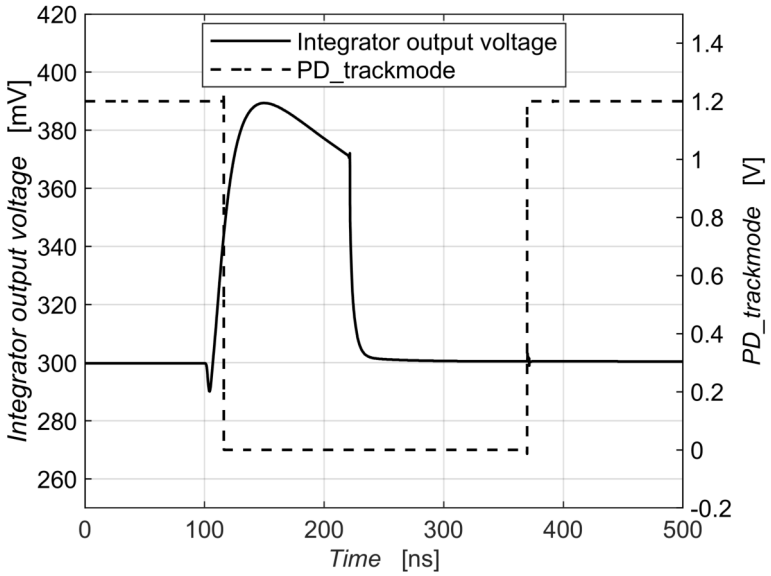

Figure 2.24: Input signals of the peak detector.

been identified, while PH switches to high only when the voltage across the output capacitor is stable and ready to be read. The circuit output voltage settles down to a value around 393 mV in response to a peak of approximately 390 mV. It has been observed that this offsets of 3 mV constantly recurs, regardless of the amplitude of the input signal.

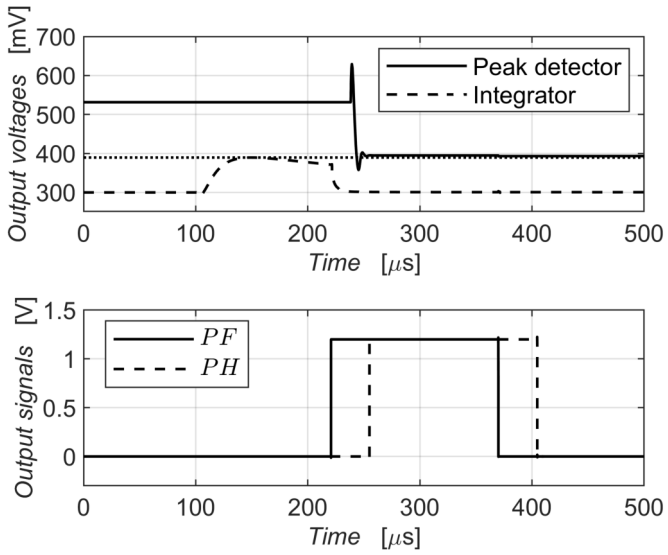

Figure 2.25: Output signals of the peak detector.

Finally, it is worth mentioning that, after the process is completed, the internal capacitor C_H needs a variable amount of time to discharge through the current sink I_B; the higher the input charge, the longer this time. Therefore, during this recovery

phase the peak detector cannot accept another pulse and the whole channel must wait until it becomes operational again. For this reason, an auxiliary digital output, called *PD_idle*, has been added to the circuit and it is monitored by the digital state machine to keep the channel from starting another conversion when the peak detector is not yet available.

As it has been done for the integrator, the whole measurement chain of the energy of the detected event has been tested and the linearity of the slow path has been assessed for each of the ranges mentioned above; even in this case, a non-linearity error less than 10 % has been derived for all the ranges.

2.5 Other circuits

Aside from the circuit in the fast and slow paths, three additional circuits have been deployed in the channel to provide all the necessary for the proper operation of the two measurement chains. These blocks are the biasing circuit, which supply all the required bias current, a linearity extender to prevent the input MOSFET of the right branch of the front-end from turning off in case of high amplitude input signals, and finally a further voltage discriminator has been implemented in the slow path in order to add to the channel the capability of filtering out all the undesired events.

2.5.1 Biasing circuit

Each of the blocks described in this section requires a biasing current of an accurate value to run properly. For this reason, a biasing circuit that generates all the required current has been designed and implemented in the channel. The schematic of this system is reported in figure 2.26 and it is fundamentally a V-to-I converter, whose output current is copied and conveniently scaled into as many branches as the circuits that need a polarization.

The input voltage, attached to the non-inverting terminal of the operational amplifier, must be stable and independent of temperature. Therefore, it has been decided to generate this reference signal using a bandgap voltage reference circuit, whose value can be adjusted from 50 mV to 200 mV with a 50 mV step by means of two configuration bits. Moreover, in order to have a reference current of 10 µA, the MOSFET has been provided with a source degeneration resistor of 20 kΩ. Actually, in the bias circuit two V-to-I converters have been implemented where the only difference is related to the resistors. In the first circuit, the resistor is untrimmed and its value is affected by the process variation, as well as the current that flows through; thereby, such current has been copied and then used to generate the fixed reference voltage by employing

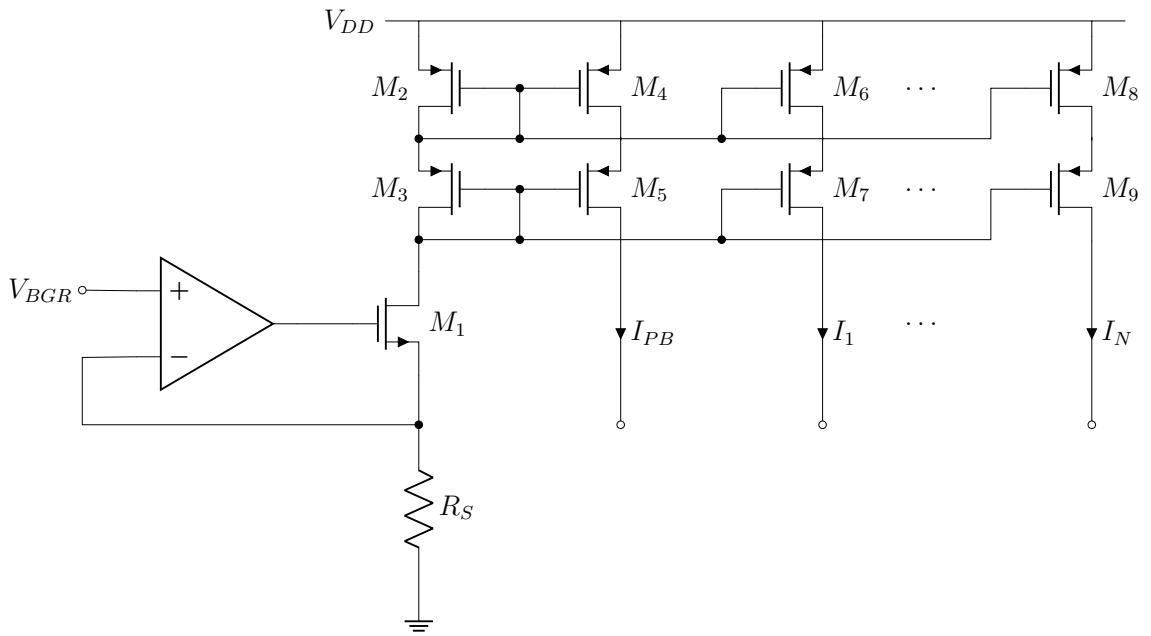

Figure 2.26: Schematic of the biasing circuit.

the same kind of resistors, which thus depend on the process in the same way of the degeneration one. On the other hand, the other circuit is equipped with a trimmable resistor, through which it is possible to compensate the deviation of the reference current from its nominal value due to PTV variations. This resistor can be varied from 15.5 kΩ to 23.5 kΩ with a step of 500 Ω by means of four further configuration bits. The drain current of M_1 is copied in several branches by means of several cascoded current mirrors, each of which exhibits a high output resistance that ensures a lower fluctuation of the current with the output voltage.

In order to monitor the reference current in the trimmable stage and hence adjust its value, an auxiliary current mirror with an aspect ratio 1:1 has been included in the circuit, so that the drain terminal of the transistor M_5 can be brought externally to the circuit and attached to some sort of current probe that allows the assessment of I_{PB}.

As regards to the operational amplifier included in this block, it has been designed with an intrinsic bias generation system, whose fluctuations of the parameters due to PVT variations are contained.

2.5.2 Linearity extender

Taking up the analog front-end circuit in figure 2.6, it is important to bring out an crucial issue that the circuit would deal with if it operates with input pulses of several milliamperes of amplitude. Since the cathode and the anode of the SiPM are attached

respectively to the left and right branches of the differential structure, the signal current flows (starting from the power supply) from the left load resistor to the sink current source I_{B3}, passing through the MOSFET M_1 and naturally the SiPM. Whilst the diode-connected transistor M_5 prevents the input MOSFET to work in triode region in case of high amplitude pulses, with the increase of the signal intensity, the drain current in the right-side input device M_3 decreases gradually till it shuts down. In these conditions, altering the operating point of this transistor, the input resistance seen by the photomultiplier would grow, resulting in a deterioration of the parameters of the signal in terms of slope and peak value.

Since a dynamic range up to 8000 photoelectrons is required, which may appear as a current with a peak of several tens of milliamperes, a linearity extender has been designed and connected to the front-end to overcome this problem. This circuit is made up of an unbalanced operational transconductance amplifier that senses the voltages on the input node on the front-end and drives the gate of a current source placed in parallel to I_{B3}. This solution allows the input MOSFET M_3 to be always biased with its nominal current, without causing problems to the functioning of the SiPM, and the supplementary current mirror to sink the signal current coming from the sensor.

2.5.3 "Slow" voltage discriminator

In certain applications, when low-light levels are not of interest, is often worthwhile to halt the measurement process if the amplitude of the input signal is less than a desired threshold. As a result, it is possible to release the channel earlier, before that the whole measurement process is accomplished in vain and make the circuit available for another acquisition as soon as possible. One way to accomplish this is to furnish the slow chain with a voltage discriminator, whose inverting input is attached to the node A (cf. figure 2.20). The other terminal, instead, is the output of another DAC (the same employed for the Threshold Adjustment System), which acts as reference signal and can be configured according to the requirements. The output signal of this discriminator, named "slow" after the path in which it is placed, switches to the high logic level only if the input current features an amplitude greater than the threshold and then it will be the task of the digital state machine to handle this digital signal, as well as the output of the LED, and continue or arrest the processes based on its value.

The architecture chosen for the design of this voltage discriminator is the one presented by Allen and Holberg in [35]. Its complete schematic is reported in figure 2.20. The peculiarities of this circuit is the presence of a internal positive feedback and the possibility of offering an intrinsic hysteresis by properly selecting the aspect ratios of the diode and the cross-connected MOSFETs M_3, M_4, M_5 and M_6. Moreover, this

comparator is equipped with a class AB output stage that is not only able to provide a good value of the output resistance and voltage swing, but it is also intended for the conversion from differential to single-ended.

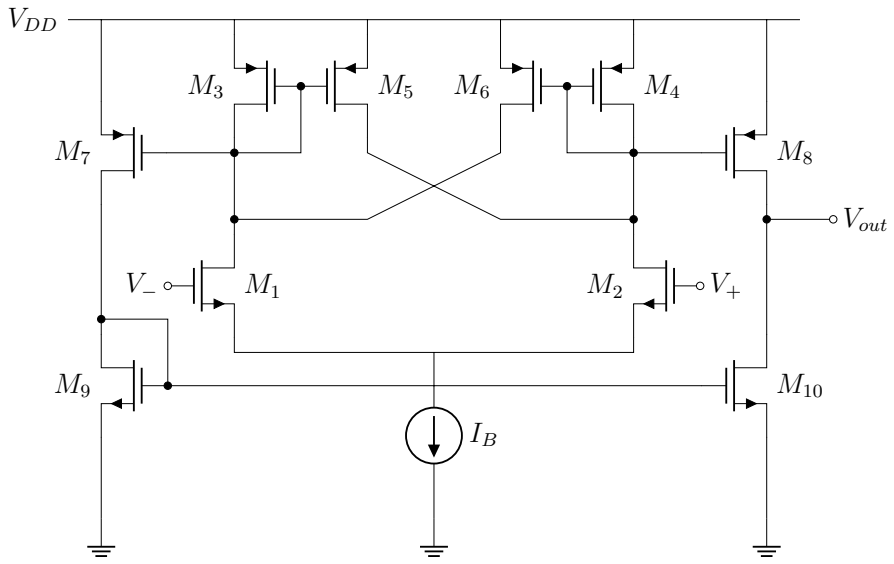

Figure 2.27: Schematic of the "slow" voltage discriminator.

2.6 Digital state machine

The last circuit block that has been designed in the SiPM channel is a *digital state machine*, whose main roles are to interface the front-end with the external logic and to handle the internal reset signals. Wanting to avoid a fast clock signal in the channel, it has been decided to implement an asynchronous digital macro, which changes state in correspondence of the rising or falling edge of certain kind of signals.

As already seen previously, the "slow comparator" is employed to check if the amplitude of the current pulse is larger than a certain value. In order to do that, a further feature needs to be added in this machine, so that any event whose amplitude is lower than the desired threshold, and thus not relevant, can be filtered out.

Before starting the design of the machine, the set of input and output signals has been identified:

- Inputs:

  1. LED: It is the active-high output signal from the leading-edge discriminator, which also goes through two identical buffers, with a variable delay, to generate the signals d_LED and $2d\_LED$.
  2. SC: It is the active-high output logic state from the "slow" comparator.

3. PF: It is the active-high digital signal, called *peak formed*, sent by the peak detector to warn the FSM that the peak has been found.
4. PH: It is another active-high digital signal, called *peak in-hold*, coming from the peak detector few nanoseconds after PF to inform the FSM that the analog signal is ready to be converted.
5. EOC: It is the *end-of-conversion* active-high signal coming from the external logic that notifies the FSM that the conversion has been completed.

- Outputs:

1. TDC_start: It is an active-high signal sent to the Time-to-Digital Converter (TDC) to start the conversion.
2. TDC_reset: It is an active-low signal sent to the TDC to bring it back to the initial state;
3. INT_reset: It is an active-high signal sent to the integrator whenever its capacitor needs to be discharged.
4. $PD\_trackmode$: It is an active-low signal sent to the peak detector to start the tracking of the peak.
5. $ADC\_request$: It is an active-high signal aimed at warning the external logic that the analog signal is ready for the conversion.

First of all, the state diagram of the digital macro has been drawn implementing a Moore machine, whose outputs only depends on the current state and not on the input signals. The diagram, shown in figure 2.28, presents seven different states:

1. A: In this state, the digital machine waits for the rising-edge of the LED to start the whole conversion process and to bring the FSM in the state B. Here, all the output signals are in their steady-state conditions. Moreover, any other rising-edge signals from the LED occurring while the FSM is not in this state are just discarded.

2. B: As soon as the FSM enters in this state, TDC_start switches from low to high, in order to trigger the time conversion. In this state the machine waits for the delayed rising-edge of the LED, d_LED. During this phase, the slow comparator output may switch from low to high whether the slow path input signal is larger than the selected energy threshold, or not. If that is the case, the conversion proceeds and the machine enters in the C state, otherwise it directly passes in the F state.

3. C: Once the TDC starts converting the information of the event occurrence time, it is necessary to reset the TDC_start signal (TDC conversion already started) as well as the $PD\_trackmode$ output, in order to begin the tracking of the maximum value of the integrator output. When the rising edge of the PF signal is detected, the machine can evolve in the state D.

4. D: During this phase, instead, both the TDC and the integrator are reset, because they have already completed their tasks. Moreover, the channel logic waits for the PH signal of the peak detector and, once it is arrived, the machine moves to the E state.

5. E: As soon as the FSM enter in this state, the $ADC\_request$ output is set so that the external logic can start the conversion of the energy information. The machine waits for the EOC signal from the external logic, which means that the conversion has been carried out and the channel can return to the original state. Nonetheless, before that, the LED output must be checked. If it is low, it means that the SiPM recovery phase has ended and the FSM can pass in the A state again without any issue. On the contrary, if the SiPM is still in the recovery phase, it is appropriate to wait for it to finish in a further state G.

6. F: When the event is not valid, i.e. when the slow comparator output is low in correspondence of d_LED, the conversion process needs to be halted, because it means that the input current is not of the minimum required amplitude. Here, the TDC is reset and the FSM waits for the $2d\_LED$ signal to go back in A or in G, whether the SiPM recovery phase is over, or not (same situation of E).

7. G: If the conversion process has been already concluded but the SiPM is still in the recovery phase (It may happen when several SiPM pixel are fired at the same time), it is not possible to go back to the initial state yet, otherwise the tail of the input signal would be integrated anyway in the slow path. This excess amount of charge would affect the following conversion because it would be stored in the capacitor of the integrator without any possibility of being discharged. All the signals are restored to their initial state with the exception of INT_reset, which is kept low until the machine goes in A. The signal that leads the FSM to the initial state is the falling-edge of the LED.

The LED output voltage is delayed (two times) by means of a programmable delay circuit. Indeed, through two configuration bits, it indeed is possible to vary the time window between LED and d_LED from 15 ns to 30 ns, while the delay between LED and $2d\_LED$ can be between 30 ns to 60 ns.

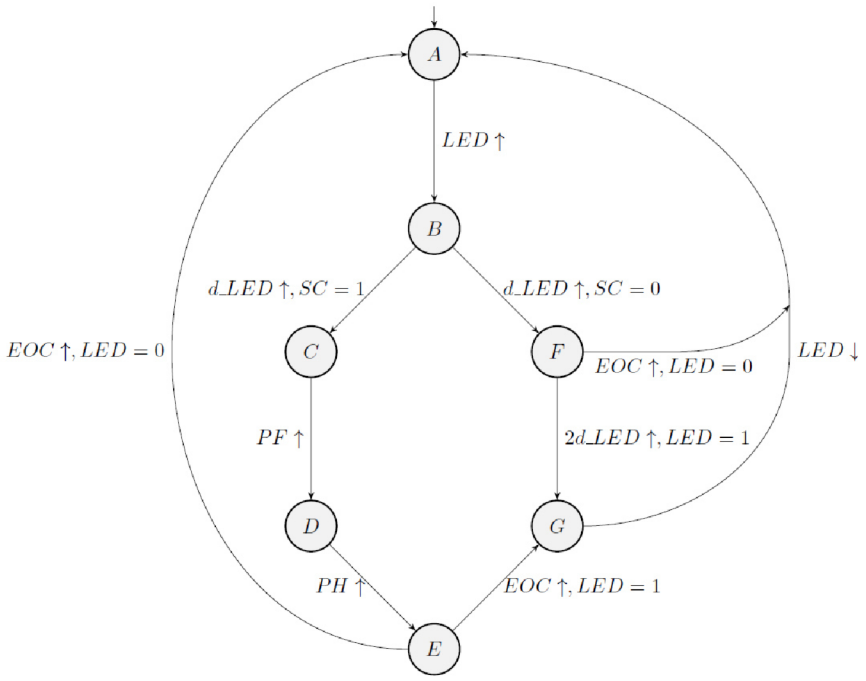

Figure 2.28: FSM state diagram (↑ identifies a rising-edge, while ↓ identifies a falling-edge).

Since the number of the states is seven, three bits must be used for the assignment, which has been done trying to have, at least in the main conversion sequence, state with a code distance equal to one. The assignment reported in table 2.5 has been adopted.

Table 2.5: FSM assignment.

State	Decimal	Code
A	0	000
B	1	001
C	3	011
D	2	010
E	6	110
F	5	101
G	4	100
H	7	111

The state H, included in table but not in the previous description, is *not used*. If, for any reason, the FSM enters in this state, an *autoreset* signal in generated and after 20 ns the machine is reset. Moreover, another output signal called $SM\_monitor$ is generated in this circuit; it can be useful to check if after the start of the conversion it goes back to the initial state A or if, at some point, it is stuck in one of the intermediate states. An additional input that has not been mentioned before, is *reset*, which is a

global asynchronous signal used to force the FSM to the initial state. In table 2.6 the values of all the output signals of the FSM are reported for each state.

After implementing the algorithm in *VHDL*, the RTL netlist has been implemented in the circuit simulator and all its features have been successfully simulated together with all the other analog blocks.

Table 2.6: FSM output signals.

State	TDC_start	TDC_reset	INT_reset	$PD\_trackmode$	$ADC\_request$
A	0	1	0	1	0
B	1	1	0	1	0
C	0	1	0	0	0
D	0	0	1	0	0
E	0	1	1	0	1
F	0	0	1	1	0
G	0	1	1	1	0
H	0	0	1	1	0

2.7 Channel layout

In parallel to the design of the channel, its layout has been drawn, attempting to optimize the position and the shape of each block to improve the circuit performances. Two different versions of the layout have been produced [36].

2.7.1 First version

The first version, presented in [29], has been realized considering as maximum height the one of the TDC, which is the biggest block in the entire channel. This layout is reported in figure 2.29, where the analog inputs are located on the left, while all the outputs (peak detector output voltage and TDCs signals) and the configuration bits are directed toward the right side of the channel, to facilitate the connections with the external circuits.

Among all the blocks, those belonging to the fast path required the most effort because it has been observed, through several post-layout simulations, that the time jitter is strongly dependent on the placement and routing of the different blocks. Therefore, different precautions have been put in place to avoid the deterioration of the time resolution. First of all, the TDC has been placed at the right edge of the channel, for reason of its size, while the other blocks of the fast path (preamplifier, leading-edge discriminator and finite-state machine) have been originally located on the left side, as

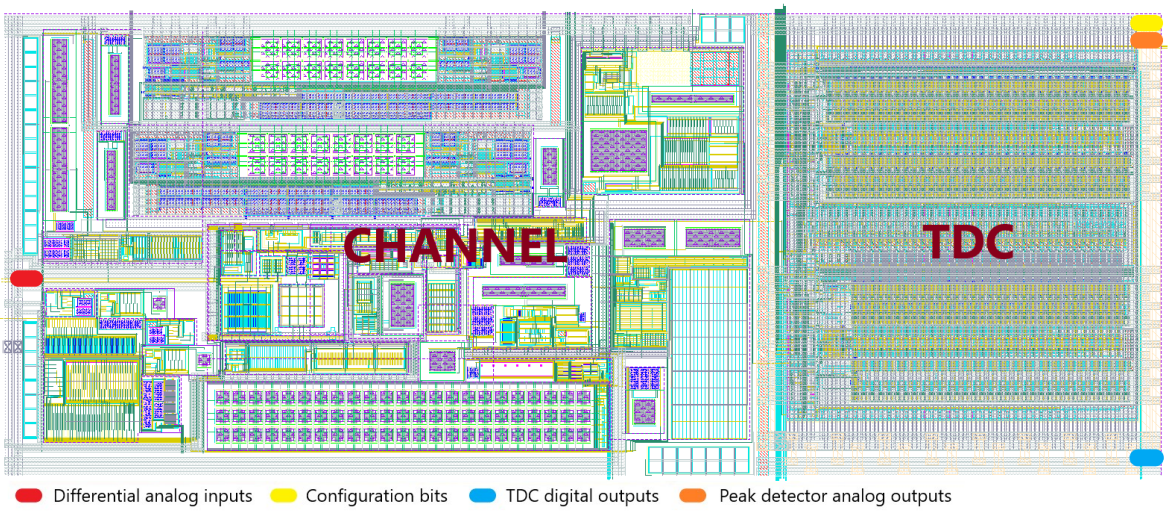

Figure 2.29: First version of the readout channel for SiPMs (Metal 6,7 and 8 are not displayed).

close as possible to each other. However, with this arrangement a long wire between the FSM and the TDC was required, which would have introduced a further degree of uncertainty on the time measurements, because of the presumable couplings with other signals along the path. Since the TDC could not be moved, all the three other circuits have been shifted toward the center of the channel. Obviously, a longer wire was needed between the input pad and the preamplifier, but, given the fact that the preamplifier has a low-impedance inputs, it does not imply any problem on the input signal, other than a small increase on the parasitic series inductance. Nevertheless, this increment is negligible if compared with the wire bonding and the PCB tracks.

2.7.2 Second version

After some meetings with the IC group of SLAC for planning the development a multichannel ASIC, it became necessary to change the aspect ratio of the front-end, so as to fit more channels into the same chip. Moreover, after several simulations, a couple of small bugs has been found in the digital machine and in the peak detector, which led to the design of an auxiliary digital circuit that ensures the proper operation of the channel even with high-current pulses coming from the SiPM.

In figure 2.30, the layout of two channels is reported, whose length and width are respectively 2800 µm x 520 µm. Thus, it is possible to observe that now the width of a single channel is half compared to the first version, so that two of them can be placed at the left of the TDC, one above the other. Since for each front-end a TDC is required, the other TDC has been implemented side by side with the first one, vertically flipped. Such expedient allows the leading-edge discriminators and the digital state machines

of both the readout circuits to have the same distance from their own TDC, so as to reduce the imbalances among the channels.

Even in this second revision, the fast path circuits have been shifted toward the center of the channel for the same reason mentioned before. Nonetheless, this time the channel is longer and the distance that the metal tracks cover from the pads to the preamplifier is greater than the first version. Therefore, it has been decided to insert metal shields connected to ground surrounding the input wires, so as to prevent mismatches between the cathode and anode terminal and limit the coupling with other signals.

As far as the interconnections between the finite-state machine and the TDC are concerned, it has been chosen to implement not only a metal shield, but also a couple of digital buffers for regenerating the signal along the way. For the design of these structures, multiple post-layout simulations have been run to ensure that the additional capacitive load did not affect the slope of the signal and that the time jitter introduced by the buffers was negligible with respect to that of the preamplifier.

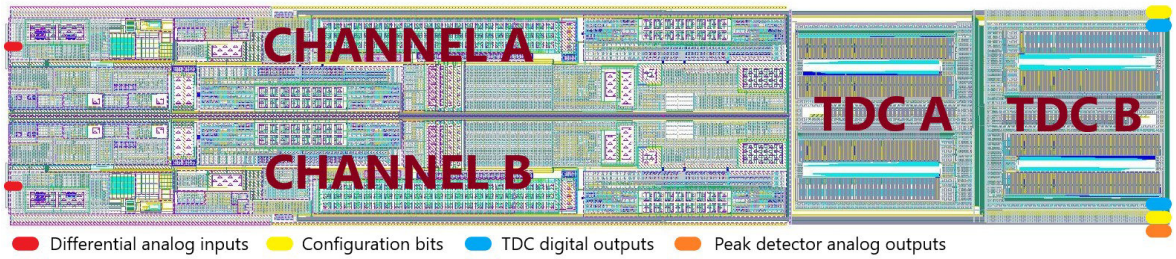

Figure 2.30: Two readout channels for SiPM (Metal 6,7 and 8 are not displayed).

2.8 Post-layout simulations

During the design of the channel, several simulations have been performed to assure the proper functioning of the system and that the requirements for both time and energy measurements were fulfilled. In order to test the whole measurement process, including the FSM and the peak-detector, a *Verilog-A* script has been developed, which is able to simulate the behavior of the external logic, reading the *ADC_request* signal from the channel, digitizing the analog output and sending back the *EOC* signal.

As already stated in the previous paragraphs, the simulated time jitter of the front-end circuit is around 21 ps rms for a single photon signal, while the non-linearity error for the charge measurements is lower than 10 % in the whole range, until 1280 pC. Nonetheless, since the preamplifier is very sensitive to parasitic components, post-layout simulations must be carried out to validate the good performances of the system.

For these purposes, the extraction of the parasitic components of the preamplifier and the leading-edge discriminator has been executed to take them into account in the post-layout simulations.

In table 2.7 the post-layout parameters of the system are reported. The time jitter appears higher than the ideal case, but still its value is compliant with the specifications.

Table 2.7: Simulation results.

Circuit parameters	Post-layout simulation
Dynamic range	1300 pC
Peak value	12 mV
Rise time	0.8 ns
Max slope	18 MV/s
Output noise	600 µV rms
Time jitter	33.5 ps rms

Chapter 3

Data conversion and power management

3.1 Introduction

Within this PhD program, a period of internship at SLAC has been foreseen, during which several ancillary activities have been carried out, besides the development of the SiPM analog channel. Among them, the two main activities have been the optimization of a 12-bit Analog-to-Digital Converter, which has been used in the ASIC for testing the front-end channel (cf. chapter 4), and the design of a capacitorless low drop-out regulator.

As far as the ADC is concerned, a previous version of the circuit was integrated in another ASIC, called CRYO and the characterization measurements resulted in an ENOB *Effective Number Of Bits* lower than 10. The motivations which caused such ENOB lower than expected have been studied and solutions useful to make the circuit more effective have been proposed and implemented.

Another limitation that has been observed in the first version of CRYO was related to the amount of external components required to make the chip work. In fact, because of the number of LDO regulators implemented, the ASIC required multiple external capacitors to achieve good stability and transient response of its outputs. However, these components, whose value can be of several microfarads, take up a lot of external PCB area and, in certain kind of applications where the area occupation is important and multiple ASICs have to be placed one next to each other, this issue is not negligible. For these reasons, a Capacitorless Low Drop-Out regulator has been designed, trying to achieve the same performances of the one already implemented without increasing the silicon area.

These two projects will be described in this chapter, wherein it seemed appropriate

to append a brief overview on the Time-to-Digital Converter (TDC) used in the ASIC described afterward.

3.2 SAR-ADC optimization

3.2.1 Overview

In the CRYO family of ASICs, the digitization of the output signals, coming from the analog front-ends, is performed by means of a 12-bit *Successive Approximation Register* (SAR) *Analog-to-Digital Converter* (ADC) with a throughput of 8 MS/s. Actually, this converter is included in a more complex system, called *SubBank*, which comprises also four *Sample And Hold* (SAH) circuits, an analog multiplexer and a fully-differential signal driver, as shown in the block diagram in figure 3.1 [37] [38].

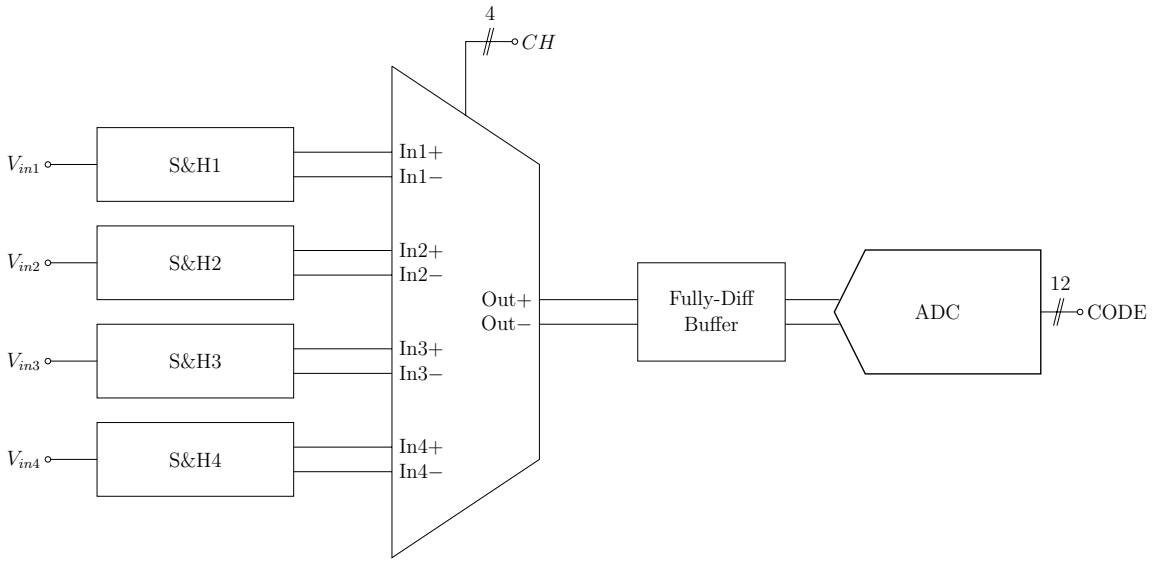

Figure 3.1: Block diagram of the SubBank.

Each of the SAHs is connected to the single-ended analog output of a readout channel, which is sampled, converted into a differential signal and then held as long as necessary. The differential outputs of these four SAHs are then applied to an analog multiplexer, which selects the right input to be conveyed to the signal driver and thus to the ADC for the conversion. This operation is handled by four internal non-overlapping signals, called CHx (x from 1 to 4), which are generated from the main clock of 112 MHz. These timing signals not only manage the switching of the multiplexer, but also handle the sampling process inside the SAHs; indeed, all the SAHs sample the analog signals at the same time, when $CH1$ is high, and hold the information until the multiplexer allows them to transmit the signal to the ADC [39].

As far as the ADC is concerned, its block diagram is similar to the one in figure 3.2. The core of this circuit is the differential capacitor network CDAC (*Capacitor Digital-to-Analog Converter*), whose outputs are connected to the input terminals of a multistage comparator though a bunch of switches. The result of this comparator (also fully-differential) is thus sent to the SAR logic that oversees the conversion process by opening and closing the switches of the CDAC and refreshes the output bus *ADC_C*.

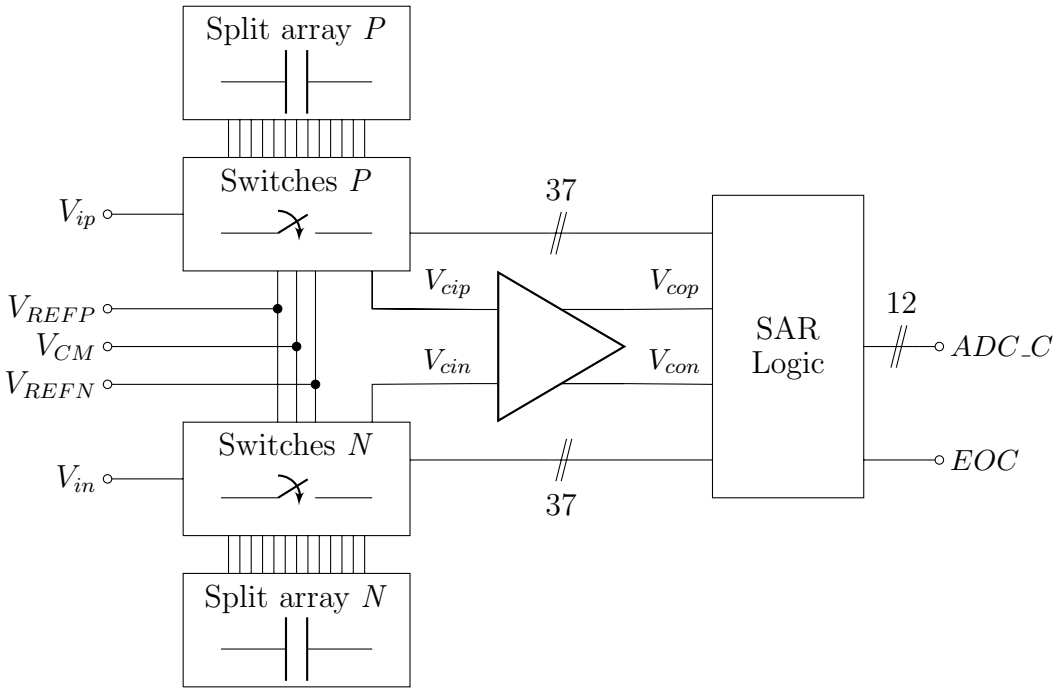

Figure 3.2: Block diagram of the 12-bit SAR ADC.

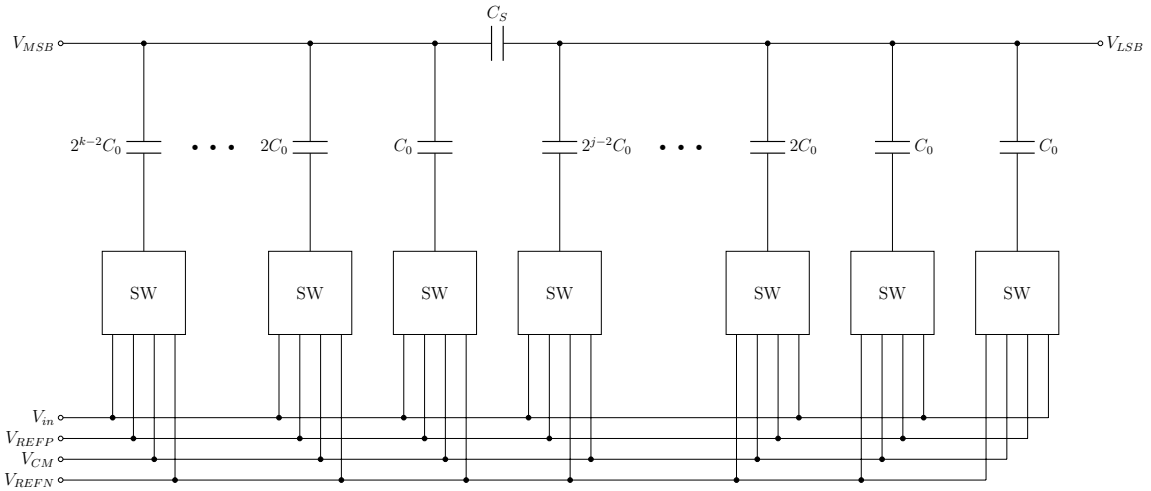

Figure 3.3: Schematic of a generic n-bit V_{CM}-based split array ($n = j + k$).

The operation of this circuit is beyond the purpose of this work and more details can be found in [40]. In this section, instead, the optimization process of the ADC

designed and implemented in the CRYO family of chips is briefly presented, which mainly concerns the improvement of the capacitor network. In figure 3.3, the generic schematic of a n-bit CDAC is reported, which is based on a $k-j$ split array configuration ($n = k+j$), where the ratio between each capacitor and the basic element C_0 is binary-weighted while the cross-capacitor C_S must be $2^{k-1}/\left(2^{k-1}-1\right)$ times C_0. This scheme allows a relevant reduction not only of the silicon area, but also of the switching energy and the settling time with respect to the standard architecture. Moreover, as is it possible to observe in the figure, the structure requires three different reference voltages and not two as usually happens in SAR ADC [41]; in particular, besides the upper and the lower voltages V_{REFP} and V_{REFN}, another reference V_{CM} has been considered, which corresponds to the mean value of the dynamic range of the converter. This V_{CM}-based approach allows the elimination of the biggest capacitor (2^{k-1} times C_0) that a conventional split array CDAC would require, with a further reduction of area and power dissipation. The output of this circuit is the node V_{MSB}, which is connected to one of the inputs of the comparator though a switch, while V_{LSB} is left open. In figure 3.4, instead, the 12-bit ($k = 7$ and $j = 5$) split array implemented in the concerned converter is reported.

Nevertheless, for this architecture the quality of the conversion strongly depends on the relative value of the capacitors, since even a small variation might result in sluggish performances, especially if it affects C_S. These changes may be caused not only by the process variations, but also by different parasitic capacitances that insist on the numerous nodes of the array. This is precisely the issue that has been experienced in the SAR ADC implemented in the ASICs developed for CRYO experiments, where the circuit parameters between simulations and measurements substantially differ from each other, because of the presence of multiple parasitic components that have a negative impact on the values of the capacitors. Indeed, in the simulations carried out during the design phase, the ADC showed an *Effective Number Of Bits* (ENOB) above 11.9 bits and thus close to the resolution of the ASIC, while during the measurements an ENOB slightly lower than 10 has been obtained. Therefore, it has been required to investigate the cause of this problem, analyzing the layout and the corresponding schematic with the parasitic components so as to find a remedy without revolutionizing the whole design.

Before starting with the description of the steps that have been carried out during the optimization process, it is important to provide few other details about the ADC. In particular, it requires a 2 V as supply voltage, while the two references V_{REFP} and V_{REFN} are respectively 1.8 V and 0.2 V, which define the full scale range of $FSR =$ 1.6 V. For the V_{CM}-based approach, the common-mode voltage is also required, equal to $V_{CM} = 1$ V for this converter. At this point, it is also possible to derive the voltage

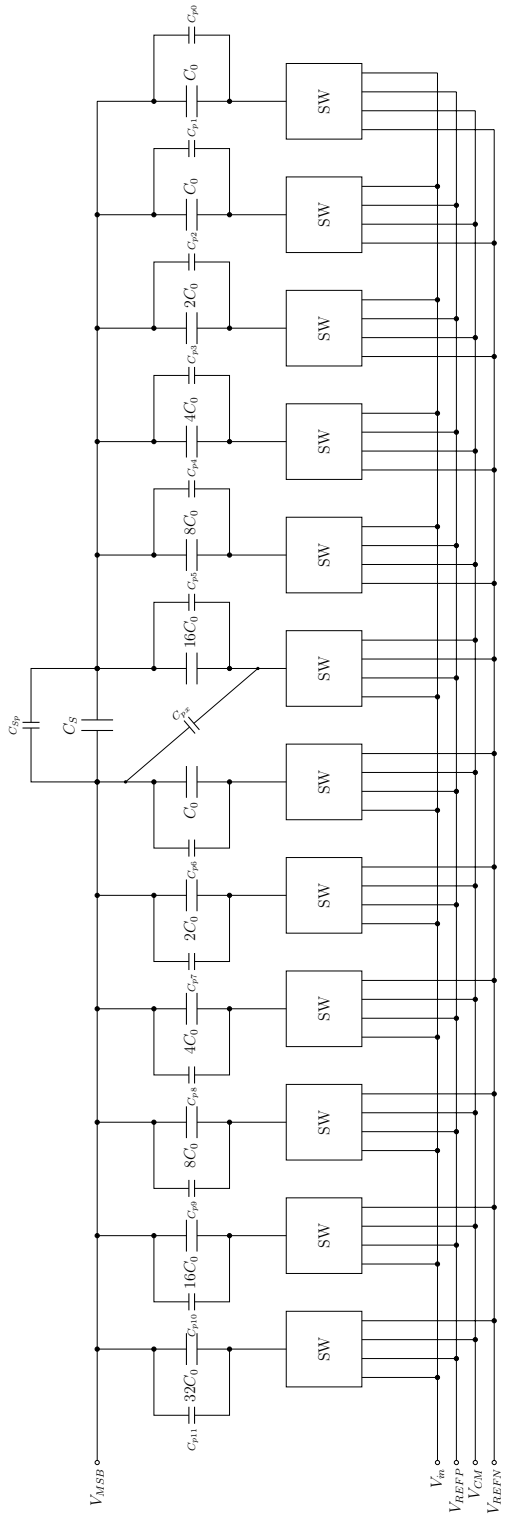

Figure 3.4: Simplified schematic of the V_{CM}-based split array ($k = 7$ and $j = 5$) with parasitic components.

Least Significant Bit LSB (or resolution) of the ADC by simply dividing the span (1.6 V) by the total number of intervals:

$$LSB = \frac{FSR}{2^n} = \frac{1.6\,\mathrm{V}}{2^{12}} = \frac{1.6\,\mathrm{V}}{4096} = 390.625\,\mu\mathrm{V} \tag{3.1}$$

Therefore, in order to obtain the output voltage value on the basis of the result of the conversion, it is possible to apply the following formula:

$$V_{out} = V_{REFN} + LSB \cdot ADC\_C \tag{3.2}$$

where ADC_C is the output code of the converter that ranges from 0 to 4095. Furthermore, the ADC provides an ancillary digital output, called EOC (*End Of Conversion*) that is set by the SAR logic whenever a conversion has been accomplished, regardless of the channel that is being processed.

3.2.2 Optimization procedure

In order to test the performances of this system in terms of ENOB, a transient simulation has been set up by feeding the ADC with a differential sinusoidal waveform with a frequency of 242.187 50 kHz, an amplitude of 0.7728 V and a DC offset of 1 V. The evolution of such signal does not reach neither the upper boundary nor the lower one, so as to not achieve the saturation of the ADC. This test-bench has been run using different entry netlist (schematics and parasitic extractions) and then analyzing the value of the digital output bus over time and most importantly the Fast Fourier Transform of these data. Consequently, it has been possible to derive the information about the ENOB by using a custom MATLAB script.

The first step of the optimization process has been to perform the transient simulation with the schematic entry, configuring the simulator with the most accurate mode available; in this way, the target value of ENOB has been obtained, which turns out to be 11.95. Afterwards, the extraction of the schematic entry comprising the parasitic components has been carried out using different settings in terms of capacitance filter, beginning with the default value of 10 fF, until 0.001 fF; this option allows the removal of all the parasitic capacitance below the specified value. It has been observed that, running the simulation with the extracted entries produced with the 0.01 fF filter, the ENOB approached the one achieved during the measurement; the use of tighter filters did not guarantee any advantage, but instead a relevant lengthening of the simulation time. A summary of the results obtained during this phase is reported in table 3.1.

After identifying the right PEX schematic to use during simulations, a deep analysis regarding the influences of the parasitic capacitances of the split array on the

Table 3.1: Summary of the ENOB obtained during the preliminary simulations.

Netlist	Simulator model	ENOB
Schematic	*SPICEXD*	11.618
	SPICEAD	11.952
PEX with 10 fF	*SPICEXD*	7.818
	SPICEAD	7.884
PEX with 1 fF	*SPICEXD*	9.661
	SPICEAD	9.950
PEX with 0.01	*SPICEXD*	9.882
	SPICEAD	9.889

performances of the ADC has been done. In particular, three main types of parasitic components have been found: the parasitic capacitances in parallel to the actual elements of the structure (C_p in figure 3.4), the capacitances between the output nodes and ground (not shown in the schematic) and last the one between V_{MSB} (both N and P) and the bottom plate of $16C_0$ in the right side of the array. First of all, as expected, it turned out that, owing to the presence of the parasitic components, all the capacitor values differed from their nominal ones, thus jeopardizing the binary-weighted structure. However, as it is possible to notice from table 3.2, where the nominal and the parasitic capacitances are reported, the elements that most influence the performances of the converter is C_{Sp}, which diverge from the target value of more than 70 %, and C_{px} that instead is not supposed to be there. The percentage error mentioned in table 3.2 has been derived by means of the following equation:

$$E = \frac{|C_{ip} - RC_{0Dp}|}{RC_{0Dp}} \cdot 100\,\% \tag{3.3}$$

where C_{ip} is the parasitic capacitance of the i-th capacitor, R is the corresponding ratio and C_{0Dp} is the parasitic capacitance of the dummy capacitor C_{0D}.

It has been proven that, in this particular case, the capacitances towards ground do not have a significant impact on the value of ENOB, in contrast to the other parasitic components already mentioned. Indeed, before revising the layout, further simulations have been performed with a customized version of the PEX (Parasitic Extracted) netlist, balancing all the C_p components with a binary-weighted configuration and removing the cross-capacitor C_{px}, thus achieving a major improvement of the ENOB by slightly less than 1 bit. Starting from these results, it has been decided to understand the causes of such mismatches among the parasitic capacitances by directly analyzing the layout and consequently adjust it.

The layout of the split array of capacitors is reported in figure 3.5, where several

Table 3.2: Initial values of the parasitic capacitances and their relative errors with respect to the dummy capacitance C_{0D}.

Device	Ratio	Nominal value	Parasitic C	Difference
C_{0D}	1	62.074 fF	2.329 fF	-
C_1	1	62.074 fF	2.316 fF	0.56 %
C_2	2	124.148 fF	4.587 fF	1.31 %
C_3	4	248.296 fF	9.581 fF	2.84 %
C_4	8	496.592 fF	17.131 fF	8.05 %
C_5	16	993.184 fF	35.616 fF	4.42 %
C_6	1	62.074 fF	2.399 fF	3.01 %
C_7	2	124.148 fF	4.537 fF	2.60 %
C_8	4	248.296 fF	9.191 fF	1.34 %
C_9	8	496.592 fF	17.660 fF	5.21 %
C_{10}	16	993.184 fF	34.700 fF	6.89 %
C_{11}	32	1986.37 fF	74.483 fF	0.06 %
C_S	32/31	64.076 fF	0.627 fF	73.91 %

instances of the elementary unit equal to C_0 have been arranged with a common-centroid architecture for both the left and the right side of the structure. The V_{MSB} (in red) and V_{LSB} (in orange) nodes are shared among the capacitors by means of the seventh metal and connected to the top plates through *M67* vias placed in the middle of the devices. On the other hand, although the connections to the bottom plates are always realized with *M7* and the same kind of vias, these latter have been implemented in the four corners of the capacitor, while the routing to the switches has been carried out by using the lower metals (from the second to the fourth).

As regards to the parasitic capacitances C_p that such design exhibits, it has been observed that they are not related to the wires that cross the entire structure to reach the switches, but they are due to the couplings between V_{MSB} and V_{LSB} nodes with their respective bottom plates. Their values not only depend on the number of elements of which each capacitor consists, but also on their locations in the structure. Therefore, the easiest way to solve this problem and counterbalance the mismatches among the parasitic capacitances has been to develop a custom metal patch with *M7* to be placed on the split array, aimed at increasing where necessary the capacitive load and restoring the binary-weighted architecture. The layout of this patch is reported in figure 3.6, which allowed an improvement in the performances of the ADC without changing its layout.

As pointed out before, the critical part is represented by the junction between the left and the right sides of the split array. Indeed, the capacitance of C_S between the nodes V_{MSB} and V_{LSB} must be precisely 32/31 times the value of the basic element

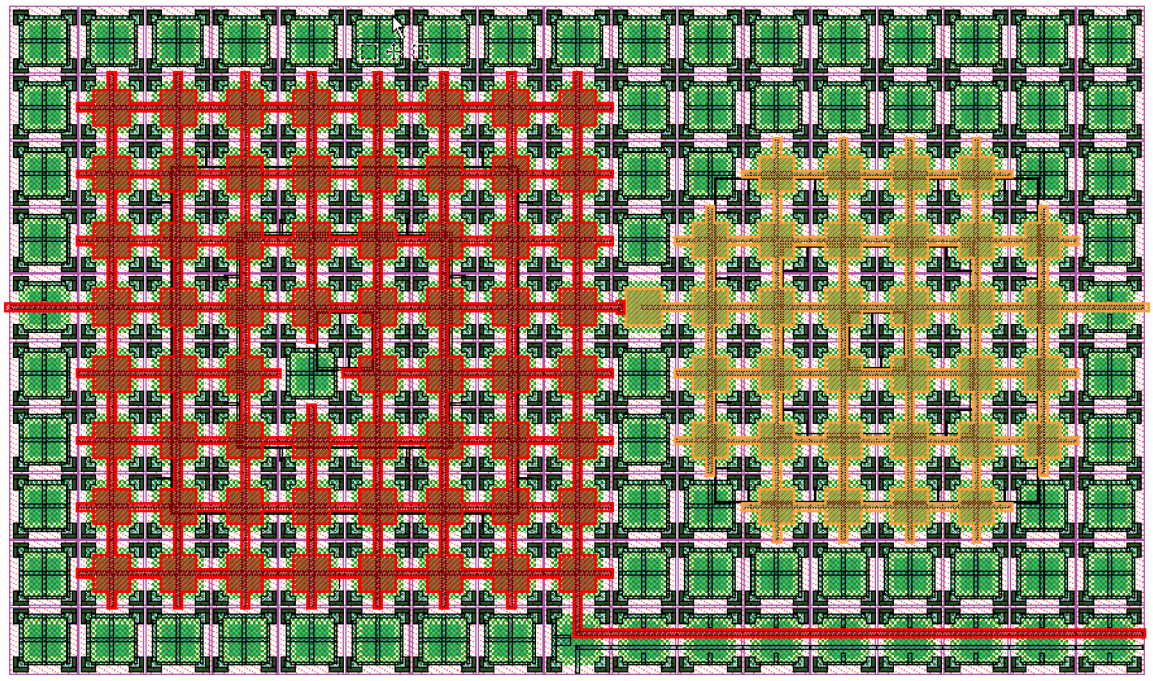

Figure 3.5: Layout of the split array of capacitors. The node highlighted in red is V_{MSB}, while the one in orange is V_{LSB}. Between them, the capacitor C_S can be noticed.

C_0 otherwise the quality of the conversion would be compromised. This relation must remain valid also taking into account the parasitic components, but, as already noticed in the extracted netlist, the error between C_{p0} and C_{Sp} appears to be higher than 70 %. To solve this issue, a forked frame has been added to the metal patch so as to reduce this discrepancy and make the structure fully balanced. In figure 3.7 the zoom of the split array nearby the capacitor C_S is reported, where this junction is noticeable; it has been designed by adjusting its length and clearance with the V_{LSB} in such a way that the parasitic capacitance C_Sp was the one desired and the error minimized. During this optimization phase, the capacitances and percentage errors in table 3.2 have been repeatedly verified, till obtaining the final values reported in table 3.3 where the error never exceeds 0.1 %.

The last step that has been performed in the optimization of the CDAC concerns the location and the suppression of the cross-capacitance C_{pX}, whose value should be the lowest possible. It acts between the V_{MSB} node and the bottom plate of $16C_0$ (in the right side of the split array) and it has been observed during simulations that, even if its value is equal to just 1.6 fF, it heavily affects the performances of this device. After a careful analysis of the layout and a "trial and error" procedure, it has been possible to cut down this parasitic effect by simply removing the vias that connect *M6* and *M7* from the corners on the left side of the capacitor contiguous to C_S (cf.

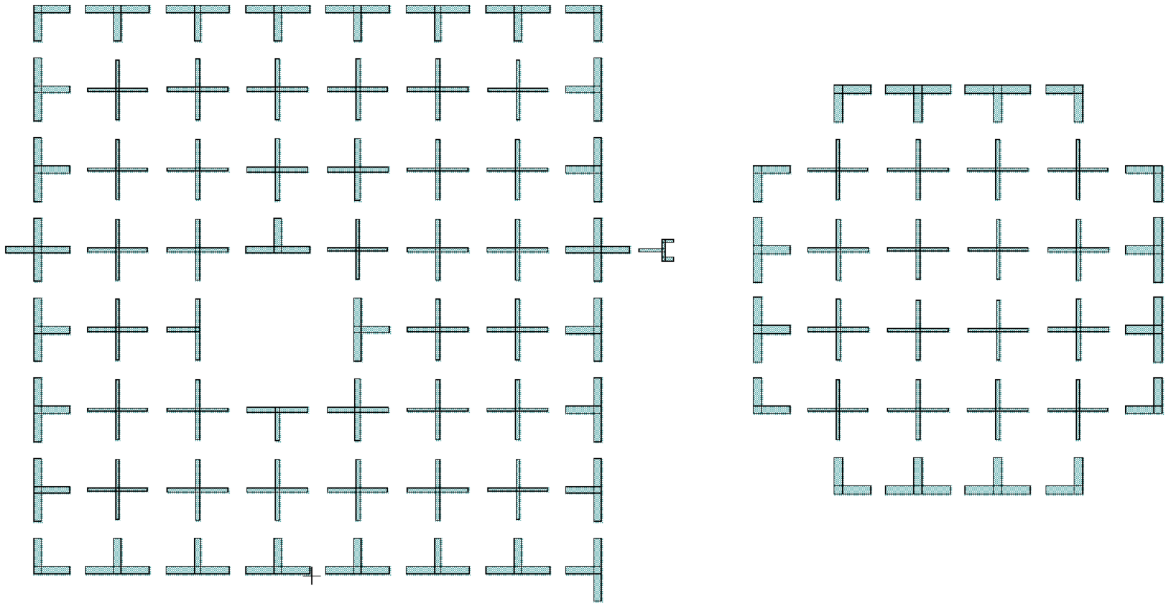

Figure 3.6: Layout of the custom metal patch.

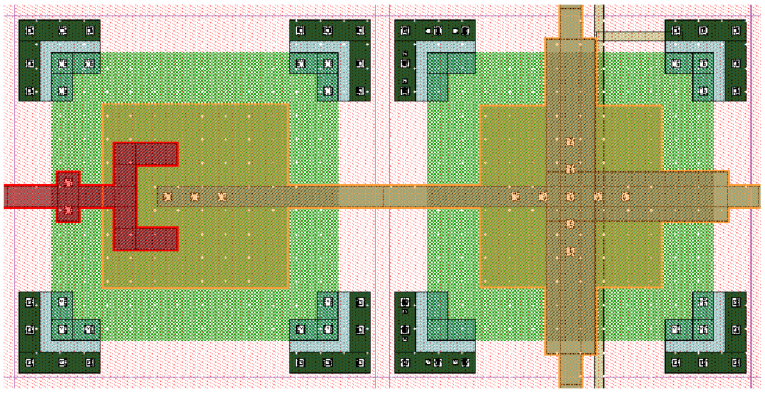

Figure 3.7: Zoom of the split array nearby the capacitor C_S. The node highlighted in red is V_{MSB}, while the one in orange is V_{LSB}.

figure 3.7). This operation has enabled the reduction of such capacitance from 1.6 fF to just 0.07 fF, allowing the ADC to enhance its performances; indeed, at the end of this optimization process on the split array of capacitors, the converter exhibited an ENOB of 10.60 bits, approximately 0.70 bits higher than the initial one.

In addition to the improvement of the CDAC, it has also been deemed appropriate to examine the impact of the comparator on the ENOB. In particular, the analysis has been performed exclusively by increasing the bias current of the two amplification stages, without altering neither its schematic nor the layout. Actually, the capability of changing the bias current of the comparator has been already implemented in the first version of the SubBank, but it could only range from 10 µA to 25 µA in 5 µA steps; for all the previous simulations, the bias circuit has been configured so as to deliver

Table 3.3: Final values of the parasitic capacitances and their relative errors with respect to the dummy capacitance C_{0D}.

Device	Ratio	Nominal value	Parasitic C	Difference
C_{0D}	1	62.074 fF	2.502 fF	-
C_1	1	62.074 fF	2.503 fF	0.04 %
C_2	2	124.148 fF	5.003 fF	0.02 %
C_3	4	248.296 fF	10.001 fF	0.07 %
C_4	8	496.592 fF	19.994 fF	0.10 %
C_5	16	993.184 fF	40.015 fF	0.04 %
C_6	1	62.074 fF	2.500 fF	0.08 %
C_7	2	124.148 fF	5.003 fF	0.04 %
C_8	4	248.296 fF	10.009 fF	0.01 %
C_9	8	496.592 fF	20.020 fF	0.02 %
C_{10}	16	993.184 fF	40.004 fF	0.07 %
C_{11}	32	1986.37 fF	80.076 fF	0.01 %
C_S	32/31	64.076 fF	2.581 fF	0.06 %

the maximum available current to the comparator. During this further optimization phase, the maximum bias current at schematic level has been increased up to 45 µA and the ENOB has been assessed for all the configurations. Specifically, setting it up to 35 µA, a considerable improvement in the ENOB has been obtained, achieving 11.66 bits; further increases of the bias current have resulted in minor enhancement in the ENOB of just few tenths of bits. Therefore, it has been decided to extend the capability of adjusting the current by means of two additional configuration bits, thought which it is possible to reach 45 µA.

Before finalizing this section, it is worth reporting, as an example, a couple of plots obtained with the result of the last transient simulation that has been carried out. In particular, in the first subplot of figure 3.8, the behavior of the bus ADC_C is shown in decimal notation, while in the second graph its FFT is presented. For each iteration, the FFT waveform has been used as input variable of a custom MATLAB script aimed at evaluating the ENOB. Eventually, as already said, an ENOB of 11.66 bits has been obtained, 1.77 bits higher than the original one. With regard to this latter, in the second plot of figure 3.8, the FFT of the output waveform obtained with the original PEX netlist has also been reported dotted. In this case, it is clear how the odd harmonics have a not-negligible amplitudes, which affect the THD, the SINAD and of course the ENOB. The optimization process has allowed the suppression of such harmonics and thus the increase of the ENOB.

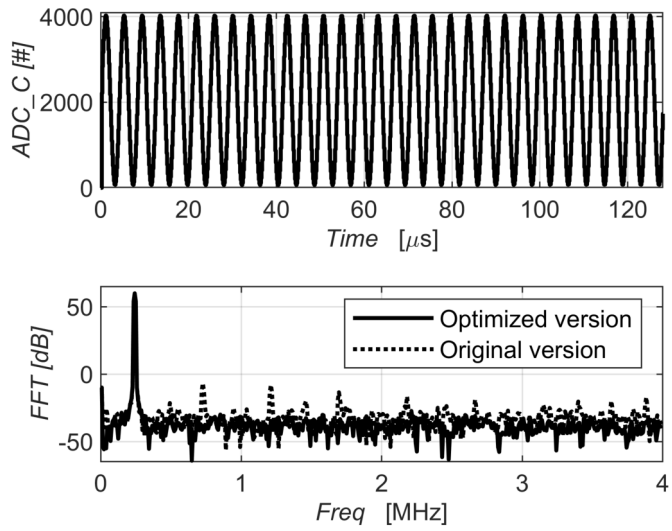

Figure 3.8: Behavior of the *ADC_C* output bus (above) and comparison between the FFT of the output waveforms of the optimized version and the original one (below).

3.3 Time to Digital Converter

Each of the eight channels implemented in the ASIC comprises a *Time to Digital Converter* (TDC), whose purpose is to measure the time difference between the output of the leading-edge discriminator and a reference signal, named *TDC_Clock*, and to convert this information in digital data. In this section, some information about its architecture and the way it works will be provided. In the framework of this project, in order to obtain the best resolution possible for the time information on the occurrence time of the detected event, it has been deemed appropriate to employ a TDC with a 10 ps binning, already developed by the IC group of SLAC [42].

The key component of this block is the differential shunt-capacitor voltage controlled delay cell, whose schematic is reported in figure 3.9. The main feature of this circuit is that the time delay introduced by the circuit can be set to the desired value by adjusting the *Vctrl* signal. Indeed, the change in the control voltages of the cell causes a variation of its capacitive load and thus of its delay time.

The TDC chosen for this project has been designed according to the Vernier architecture (cf. figure 3.10), which involves two parallel delay lines made up of the cascade of multiple instances of the delay element just introduced. The only differences between the two lines concern the delay time of each single cell and the input signals. In particular, the control voltage of the delay element in the first line, also named "fast", ensures that the propagation time of each cell is 120 ps, while it is 130 ps for the elements of "slow" line. This establishes the resolution of the TDC, which is therefore able to distinguish time differences as low as 10 ps. As far as the inputs are

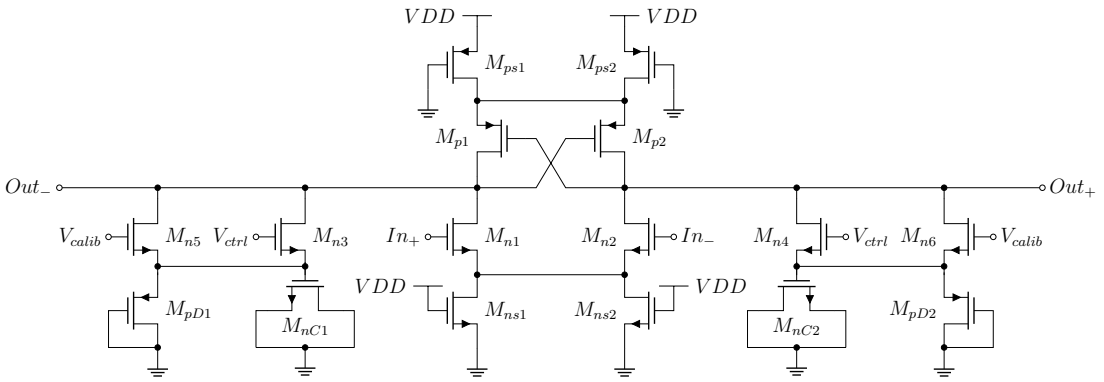

Figure 3.9: Schematic of the shunt-capacitor voltage-controlled delay cell [42].

concerned, the trigger signal of the "slow" line is called *START* and it is connected to the output terminal of the leading-edge discriminator, while the "fast" line is triggered by the *TDC_Clock* signal, i.e. the *STOP* input of the TDC. When an event is detected by the analog front-end, the comparator output signal enters the Vernier structure and begins propagating thought the "slow" delay line, while the internal logic waits for the subsequent pulse of the *STOP* signal. As soon as it is picked up, this signal starts propagating along the "fast" line and, whenever it passes through a delay cell, it approaches the other signal of an amount of time equal to the resolution of the Vernier system. The number of elements that the *STOP* signal needs to overtake the *START* signal establishes the result of the conversion.

In order to optimize the area occupation, the Vernier has been designed as a cyclic structure, so as to employ one fourth of the cells required by a linear delay chain, while preserving the same dynamic range. For example, for the concerned device, which exhibits 1.28 ns of range, it has been possible to implement only 32 delay cells, instead of 128, with a saving of a considerable amount of silicon area. However, to ensure the proper operation of such structure, an auxiliary digital machine has been included in the Vernier as well.

As a further feature, this TDC offers the possibility to also evaluate time differences greater than 1.28 ns. Indeed, a "coarse" delay line with a bin of 160 ps has been added right before the Vernier circuit just described, according to the block diagram in figure 3.11, so as to reach a dynamic range of 20.48 ns. Although the Vernier and the "coarse" delay line can work as a standalone 7-bit TDCs, with the appropriate settings they can also operate concurrently to obtain a time measurement with a resolution of 10 ps and a range of 20.48 ns. In particular, it is necessary to make the "coarse" line work as a 4-bit TDC with a resolution of 1.28 ns and then exploit the "fine" TDC to obtain the residue of the conversion, i.e. the least significant bits of the whole digital word. The time difference Δt between the rising edges of *START* and *STOP* can be derived with

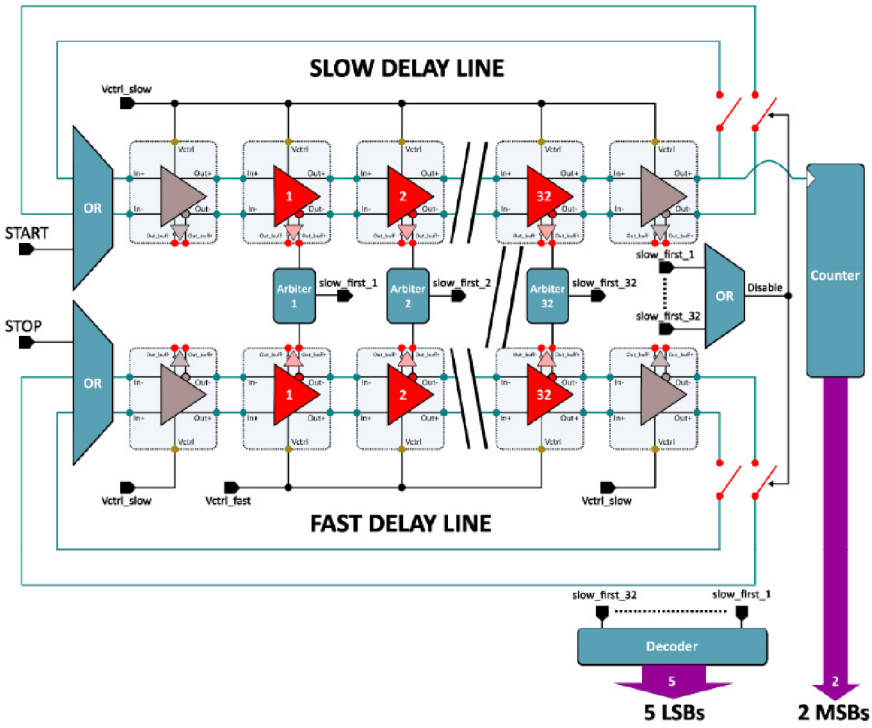

Figure 3.10: Block diagram of the cyclic Vernier architecture that includes the "slow" and the "fast" delay lines [42].

the following equation:

$$\Delta t = (TDC\_C < 3:6 > +1) \cdot 1.28\,\text{ns} - TDC\_F < 0:6 > \cdot 10\,\text{ps} \tag{3.4}$$

where $TDC\_C < 3:6 >$ are the four most significant bits of the coarse conversion, while $TDC\_F < 0:6 >$ is the result of the fine measurement. Instead, considering the only coarse conversion, it is possible to derive the time difference between $START$ and $STOP$ by means of the following expression:

$$\Delta t = TDC\_C < 0:6 > \cdot 160\,\text{ps} \tag{3.5}$$

Before concluding this last section and the third chapter, it is worth showing an example of conversion. In particular, after sending the reset signal to the TDC, as shown in the first subplot in figure 3.12, two pulses have been delivered respectively to the $START$ and $STOP$ terminals in order to initiate the TDC and assess the time difference between their rising edges. The voltage sources have been set so that the pulses are generated with a time difference $\Delta t = 8.48\,\text{ns}$ one from the other, as reported in the second subplot of figure 3.12.

As soon as the rising-edge signal of the $START$ signal is detected, the TDC starts

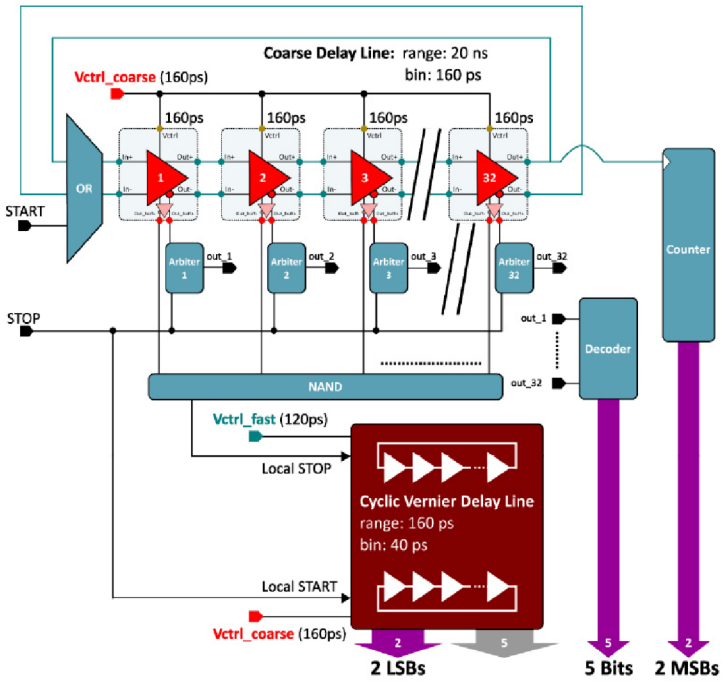

Figure 3.11: Block diagram of the TDC with the coarse delay line [42].

the conversion and raises the *Busy* output pin; this terminal is eventually reset when the device identifies the value of Δt and refreshes its output data. The process may also be monitored by means of the *DataReady* output, through which a pulse signal is generated whenever the output data of the TDC are ready to be collected and then analyzed. The plots of the *Busy* and *DataReady* signals of the concerned example are shown in figure 3.13.

It is important to point out that the conversion time is not constant, but it is function of Δt. Moreover, the updates of the output bus (coarse and fine) do not occur at the same time, but the coarse TDC starts working before the Vernier structure and thus makes data available earlier. This can be observed in figure 3.14, where the evolution of both *TDC_C* (above) and *TDC_F* (bottom) buses are reported along with their digital values. The final value of the fine bus is 48, while the coarse data settles down to 53, which correspond to 0110101 in the binary number system (with 7 bits). In order to apply equation (3.7), the 4 most significant bits of *TDC_C* must be counted ($0110_2 = 6$), so that it becomes:

$$\Delta T = (6 + 1) \cdot 1.28\,\text{ns} - 48 \cdot 10\,\text{ps} = 8.48\,\text{ns} \tag{3.6}$$

Actually, the TDC offers two further outputs called *Overflow*, one for each type of conversion. They allow the external logic to determine when an overflow takes place during the conversion. During normal operation, this error should not occur, but it

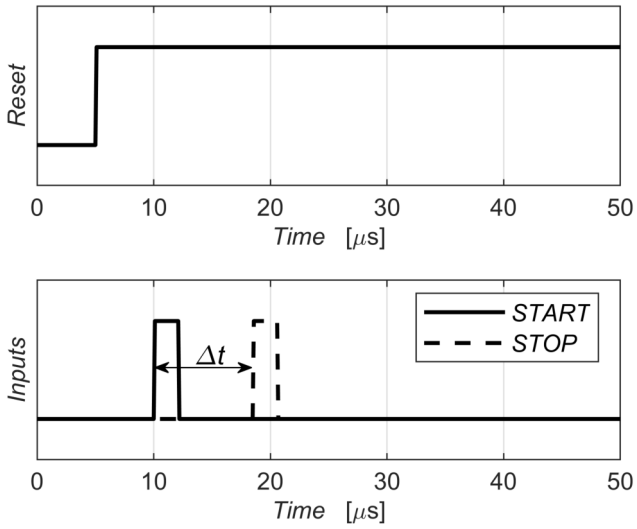

Figure 3.12: Input signals of the TDC.

may emerge when the period of the timing signal is larger than the range of the coarse TDC. Specifically, for all the simulations a clock frequency of 50 MHz has been chosen, which is greater than the minimum recommended value of 1/20.48 ns.

3.4 Capacitorless LDO design

3.4.1 Overview

During the design phase of both the ASICs for the CRYO experiments (DUNE and nEXO), it has been decided to lower the nominal power supply voltages of the employed CMOS technology in order to guarantee the smooth functioning of the chips for several years. Indeed, the core voltages have been reduced respectively from 2.5 V and 1.2 V to 2 V and 1 V, by means of multiple Low Drop-Out (LDO) voltage regulators. These circuits are able to adjust the external power supply of 2.5 V and obtain a stable voltage of the desired value (2 V or 1 V), based on the type of MOSFETs employed in the circuit powered by the LDO.

Each LDO regulator has been designed using the conventional architecture, which is reported in figure 3.15, with a pass transistor of type p, an error amplifier, a feedback network of resistors ($RF1$ and $RF2$) and a load capacitor C_L. The latter can be of several tens of µF to ensure good stability and transient response and it is always located outside the chip, on the printed circuit board (PCB) on which the ASIC is mounted, since such large capacitance is impractical in integrated technology. However, when the number of LDOs inside the chip is significant, as in the case of both

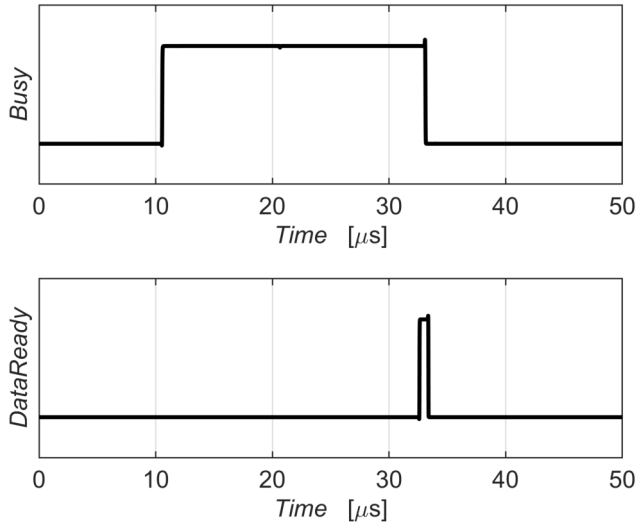

Figure 3.13: Auxiliary output signals provided by the TDC.

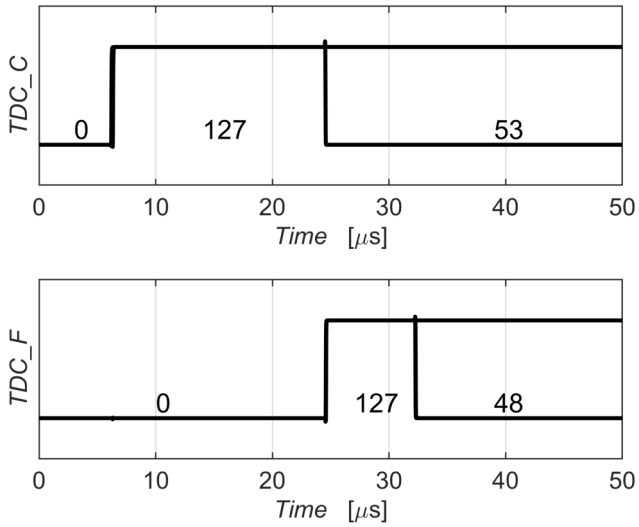

Figure 3.14: Values of the *TDC_C* (above) and *TDC_F* (bottom) buses during the conversion process.

these ASICs, the area occupation on the PCBs due to the external components may become relevant, especially in those applications where the space management of the whole electronic system is crucial. For this reason, the design and development of a CapacitorLess-LDO (CL-LDO) has been demanded, so as to replace the standard LDOs already implemented in these System on Chip (SoC) [43].

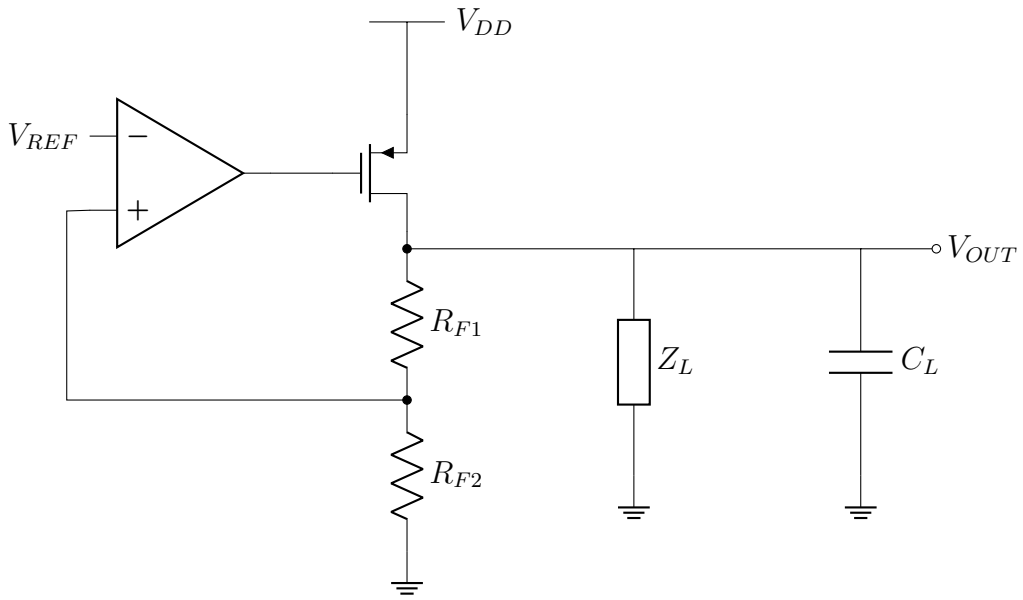

Figure 3.15: Schematic of a conventional LDO.

First, the CL-LDO for the CRYO analog channels inside the ASIC has been designed, which must provide an output voltage of 2 V with the lowest possible noise contribution. In particular, the specifications of a commercial low-noise device from *Analog Devices* have been considered as reference, which has been already used for experiments with Liquid Argon (LAr), where the operating temperatures of the system are usually really low (less than 100 K or −173 °C). This is also the case of nEXO and DUNE experiments, where the electronic systems work in a temperature-controlled environment inside a cryostat, respectively at −113 °C and −186 °C. Specifically, the reference LDO has a spectral density of $5\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$ at 1 kHz, where the flicker noise is the main contribution, while for frequencies greater than 100 kHz the spectral density is around $1.4\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$. Moreover, the data sheet of this device reports a total integrated noise from 10 Hz to 100 kHz of 1.6 µV rms. As far as the other specifications of this circuit are concerned, a maximum output current of 50 mA is requested, with the possibility of scaling this value up by simply connecting multiple LDOs in parallel. Furthermore, in order not to affect the good features of the front-end channels, it is important to keep the line regulation less than 2 %, the load regulation lower than $20\,\mathrm{mV\,A^{-1}}$ and the $PSRR$ greater than 60 dB at least up to 1 MHz.

The load regulation denotes how the output voltage varies in correspondence of a change of the current drawn by the load, accordingly to the equation:

$$\text{Load regulation} = \left.\frac{\Delta V_{OUT}}{\Delta I_L}\right|_{\Delta V_{IN}=0} \tag{3.7}$$

while the line regulation is defined by the following expression:

$$\text{Line regulation} = \left.\frac{\Delta V_{OUT}}{\Delta V_{IN}}\right|_{\Delta I_L=0} \tag{3.8}$$

which measures the variation of V_{OUT} when the power supply changes. Regarding the *Power Supply Rejection Ratio* ($PSRR$), instead, it is a parameter that estimates the ability of a voltage regulator to reject any ripple on the power supply. In table 3.4 the specification required for the CL-LDO are summarized; they must be fulfilled for both nEXO and DUNE temperatures, but it is important that the device also operates at room temperature for testing and debugging purposes.

Table 3.4: Specs for the capacitorless LDO regulator.

Parameter	Condition
V_{IN}	2.5 V
V_{OUT}	2 V
$\overline{v_n}$ @ 1 kHz	$5\,\text{nV}/\sqrt{\text{Hz}}$
$\overline{v_n}$ @ 100 kHz	$1.4\,\text{nV}/\sqrt{\text{Hz}}$
v_n [1]	1.6 µV rms
$PSRR$ until 1 MHz	> 60 dB
Load regulation	$< 20\,\text{mV A}^{-1}$
Line regulation	$< 2\,\%$

3.4.2 Design and simulations

For the design of the CL-LDO regulator, several architectures have been examined and studied in order to pick the one that allows the achievement of the required specifications. Eventually, the any load stable LDO with a multiloop structure proposed by Vadim Ivanov made the cut, because of its excellent features in terms of stability, noise and $PSRR$. The simplified schematic of the regulator that has been implemented is reported in figure 3.16 [44].

This circuit comprises, as error amplifier, a simple differential pair with active load, whose non-inverting input is connected to the reference voltage V_{REF}, while the other input is connected to the output V_{OUT} through a voltage divider R_{F1} –

[1] Total integrated noise from 10 Hz to 100 kHz

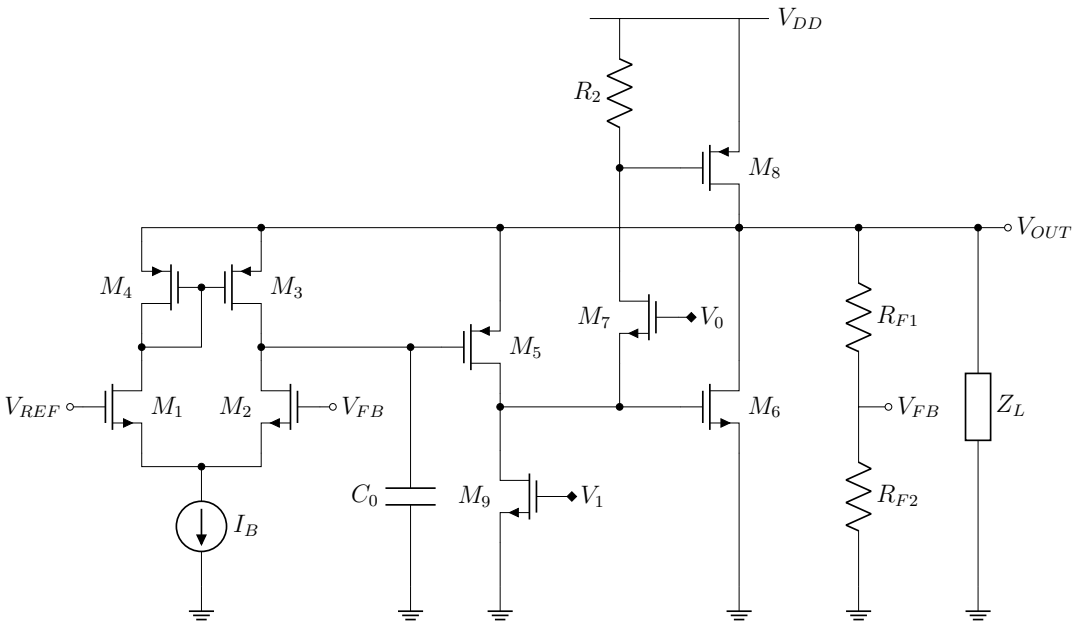

Figure 3.16: Schematic of the multiloop any load stable LDO regulator.

R_{F2}. The distinctive feature of this circuit is that, in contrast to what happen with the conventional LDO, the error amplifier is biased with the output voltage V_{OUT} of the regulator itself; this structure allows the circuit to boost its $PSRR$, which is theoretically infinite [45].

In the reported schematic, three different loops can be identified. The main feedback includes the MOSFET M_2 of the error amplifier and the voltage follower M_5, the second one is given by M_5 and M_6, while the last one is provided by the cascade of M_5, M_7 and M_8. This latter is the pass-transistor of the LDO, which ensures the pull-up capability required in this kind of circuits. The gate voltage of M_7 establishes the current through M_8 and M_6 when no current is absorbed by the load.

However, the presence of these three feedback loops inside the circuit makes the design particularly complicated, because it is necessary to guarantee for all of them a reasonable phase margin so that the output voltage V_{OUT} is stable for any load C_L. Nevertheless, as reported in [45], the design of such circuit may be split up into three different steps, at first considering the error amplifier and $M5$ only (without the capacitor C_0) and hence adding the other two loops, in turns. This procedure relies on two facts: the first one is that a 2-stage amplifier can be made stable for any capacitive load by reducing the gain of the first stage with respect to the one of the second stage [46] [47], while the other thing is that the stability of each loop is a sufficient condition (but not necessary) for the stability of a multi-feedback system. This approach has been adopted for the design of all the three loops, as suggested by Ivanov himself, also trying to limit the three loop transfer functions to the second order. In particular, as

far as the main feedback is concerned, the first stage consists of the error amplifier, whose voltage gain is given by:

$$A_1 = g_{m1} R_{p5} \tag{3.9}$$

where g_{m1} is the transconductance of M_1, while R_{p5} is the equivalent resistance seen from the gate of M_5, which is given by the parallel of the output resistances r_{ds} of M_2 and M_4. Hence, A_1 can be decreased either by reducing the transconductance of the differential pair or by making the channel of these transistors shorter. Regarding the loop $M_5 - M_6$, instead, which enhances the gain of the regulator while improving the load regulation as well, the gain of the first stage is:

$$A_2 = g_{m5} R_1 \tag{3.10}$$

Again, g_{m5} is the transconductance of M_5, while R_1 is the resistor on the drain of M_5. Moreover, in order to avoid the introduction of further poles in this loop and avoid further stability issues, this architecture also contains the capacitor C_0, which allows M_5 to work as current-follower at high frequencies. Finally, the last loop has been designed reducing the gain of M_6, by implementing a low value resistor R_2 on its drain.

In addition to these factors, the design of this regulator has been carried out by carefully analyzing not only the stability of the system, but also the noise contribution of each transistor on the output node. Indeed, having to comply with the strict specifications mentioned before, it has been necessary to identify and optimize those MOSFETs with the greatest influence on the noise performances of the circuit. In this regard, it is worth mentioning the devices of the differential amplifier ($M1$ - $M4$), whose size has been increased in order to decrease their noise contributions at low frequencies (flicker noise).

As far as the stability assessment is concerned, it is worth providing further information for multiloop systems. Unfortunately, the standard procedure for the stability analysis with the evaluation of the phase margin by means of AC simulations cannot be applied for this specific situation because more loops are involved. Therefore, it is preferable to exclusively refer to transient simulation and ensure the functionality of the circuit by means other parameters, such as the overshoot factor M, which is given by the following expression:

$$M = \frac{V_{max} - V_{set}}{V_{set}} \tag{3.11}$$

where V_{max} is the maximum voltage that the waveform reaches before settling down to the final value V_{set}. Thus, the stability analysis has been performed by running

transient simulation and by observing the trend of the voltage at the output node and the critical point of each loop, i.e. the gate terminal of M_5, M_6 and M_8 respectively for the three feedback loop above described.

For this specific situation, a 600 µs transient simulation has been carried out, initially turning on the main power supply of 2.5 V and then switching the load current of the CL-LDO regulator so as to stress it out. The evolution of I_L over time is reported in figure 3.17, which allows the verification of the behavior of the circuit even under the harshest of operating conditions.

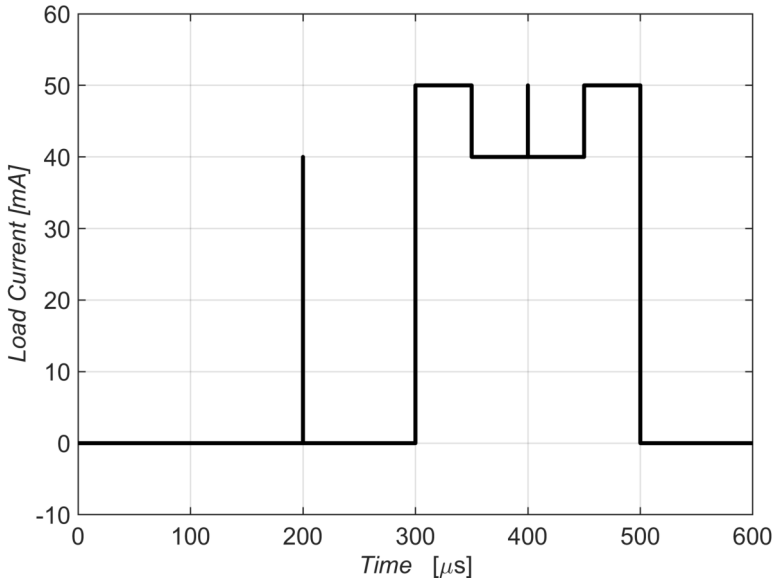

Figure 3.17: Load current I_L over time.

As an example, the plot of the output node V_{out} is shown in figure 3.18, not only at room temperature (27 °C), but also at nEXO and DUNE temperatures (−113 °C and −186 °C respectively), along with the detail of the picture in correspondence of the 40 mA current peak at 200 µs. In this case, it is possible to observe that, when a variation of the load current occurs, the output voltage settles down within few hundreds of nanoseconds to its nominal value with an overshoot factor that never exceed 2 %. Moreover, it has been derived from this simulation that the load regulation for a $\Delta I_L = 50\,\text{mA}$ is $5\,\text{mV}\,\text{A}^{-1}$, compliant with the requirements.

As regards to the noise of the CL-LDO, the power spectral densities at 1 kHz and at 100 kHz have been considered and they have been compared to the ones of the reference LDO. These values have been reported in table 3.5, where it is evident that they are higher than the target PSD at both frequencies. However, the rms noise evaluated between 10 Hz and 100 kHz is lower than the one of the reference device at cryogenic temperature and this condition has been considered sufficient for the application. Nevertheless, as already said, it is mandatory that this regulator is compliant with the

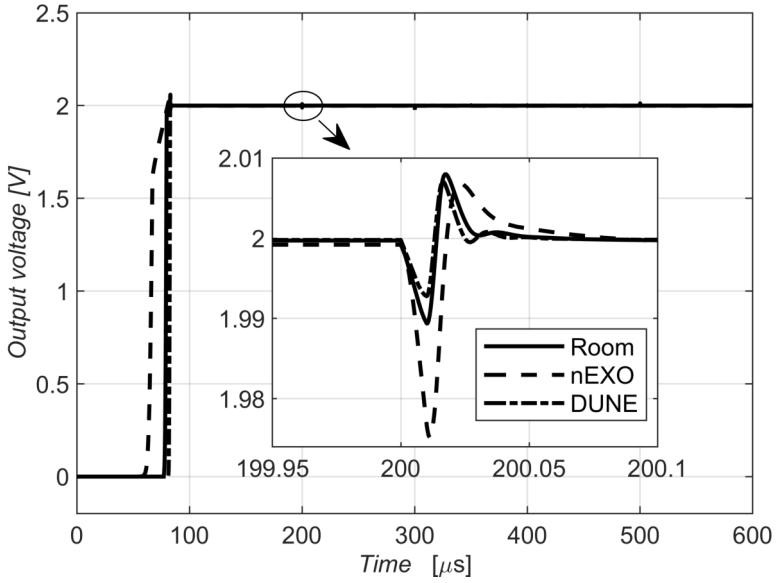

Figure 3.18: CL-LDO output voltage over time.

specification of table 3.4 only at nEXO and DUNE temperature and that it is only capable of operating at room temperature, without any particular constraint about performances.

Table 3.5: CL-LDO noise performances for different temperatures.

Parameter	Room	nEXO	DUNE	target
$\overline{v_n}$ @ 1 kHz	13.6 nV/$\sqrt{\text{Hz}}$	7.42 nV/$\sqrt{\text{Hz}}$	6.57 nV/$\sqrt{\text{Hz}}$	5 nV/$\sqrt{\text{Hz}}$
$\overline{v_n}$ @ 100 kHz	5.25 nV/$\sqrt{\text{Hz}}$	3.69 nV/$\sqrt{\text{Hz}}$	2.45 nV/$\sqrt{\text{Hz}}$	1.4 nV/$\sqrt{\text{Hz}}$
v_n	2.1 µV	1.32 µV	1 µV	1.6 µV

Eventually, the PSRR has also been estimated by performing another AC simulation, with an AC voltage source directly connected on the main power supply of 2.5 V. As it can be seen in figure 3.20, where the PSRR is plotted as a function of the frequency, the regulator is able to attenuate any ripple with a frequency until 1 MHz by more than 60 dB at all temperatures; it is actually around 70 dB if cryogenic temperatures are concerned.

All the previous analyses have also been performed considering process variations (FF, SS, FS and SF) of all the devices involved in the circuit, possible fluctuation of the power supply around its nominal value (±10 %) and different load capacitances. Under all conditions, the aforementioned performance parameters of the CL-LDO meet the requirements and make this circuit a valid candidate for replacing the standard LDO regulator already implemented in the ASICs of the CRYO family and will also be used for the regulation of the power supply in the second version of the ASIC described in

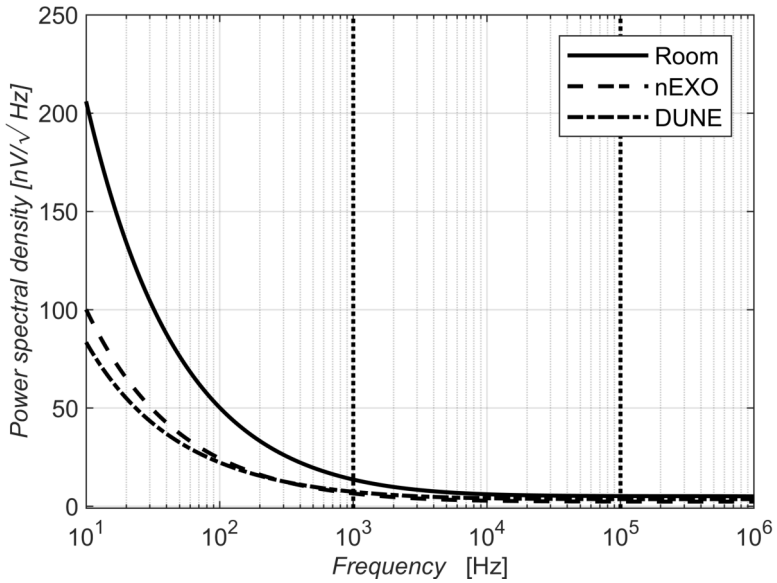

Figure 3.19: Power spectral density of the CL-LDO over frequency.

the next section.

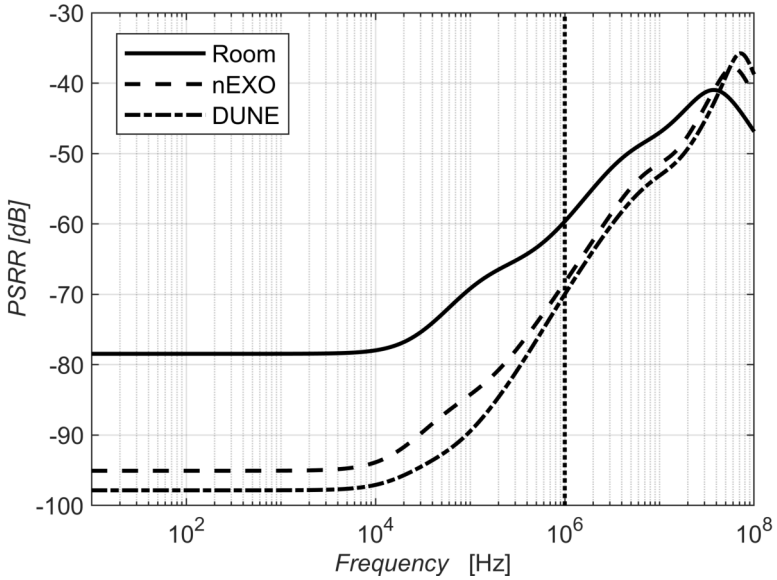

Figure 3.20: PSRR of the CL-LDO over frequency.

Chapter 4

Development of a multichannel ASIC

4.1 Introduction

The last months of this doctoral project have been dedicated to the development of a multichannel *Application Specific Integrated Circuit* (ASIC) to test the SiPM front-end described in the previous chapters. Inside this chip different other circuits have been included for the following purposes:

- Digitization of the analog output of the channels;
- Data parsing and transmission;
- Biasing of all the circuit blocks;
- Generation of the reference voltages;
- Test and debug.

Most of the blocks used in the ASIC for these purposes have been designed by the IC group of SLAC for other projects, but they have been edited and adjusted according to the needs. However, a custom digital state machine has been developed to read the digital signals from the Time to Digital Converters (TDCs) and from the Analog-to-Digital Converters (ADCs), parsing and sending them off using an ad-hoc protocol. The interface for digital signals is the standard *Low-Voltage Differential Signaling* (LVDS), which allows the communication between two points, even far from each other, reducing to a negligible amount the issues related to electromagnetic compatibility.

After introducing the architecture of the ASIC, the design of the digital-state machine and the ASIC layout will be presented. Finally, the full-chip simulations that have been performed will be briefly described.

4.2 Architecture

In this paragraph, the four different sections of the ASIC will be described, starting from the architecture of the so-called "banks", which include the front-end channels and the ADC, and concluding with all the circuitry necessary for testing and debugging the chip.

In particular, it has been chosen to implement 8 SiPM channels, with their respective TDCs, and a supplementary front-end for testing and debugging purposes. In fact, both the analog and digital output signals of this ninth channel, after being buffered respectively by means of an operational amplifier configured for unity gain and a cascade of digital inverters, are connected to the external pads and hence accessible from the outside of the ASIC. Moreover, the *Current_Probe* signal of this channel may be monitored from the outside, as well as the *Request_ADC* digital signal that is set whenever the output of the peak detector is ready to be read out. Additionally, in order to warn the internal digital machine that the peak detector output has been converted (by an external ADC), the *EOC_ADC* signal can be set by the off-chip computing system and bring the channel back to the initial state for following acquisitions.

4.2.1 Banks and biasing circuits

In order to make the development of the multichannel ASIC quicker, it has been chosen to divide the circuit into two part very similar to each other, called *top* and *bottom banks*, based on their location in the chip. Each of them are made up of the following circuits:

- 4 SiPM front-end channels;
- 4 Time-to-Digital Converters (TDCs);
- Bandgap voltage reference;
- Analog-to-Digital Converter (ADC);
- Biasing circuit for the ADC;
- Level shifter and buffers for digital signals.

As already mentioned in the previous chapters, each SiPM front-end channel includes a biasing block, which requires a 200 mV voltage reference to generate all the currents necessary for the proper functioning of the internal circuits. For this reason, the bank has been equipped with a bandgap reference, whose output voltage can be

varied between 50 mV and 200 mV (with a 50 mV step) by means of two configuration bits. Of course, during the operation of the ASIC, they must be both at high logic level, so that to provide the right reference signal to the channels. As far as the ADC is concerned, its reference currents and voltages are generated by a programmable biasing circuit, whose output levels can be conveniently adjusted even for balancing potential mismatches due to process, voltage and temperature variations. On the contrary, the three Delay Locked-Loop circuits, which provide the control voltages the TDCs of both banks, have been placed outside the bank due to area restrictions.

Instead, for what concerns the signals processed by the channels, the peak detector outputs are directly connected to sample-and-hold (SAH) circuit embedded in the ADC system and they are first sampled and then converted accordingly to the internal clocks. Meanwhile, the TDC data, after the conversion, are sent to the external logic through cascaded digital buffers, to prevent the degradation of the signals because of the long path.

4.2.2 Digital back-end

The digital data provided by the two Analog-to-Digital Converters and the eight Time-to-Digital Converters must be transmitted to the external computing system in order to be analyzed and processed. For this purpose, a custom digital state machine has been developed (described in detail in the next section) and an instance of such circuit has been implemented in each bank; it is therefore able to manage the conversion processes of four channels, acquiring the data from the four TDCs and the ADC and producing four 12-bit parallel words. Taking into consideration both banks, each of which requires a digital machine, eight 12-bit words are generated in total.

The serialization of these data is performed by the digital back-end, designed and developed by the SLAC IC group and also employed in the aforementioned CRYO family of ASICs. In this circuit block eight 12-bit inputs are available and the output data stream is delivered through a single serial output. Furthermore, this block also provides two clock signals, called *Clk_Word* and *Clk_Ser*, for the debug and the synchronization of the digital data with the timing signal. In addition to data serialization, the circuit can also be configured to encrypt the data following a protocol owned by SLAC. This can be accomplished by means of an on-board encoder that not only enciphers the data, but also computes two further parity bits for error detection, for a total of 14 bits per word; when the encoding is disabled, the circuit adds two zeros at the end of word.

This circuit has been developed so that the data stream signal can switch state on both rising and falling edges of the external clock, which has a nominal frequency of

448 MHz. In these conditions, the system reaches a throughput of 896 Mbit/s, in such a way that each of the eight inputs are updated every 125 ns, which correspond to a frequency of 8 MHz, i.e. the conversion rate of the ADCs. The clock signal forwarded to the ADCs is generated starting from the main one in the digital back-end, since it includes some clock dividers that are able to scale down the input frequency until the desired value.

The three digital signals (*Serial_Out*, *Clk_Word* and *Clk_Ser*) generated by this circuit, after passing through a level shifter block that translates the data from the 1.2 V domain to the 2.0 V one, are hence sent to the LVDS transmitters that drive the external pads.

4.2.3 ASIC configuration

In this first prototype of the chip, several configuration bits have been set up to change some parameters of the ASIC for different reasons. In the case of the SiPM front-end channel, for example, it has been deemed appropriate to include 54 configuration bits, not only for the setting of the bias currents and the reference voltages (DAC), but also to test and debug all the features implemented in the channel under different working conditions. Moreover, as already said, even the bias circuits, the DLLs and the TDCs have some programmable bits to keep constant their performances over PVT variations.

In order to provide the appropriate configuration to the channels and to the other circuits, the *SLAC ASIC Control Interface*, also called *SACI*, has been implemented in the chip, whose main task is to change the value of its internal registers accordingly to commands and data received from another digital system outside the ASIC. Indeed, it has three input terminals, along with a single output terminal. As far as the inputs are concerned, the SACI requires above all a clock signal to work, whose frequency can be as low as some kHz; then it has a pull-up selection pin that activates the communication whenever its value is low, and last a command signal, through which the external logic communicate with the SACI. On the other hand, the output terminal is used both for debug purposes and for receiving information about the status of the SACI, i.e. the data in the on-board registers.

Internally, instead, this system has numerous I/O signals, among which there are the outputs of 19 banks of registers, each consisting of 16 bits, for a total number of 304 configuration bits. Nevertheless, the first four registers are employed by SACI to handle some internal processes and they cannot be used in the ASIC. Therefore, the total amount of configuration bits available for the setup of the ASIC is 240. In order to allow the SACI to individually configure the front-ends, a set of 54 memories have been implemented in each channel, whose inputs signals are connected to 54 registers

of the SACI and shared among all the channels. The input states are transferred to the outputs (or vice versa) accordingly to the value of three further signals coming from the SACI, named *rw*,*latch* and *addr*. When the writing of the internal register of a specific channel is required, the following three steps are executed by the SACI:

- The corresponding address signal, *addr*, is set to enable the data transfer;
- The read-write signal, *rw*, is set to initialize the write operation;
- The signal *latch* is then raised and, in correspondence of its rising edge, the data are transferred inside the channel.

The memories implemented allow the reading of their internal status by setting the *rw* signal low. Consequently, during the rising-edge of the *latch* signal, the data are shifted from the inside of the channel to the external shared bus attached to SACI and then sent out the ASIC, when required.

In order to finalize the implementation of the SACI, a register map has been defined by associating each configuration bit of the ASIC with a specific memory element of the SACI. Eventually, it turned out that the number of registers, employed for the setup of the chip, is 238 out of 240 available.

4.2.4 Test and debug

The first version of the ASIC also includes specific circuits for testing and debugging all its features and functionality.

First of all, it should be pointed out that the SACI disposes of two on-board digital monitors (*DM1* and *DM2*), each of which made of 16 inputs and one output terminal. By sending the appropriate set of command and data, it is possible to activate the desired DM and choose the digital signal to be monitored externally. In this case, this block has been used to track the monitor signal of the internal logic of the channels, named *SM_Monitor*, which allows us to understand when there is an error in the machine and if, for any reason, it is stuck in one of the states. In this situation, it is therefore possible to make the ASIC return to its initial state by sending a reset to the SACI, which takes care of streaming the same signal to all the other digital systems.

As far as the analog signals are concerned, an auxiliary circuit, developed in SLAC and called *Analog Monitor* (AM), has been implemented in the ASIC in order to supervise some reference signals and adjust their values accordingly to the working conditions. This block comprises an analog multiplexer that selects one out of its 16 inputs on the basis of the value of four selection bits. The signal is then buffered and sent off the chip so that it can be monitored. Aside the several configuration bits,

the analog monitor requires only one further external reference voltage of 1 V, which must be provided externally. The analog signal is then transferred outside the ASIC by means of a voltage buffer and a single-ended to differential circuit, to which the two analog outputs of the AM, *vop* and *von*, are connected. Moreover, this circuit offers the possibility to bypass the last block and read outside the single-ended version of the desired signal on the terminal *vop*; under this condition the other terminal, *von*, is connected to the reference signal at 1 V.

In order to configure the circuit with the proper settings, one of the 16 inputs is attached to an external signal for the calibration process. In addition, it has been deemed appropriate to allocate four of the input terminals to check the reference voltages of the sample-and-hold and of the ADC used in the bottom bank, in such a way that their value can be properly tuned by changing the setup of the bias circuit. Moreover, 8 of the 11 remaining inputs have been attached to the *Current_Probe* signal of the 8 channels. These nodes are also connected to a bunch of resistors, which allow the current to be converted in voltage signals. These resistors have a nominal value of 50 kΩ, whereby in steady-state the voltage drop across them is 500 mV. By monitoring the potentials on these nodes, it is possible to verify the value of the bias current in each channel and, where appropriate, tune it so that it can always be 10 µA during operation. Furthermore, an ancillary circuit has been implemented in this section aimed at obtaining a reliable estimate of the value of the resistors, since they may be affected by temperature and process variations.

4.3 Digital data parser

As already said in the previous sections, when a valid event is detected, each channel generates two different signals: the digital one generated by the leading-edge discriminator and the analog output of the peak detector. The former triggers the TDC, whose output digital data identifies the occurrence time of the detected event, while the analog signal coming from the peak detector, which is proportional to the energy of the event, is directly connected to one of the four sample-and-hold circuits of the ADC system. When the peak detector output signal is ready for the conversion, the internal finite-state digital machine of the channel sets its *ADC_request* signal and waits for the *EOC* signal to be raised by the external logic. In this case, it is not possible to use the *EOC* signal of the ADC, because it is set every time a conversion is completed, regardless of which channel (among the four) is being processed and regardless of the state of the *ADC_request* signals. For these reasons, in order to handle the digital signals of both the ADC and the four channels connected to it, an auxiliary digital

state machine has been designed. Moreover, a data parser and a custom transmission protocol have been implemented in the same circuit, to handle, store and then convey the digital data coming from the TDCs and the ADC to the ASIC back-end. The diagram of the whole circuit is reported in figure 4.1; the four different blocks that can be identified in this figure are described in the following subsections.

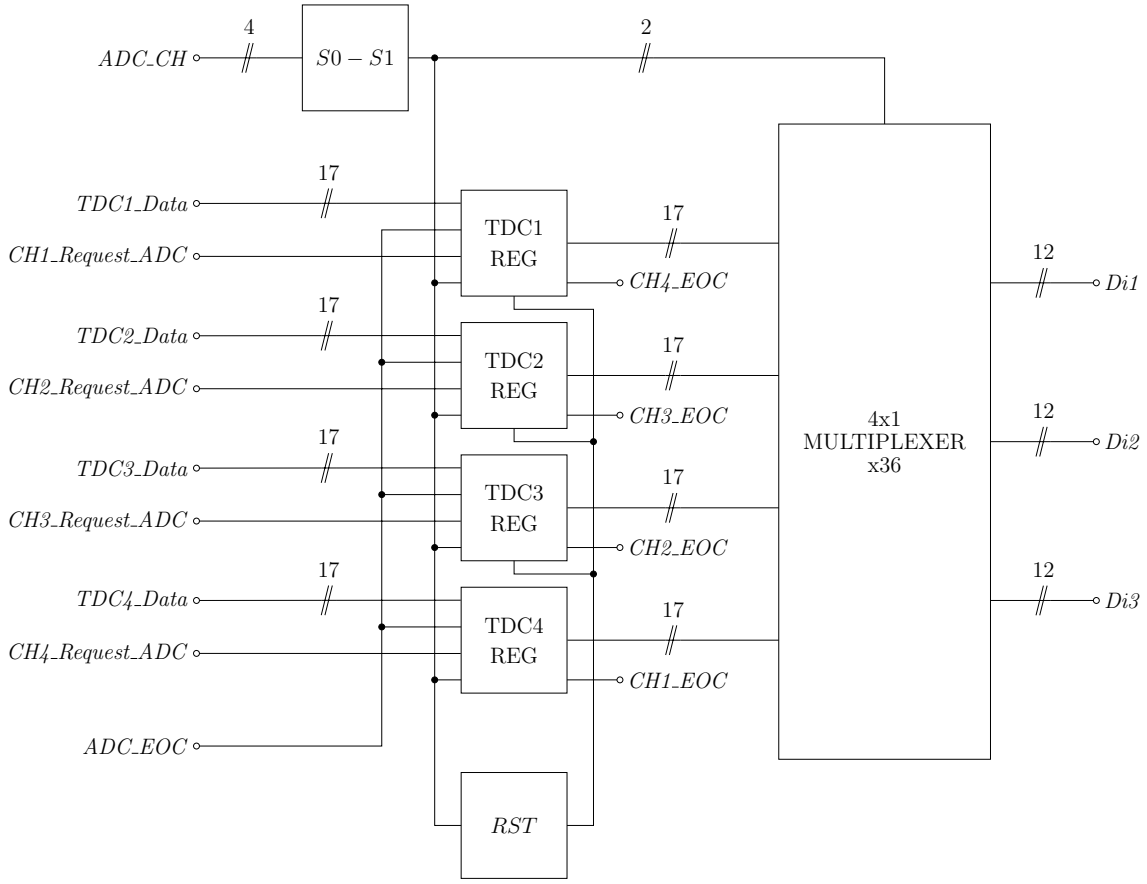

Figure 4.1: Block diagram of the digital data parser.

4.3.1 Custom transmission protocol

Whenever a SiPM generates a valid current signal, the data from the TDC and the ADC need to be sent out of the ASIC. The TDC provides two different kind of data, fine and coarse, each of which consists of 7 bits plus an additional one that is set as soon as an overflow occurs. As far as the ADC is concerned, its output data are made up of 12 bits. Since the ASIC back-end circuit has 8 inputs of 12 bits, it has been decided to employ 4 of them for each bank according to the protocol reported in table 4.1, where:

- *Dix*, with $x = 1 \dots 4$, are the first four inputs of the back-end circuit (for the second bank $x = 5 \dots 8$).

- *BK1* and *BK0* are two constant bits that allow the identification of the bank to which the data frame is associated (i.e. for the first bank it is 0, while it is 1 for the second bank).
- *Vx*, with $x = 1 \dots 4$, are the validation bits that allow the identification of the channel to which the data frame is associated; therefore, for each data frame one of them can be set at most.
- *Cx*, with $x = 0 \dots 7$, are the coarse data of the TDC; *C7* is the overflow bit and *C0* is the Least Significant Bit (LSB).
- *Fx*, with $x = 0 \dots 7$, are the fine data of the TDC; *F7* is the overflow bit and *F0* is the LSB.
- *Ax*, with $x = 0 \dots 11$, are the ADC data where *A0* is the LSB.

Table 4.1: Transmission protocol.

Di	b11	b10	b9	b8	b7	b6	b5	b4	b3	b2	b1	b0
Di1	1	0	1	0	1	0	BK1	BK0	V4	V3	V2	V1
Di2	1	0	1	0	C7	C6	C5	C4	C3	C2	C1	C0
Di3	1	0	1	1	F7	F6	F5	F4	F3	F2	F1	F0
Di4	A11	A10	A9	A8	A7	A6	A5	A4	A3	A2	A1	A0

Even if this protocol has been designed for an ASIC with only 8 front-end channels, it also may be extended for a chip with up to 256 channels, by simply using the unused bits of *Di1* as indexes for other banks.

For each of the three input of the digital back-end, aimed at the serialization and the transmission of the data, 12 4x1 multiplexers have been implemented. The inputs have been connected in accordance with the transmission protocol just described, while the selection signals are $S1$ and $S0$, which are generated by the circuit presented in the following section.

4.3.2 Auxiliary internal signals

Before starting to talk about the way the data are sent out, it is important to briefly describe an auxiliary asynchronous state machine that has been developed to define the timing of the data parser on the basis of the ADC state. Indeed, as already pointed out in the third chapter, the analog input of the ADC depends on the state of four not-overlapping 8 MHz clocks, from which two internal additional signals, $S0$ and $S1$, have been generated. These two signals are useful for three reason: the generation of

the validation bits, the transmission of the right set of data and the production of the reset signals for the internal registers.

After sketching the state diagram of the machine and minimizing it, four different states have been found, one for each channel of the ADC system; it has been deemed appropriate to implement a Gray code for the assignment, according to table 4.2. In this way, two adjacent codes differ in just one bit so that any potential problem of logic hazard is avoided. The output functions for $S0$ and $S1$ are:

$$S0 = CH2 + \overline{CH4} \cdot y1 \qquad\qquad S1 = CH3 + \overline{CH1} \cdot \overline{y0} \tag{4.1}$$

where y are the internal states; the functions of the excitation variables Y are instead:

$$Y0 = CH1 + \overline{CH3} \cdot y0 \qquad\qquad Y1 = CH2 + \overline{CH4} \cdot y1 \tag{4.2}$$

The simplified schematic implemented for the generation of $S0$ and $S1$ is reported in figure 4.2.

Table 4.2: Truth table of $S1$ and $S0$.

CH	$S1$	$S0$
1	0	0
2	0	1
3	1	1
4	1	0

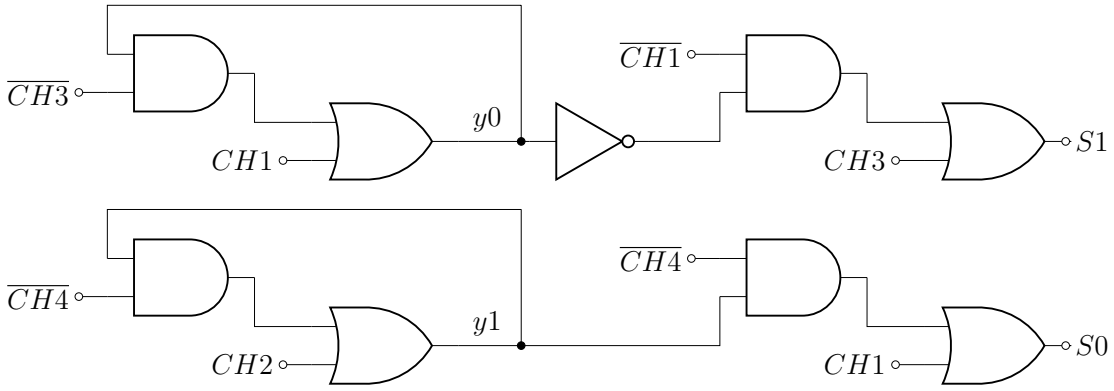

Figure 4.2: Schematic of the circuit for the generation of $S0$ and $S1$.

In figure 4.3 the waveform of the bus CH from the ADC system is reported, along with the one of the bus S.

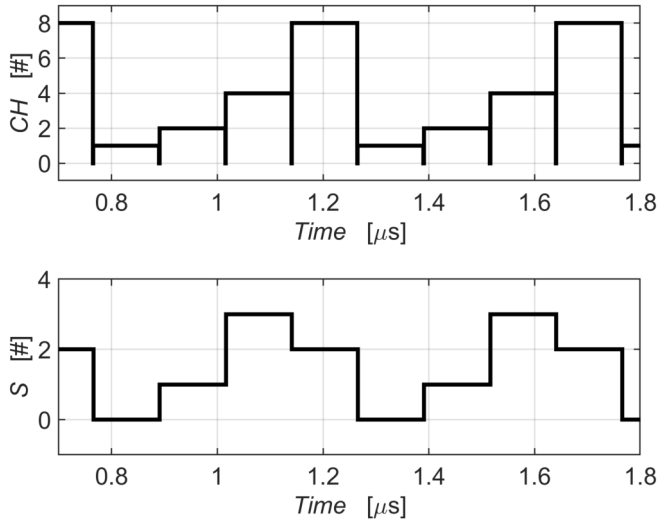

Figure 4.3: Trends of the input (CH) and output (S) buses of the circuit in figure 4.2.

4.3.3 TDC data registers

The data coming from the TDC are provided few tens of nanoseconds after the SiPM is triggered by an impinging radiation, while the generation and the subsequent conversion of the peak detector analog output can take up to 1 µs from the moment of the detection of the event.

Since we must provide the time and energy information of the event in the same data frame, it has been necessary to implement 16 D-type registers for each TDC, so as to maintain the data for as long as required. The storage of these data occurs in correspondence of the falling-edge of the *busy* signal of the TDC, which notifies the end of the conversion. As far as the validation bit Vx of the channel is concerned, it is the result of an AND operation between two further internal signals called *DataReady_TDC* and *DataReady_ADC*, which respectively indicate that the TDC and ADC data are ready to be processed. The former is also set in correspondence of the falling-edge of the *busy* signal of the TDC, while *DataReady_ADC* is set by the circuit shown in figure 4.4.

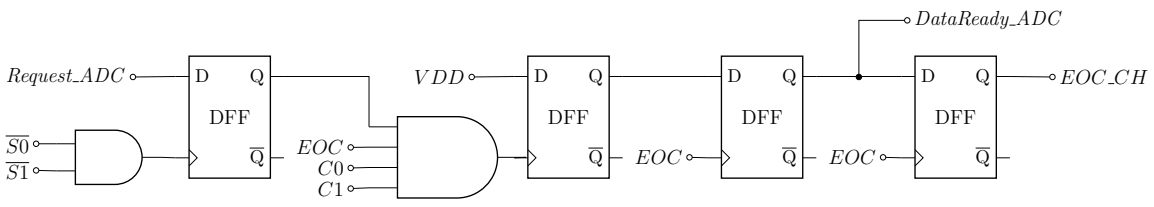

Figure 4.4: Schematic of the circuit for the generation of *DataReady_ADC* and *EOC_CH*.

The *DataReady_ADC* signal is set after three different steps. First of all, the first condition that must occur is that the corresponding *Request_ADC* signal is set by the front-end channel. Therefore, this signal has been connected to the input terminal of a D flip-flop, whose clock signal instead is $\overline{S1} \cdot \overline{S0}$. In this way, the input data of the register is transferred to the output only when both $S1$ and $S0$ are low, which is the instant when all the four SAHs of the ADC system sample the input signals. This ensures that the *DataReady_ADC* is set only when the analog signal is ready for the conversion and it is properly sampled by the ADC system. The second phase for the raising of *DataReady_ADC* regards the identification of the moment when the conversion of the corresponding channel starts. This has been done by means of an AND operation between the output of the first flip-flop, the end-of-conversion signal coming from the ADC, $C1$ and $C0$. These last two variables are directly connected to $S1$, $S0$ or their complementary, based on the channel that they are associated to. Therefore, as regards to the first channel $C1 = \overline{S1}$ and $C0 = \overline{S0}$, for the second channel $C1 = \overline{S1}$ and $C0 = S0$, and so on. The rising edge of the result of this AND operation causes a further flip-flop, whose input terminal is connected to the power supply, to toggle. Finally, in order to detect the end of the conversion, the output of the second flip-flop becomes the input of a third one, placed in cascade, which is sensible to the rising edge of the *EOC* signal. Therefore, as soon as *EOC* is sent by the ADC, *DataReady_ADC* is set. Whenever this signal is high, the machine waits for a further *EOC* pulse to generate the *CH_EOC* that the internal logic of the channel requires to go back to the initial state.

When both *DataReady_ADC* and *DataReady_TDC* are high, the validation bit V is set and the data frame associated to the channel of interest becomes valid.

4.3.4 Internal reset generation

After parsing, validating and transmitting the data frame, all the internal registers in this circuit must be reset to allow the machine to parse the data of the following detected event. An additional asynchronous state machine has been therefore designed and implemented in the circuit to provide an active-low reset signal RST. Nonetheless, this reset signal cannot be sent immediately after the setting of V, otherwise the digital back-end would not have enough time to sample and transmit the data, but it is necessary to wait an entire cycle of the CH, so as to be sure that the transmission has been carried out successfully.

This machine needs only one internal variable y, in order to keep track of the behavior of the bus S, while it requires three different inputs: the validation bit V, $C0$ and $C1$. The latter two are always connected to $S1$, $S0$ or their complementary,

as before, but this time the code corresponds to the one of the following channel. Therefore, as regards to the first channel $C1 = \overline{S1}$ and $C0 = S0$, for the second channel $C1 = S1$ and $C0 = S0$, and so forth. The circuit designed to generate the internal reset is reported in figure 4.4, where the RST output function is given by:

$$RST = \overline{y} + \overline{V} + C0 + C1 \tag{4.3}$$

In the previous equation, y is the state variable, whose excitation signal is:

$$Y = y \cdot (V + C0 \cdot C1) \tag{4.4}$$

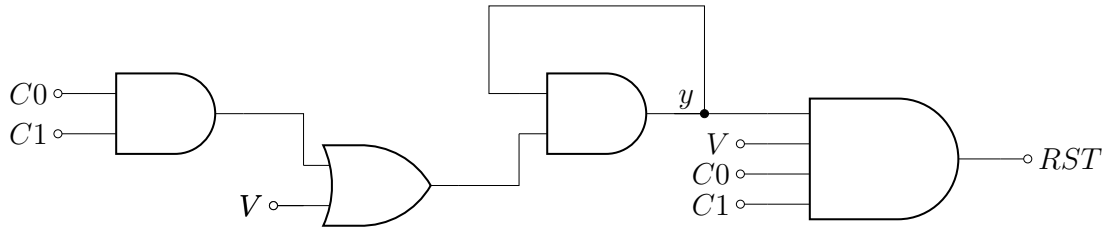

Figure 4.5: Schematic of the circuit for the generation of the RST signal.

4.4 Layout

Concurrently with the generation of the schematic, the placement of the layout of each block has been carried out. The most important constraint that had to be fulfilled concerned the total area of the chip that should not exceed 25 mm^2, due to the restriction of the foundry in *Multi-Project Wafer* runs.

Throughout this process, the topology of the ASICs of the CRYO family has been considered since several circuits implemented in this chip, especially in the back-end part, have been taken from them. Therefore, as in the case of these chips, all the front-end channels have been located at the left end side of the ASIC, the ADCs with their bias circuit in the middle, while the digital back-end on the right side. This made also possible to adequately separate the analog signals from the digital ones, so as to avoid any interference due to their coupling. Moreover, following this approach, it has also been chosen to separate the analog pads of the power supplies from the digital one by placing them along opposite side of the ASIC. Indeed, with this structure, the crosstalk on the analog voltages due to the ground bounces of the digital part may also be limited, even on the printed circuit board on which the ASIC will be mounted.

As far as the digital signals are concerned, it has been observed that the distance between the TDC circuits and the custom digital machine was considerable (more

than 3 mm in some cases) and the direct connections of these blocks with simple metal buses would have been hazardous for the integrity of the information. For this reason, multiple digital buffers have been implemented along the path between each TDC and digital circuit for the regeneration of the signals. Similarly, it was deemed important to do the same thing for the clock signal for the TDCs, but, having to provide the same timing signal to all the 8 block synchronously, it was preferred to design a clock distribution network (or clock tree) in a such a way that the reference signals run through paths with the same length.

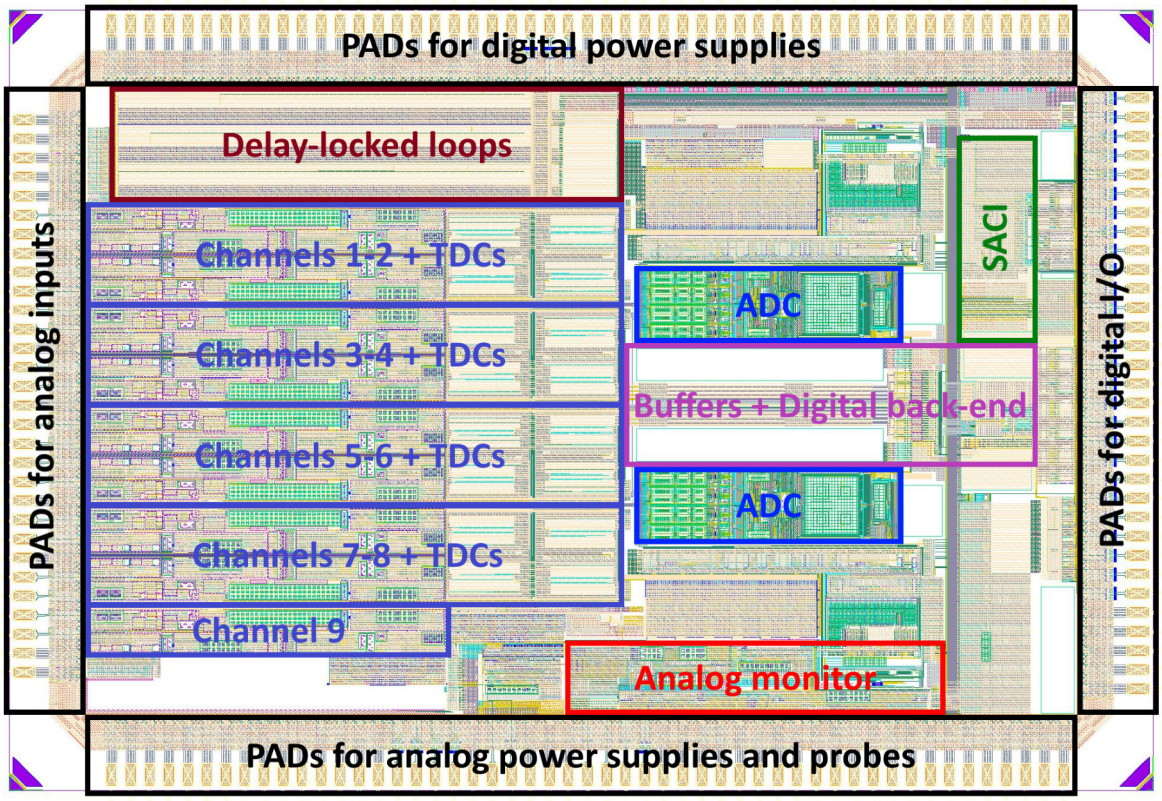

Figure 4.6: Layout of the ASIC (6 mm x 4.1 mm).

In figure 4.6 the full layout of the chip is shown, where the two upper metals (M7 and M8) have been hidden for visualization purposes. It is also possible to notice multiple blank spaces in the image, which may appear empty; they are actually filled with decoupling capacitors (MOSCAP), implemented not only with the purpose of filtering out the noise on the power supplies coming from the external voltage sources, but also to comply with the design rules about metal density, which must not be beneath a certain value. As far as the frame is concerned, these requirements have been fulfilled by placing dummy metals surrounding the pads and the corners.

Moreover, some of the blocks in the picture have been highlighted by means of colored boxes. Starting from the left, where the analog input pads have been placed,

it is possible to notice 4 adjacent boxes, each containing 2 channels and 2 TDCs circuits (cf. figure 2.30), for a total number of 8 analog front-ends. Immediately below the eighth, a further channel has been included, without TDC, for testing and debugging purposes; indeed, all its input and output signals (analog and digital) have been connected to the external pads aimed at monitoring its internal behavior. Instead, in the top-right part of the chip, the Delay-Locked loop circuits (DLLs) have been placed, which generate the three control voltages for the delay cells inside the TDCs. Each DLL provides a terminal for monitoring its output voltage; it has been decided to connect these signals to three external pads, so as to be able to check if the control signals delivered to the TDCs are the expected ones. The analog signals at the output terminals of the first eight channels are connected to the inputs of two ADC systems (in the blue boxes), which sample and convert the information in digital data. These latter are than connected to the digital back-end, together with the digital signals coming from the TDCs. On the right end side of the chip (not marked in the picture), it is possible to localize the LVDS transceivers, which handle the I/O digital signals on the associated pads. The green box puts in evidence the SACI, which manage the chip configuration and the monitoring of specific digital wires from the channels, while the analog monitor is bordered by a red rectangle.

On the completion of the layout of the ASIC, the *Design Rules Check* (DRC) and the *Layout Versus Schematic* (LVS) processes have been executed. The former verifies the design constraints imposed by the foundry for the specific technological node of interest (*TSMC* 130 nm), while the LVS checks whether the layout matches the schematic entry or not. Eventually, the issues of both procedures have been solved and the results files came up clean.

4.5 Full-chip simulations

Once completed the layout, a large set of simulations have been performed to test all the features of the ASIC, starting from the power supplies ramp-up and concluding with the analysis of the data stream. It has been considered reasonable to split the whole process into different sections both for debugging purposes and convenience, because it was necessary to provide different kind of input signals (power supplies, analog and digital) based on the time stamp of the simulation. Therefore, for each section, a transient simulation has been set up by using as initial condition the final state of the previous run. The parts in which the simulation has been divided are the following:

1. From 0 s to 100 µs - Ramp-up of the power supply sources;

2. From 100 µs to 400 µs - Configuration of the ASIC;

3. From 400 µs to 410 µs - Data acquisition.

Actually, before starting the data acquisition process, a further transient simulation between 400 µs and 500 µs has been run to test the calibration section of the ASIC; for this simulation the second state file, with the final conditions of the chip immediately after the configuration of the ASIC, has been used.

4.5.1 Power supplies ramp-up

In order to completely verify the operation of the ASIC, the first transient simulation that has been performed is the ramp-up of all the five power supplies, both analog and digital. As already said, no LDO regulators have been implemented in this prototype and there are no big capacitors that requires time to charge. Consequently, it has been decided to configure the supply sources to have a rise time of just 10 µs. During this time interval the reset signal has been held low and then released once the power supplies reached their nominal values. In figure 4.7 the evolution of three different power supplies is reported, together with the reset signal.

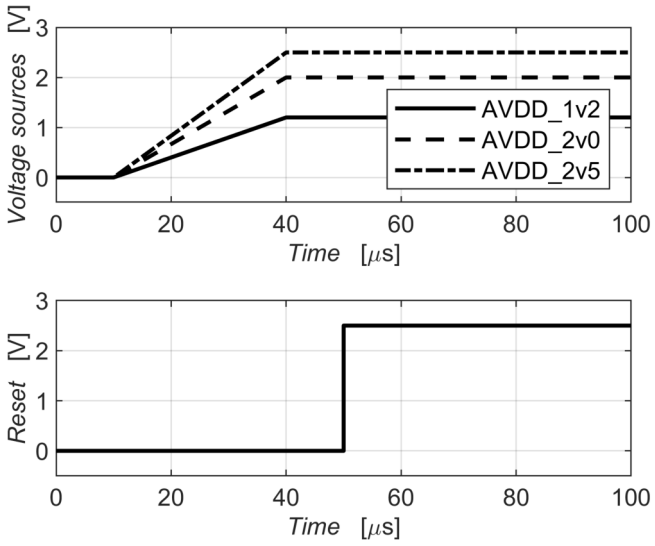

Figure 4.7: Analog power supplies (above) and global reset (below).

In this stage, two different signals have been monitored and plotted in figure 4.8: the output terminal of the bandgap circuits and the potential on the *Current_Probe* of the ninth channel, to which a 50 kΩ resistor has been connected externally to the ASIC. Nonetheless, these signals, after the initial transient, do not reach the expected values because both the bandgap circuit and the channel have a wrong set of configuration bits. These values need to be adjusted by sending the correct setting to the SACI.

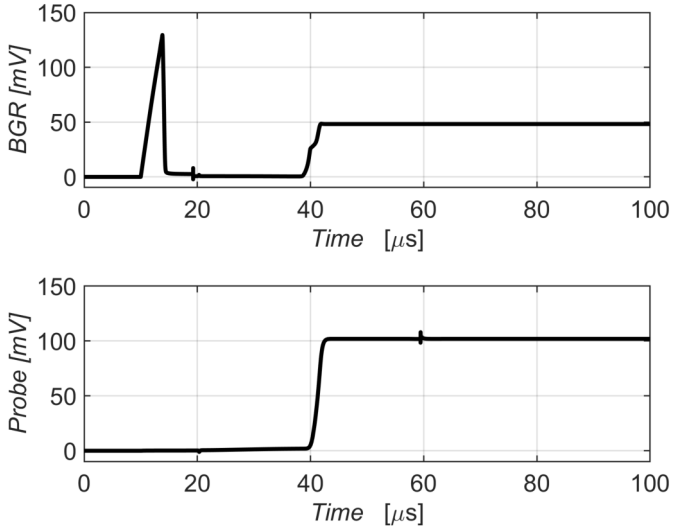

Figure 4.8: Bandgap reference voltage (above) and signal across the external probing resistor (below).

4.5.2 ASIC configuration

Before starting to check the behavior of the 9 front-end channels, the ADCs and the digital back-end, it is necessary to send the appropriate configuration to the SACI, in order to set up the right reference voltages, the bias currents and make the circuit work in the correct operative conditions.

As already mentioned, the SACI provides four terminals (three inputs and an output) and all of them, besides the *Sel* pin, are connected to LVDS transceivers, which drive the pads of the ASIC. Externally, a circuit that mimics the operation of other LVDS transceivers has been implemented for each couple of pads, in order to send the right digital signals and read back the responses of the SACI. The voltage sources connected to the input terminals of such circuits generate three differential signals whose values depends on a digital vector file. This latter has been generated by means of a MATLAB script, through which it is possible to define the clock frequency, the voltage levels (low and high), the delay and the intended command and data to be sent to the SACI. For this first configuration, three different operations have been carried out; first of all the value of all the 16-bit registers have been updated according to the needs, then the *ADDR* signals of all the channel (including the ninth one) have been raised and, last, the write command have been sent to the SACI. Such command affects two signals: *rw* and *latch*. The former is set to notify the internal registers of the channels that a write operation is required, while the *latch* signal is switched immediately afterwards, so that the desired channel configuration is transferred inside each front-end. In figure 4.9 the plots of the signals on the four I/O terminals of SACI are reported.

In this particular case, the *Rsp* signal is used for debugging purposes, considering that this digital system replies with the same pattern (command and data) that receives from outside the ASIC.

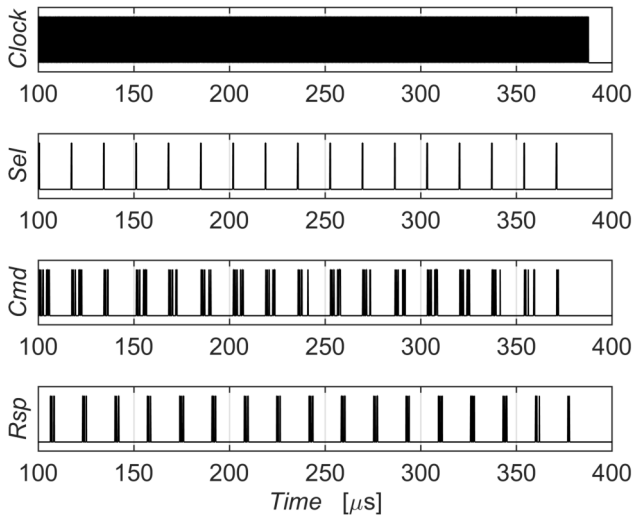

Figure 4.9: SACI signals during the configuration of the ASIC.

In figure 4.10, instead, the same signals described in the plot 4.8 are again reported, but with a different simulation time: the bandgap reference voltage, whose values goes from 50 mV to approximately 200 mV and the voltage across the external resistor, which reaches 500 mV right after the bias circuit of the channel is properly configured. Additionally, in the second subplot of figure 4.10, the signal on the *vop* terminal of the analog monitor is shown. This circuit has been configured so that it is possible to monitor the single-ended version of the *Current_Probe* signal from the first channel, which follows the same trend of the probe right after the configuration.

4.5.3 Calibration

Before checking the correct operation of the ASIC with the current signals coming from the SiPMs, the calibration circuits have also been tested by sending from the outside a sine wave signal through the corresponding pad and by setting the appropriate configuration bits. Indeed, two registers of the SACI have been earmarked for initiating the calibration processes, respectively for the analog monitor and for the bank of resistors employed for the current sensing.

As far as the analog monitor is concerned, the calibration it is useful to determine the adequate configuration to make the circuit work as effectively as possible. In fact, knowing the characteristics of the input sine wave (amplitude and frequency), it is

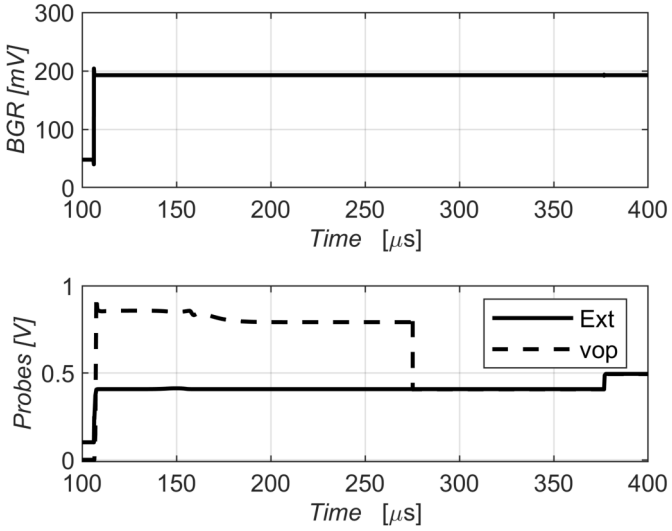

Figure 4.10: Bandgap reference voltage (above) and probe signals (below).

possible to adjust the set-up of the analog monitor to balance potential variation due to temperature, process and power supply. In this particular case, the *vop* output of the analog monitor has been tracked to verify that it follows the input signals. Regarding the bank of resistors, as already said, an auxiliary circuit has been implemented in order to obtain an accurate estimation of their actual values, against the nominal 50 kΩ. Actually, the resistance can be derived by simply fixing the voltage of the external signal and then reading the absorbed current from the outside of the ASIC, since nothing else is connected to the same node.

After sending through the input signal pad a sinusoidal waveform of 40 kHz, with a peak-to-peak amplitude and a DC offset of 1 V (cf. the first subplot of figure 4.11), the calibration circuit of the bank of resistor has been enabled by means of SACI and the current delivered by the external voltage source has been checked. When the external pad is directly connected to the test resistor of 50 kΩ, the current follows the same trend of the voltage, with a peak amplitude around 40 µA. In practice it is likely that this value differs from the nominal one because of PVT variations. Subsequently, during the same transient simulation, the other calibration system has been activated and the voltage on the *vop* pad has been observed. By observing figure 4.11, it is evident that the output of the analog monitor follows the input signal only when the calibration circuit is enabled, otherwise it is connected to ground.

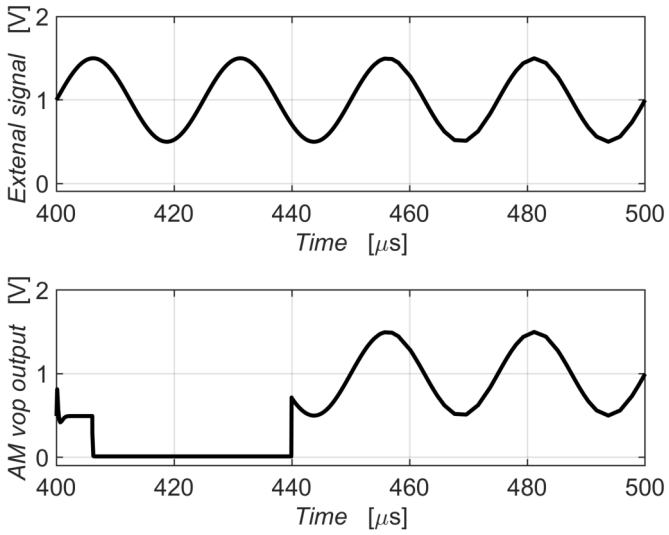

Figure 4.11: External signal (above) and voltage on the *vop* terminal of the analog monitor.

4.5.4 Data acquisition

With the same initial conditions of the previous simulation, it is possible to start the data acquisition process, by activating the timing signals and triggering the nine SiPMs with their corresponding current stimuli. As already mentioned before, the digital back-end requires a 448 MHz clock for its proper operation, but it also needs two further inputs that must be set in order to start the data transmission. The first is called *SampClkin_en*: when it is raised by the logic outside the ASIC, this signal makes the digital back-end start its normal operation. The second signal is *StartRO* (*Start Read Out*), which allows the ADCs to begin the conversion. Moreover, it is necessary to provide a further clock signal for the TDCs, named *TDCs_clocks*, that is connected (by means of a proper clock-tree network) to the eight time-to-digital converters. In this case, it has been deemed sufficient to set up its frequency to 50 MHz, but in practice it may be adjusted according to the requirements. In figure 4.12, the aforementioned digital signals are reported, where *SampClkin_en* and *StartRO* are delayed respectively of 200 ns and 400 ns from the master clock, which starts off at 401 µs.

As regards to the pulse sources that trigger the SiPM, a pseudo-random configuration has been chosen in order to fire the SiPMs with different delays and periods. In table 4.3, the parameters of each current source have been reported, whereas the total number of SiPM cells triggered at the same time is reported under the column Mx.

As an example, the waveforms of the stimuli applied to three SiPM circuits are shown in figure 4.13. In particular, for the first channel (FE1), the pulse simulates the avalanche in just one cell of the SiPM and it is 10 ps wide and 16 mA tall, so that the

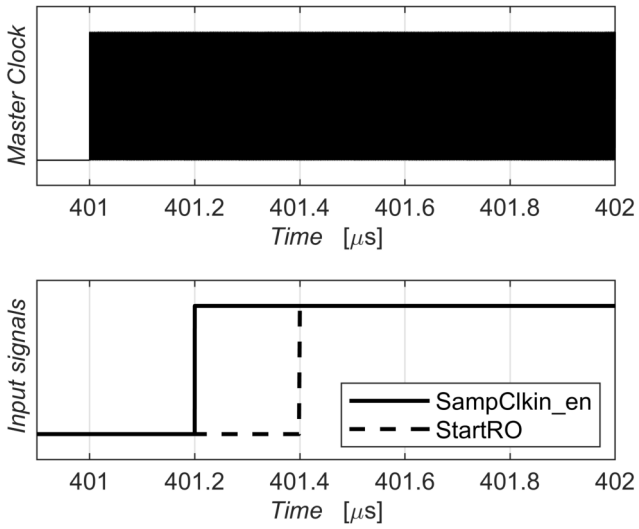

Figure 4.12: Master clock (above) and the synchronization signals of the back-end (bottom).

Table 4.3: Pseudo-random configuration of the pulse sources that trigger the SiPMs.

FE	Mx	Delay	Period
1	1	402.1 µs	800 ns
2	30	402.2 µs	700 ns
3	10	402.3 µs	200 ns
4	15	402.4 µs	1 µs
5	8	402.5 µs	800 ns
6	22	402.6 µs	200 ns
7	17	402.7 µs	350 ns
8	2	402.8 µs	400 ns
9	1	402.9 µs	350 ns

area is exactly equal to Q_0 (cf. table 1.1). The other two waveforms are related to FE5 and FE8, where the current pulses deliver a charge equal to the one generated simultaneously by eight and two micro-cells respectively; therefore, they present the same width of the previous one, but amplitudes eight and two times bigger. Actually, as already said, the system must sustain an event rate up to 100 kHz, whose period (10 µs) is far larger than the ones reported in table 4.3. This has been done in order to assess the robustness of whole system and ensure that it does not get stuck in some undefined state because of the detection of two events way too close to each other. Indeed, the front-end channel has been designed so that it starts the conversion process only when the previous cycle has been completed.

As soon as a valid event is detected, the channel sends a digital pulse to the cor-

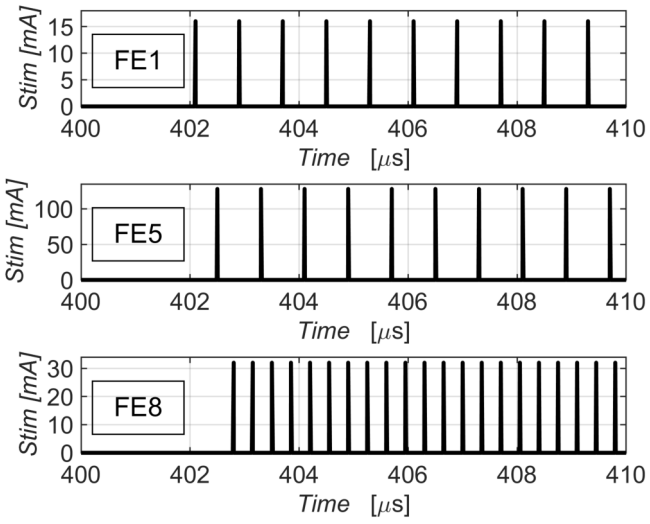

Figure 4.13: Stimuli current signals of FE1, FE5 and FE8.

responding TDC, which converts the time information into digital data, and generates an analog voltage signal which is proportional to the energy of the detected event. For each of the three channels mentioned (1,5 and 8), the $START$ signal of the TDC and the output of the peak detector are reported respectively in figure 4.14 and 4.15. In the first plot, it is possible to notice that not all the input pulses (cf. figure 4.13) are processed internally, but only few of them for the reasons set out above. Instead, in the other picture it is possible to notice the behavior of the output signals as a function of time, which settles down to a voltage proportional to Mx. In fact, the lowest value is related to the FE1 that is triggered with the smallest pulse, while FE5 produce the highest signal among the three. It should also be noticed that these three signals exhibit some ripples, which appears to be periodic, that may affect the proper conversion of the analog voltage. Actually, these phenomena are due to the sample and hold circuits of the ADC system and in particular to the charge injection effect caused by the opening and the closing of the switches implemented in these blocks. Furthermore, it has been verified that this effect does not have impact on the result of the conversion, since it does not occur when the signal is being sampled.

Subsequently, when both time and energy data are ready to be sent, the custom digital machine parses this information, brings them together accordingly to the transmission protocol already presented and sends them, through parallel buses, to the digital back-end. Thus, this latter block handles above all the serialization of these data, their eventual encryption and then the transmission to the LVDS transmitters, connected to the external acquisition system. Outside the ASIC, as said previously, various auxiliary circuits that act as LVDS receivers have been implemented in the

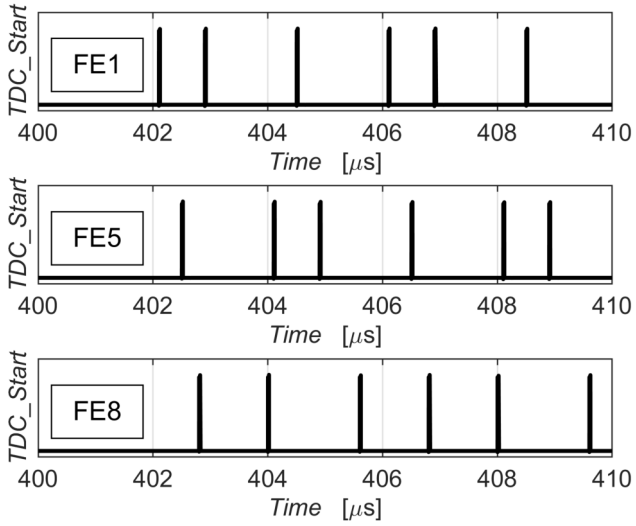

Figure 4.14: *TDC_Start* signals of FE1, FE5 and FE8.

test bench, aimed at converting the differential digital signals into a single-ended ones. In order to check the functionality of these sections, two different scripts have been developed:

- *Verilog-A* script: it is executed throughout the transient simulations and it captures the value (0 or 1) of the single-ended terminal of the LVDS receiver for the *DataStream* signals in correspondence of each rising or falling edge of *MasterClock*. These data are hence saved in a *.csv* file, which is afterwards post-processed by means of a Python script.

- *Python* script: It parses the information contained in the *.csv* file in accordance with the custom transmission protocol employed in this project. This script is not only able to detect potential errors in the transmission, but it also allows the user to filter out the words where none of the validation bits Vx is set.

After performing the second script, the result is a further text file in which all the information about the valid messages is reported. Specifically, four different properties are listed:

- Iteration: the identification number of the current data (even the invalid words are counted).

- Channel: the identification number of the concerned channel. In this ASIC, it ranges from 1 to 8.

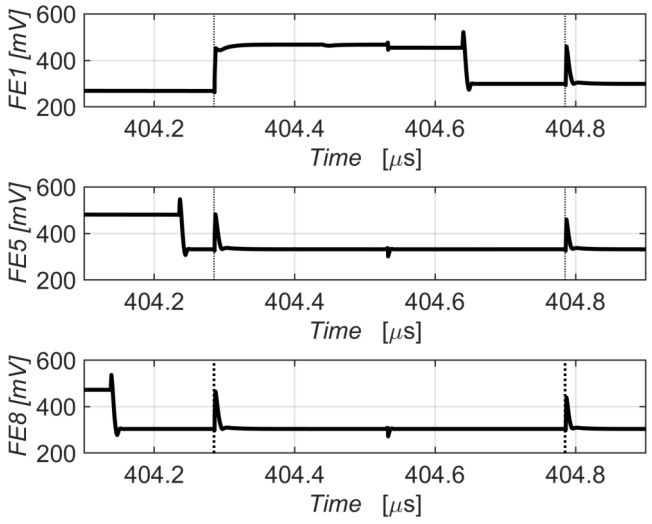

Figure 4.15: Analog output signals of the peak detectors in FE1, FE5 and FE8.

- TDC data: the values of both the output buses (*TDC_F* and *TDC_C*) of the TDC are given. Moreover, the equation (3.4) and (3.5) are applied to the data in order to obtain a fine and a coarse estimation of the time difference between the *START* and the *STOP* signal of the event under consideration.
- ADC data: the result of the conversion of the analog output voltage of the specified channel is given in the decimal notation. Moreover, the corresponding voltage derived using the equation (3.2) is reported.

An example of the information comprised in the file is the following:

```
ITERATION 16 - Channel 1

TDC Data:
        - TDC_C = 118
        - TDC_F = 41
        - Overflow  = 0
        - Overflowf = 0

ADC Data:
        - ADC_C = 247

Event information:
        - dt_C = 18.880 ns
```

- dt_F = 18.79 ns
- Vout = 296.48 mV

Since the input pulses feature the same amplitude, the results of the ADC should always be approximately the same; therefore, it has been verified that the conversion code does not fluctuate a lot around this value, but it actually ranges between 244 and 247, with a maximum variation slightly greater than 1 mV. Checking the value of signal at the sampling instant of the conversion, it settles down to around 296 mV, compliant with the data provided by the ADC.

As far as the TDC conversion is concerned, the time difference between the *START* (*TDC_Start* from the channel) and the *STOP* (*TDC_Clock*) terminals has been monitored. Indeed, in figure 4.16 both signals have been plotted and the time difference between their rising-edge has been evaluated halfway between the low and the high logic level, i.e. 600 mV. The time difference Δt is equal to 18.20 ns, which is not that different from the one derived from the combination of the fine and coarse data of the TDC. It has been verified that the difference between the value of Δt obtained by using the TDC data and the one derived by analyzing the waveforms in figure 4.16 is consistent.

Even if the conversion results are not perfectly accurate, both processes can be improved by tuning the configuration bits of their bias circuits. This procedure may also be useful to compensate any potential mismatches between the channels, not only due to technological process of production but also to their different locations within the ASIC.

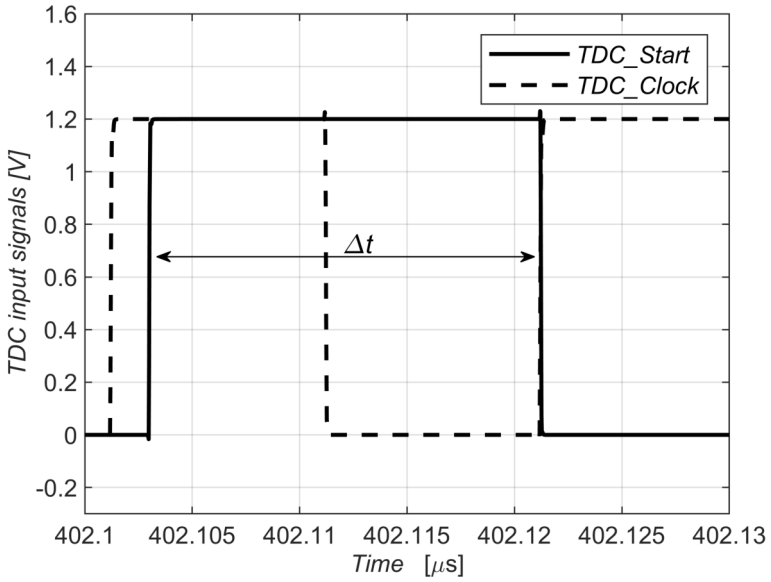

Figure 4.16: Input signals of the TDC in FE1: *TDC_Start* (solid line) coming from the FSM and *TDC_Clock*.

Conclusions and future work

After studying and analyzing different techniques for reading out the signals of silicon photomultipliers, a new current-mode front-end channel for SiPMs has been developed and implemented in a standard 130 nm CMOS technology. The preliminary performances of the readout electronics have been assessed by means of extensive post-layout simulations; as far as time measurements are concerned, a single-photon jitter σ_t of 33.5 ps rms has been estimated. As regards to the measurement of the energy of the detected event, the circuit exhibits a good non-linearity error up to 1280 pC, which correspond to 8000 photons if a SiPM with a 10^6 gain is considered.

Furthermore, a 6 mm x 4 mm multichannel ASIC has been designed by integrating eight of the above-mentioned SiPM readouts, with all the circuit blocks necessary for the conversion, parsing and transmission of the information provided by each channel. An additional analog front-end has also been implemented for testing and debugging purposes, by connecting all its input and output signals to the external pads of the chip. All the features of this system have been verified by performing a large set of full-chip simulations, starting from the ramp-up of the power supplies until the acquisition of the data from the channels, also including the configuration and calibration processes.

This integrated circuit has been submitted to the foundry for production in February 2021 and the fabrication process should be finalized within May 2021. The Printed Circuit Board (PCB) on which the ASIC will be mounted is currently being designed, along with the measurement setup necessary for checking the operation of the whole chip. For this prototype, the photomultipliers soldered on the PCB will be coupled with a synchronous pulsed laser source in order to test all the functionality implemented in the ASIC, by controlling the energy of the event and the instant in which the light signal is detected by the sensor. If the results of measurements will confirm the performance parameter derived from simulations, this front-end channel could be employed in those detection systems with SiPMs where very good temporal resolution are demanded.

Concurrently with the project just described, innovative methods for the design of front-end circuits for silicon photomultipliers are being studied. Specifically, beyond

the analytic approximation of the slope of the output signal already presented in the first chapter, it is worth mentioning the *gm-over-Id* technique that may be applied for this kind of readout circuit, in order to optimize the design of the devices, also taking into account the parasitic components that may affect the acquisition, and further improve the time jitter of the electronic system.

Bibliography

[1] A. B. Arons and M. B. Peppard, "Einstein's proposal of the photon concept-a translation of the Annalen der Physik paper of 1905," *American Journal of Physics*, vol. 33, no. 5, pp. 367–374, 1965.

[2] S. Morel and S. K. Saha, "100 years of photon-counting: the quest for the perfect eye," *21st Century Astrophysics*, pp. 237–257, 2005.

[3] J. Wilson and J. Hawkes, *Optoelectronics An Introduction*. Prentice Hall Europe, 1998.

[4] D. A. Neamen, *Semiconductor Physics and Devices: Basic Principles*. McGraw Hill, 2011.

[5] G. Knoll, *Radiation Detection and Measurement*. John Wiley & Sons, Inc., 1999.

[6] S. Flyckt and C. Marmonier, "Photomultiplier tubes, principles and applications," 2002.

[7] H. W. Ruegg, "An optimized avalanche photodiode," *IEEE Transactions on Electron Devices*, vol. 14, no. 5, pp. 239–251, 1967.

[8] K. G. McKay, "Avalanche breakdown in silicon," *Phys. Rev.*, vol. 94, pp. 877–884, May 1954.

[9] F. Licciulli, I. Indiveri, and C. Marzocca, "A novel technique for the stabilization of SiPM gain against temperature variations," *IEEE Transactions on Nuclear Science*, vol. 60, no. 2, pp. 606–611, 2013.

[10] C. Marzocca, F. Ciciriello, F. Corsi, F. Licciulli, and G. Matarrese, *Front-End Electronics for Silicon Photomultipliers*, ch. 9, pp. 203–235. Taylor & Francis Group, 2016.

[11] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, "Avalanche photodiodes and quenching circuits for single-photon detection," *Appl. Opt.*, vol. 35, pp. 1956–1976, Apr 1996.

[12] F. Corsi, A. Dragone, C. Marzocca, A. D. Guerra, P. Delizia, N. Dinu, C. Piemonte, M. Boscardin, and G. D. Betta, "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*, vol. 572, no. 1, pp. 416 – 418, 2007.

[13] SensL, "An introduction to the silicon photomultiplier," 2011.

[14] V. Savuskan, I. Brouk, M. Javitt, and Y. Nemirovsky, "An estimation of single photon avalanche diode (SPAD) photon detection efficiency (PDE) nonuniformity," *IEEE Sensors Journal*, vol. 13, no. 5, pp. 1637–1640, 2013.

[15] P. Buzhan, B. Dolgoshein, A. Ilyin, V. Kantserov, V. Kaplin, A. Karakash, A. Pleshko, E. Popova, S. Smirnov, Y. Volkov, L. Filatov, S. Klemin, and F. Kayumov, *The Advanced Study Of Silicon Photomultiplier The Advanced Study Of Silicon Photomultiplier*, pp. 717–728. 2013.

[16] M. Ward and A. Vacheret, "Impact of after-pulse, pixel crosstalk and recovery time in multi-pixel photon counter response," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 610, no. 1, pp. 370 – 373, 2009.

[17] R. Pagano, D. Corso, S. Lombardo, G. Valvo, D. N. Sanfilippo, G. Fallica, and S. Libertino, "Dark current in silicon photomultiplier pixels: Data and model," *IEEE Transactions on Electron Devices*, vol. 59, no. 9, pp. 2410–2416, 2012.

[18] D. Renker, "Geiger-mode avalanche photodiodes, history, properties and problems," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 567, no. 1, pp. 48 – 56, 2006.

[19] W. J. Kindt and H. W. Van Zeijl, "Modelling and fabrication of geiger mode avalanche photodiodes," *IEEE Transactions on Nuclear Science*, vol. 45, no. 3, pp. 715–719, 1998.

[20] P. A. P. Calò, F. Ciciriello, S. Petrignani, and C. Marzocca, "SiPM readout electronics," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 926, pp. 57–68, 2019. Silicon Photomultipliers: Technology, Characterisation and Applications.

[21] 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), pp. 1–6, 2014.

[22] G. Matarrese, P. A. P. Calò, F. Corsi, C. Marzocca, and S. Petrignani, "Selecting and designing the front-end amplifier for high-gain photomultiplier detectors with optimal timing performance," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 936, pp. 360–362, 2019. Frontier Detectors for Frontier Physics: 14th Pisa Meeting on Advanced Detectors.

[23] P. A. P. Calò, S. Petrignani, M. D. Gioia, and C. Marzocca, "Analytical study of front-end circuits coupled to silicon photomultipliers for timing performance estimation under the influence of parasitic components," *Sensors*, vol. 20, no. 16, 2020.

[24] P. A. P. Calò, F. Ciciriello, F. Corsi, C. Marzocca, G. Matarrese, and M. Bisogni, "BASIC64: A new mixed-signal front-end ASIC for SiPM detectors," *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*, 2016.

[25] M. Bisogni, P. A. P. Calò, F. Ciciriello, F. Corsi, C. Marzocca, G. Matarrese, and S. Petrignani, "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*, vol. 936, pp. 313–315, 2019.

[26] S. Ziegler, "Positron emission tomography: Principles, technology, and recent developments," *Nuclear Physics A*, vol. 752, pp. 679 – 687, 2005.

[27] J. Sæterstøl, "Characterization of scintillation crystals for positron emission tomography," 2010.

[28] D. C. Dias, F. S. de Melo, L. B. Oliveira, and J. P. Oliveira, "Regulated common-gate transimpedance amplifier for radiation detectors and receivers," *2014 Proceedings of the 21st International Conference Mixed Design of Integrated Circuits and Systems (MIXDES)*, pp. 540–545, 2014.

[29] P. A. P. Calò, S. Petrignani, C. Marzocca, B. Markovic, and A. Dragone, "A CMOS front-end for timing and charge readout of silicon photomultipliers," *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pp. 1–5, 2019.

[30] J. Baker, *CMOS: Circuit Design, Layout, and Simulation.* Wiley-IEEE Press, 2019.

[31] 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.

[32] G. De Geronimo, P. O'Connor, and A. Kandasamy, "Analog CMOS peak detect and hold circuits. part 1. analysis of the classical configuration," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 484, no. 1, pp. 533–543, 2002.

[33] G. De Geronimo, P. O'Connor, and A. Kandasamy, "Analog CMOS peak detect and hold circuits. part 2. the two-phase offset-free and derandomizing configuration," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 484, no. 1, pp. 544–556, 2002.

[34] G. De Geronimo, A. Kandasamy, and P. O'Connor, "Analog peak detector and derandomizer for high rate spectroscopy," *2001 IEEE Nuclear Science Symposium Conference Record (Cat. No.01CH37310)*, vol. 1, pp. 147–150, 2001.

[35] P. Allen and D. Holberg, *CMOS Analog Circuit Design.* Oxford University Press, 2011.

[36] A. Hastings, *The Art of Analog Layout.* Prentice Hall, 2001.

[37] A. Gupta, A. Pena-Perez, B. Markovic, D. Doering, B. Reese, C. Tamma, H. Ali, P. Caragiulo, S. Petrignani, L. Rota, U. Kamath, X. Xu, F. Abu-Nimeh, and A. Dragone, "Read-out architecture of CRYO system-on-chip ASIC for noble liquid tpc detectors," *2020 IEEE 63rd International Midwest Symposium on Circuits and Systems (MWSCAS)*, pp. 611–614, 2020.

[38] A. Gupta, A. Pena-Perez, B. Markovic, C. Tamma, D. Doering, H. Ali, U. Kamath, P. Caragiulo, L. Rota, S. Petrignani, X. Xu, F. Abu-Nimeh, B. Reese, and A. Dragone, "Read-out architecture of CRYO system-on-chip ASIC for noble liquid tpc detectors," *2020 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*, 2020.

[39] A. Pena-Perez, D. Doering, A. Gupta, C. Tamma, B. Markovic, H. Ali, P. Caragiulo, L. Rota, U. Kamath, S. Petrignani, X. Xu, F. Abu-Nimeh, P. Breur, P. Tsang, M. Convery, and A. Dragone, "CRYO: A system-on-chip ASIC for noble liquid

tpc experiments," *2020 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*, 2020.

[40] Y. Zhu, C. Chan, U. Chio, S. Sin, S. U, R. P. Martins, and F. Maloberti, "Split-SAR ADCs: Improved linearity with power and speed optimization," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 22, no. 2, pp. 372–383, 2014.

[41] Y. Zhu, C. Chan, U. Chio, S. Sin, S. U, R. P. Martins, and F. Maloberti, "A 10-bit 100-MS/s reference-free SAR ADC in 90 nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 6, pp. 1111–1121, 2010.

[42] B. Markovic, S. Conforti, C. de La Taille, G. Martin-Chassard, N. Seguin-Moreau, C. Agapopoulou, N. Makovec, L. Serin, P. Caragiulo, A. Dragone, K. Koua, A. G. Schwartzman, D. Su, D. Gong, Q. Sun, J. Ye, W. Zhou, and R. Casanova, "ALTIROC1, a 20 ps time-resolution ASIC prototype for the ATLAS high granularity timing detector (HGTD)," *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*, pp. 1–3, 2018.

[43] J. Torres, M. El-Nozahi, A. Amer, S. Gopalraju, R. Abdullah, K. Entesari, and E. Sanchez-Sinencio, "Low drop-out voltage regulators: Capacitor-less architecture comparison," *IEEE Circuits and Systems Magazine*, vol. 14, no. 2, pp. 6–26, 2014.

[44] V. Ivanov and K. Kunz, "Low drop voltage regulator with instant load regulation and method," U.S. Patent 7 633280, Dec. 2009.

[45] V. Ivanov, *Design Methodology and Circuit Techniques for Any-Load Stable LDOs with Instant Load Regulation and Low Noise*, pp. 339–358. Dordrecht: Springer Netherlands, 2009.

[46] J. Hu, J. H. Huijsing, and K. A. A. Makinwa, "A three-stage amplifier with quenched multipath frequency compensation for all capacitive loads," *2007 IEEE International Symposium on Circuits and Systems*, pp. 225–228, 2007.

[47] R. J. Reay and G. T. A. Kovacs, "An unconditionally stable two-stage CMOS amplifier," *IEEE Journal of Solid-State Circuits*, vol. 30, no. 5, pp. 591–594, 1995.

Relazione sull'attività complessivamente svolta dal dottorando nel triennio di Corso di Dottorato in Ingegneria Elettrica e dell'Informazione (ciclo XXXIII)

Il sottoscritto Ing. Savino Petrignani, nato a Trani il 4/12/1993, ha conseguito la laurea in Ingegneria Elettronica presso il Politecnico di Bari in data 10/10/2017, con la votazione di 110/110 con lode, discutendo la tesi dal titolo "Elettronica di front-end per rivelatori SiPM in applicazioni di imaging medicale", Relatore: Prof. Cristoforo Marzocca.
Durante i tre anni del corso di Dottorato in epigrafe, il sottoscritto ha frequentato i seguenti percorsi formativi organizzati dal Corso di dottorato o offerti da Università:

1. *How to write a technical paper and to present it effectively to an educated audience.* Docente: Prof. Michele Napolitano - 3 CFU
2. *Optimization and control of complex system.* Docente: Prof. Agostino Marcello Mangini - 3 CFU
3. *Application of MATLAB.* Docente: Prof. Mariagrazia Dotoli - 3 CFU
4. *Elements of probability.* Docente: Prof. Pietro Camarda - 3 CFU
5. *Antenna technology for 5G commnications: propagation, arrays and integration.* Docente: Prof. Marco Grande - 3 CFU
6. *Preparing brief scientific presentation in English.* Docente: Prof. Carmela Mary White - UNIBA, 2 CFU
7. *Introduction to Python.* Docente: Prof. Domenico Diacono - UNIBA, 2 CFU

Il sottoscritto ha altresì partecipato ai seguenti seminari e convegni scientifici su tematiche afferenti al dottorato in oggetto:

1. *Fundamentals of Network Analysis.* Relatore: Ing. Alessandro Tritta - 21/11/2017
2. *Soluzioni e tecnologie per la prototipazione di PCB.* Relatore: Ing. Angel Martinez - 9/4/2018
3. *Machine learning and condition-based monitoring.* Relatore: Ing. Anselm Haselhoff - 17/7/2018
4. *Neural networks and state estimation.* Relatore: Ing. Anselm Haselhoff - 19/7/2018
5. *Systematic design of analog CMOS circuits using gm-over-Id-based lookup tables.* Relatore: Prof. Claudio Talarico

Nel corso dei tre anni, ha svolto la seguente attività all'estero:

1. Internship da aprile a ottobre 2019 presso SLAC National Accelerator Laboratory, Menlo Park, California, Stati Uniti D'America. Supervisore: Dott. Ing. Angelo Dragone

Il programma di ricerca individuale assegnato dal Collegio dei Docenti svolto nell'ambito del SSD ING-INF/01 ha per titolo: "Sviluppo di modelli circuitali e di elettronica integrata innovativa per rivelatori di tipo Silicon Photomultiplier" con il relatore Prof. Ing. Cristoforo Marzocca.
Descrizione sintetica dell'attività di ricerca: Nell'ambito di una collaborazione con SLAC National Accelerator Laboratory è stato sviluppato un nuovo front-end, in tecnologia TSMC 130nm, per fotomoltiplicatori al silicio con una risoluzione temporale allo stato dell'arte, anche in corrispondenza dell'assorbimento di un singolo fotone. Successivamente, è stato progettato un ASIC multicanale per il test del front-end analogico, in cui sono stati implementati non solo 8 canali analogici, ma anche tutti i sistemi elettronici necessari alla conversione e trasmissione dei dati. Inoltre, nell'ambito dell'intership condotta presso i laboratori di SLAC, è stato progettato un *Capacitor-less LDO regulator* a basso rumore, ottimizzato per funzionare a temperature criogeniche, ed è stato ottimizzato il SAR-ADC da 12 bit utilizzato nell'ASIC per SiPM, portando il suo ENOB da 9.8 a 11.6.

Titolo della tesi di dottorato: High timing resolution front-end electronics for Silicon Photomultiplier detectors

Nell'ambito dell'attività di ricerca sono state prodotte le seguenti pubblicazioni scientifiche:

1. M.G. Bisogni, P.A.P. Calò, F. Ciciriello, F. Corsi, C. Marzocca, G. Matarrese, S. Petrignani, *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, 2019, Volume 936.
2. G. Matarrese, P.A.P. Calò, F. Corsi, C. Marzocca, S. Petrignani, *Selecting and Designing the Front-end Amplifier for High-gain Photomultiplier Detectors with Optimal Timing Performance*, Nuclear Instruments and Methods in Physics Research Section A, 2019, Volume 936.
3. P.A.P. Calò, F. Ciciriello, S. Petrignani, C. Marzocca, *SiPM readout electronics*, Nuclear Instruments and Methods in Physics Research Section A, 2019, volume 926.
4. P.A.P. Calò, S. Petrignani, M. Di Gioia, C. Marzocca, *Analytical study of front-end circuits coupled to silicon photomultipliers for timing performance estimation under the influence of parasitic components*, Sensors, volume 20, art. no. 4428, 2020.
5. V. Basile, M. Grande, V. Marrocco, D. Laneve, S. Petrignani, F. Prudenzano, I. Fassi, *Design and Manufacturing of Super-Shaped Dielectric Resonator Antennas for 5G Applications Using Stereolithography*, IEEE Access, Volume 8, 2020.
6. P.A.P. Calò, S. Petrignani, C. Marzocca, B. Markovic, A. Dracone, *A CMOS Front-End for Timing and Charge Readout of Silicon Photomultipliers*, 2019 IEEE NSS-MIC, Manchester (Regno Unito), 26 ottobre – 2 novembre 2019.
7. S. Petrignani, A. D'Orazio, M. Grande, V, Marrocco, V. Basile, I. Fassi, *Supershaped dielectric resonator antenna for 5G applications*, Antenna and Propagation Conference 2019 (APC 2019), Birmingham (Regno Unito), 11 – 12 novembre 2019.
8. A. Gupta, A. Pena-Perez, B. Markovic, D. Doering B. Reese, C. Tamma, H. Ali, P. Caragiulo, S. Petrignani, L. Rota, U. Kamath, X. Xu, F. Abu-Nimeh, A. Dragone, *Read-out Architecture of CRYO System-On-Chip ASIC for Noble Liquid TPC Detectors*, 2020 IEEE 63rd International Midwest Symposium on Circuits and Systems, Springfield (MA – Stati Uniti d'America), 9 – 12 agosto 2020.

9. A. Gupta, A. Pena-Perez, B. Markovic, D. Doering B. Reese, C. Tamma, H. Ali, P. Caragiulo, S. Petrignani, L. Rota, U. Kamath, X. Xu, F. Abu-Nimeh, A. Dragone, *Read-out Architecture of CRYO System-On-Chip ASIC for Noble Liquid TPC Detectors*, 2020 Virtual IEEE NSS-MIC, Boston (MA - Stati Uniti d'America), 31 ottobre – 7 novembre 2020.

10. A. Pena-Perez, D. Doering, A. Gupta, C. Tamma, B. Markovic, H. Ali, P. Caragiulo, L. Rota, U. Kamath, S. Petrignani, X. Xu, F. Abu-Nimeh, P.A. (Sander) Breur, P. Tsang, M. Convery, A. Dragone, *CRYO: A System-On-Chip ASIC for Noble Liquid TPC Experimentss*, 2020 Virtual IEEE NSS-MIC, Boston (MA - Stati Uniti d'America), 31 ottobre – 7 novembre 2020.

Il dottorando Savino Petrignani

Il tutor Prof. Cristoforo Marzocca