

Master thesis

Design of an analogue front-end for Silicon Photo Multipliers

Corentin Delacour

02/2020 - 06/2020

Supervised by Matthieu Heller and Adil Koukab

A thesis submitted for the degree of Master of Science in
Electrical and Electronic Engineering at EPFL

EPFL



**UNIVERSITÉ
DE GENÈVE**



Acknowledgments

I would like to express my sincere gratitude to Prof. Teresa Montaruli for this opportunity to join the fascinating CTA project within the Geneva DPNC research team during my 4-months thesis. Working with the DPNC for such an outstanding application was indeed a very exciting experience and I hope there will be more collaborations between EPFL and UniGe in this domain.

I also would like to thank my thesis supervisor Dr. Matthieu Heller for his constant support and feedback. Matthieu has taught me many new concepts related to Imaging Atmospheric Cherenkov Telescopes from an enlightening system point of view and through deep hardware considerations. I would not have been able to optimise the design without his expertise in the domain. Also, thank you Matthieu for my good integration in the team and for the opportunities to present my work despite the particular Covid-19 situation.

Last but not the least, I would like to thank my thesis supervisor Dr. Adil Koukab for his great teaching in the Analog Design field during the last two years. His expertise and clear explanations helped me to better understand the Analog world and guided me in that direction. Thank you Adil for the confidence you have shown during my master's projects and especially during my thesis, for which your help was precious at each design step.

Table of contents

1. Introduction	7
1.1. The Cherenkov Telescope Array project	7
1.2. PMTs vs SiPMs	9
1.3. Frame and goal of the thesis	13
1.4. State-of-the-art analogue front-ends and specifications	14
2. SiPM electrical model	16
2.1. Corsi model	16
2.2. Equivalent input source	19
2.3. Equivalent input source loaded by Z_L	20
2.4. Ways of decreasing the pulse falling time	21
3. Study of different input-stages dedicated to a SiPM	22
3.1. Regulated-drain current mirror	22
3.1.1. Implementation of the regulated-drain current mirror with an OTA	23
3.1.2. Regulated-drain current mirror with a common-gate amplifier . .	23
3.1.3. Regulated-drain current mirror with a compact common-gate amplifier	26
3.2. Boosted Common-gate architecture	27
3.3. Comparison of all the studied input-stages	27
4. Shape optimisation of the input current pulse	27
4.1. Desired case: compensation of the zero $R_q C_q$	29
4.2. Undesired case: bad compensation of the zero $R_q C_q$	30
4.3. Illustration with Matlab	30
5. Final architecture modelling	32
5.1. General overview of the final architecture	32
5.2. Intrinsic parameters of the LCT5 sensor	33
5.2.1. Corsi parameters	33
5.2.2. Intrinsic pole, zero and output capacitance of the sensor	34
5.2.3. Computation of the Night Sky Background DC current	34
5.3. Description of the input-stage	36
5.3.1. Trans-impedance gain	37
5.3.2. Voltage gain	37
5.3.3. Input Impedance	38
5.3.4. Pulse shape optimisation	38
5.3.5. Implementation of 4 different configurations	39
5.3.6. Signal to Noise Ratio	40
5.3.7. Biasing and sizing of the devices	41
5.4. Description of the output stage - Zero/Pole/Cancellation	41
5.4.1. Transfer function	42

5.4.2.	Noise analysis	43
5.4.3.	Biasing and sizing of the devices	43
5.5.	Biasing circuits	43
5.5.1.	Current reference	44
5.5.2.	Voltage reference for the differential amplifier	45
5.5.3.	Temperature dependence of the biasing circuits	46
6.	Physical implementation of the front-end and post-layout simulations	48
6.1.	Layouts of the circuits	48
6.1.1.	Layout of the input stage	49
6.1.2.	Layout of the differential amplifier	50
6.1.3.	Layout of the biasing circuits	51
6.2.	Post-layout results	52
6.2.1.	Front-end signals	52
6.2.2.	Gain, FWHM and SNR	54
6.2.3.	Linearity and Dynamic range	57
6.2.4.	Gain and FWHM variations with temperature	59
6.2.5.	Gain and FWHM variations with I_{NSB}	59
7.	Limitations and possible improvements	61
7.1.	Linearity	61
7.2.	New biasing circuits to decrease the FWHM and Charge variations	61
7.2.1.	Resistor-less Beta-Multiplier - current reference	62
7.2.2.	Resistor-less Beta-Multiplier - voltage reference V_{REF}	63
7.2.3.	Temperature dependence of the two new biasing circuits	63
7.2.4.	FWHM, Gain and SNR variations with the Resistor-less biasing circuits	65
8.	Conclusions	65
8.1.	Pros and Cons of the proposed front-end	65
8.2.	Comparison with the state-of-the-art front-ends	66
A.	Derivation of the SiPM equivalent circuit	67
B.	Derivations - noise analysis	68
B.1.	Input-stage	68
B.1.1.	Transimpedance Z_m when the sensor is connected	69
B.1.2.	Voltage output-referred power spectral density of noise	70
B.1.3.	Current input-referred power spectral density of noise	70
B.2.	Differential amplifier - ZPC	70
B.2.1.	Voltage gain G	71
B.2.2.	Output-referred power spectral density of noise	71
B.2.3.	Input-referred power spectral density of noise	72

List of Figures

1.	Crab Nebula in multiple wavelengths	7
2.	Detection principle of the Cherenkov Telescope Array. Credit: [1]	8
3.	Layout of the CTA in both hemisphere. Credit: [1]	9
4.	Basic operation of a PMT. Credit: [4]	10
5.	2D-cut of an Avalanche Photo Diode. Credit: [5]	11
6.	Photography of a SiPM with its corresponding electrical diagram. Credit: https://www.ketek.net/sipm/technology/working-principle/	11
7.	Basic 2-steps operation of a SiPM. Adapted from [5]	12
8.	SST-1M camera overview: the analogue signals are digitised by DigiCam [6]	14
9.	GAPDs matrix forming a SiPM from [10]	17
10.	SiPM Corsi electrical model from [10]	18
11.	Final Corsi model used to derive an equivalent small-signal input source. Adapted from [10]	19
12.	equivalent small-signal input source	20
13.	equivalent small-signal input source loaded by Z_L	20
14.	Regulated-drain current mirror schematic	22
15.	Regulated-drain current mirror with an OTA	23
16.	Regulated-drain current mirror implementation using a Common-gate amplifier	24
17.	Transient simulation of the regulated-drain current mirror followed by a differential ZPC stage - 1 p.e	25
18.	Regulated-drain current mirror with a passive load common-gate amplifier	26
19.	Boosted common-gate amplifier schematic	27
20.	Typical output pulse of a SiPM simulated with Virtuoso, showing the two decaying time constants	28
21.	Representation of a good zero compensation in the frequency domain	29
22.	Representation of a bad zero compensation in the frequency domain	30
23.	Zero compensation in frequency domain - Matlab simulation showing $ I_L $	31
24.	Zero compensation in time domain - Matlab simulation	31
25.	General front-end overview	32
26.	Signals and Transfer functions of the front-end in the frequency domain	33
27.	NSB signal with its DC component	35
28.	Chosen input stage: Regulated-drain current-mirror with a boosted-common- gate amplifier	36
29.	Transistor-level schematic of the input-stage	37
30.	Post-layout Bode-plot of the first stage. The dashed lines are the phases.	39
31.	4 different input-stage configurations	40
32.	Schematic of the output stage implementing a Zero	42
33.	Schematic of the Beta-multiplier - current reference	44
34.	Schematic of the Beta-multiplier - voltage reference	45
35.	Front-end overview	46

36.	Front-end temperature dependence - post-layout simulation	47
37.	Layout of the input stage	49
38.	Layout of the differential amplifier	50
39.	Layout of the current reference	51
40.	Layout of the voltage reference	52
41.	Output voltage for 1 photoelectron and for the 4 versions	53
42.	Input current (top), output voltage of the first stage (middle) and front-end output voltage (bottom) - Low gain case (V4)	53
43.	Representation of MC process variations (left) and MC process & mismatch variations (right). Credit: https://www.eetimes.com/overcoming-32-28-nm-ic-implementation-challenges/	54
44.	Resistance corner values given by the foundry. They are defined as a $\pm 3\sigma$ deviation from the nominal value	55
45.	Simulation results of MC process & mismatch variations for the three resistance corner cases and for the 4 versions	55
46.	Simulation results of MC process & mismatch variations for the 4 versions - Standard deviation estimated thanks to the 3 resistance corner cases . .	56
47.	Front-end linearity for the 4 configurations - The error bars correspond to the maximum error	57
48.	Output pulse integral for the 4 versions - The error bars correspond to one standard deviation	58
49.	Gain and FWHM variations with temperature - Low gain case (V4) . . .	59
50.	Gain and FWHM variations with I_{NSB} - Low gain case (V4)	60
51.	Front-end gain in frequency domain for $0 \leq I_{NSB} \leq 1 \text{ mA}$	60
52.	Front-end linearity with ZPC (left) and without ZPC (right) - V4	61
53.	Upgrade of a classic Beta-Multiplier (left) to a Resistor-less Beta-multiplier (right) [13] - Startup circuits are not shown	62
54.	Resistor-less Beta-multiplier used as a voltage reference - Startup circuit is not shown	63
55.	DC voltages shift with temperature. Classic and Resistor-less biasing are compared	64
56.	Gain & SNR vs FWHM - classic biasing (left) and resistor-less biasing (right). Error bar = one standard deviation	65
57.	1st step in the computation of the SiPM equivalent circuit. Left-sided schematic adapted from [10]	67
58.	Last step in the computation of the SiPM equivalent circuit	68
59.	Noiseless input-stage with its equivalent input noise source	69
60.	Input-stage schematic with noise sources	69
61.	Noisy output stage: equivalent schematic	70
62.	Output stage schematic with the noise sources	71

1. Introduction

1.1. The Cherenkov Telescope Array project

The Cherenkov Telescope Array project [1] is an international initiative to develop the greatest ever-made ground-based observatory for gamma-ray astronomy. It will be composed of more than 100 telescopes distributed on the two hemispheres (Chile and La Palma, Spain), achieving a full-sky coverage. The observatory will detect photons with energy from 20 GeV up to 300 TeV, pushing the limits of the known electromagnetic spectrum and thus giving new views of the universe. The project, gathering more than 200 institutes in 31 countries, has a large number of ambitious scientific goals that can not be listed exhaustively. However, we can cite three key science projects in which CTA will address outstanding challenges:

- Better understand the origin and role of relativistic cosmic particles
- Probing extreme environments such as neutron stars and black holes
- Exploring frontiers in physics such as the nature of dark matter

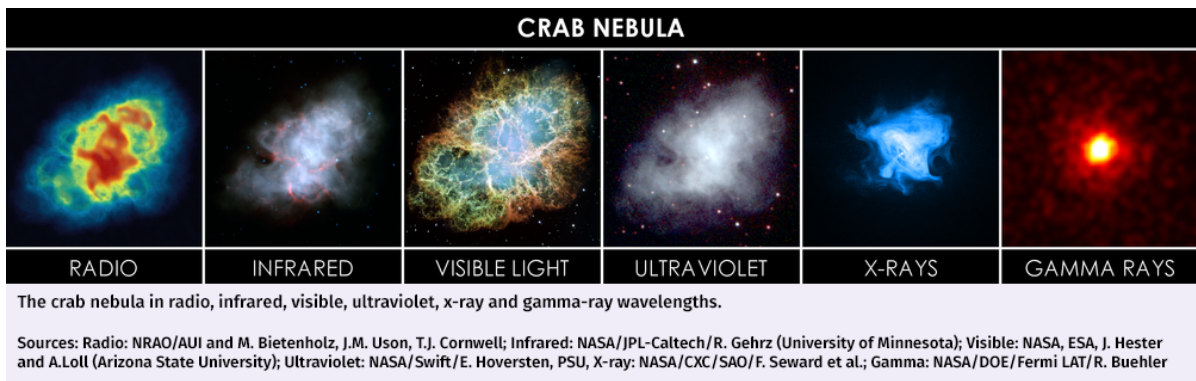


Figure 1: Crab Nebula in multiple wavelengths

The gamma-rays are not directly detectable from Earth because they can not go through the atmosphere. Gamma-rays photons rather interact with the air to create cascades of subatomic particles called "particle showers". In the atmosphere medium, those energetic particles can travel at a higher speed than light, creating blue/UV light cones by Cherenkov effect.

The telescopes are composed of large mirrors which collect and concentrate light on a high-speed camera. The faint Cherenkov flash typically lasts 10 ns; therefore, this kind of camera processes electronic short pulses triggered by the Cherenkov punctual photons. These electronic pulses are generated by single photon sensitive sensors. Two types of photo sensors have been used in gamma-ray astronomy: Photo Multiplier Tubes (PMTs)

or Silicon Photo Multiplier devices (SiPMs). The basic operation of these two sensors is described in the section 1.2. To get a reliable record of a typical 10 ns Cherenkov event (shape, dynamic and amplitude), the camera needs to have:

- a large amount of pixels (1855 PMTs for the actual LST)
- a fast response
- a sufficient Signal over Noise Ratio

Indeed, the signal to detect is already polluted by many external light sources (stars, moon, cities lights) so it is challenging to extract a single photon information.

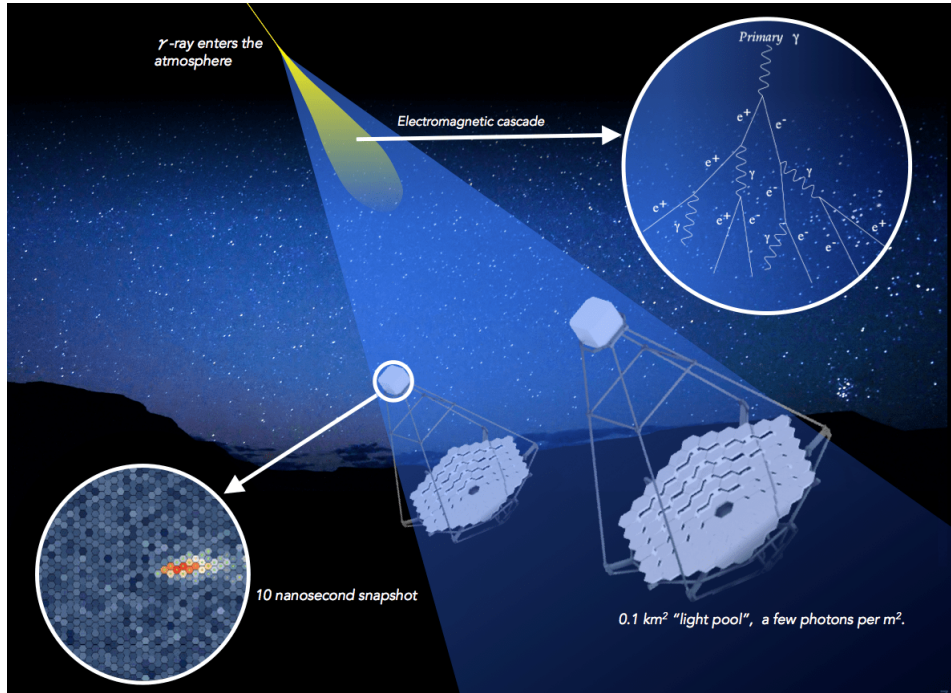


Figure 2: Detection principle of the Cherenkov Telescope Array. Credit: [1]

In the southern hemisphere, the array will have 3 types of telescopes, each one dedicated to a gamma-ray energy range and laid out specifically. For example, because the low energy gamma-rays produce a smaller amount of Cherenkov light, the Large-Sized Telescope has a large collection area and 4 of them will be closely placed (relative distance < 250 m) in each hemisphere. The LSTs will then lower the detection energy threshold of the whole array.

	SST	MST	LST
Number of telescopes	70	40	8
Energy range [TeV]	10 -300	0.1-10	≤ 0.1
Telescope diameter [m]	4	12	23

Table 1: Cherenkov Telescope Array overview

Moreover, as the flux of very high-energy photons is low, a big array area is required to achieve a good sensitivity at high energies. For example, in the southern hemisphere site, the array will be laid-out circularly for a total area of 4.5 km^2 ^[2].

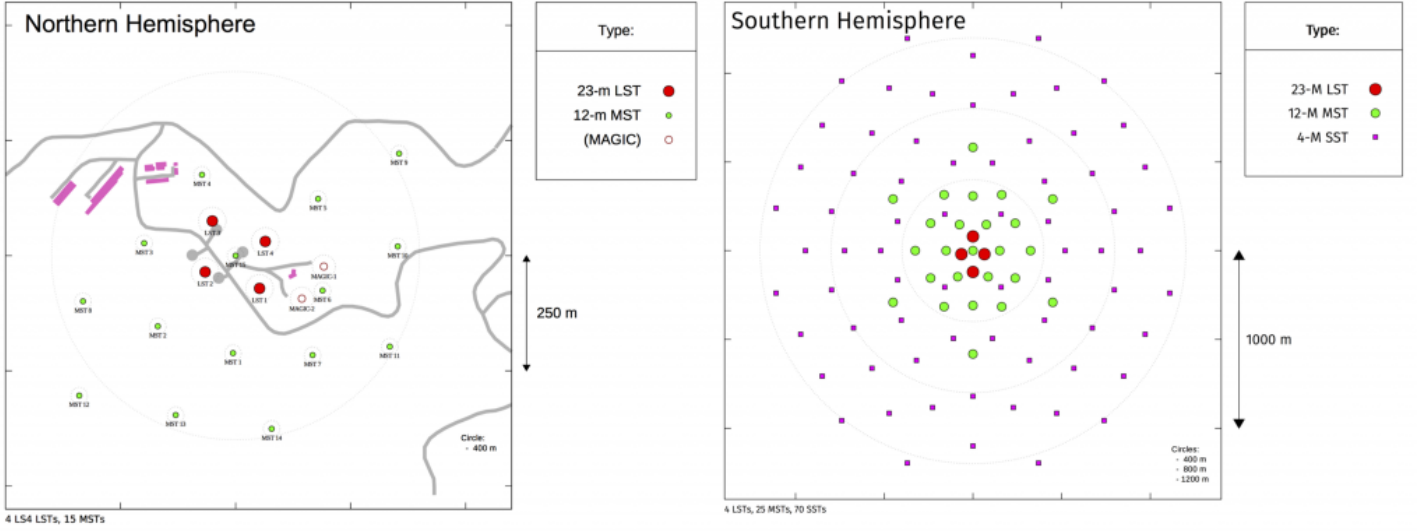


Figure 3: Layout of the CTA in both hemisphere. Credit: ^[1]

1.2. PMTs vs SiPMs

Photo Multiplier Tubes and Silicon Photo Multipliers are two light sensors used in different telescopes of CTA and have their respective advantages and drawbacks. The goal of this section is to compare the two technologies in order to understand why it is interesting to use SiPMs in our application.

These two sensors are devices that detect photons with a given efficiency: Quantum Efficiency (QE) for PMTs and Photo Detection Efficiency (PDE) for SiPMs. Once detected, an electron-hole pair is created in the device and accelerated by an intense electric field. Then for both devices, a single charge carrier is multiplied by a huge gain (up to 10^9 for PMTs and 10^7 for SiPMs) by different mechanisms presented in the next paragraphs. From an electrical point of view and for both devices, the signal to process at the output is a current generated by charge multiplication and the two sensors can be viewed as current sources ^[3].

A Photo Multiplier Tube [4, 3] is a vacuum tube presenting a photosensitive part, the photocathode (a semiconductor material), that will interact with a photon to emit an electron into the surrounding vacuum. This electrode is biased with a high negative voltage (up to -1000 V) referenced to the anode electrode which collects the signal, to create the intense electric field along the tube. To multiply the photoelectron, several dynodes are cascaded to induce secondary emissions for every electron striking the dynode surface. The dynodes are biased in such a way that every dynode attracts the flying electrons with a higher potential than the previous one. To do so, a resistive divider is implemented from the high voltage power supply. One drawback of PMTs is that to ensure a linear device response (linear output current with respect to the detected photons), the current flowing in the resistive divider needs to be much higher than the photocurrent flowing in the tube to minimise the voltage fluctuations of the dynodes with respect to the photocurrent [4]. Therefore, there is a high power consumption at steady state. The final element is the anode electrode which collects the resulting current pulse. In the case of a PMT, the electron multiplication stops by itself after few nano seconds and the resulting signal is extremely fast.

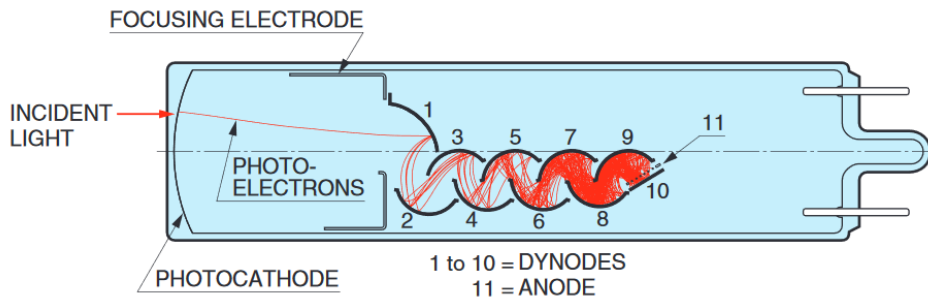


Figure 4: Basic operation of a PMT. Credit: [4]

A Silicon Photo Multiplier is a matrix of micro-cells, each one composed of an Avalanche Photo Diode (APD) in series with a quenching resistor. Contrary to the PMT, a SiPM is a solid-state device (typical dimensions of a micro-cell: $50\mu m \cdot 50\mu m$). Like any photodiode, the diode is biased in reversed mode (Positive voltage on the N-doped region, i.e the cathode, with respect to the P-doped region, i.e the anode). However, the bias voltage is high enough to induce an avalanche effect: the electric field is so intense that the kinetic energy of charge carriers is greater than the bandgap energy [5]. Those carriers can then ionise lattice atoms, creating new electron-hole pairs by avalanche effect.

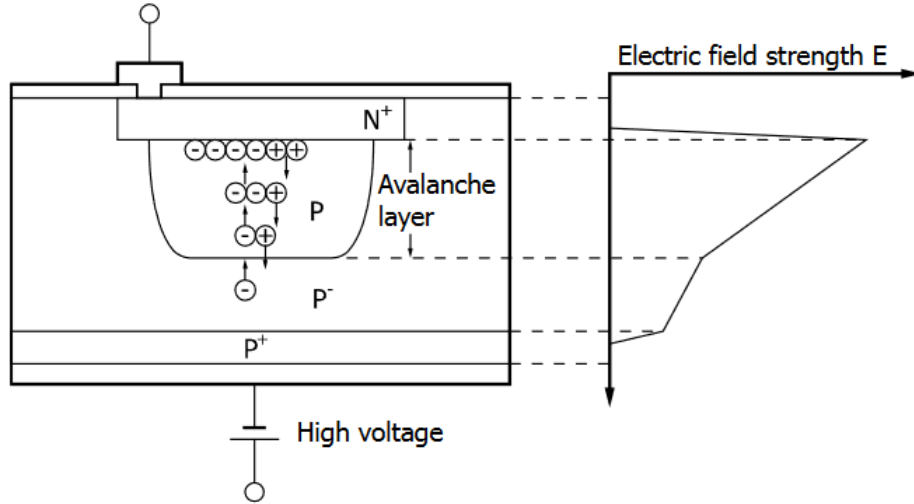


Figure 5: 2D-cut of an Avalanche Photo Diode. Credit: [5]

Like for PMTs, APDs are characterised by their multiplication coefficient or gain which depends on the device bias voltage. However for a SiPM, the APDs operate in Geiger mode (GAPD), i.e the bias voltage is greater than the break-down voltage (V_{BD}) of the photodiode (typical overdrive voltage = 3V). In this mode, there is no linearity between the initial number of photoelectrons and the total charge collected at the anode electrode. It means that a GAPD can be seen as digital photodetector [5]: it generates an output current pulse if at least one photon is detected. Nevertheless, if we place many GAPDs in parallel (=SiPM) to sum their current contributions, we have a resulting current which is an estimator of the number of photoelectrons.

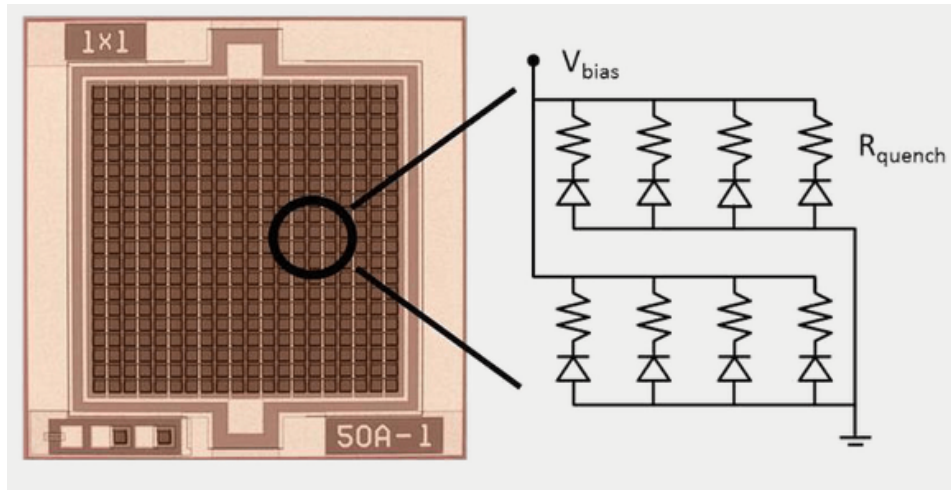


Figure 6: Photography of a SiPM with its corresponding electrical diagram. Credit: <https://www.ketek.net/sipm/technology/working-principle/>

If the bias voltage is kept above the breakdown voltage, the avalanche does not stop by itself like for PMTs. A quenching resistor is then integrated in series with every GAPD, to automatically limit the current flowing through the device and get back to a steady-state after having generated a current pulse. The principle of a SiPM can be resumed in two phases thanks to the simple model [5] shown below:

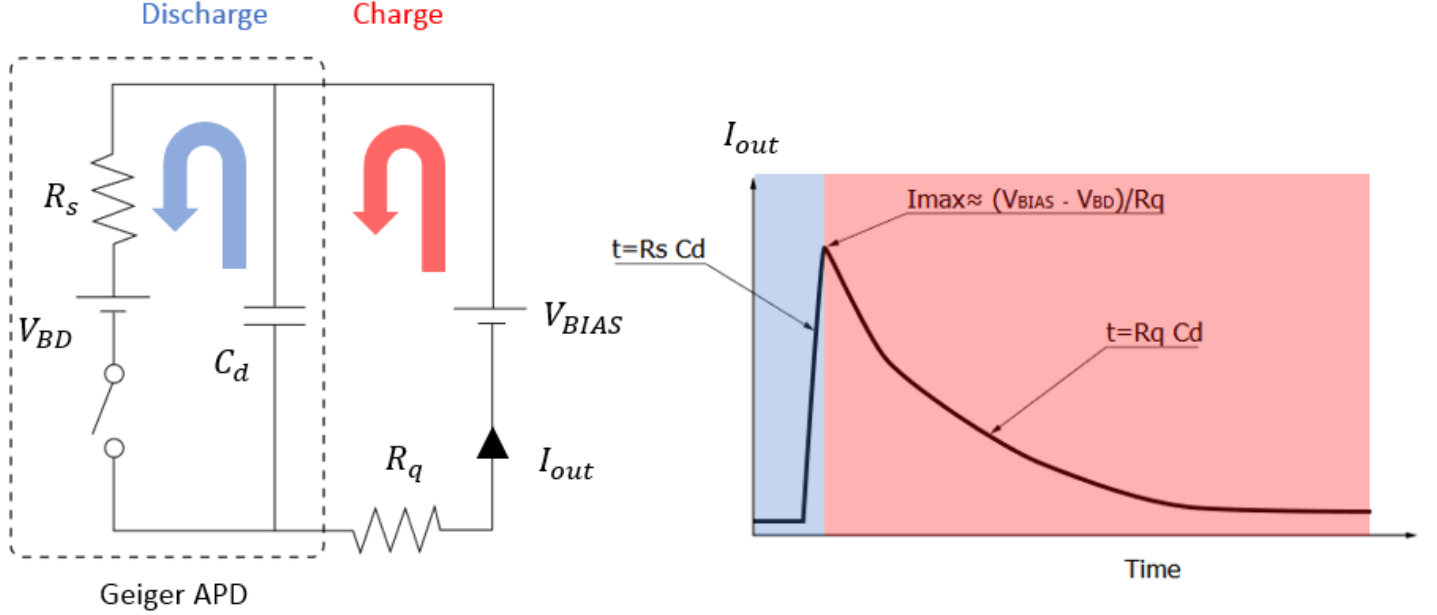


Figure 7: Basic 2-steps operation of a SiPM. Adapted from [5]

- When a photon carrier is generated, it triggers an avalanche effect modeled by the closing switch. The depletion capacitance C_d of the photodiode, initially full of charge and biased at $V_{BIAS} > V_{BD}$, is starting to fastly **discharge** through the low depletion resistance R_s . Once the voltage across C_d reaches approximately V_{BD} , the avalanche stops (the switch opens) and the output current reaches its maximum value: $I_{max} = (V_{BIAS} - V_{BD})/R_q$.
- The output current has been quenched thanks to the voltage drop across R_q . Then, the depletion capacitance **charges** through R_q with a much higher time constant than for the discharge ($R_q/R_s \approx 10^3$). The SiPM is finally ready for another detection.

PMTs and SiPMs are fascinating devices and their characteristics are well detailed in the literature. This section is not aimed to study in details the two devices but rather to compare them in the frame of our application, i.e for Cherenkov telescopes:

	PMT	SiPM
Gain	$10^5 - 10^9$	$10^5 - 10^7$
QE or PDE	Up to 40% at 415 nm	Up to 50 % at 415 nm
Dynamic Range	Limited by the divider current	Limited by the number of micro-cells
Bias voltage	Up to 1000 V	Up to 60 V
Dark output	Up to 500 cps	Up to 1000 keps
Design Complexity	High voltage	Temperature compensation
Mechanical Robustness	Medium	High
Magnetic Field	Need protection	Immune
Timing	Fast/Very Fast	Fast
Temperature Sensitivity	Low	Medium
Ambient Light Exposure	Can be damaged	No damage
Hysteresis	Yes	No
Warm Up Time	Required (minutes)	No
Price (per unit area)	High (low)	Low (high)

Table 2: PMT and SiPM comparison. Credit: [3]

The main differences between PMTs and SiPMs that are important to point out for our application are colored in the precedent table. One important drawback of PMTs is the ageing of the devices: PMTs responses will evolve in time (Hysteresis behavior) and can be damaged because of ambient light exposure. In a Cherenkov Telescope application there is a background light that can be important, for example, due to the moon [6]. Therefore, using SiPMs assure a long camera life. We can also see from the table that SiPMs are much noisier than PMTs (dark count rate for a given detector surface). This is why PMTs are used in very low light applications such as bio-luminescence [3] whereas SiPMs need to be cooled-down to decrease their dark count rate. In our application, this sensor dark count rate is negligible regarding the Night Sky Background (undesirable photons) which can reach 1600 MHz/pixel for the LST (detailed later in the section 5.2.3) and inevitably degrades the SNR at the input of any sensor.

1.3. Frame and goal of the thesis

The CTA research team from the *Département de physique nucléaire et corpusculaire* in the University of Geneva is currently working on the LST camera upgrade, relying on the previous SST-1M camera they have successfully designed [6, 7] and by using smaller SiPMs pixels. For the LST, we propose to replace the PMTs by SiPMs. Using a similar hexagonal SiPM pixel as the one developed for the SST-1M camera, one PMT pixel will be replaced by four SiPM pixels to keep the same detection area, for a total of 7420 pixels. We see here the importance of low power consumption due to the high number of pixels. Based on the SST-1M camera design, an overview of the processing chain is represented in the next figure:

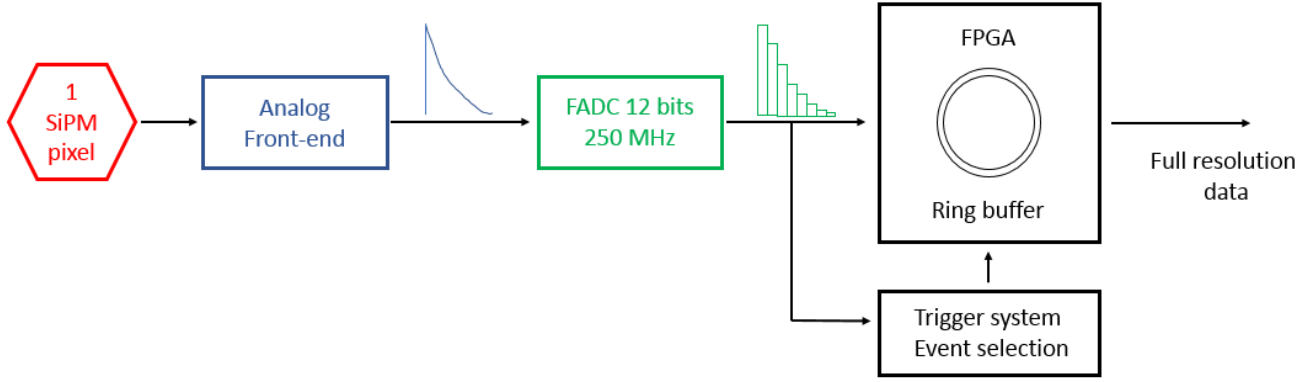


Figure 8: SST-1M camera overview: the analogue signals are digitised by DigiCam [6]

Every sampling period (4 ns in this example), the analog signals are digitised, pre-processed by a high-speed FPGA and memorised in local ring buffers. In parallel, a trigger system selects a probable event based on a lower resolution copy of the data and stored in another memory. Several selection algorithms are implemented to recognise specific patterns corresponding to a certain type of physical event. The ring buffer memory architecture allows the trigger system to analyse the previous images while keeping sampling data in real-time. The first processing stage, called "analogue front-end" has a major contribution to noise and needs to be carefully designed if we want to achieve single photoelectron detection.

Supervised by Dr.Heller from the Geneva CTA team and Dr.Koukab from EPFL, my thesis project consists of the design of an analogue Integrated Circuit dedicated to the SiPM sensor. This study has been done in the frame of the hardware researches presented before, i.e consisting in upgrading the actual LST camera.

The first motivation for designing an IC is to reduce the power consumption. Indeed, the upgraded LST version will have 7420 pixels, each one connected to its own analogue front-end. If this first processing stage is implemented with discrete components, the total power consumption reaches several kW [7]. By choosing an IC solution, we can at least divide this power consumption by 3 thanks to the supply voltage scaling. Moreover, we can expect general better performances due to the specific design for the chosen sensor, which is the definition of an ASIC (Application Specific Integrated Circuit).

1.4. State-of-the-art analogue front-ends and specifications

The Geneva CTA team is currently considering three analogue front-ends to upgrade the LST camera. The first one is the custom discrete-components front-end from SST-1M that would be adapted for the LST. Another one is also based on discrete-components and developed by the *Istituto Nazionale di Fisica Nucleare* in Torino and the last one

is an ASIC called *Multiple Use SiPM Integrated Circuit* (MUSIC) [8] designed by the *Institute of Cosmos Sciences* from the University of Barcelona.

The three front-ends have different approaches: the SST-1M front-end uses 2 OPamps in trans-impedance mode to amplify independently 2 pairs of channels (4 channels per pixel) to have the shortest output pulse (capacitance of the sensor divided by 2). Then, the 2 signals are summed with a last OPamp stage. The front-end is also DC-coupled, giving information about the NSB. (detailed in the section 5.2.3).

The INFN circuit is based on a common emitter voltage amplifier which is AC-coupled to the SiPM sensor. The output voltage is then the derivative of the input current pulse and the front-end achieves very fast responses (the pulse Full Width at Half Maximum (FWHM) ≤ 3 ns).

The ASIC MUSIC [8] implements a low-impedance current conveyor which sinks and amplifies the SiPM current. The chip is designed to satisfy a large number of SiPM applications and present several tunable functions such as a Zero/Pole/Cancellation to shape the output pulse.

The three different solutions are good candidates for the new LST analogue front-end so the Geneva CTA team is currently performing measurements that can be resumed in the next table. The SNR is the output pulse amplitude for one photoelectron (p.e) divided by the output integrated noise (one standard deviation) on a given bandwidth (1 GHz or 20 MHz in this measurement).

	Power/pixel [mW]	Total power (7420 pixels) [kW]	Gain [mV/pe]	SNR	FWHM [ns]	Dynamic Range [pe]
SST-1M front-end	400	2.97	2.4	10 (@ 1 GHz)	12	< 600
INFN front-end	800	5.94	2.0	4.3 (@ 1 GHz)	2.5	< 400
MUSIC ASIC	100	0.742	[0.63;1.25]	[2.25;2.9] (@ 20 MHz)	[12;20]	[500;1000]

Table 3: Performances of the state-of-the-art analogue front-ends that are candidates for the LST camera update

We can extract two important information from this table: as expected, a high- performances front-end based on discrete-components is power-hungry and we can save a lot by designing an IC. However, the measured SNR for the ASIC MUSIC is low and this is mainly due to the fact that the chip is not specifically designed for the SiPM sensor used in that experiment (LCT2 in that case).

In parallel to R&D on the sensor, the front-end and the digital readout, the group is also simulating the camera response to understand if the studied electronics would meet the performance requirements of the LST camera. Several configurations have been

simulated and preliminary results tend to show that the SiPM camera out-performs the PMT camera in the 100-800 GeV energy range.

The rest of the thesis will now present a full-custom IC design using the UMC 0.18 μm technology and dedicated to the LCT5 (Low Cross Talk) sensor, which is the SiPM sensor selected by the Geneva CTA team for the LST camera upgrade.

The main design specifications are:

- ASIC for the LCT5 SiPM sensor
- The camera power consumption < 1 kW
- The gain should be in the order of 3 mV/p.e
- $SNR = gain/\sigma > 10$ for one photoelectron
- Pulse width requirements: FWHM < 5 ns, Recovery time < 30 ns.
- Single gain
- Differential output

Optional specifications:

- Predictable behaviour in saturation
- DC information for NSB determination

Regarding the results obtained with the measured front-ends (Table (3)), the most challenging specification is the SNR. It has indeed been the bottleneck of many designs I have investigated.

2. SiPM electrical model

The goal of this section is to derive a simple equivalent Norton source of the SiPM sensor in order to design the best suited front-end.

2.1. Corsi model

The Corsi model is an electrical model of a SiPM using linear components [9]. It is very useful to model the SiPM current during Spice simulations because it predicts the amplitude and shape of the input signal based on several parameters given in the data-sheet of the sensor.

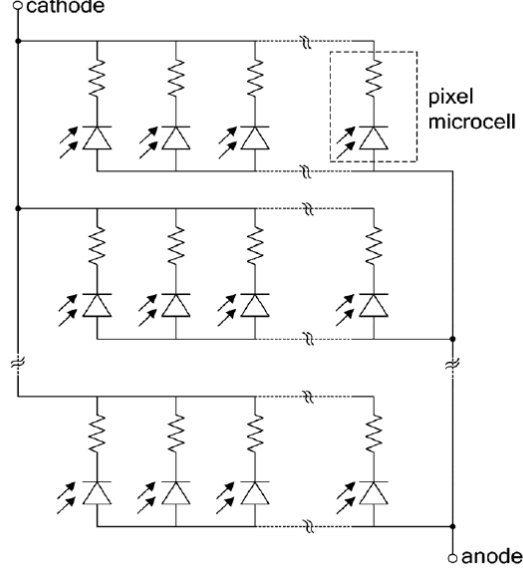


Figure 9: GAPDs matrix forming a SiPM from [10]

The Corsi model uses the following parameters:

- The quenching resistance R_q
- The parasitic capacitance C_q in parallel with R_q
- The depletion capacitance C_d in parallel with every photodiode
- The parasitic grid capacitance C_g in parallel with the entire matrix
- The number of photodiode N
- The number of firing photodiode N_f
- The photodiode breakdown voltage V_{BD}
- The resistance R_d in series with the photo-diode

Every photodiode operates in Geiger mode, meaning that they are biased above their breakdown voltage V_{BD} thanks to the voltage source V_K . When a Geiger Avalanche Photo Diode (GAPD) fires after having been triggered by a photon, its depletion capacitance C_d and C_q discharge through R_d , causing a very fast output current step. This fast rising time constant is given by:

$$\tau_r = R_d(C_d + C_q)$$

The amplitude of this current step is automatically limited by the quenching resistance R_q which causes a voltage drop and stops the avalanche. For N_f firing micro-cells at the same time, the peak current is approximately given by:

$$I_{peak} = N_f \frac{V_K - V_{BD}}{R_q} \quad (1)$$

During the quenching operation, the photodiode returns slowly in steady-state (biased to V_K) before being ready to produce another current pulse. This very slow charging time constant is given by:

$$\tau_q = R_q(C_d + C_q) \quad (2)$$

The Corsi electrical model separates the SiPM in two parts: one containing all the firing micro-cells and one modeling the $N - N_f$ micro-cells in steady-state. Because the model uses linear components, we can apply the superposition theorem on the equivalent circuits of the micro-cells to get the following SiPM equivalent electrical schematic:

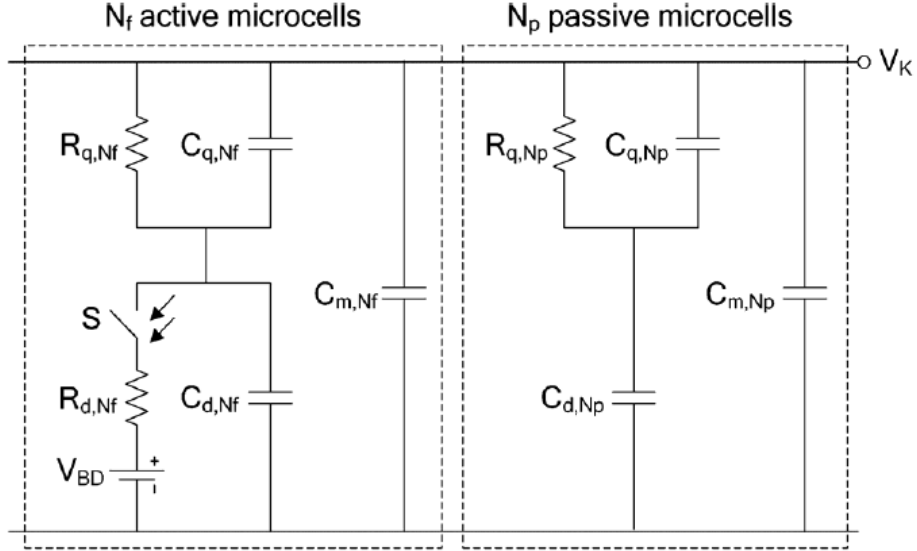


Figure 10: SiPM Corsi electrical model from [10]

With:

$$\begin{aligned} R_{q,N_f} &= \frac{R_q}{N_f} & R_{q,N_p} &= \frac{R_q}{(N - N_f)} \\ C_{d,N_f} &= C_d N_f & C_{d,N_p} &= C_d (N - N_f) \\ C_{q,N_f} &= C_q N_f & C_{q,N_p} &= C_q (N - N_f) \end{aligned}$$

We can either model the firing operation by closing an ideal switch S for a time τ_r , thus having the information about the rising time, or model the discharge with a Dirac current pulse while losing the rising time information [10]. By choosing the Dirac model, we can suppress the voltage source V_{BD} , knowing that the charge delivered by N_f micro-cells is:

$$Q = N_f(C_d + C_q)(V_K - V_{BD}) \quad (3)$$

I chose this last version in order to derive the equivalent small-signal model of the sensor because we can then easily transform the electrical diagram in an equivalent input current source. The information about the rising time is however lost, but it is not detrimental in our application because we are limited by the quenching time.

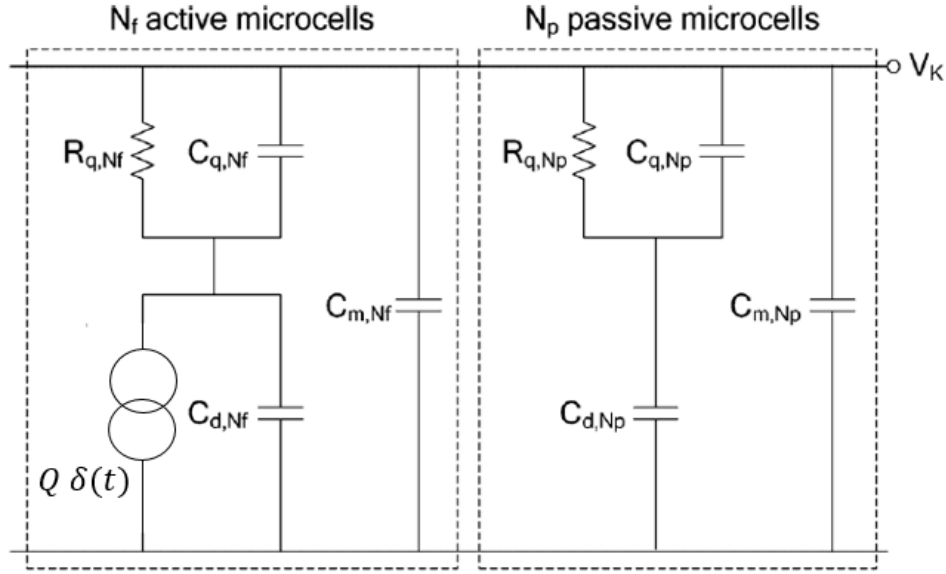


Figure 11: Final Corsi model used to derive an equivalent small-signal input source. Adapted from [10]

2.2. Equivalent input source

By using the Norton/Thevenin theorems (detailed in the appendix A), I could reduce the precedent electrical circuit to the following equivalent input source in Laplace's domain:

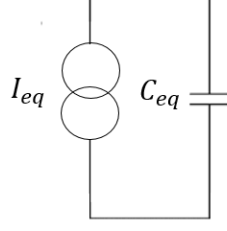


Figure 12: equivalent small-signal input source

With:

$$I_{eq} = Q \frac{1 + R_q C_q s}{1 + R_q (C_q + C_d) s} \quad (4)$$

$$\begin{aligned} \omega &<< \frac{1}{R_q (C_d + C_q)} & \Rightarrow & C_{eq} = C_G + N C_d \\ \omega &>> \frac{1}{R_q C_q} & \Rightarrow & C_{eq} = C_G + N \frac{C_d C_q}{C_d + C_q} \end{aligned} \quad (5)$$

We clearly see that the intrinsic quenching time $\tau_q = R_q (C_d + C_q)$ limits the signal bandwidth. It will get worse by loading this current source as shown in the next paragraph because of the big output capacitance C_{eq} . Moreover, if not cancelled with a pole, the zero $\tau = R_q C_q$ will degrade the pulse shape as shown later in the section 4.

2.3. Equivalent input source loaded by Z_L

The goal of the front-end is to present a load to the sensor in order to sense its output current. Because we are dealing with a current source, we need a very low impedance load. We then have two choices: processing the voltage across the load or the current flowing through it.

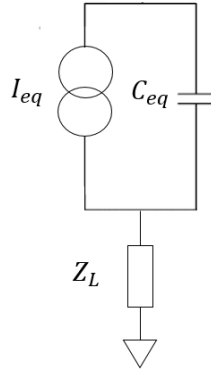


Figure 13: equivalent small-signal input source loaded by Z_L

The voltage across the load is low-pass filtered by C_{eq} :

$$V_L = I_{eq} \frac{Z_L}{1 + Z_L C_{eq} s} \quad (6)$$

If $Z_L = R_L$, another pole is introduced at the pulsation $1/R_L C_{eq}$, which is generally lower than the intrinsic cut-off pulsation $1/\tau_q$.

If we take the example of the LCT5 sensor which is the chosen sensor: in order to have $1/R_L C_{eq} > 1/\tau_q = 1/30.5 \text{ ns} = 5.2 \text{ MHz}$, we need $R_L < 3.8 \Omega$. By doing so, we will anyway be limited by the intrinsic cut-off frequency at 5.2 MHz and we will not meet the specification on the pulse falling time:

$$t_{fall} < 30 \text{ ns}$$

Indeed in this example, even if $R_L < 3.8 \Omega$

$$t_{fall} \approx 3\tau_q = 91.5 \text{ ns}$$

2.4. Ways of decreasing the pulse falling time

One way of increasing the speed is to cancel the intrinsic pole by inserting a zero. This can be done with two methods:

- Inserting a zero in the next processing stage (voltage amplifier for example)
- Inserting a zero directly in the load impedance Z_L

The first method is the most straightforward but there is also a way to cancel the intrinsic pole directly within the load impedance.

If we have:

$$Z_L = \frac{R_L(1 + \tau_z s)}{1 + \tau_p s} \quad (7)$$

Then

$$V_L = I_{eq} \frac{R_L(1 + \tau_z s)}{1 + s(R_L C_{eq} + \tau_p) + s^2 \tau_z R_L C_{eq}} \quad (8)$$

We see that if $\tau_z = \tau_q$, we cancel the intrinsic pole in (4) and the new cut-off pulsation is given by the second-order function:

$$\omega_0 = \frac{1}{\sqrt{R_L C_{eq} \tau_z}}$$

The choice between the two methods to cancel the intrinsic pole will depend on the SNR study. Indeed, the precedent proposition (8) supposes that we process the voltage across Z_L , i.e the SiPM anode voltage. In this case, we do not use at all the amplification of the first stage and this will lead of course to a worse SNR. Therefore, I decided to implement the first method and cancel the intrinsic pole τ_q later in a second stage (described in the section 5.4).

The next section presents the studies of different input-stages to amplify the SiPM current and maximise the SNR.

3. Study of different input-stages dedicated to a SiPM

All the following circuits have been considered to amplify the SiPM current. They have been studied analytically and simulated with the Cadence's tools in order to identify the best architecture for our application. To be concise, only the important results allowing us to choose or reject one type of architecture are explained.

3.1. Regulated-drain current mirror

The idea of this circuit is to sink the sensor current thanks to a current mirror with a very low input impedance. The input impedance is reduced by using a shunt/shunt positive feedback structure [11] as shown below. A very nice feature is the control of the input DC voltage with the feedback amplifier, which allows us to well control the sensor bias voltage.

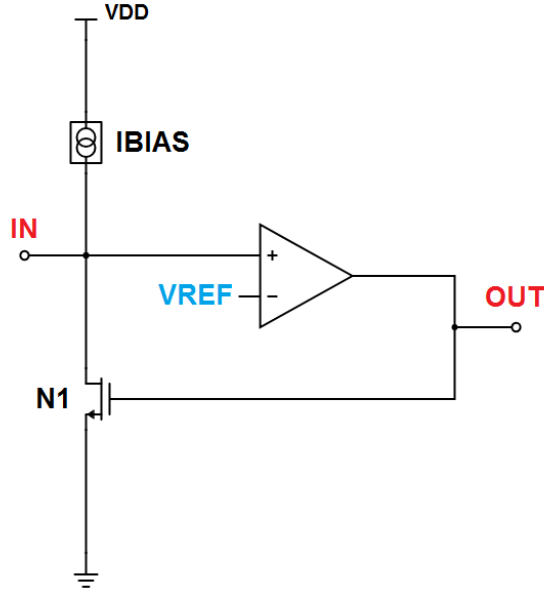


Figure 14: Regulated-drain current mirror schematic

The input impedance of this circuit is the load of the SiPM and is given by:

$$Z_L = \frac{1}{1/r_{o1} + g_{m1}A_v} \approx \frac{1}{g_{m1}A_v}$$

with A_v the positive voltage gain of the amplifier. We see that any pole from the amplifier transfer function A_v will be translated as a zero for the input impedance Z_L as it will be discussed in the section 3.1.2.

3.1.1. Implementation of the regulated-drain current mirror with an OTA

The most direct implementation we could think of is to design an Operational Transconductance Amplifier and use it in a non-inverting configuration to set precisely the positive gain A_v .

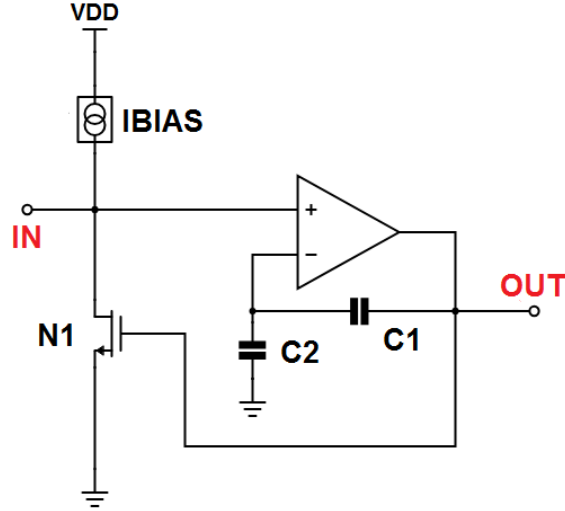


Figure 15: Regulated-drain current mirror with an OTA

The main limitation of this circuit is the noise generated by the OTA, even with large biasing currents. For the future investigations, I then focused on more compact architectures with very few transistors, in order to maximise the SNR of the input-stage.

3.1.2. Regulated-drain current mirror with a common-gate amplifier

Because we need a positive amplification of N1 drain voltage, a compact implementation is to use a common-base amplifier cascoded with N1. The input impedance is now:

$$Z_L = \frac{1}{g_{m1} g_{m2} Z_{out}}$$

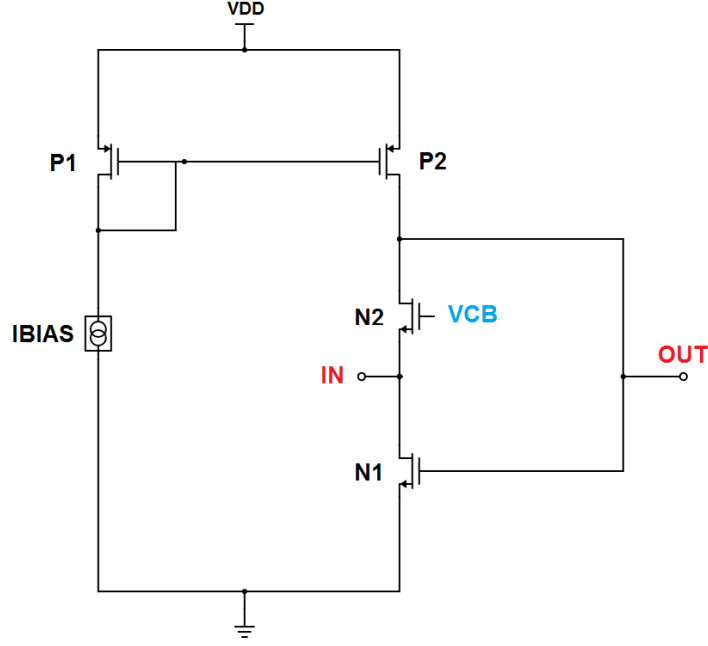


Figure 16: Regulated-drain current mirror implementation using a Common-gate amplifier

The output impedance is the small-signal output resistance $R_{out} = r_{on2}/r_{op2}$ in parallel with the output capacitance C_{out} :

$$Z_{out} = \frac{R_{out}}{1 + R_{out}C_{out}s}$$

Therefore we have:

$$Z_L = \frac{1 + R_{out}C_{out}s}{g_{m1}g_{m2}R_{out}} \quad (9)$$

As mentioned before (8), the zero given by $R_{out}C_{out}$ can easily be adjusted to cancel the sensor intrinsic pole, having:

$$R_{out}C_{out} = \tau_q$$

However, I did not use this feature because of SNR considerations and I rather chose to process the gate voltage of N1. But we now understand why the pulse voltage V_L can be faster than the output voltage of the first stage as shown in the next plot: the top waveform is V_L , the middle one is the first stage output, the bottom waveform is the output of the second stage which implements a ZPC.

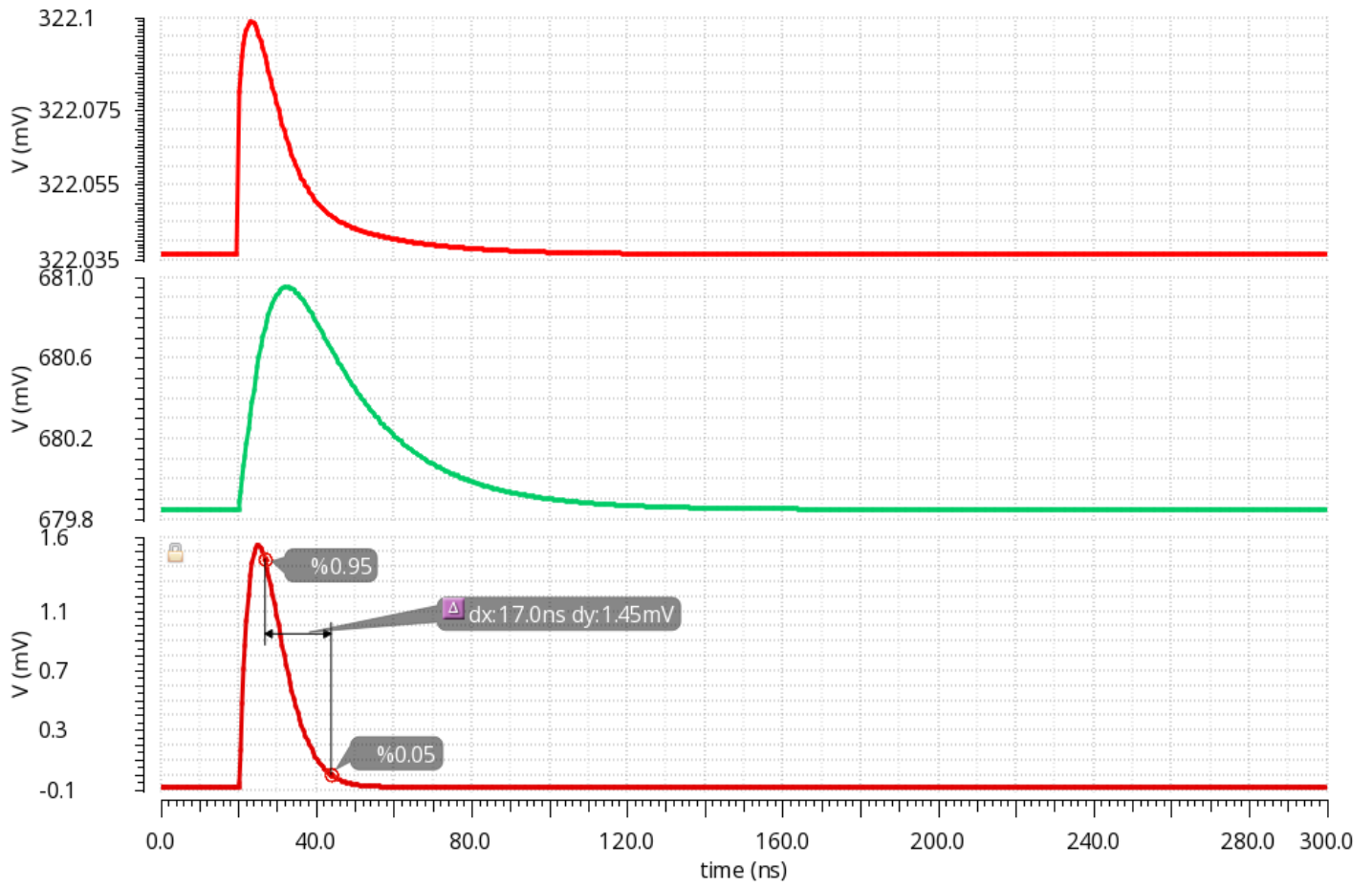


Figure 17: Transient simulation of the regulated-drain current mirror followed by a differential ZPC stage - 1 p.e

3.1.3. Regulated-drain current mirror with a compact common-gate amplifier

This circuit is the same as the precedent one excepting that a diode-connected NMOS loads the common-gate amplifier. The voltage gain set by the common-gate amplifier is now gm_2/gm_3 and is much lower than the one using an active load. The stage becomes very compact.

The main drawback of this circuit is that we do not have enough degrees of freedom to meet all the specifications. For example, we cannot set independently the output DC voltage, the biasing current and the stage output impedance. This will be detrimental during the optimisation of the front-end.

However, a nice feature deserves to be mentioned: any input positive parasitic current from the sensor will cause an increase of $V_{GS_{N1}} = V_{out}$, inducing a decrease of N3's current. Therefore, this negative feedback loop assures a robust biasing regarding the input parasitic current. The latter will be studied in details in the section 5.2.3.

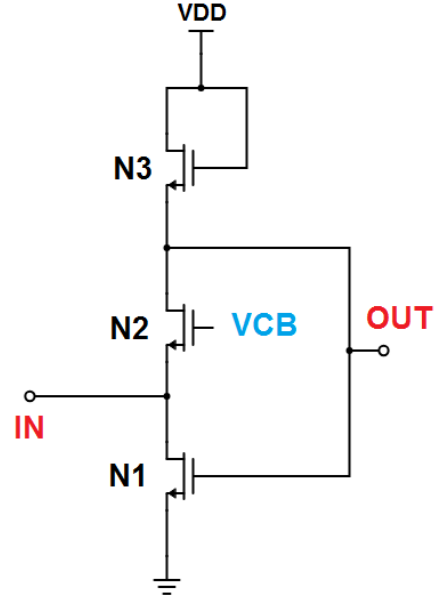


Figure 18: Regulated-drain current mirror with a passive load common-gate amplifier

3.2. Boosted Common-gate architecture

Another way of sinking the SiPM current is to use a common-gate amplifier. Used in its basic form, the input impedance is given by $1/g_{m1}$ and is too important for our application. Then, I used the idea from [12]: we need to boost the common-gate with another common-source amplifier, implemented by P2 in the following schematic. The boosted version has the advantage of not necessitating any voltage reference to bias the gate of P1. PMOS transistors are used because P1 must sink a positive parasitic DC current from the sensor.

The circuit input impedance is also given by (9), like the regulated-drain current mirror. These two circuits are very similar, but for noise and linearity considerations, the Boosted common-gate architecture has been abandoned.

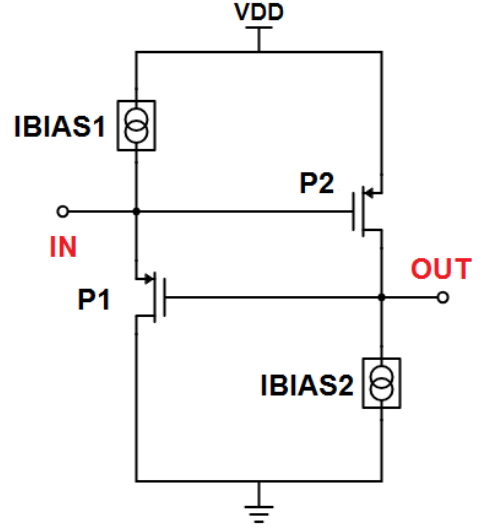


Figure 19: Boosted common-gate amplifier schematic

3.3. Comparison of all the studied input-stages

Architecture	RD with OTA	RD with active-CG	RD with passive-CG	Boosted-CG
Advantage	Gain and Z_L set precisely	Good SNR Good linearity	Good SNR Good linearity Very compact	-
Drawback	Too noisy and bulky	-	Limited number of degrees of freedom	Worse linearity Bad SNR

Table 4: Comparison of all the studied input-stages

The best studied input stage for our application is the regulated-drain current mirror with a common-gate amplifier (active load). It will be studied in details in the section 5.3.

4. Shape optimisation of the input current pulse

Before going into depth in the first-stage modeling, let's detail in this section how we can optimise the shape of the input pulse sunk by the front-end because it will directly constraint the design.

One temporal feature of the SiPM output pulse is a fast decaying exponential, followed by a slower one (quenching time). Indeed, for a short amount of time, the micro-

cells recharge through the path to ground constituted by $R_q//C_q$. This phenomenon is modeled by the zero R_qC_q in the equivalent SiPM current source expression (4).

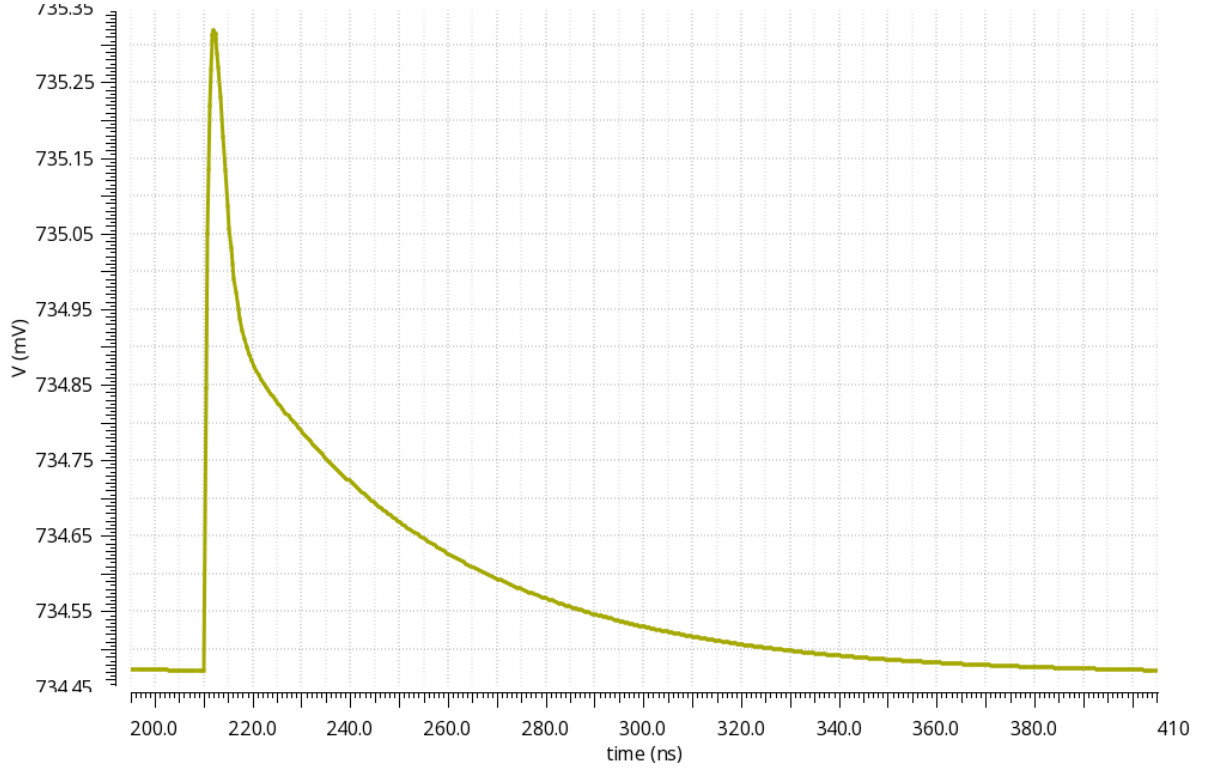


Figure 20: Typical output pulse of a SiPM simulated with Virtuoso, showing the two decaying time constants

The main problem with the presence of two distinct decaying parts is the bad shape of the output pulse after ZPC. Indeed, the ZPC stage is aimed to cancel a single pole: any additional zero will provoke undershoots and non-linearities. To have only one decaying time constant, we need to cancel this zero with a pole in the front-end before the ZPC stage. To better understand the origin of this additional fast decaying part, it is useful to look at the frequency domain. Let's recall the precedent equations from the section 2.2. I_{eq} is the equivalent current source generated by the SiPM, in parallel with the equivalent output capacitance C_{eq} and Z_L is the front-end input impedance.

$$I_{eq} = Q \frac{1 + R_q C_q s}{1 + R_q (C_q + C_d) s}$$

The current sunk by the front-end is given by:

$$I_L = I_{eq} \cdot \frac{1}{1 + Z_L C_{eq} s} = I_{eq} \cdot H$$

In our case, H will be a second-order function as seen before in (8). We can then represent this function with two poles $1/\tau_{p1}$ and $1/\tau_{p2}$ separated in the frequency domain

(assuming a large damping factor). In the section 5.3 describing the design, I used one of this pole to cancel the unwanted zero $R_q C_q$.

4.1. Desired case: compensation of the zero $R_q C_q$

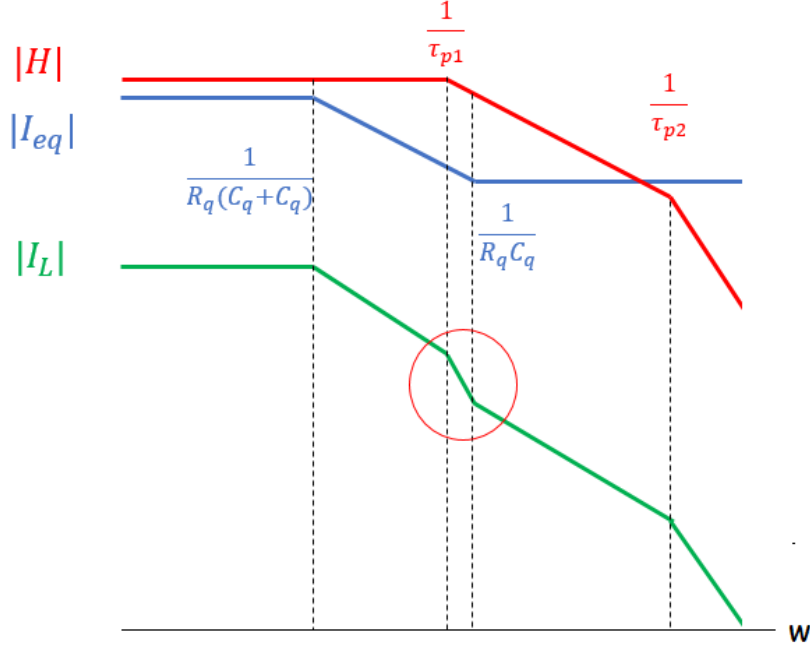


Figure 21: Representation of a good zero compensation in the frequency domain

As presented on this schematic, if we have the condition:

$$\frac{1}{\tau_{p1}} \leq \frac{1}{R_q C_q} \quad (10)$$

Then the pulse I_L will be composed of only one decaying part defined by $1/R_q(C_q + C_d)$. We need to take care about the value of τ_{p1} : if it is too big, this pole shifts to lower frequencies and creates a second-order region (circled region in the precedent schematic). In this case, $1/\tau_{p1}$ will be the new cut-off frequency of the entire system (after ZPC). Indeed, the chosen ZPC can only compensate one first-order region.

However, if we have the equality in (10), then the ZPC will not see any second-order region and the new cut-off frequency of the system will be $1/\tau_{p2}$, at a higher frequency.

The 2-decays problem happens when we place $1/\tau_{p1}$ at a higher frequency as shown in the next paragraph.

4.2. Undesired case: bad compensation of the zero $R_q C_q$

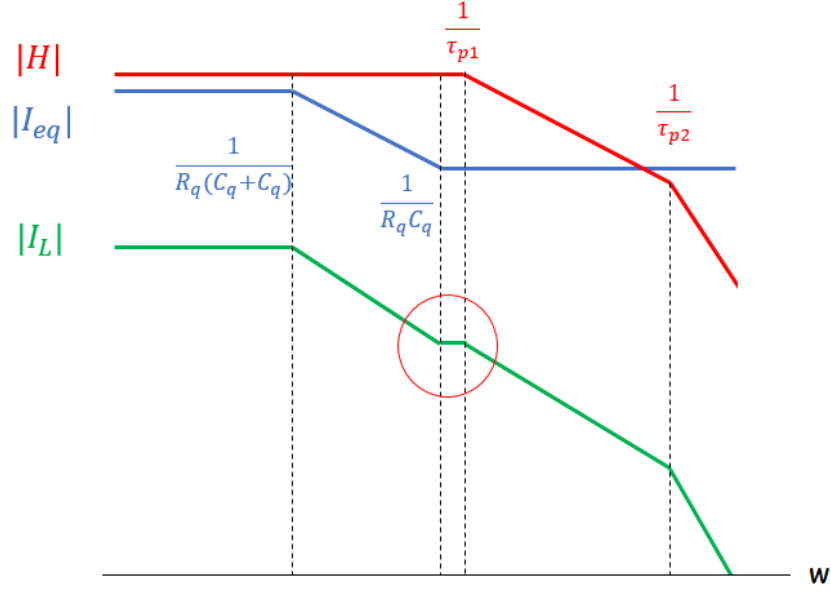


Figure 22: Representation of a bad zero compensation in the frequency domain

As presented on this schematic, when we have:

$$\frac{1}{\tau_{p1}} > \frac{1}{R_q C_q}$$

We obtain a flat-gain region (circled in the schematic) which creates this unwanted fast decaying part in time domain, defined by $R_q C_q$.

4.3. Illustration with Matlab

In the precedent subsection, we did not see the resulting current impulse response in time domain. This can be computed with Matlab:

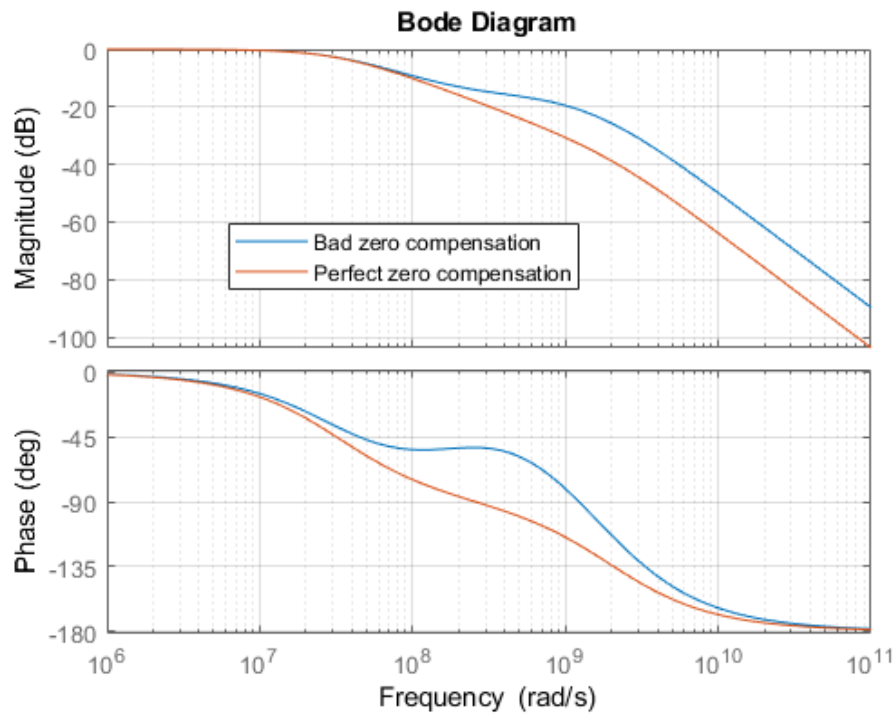


Figure 23: Zero compensation in frequency domain - Matlab simulation showing $|I_L|$

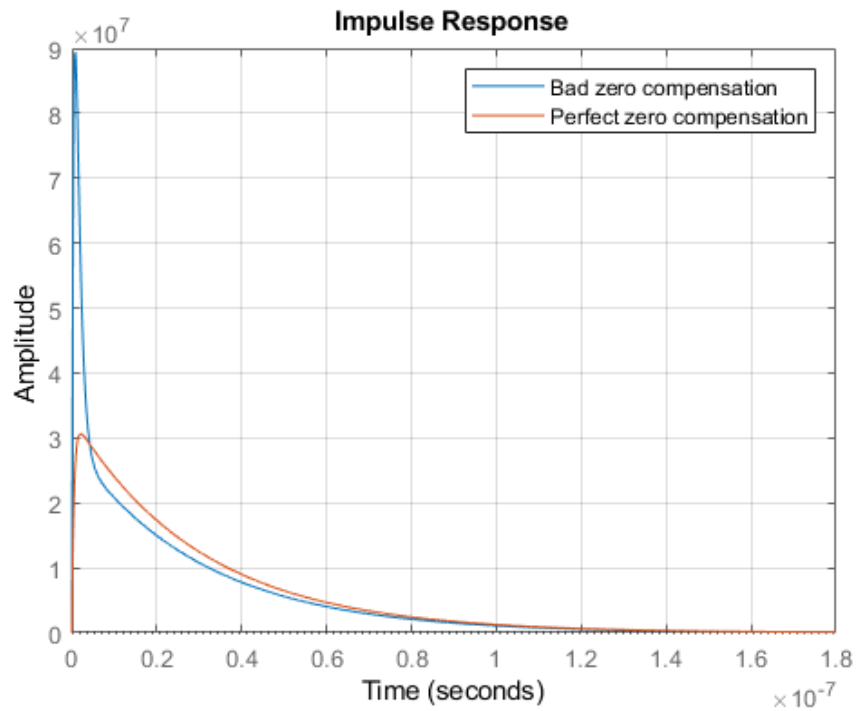


Figure 24: Zero compensation in time domain - Matlab simulation

It confirms what we discussed before: if we have the condition (10), we have only one decaying time constant which is the best shape we can have before the ZPC stage. For the next design steps, I then kept in mind the condition (10) to fulfill.

5. Final architecture modelling

In this section, the main features of the sensor, the input current stage and the differential voltage amplifier will be described using small-signals derivations.

5.1. General overview of the final architecture

The precedent studies have led to the following 2-stages front-end architecture:

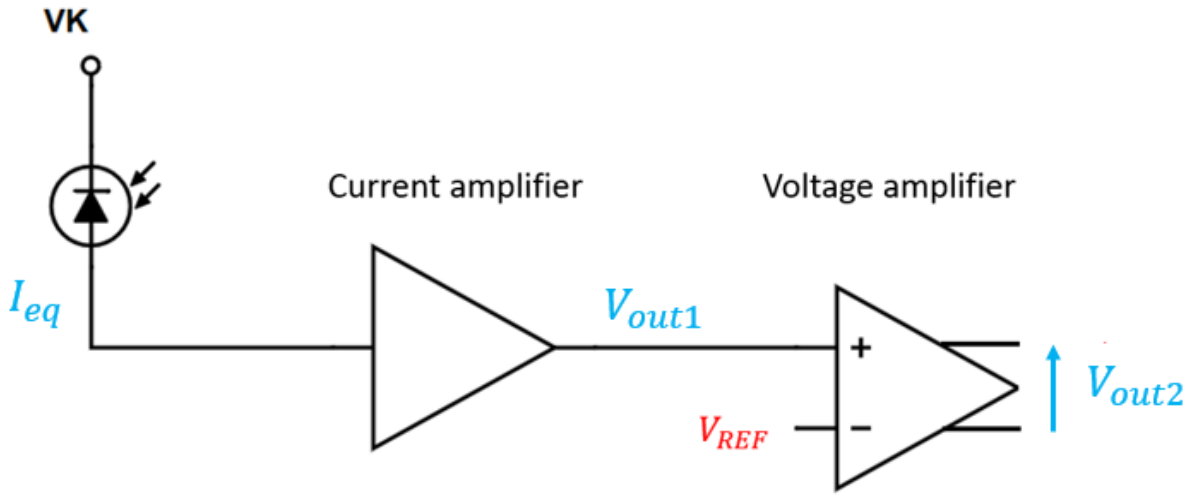


Figure 25: General front-end overview

It is composed of a Regulated-drain current amplifier, followed by a voltage differential amplifier which inserts a zero to cancel the intrinsic pole (2) of the sensor responsible for the long recovering (quenching) time. The entire front-end is DC-coupled to get the Night Sky Background information. This parasitic DC current will be computed later in the section 5.2.3.

The processing principle of the front-end can be summarised in the following frequency representation. H_1 and H_2 are the transfer functions of the first and second stage respectively.

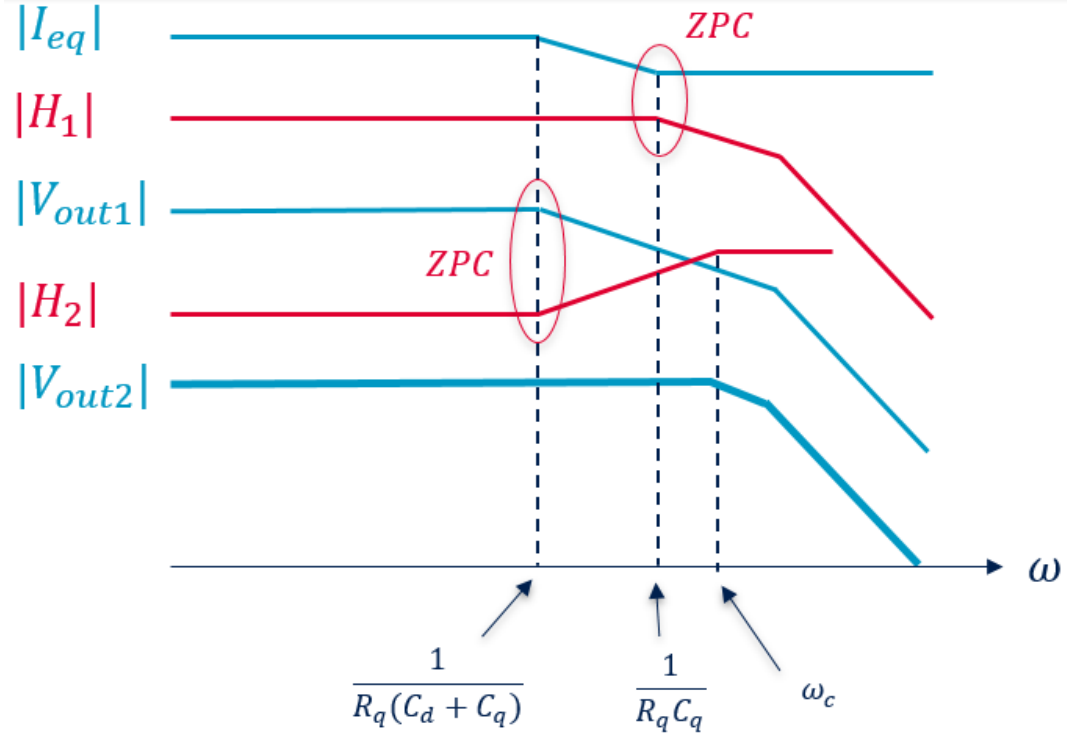


Figure 26: Signals and Transfer functions of the front-end in the frequency domain

The first stage sinks the SiPM current and converts it in voltage. As presented in the precedent section 4, one of the front-end pole is used to cancel the sensor intrinsic zero. Finally, the differential voltage amplifier cancels the remaining slow quenching pole. We can see on this representation that the final response V_{out2} shows at least a second-order low-pass behavior, with a second pole close to the first one. Therefore, the rising-time of the output pulse will not be negligible as for a first-order function (null in this latter case).

5.2. Intrinsic parameters of the LCT5 sensor

5.2.1. Corsi parameters

The Corsi parameters are given by Hamamatsu for a square-shaped sensor of 3600 micro-cells. The considered grid capacitance C_G is an estimation for an hexagonal-shaped LCT5 sensor for 4 connected channels (total number of micro-cells $N = 36840$) and is assumed to be similar to the grid capacitance of LCT2. The LCT5 Corsi parameters are confidential: therefore, the following table only presents an approximation of the right parameters that have been used in simulations.

C_d (fF)	100
C_q (fF)	20
C_G (pF)	$60 \cdot 4 = 240$
R_q (k Ω)	250

Table 5: Approximation of the LCT5 Corsi parameters for N=36840

5.2.2. Intrinsic pole, zero and output capacitance of the sensor

From the expression of I_{eq} (4):

$$\begin{aligned}
\text{Slow pole :} \quad R_q(C_d + C_q) &= 30.5 \text{ ns} & (\equiv 5.2 \text{ MHz}) \\
\text{Zero :} \quad R_q C_q &= 5 \text{ ns} & (\equiv 31.8 \text{ MHz})
\end{aligned}$$

The equivalent output capacitance of the sensor (5) is at high frequencies:

$$\omega \gg \frac{1}{R_q C_q} \Rightarrow C_{eq} = 788 \text{ pF} \quad (11)$$

In general, the output impedance of the sensor is a complex function of ω and N_f . We can use this approximated expression of C_{eq} for hand-calculations because the poles of the input stage will be located such that $1/\tau_{p1}$ and $1/\tau_{p2} \geq 1/R_q C_q$. However, the simulations using the Cadence tools include the complete Corsi model and present a more realistic sensor behavior.

5.2.3. Computation of the Night Sky Background DC current

The NSB signal is the sum of the current from all the firing micro-cells triggered by parasitic photons. Assuming that we canceled the intrinsic zero $R_q C_q$, every input current pulse is a pure decaying exponential defined by the quenching time constant τ_q . Considering $t = 0$ as the origin, the instantaneous value $i_{NSB}(t)$ is then given by:

$$i_{NSB}(t) = \sum_{n=0}^M I_{peak} \exp\left(\frac{-(t - nT)}{\tau_q}\right) \Gamma(t - nT) = \sum_{n=0}^M i(t - nT)$$

with T the average time interval between 2 NSB photons, M an integer such as $M \leq t/T < M + 1$ and $\Gamma(t)$ the Heaviside function.

To have the DC component of this signal, let's use the frequency representation. For that, we can write $i_{NSB}(t)$ using the convolution operation ($\star$) with a dirac comb:

$$i_{NSB}(t) = \sum_{n=0}^M \int_{-\infty}^{\infty} i(t') \delta(t' - (t - nT)) dt' = \int_{-\infty}^{\infty} i(t') \sum_{n=0}^M \delta(t' - (t - nT)) dt'$$

If we consider an infinite signal duration (steady-state):

$$i_{NSB}(t) = i(t) \star \sum_{n=-\infty}^{+\infty} \delta(t - nT)$$

In frequency domain, the Fourier transformation of i_{NSB} is simply the multiplication of the two Fourier transformations:

$$F\left(i_{NSB}(t)\right) = F\left(i(t)\right) \cdot \frac{1}{T} \sum_{n=-\infty}^{+\infty} \delta\left(f - \frac{n}{T}\right)$$

$F\left(i(t)\right)$ is the well-known first-order low-pass function with a DC gain of $I_{peak} \cdot \tau_q = Q$ ((1) and (3)) with Q the charge delivered to the input stage by one micro-cell. To get the DC component I_{NSB} of $i_{NSB}(t)$, we need to evaluate the precedent expression for $f = 0$:

$$I_{NSB} = \frac{Q}{T} = (C_d + C_q)(V_K - V_{BD}) \cdot f_{NSB} \quad (12)$$

To directly get this result, we could have considered a simplified shape of the input signal (green rectangle) by taking care of the charge conservation with the red pulse as represented in the next figure. The following table gives some values of this average DC current sunk by the input stage

f_{NSB} (MHz)	I_{NSB} (mA)
400	0.12
800	0.24
1600	0.48

Table 6: I_{NSB} for different f_{NSB} values and $V_K - V_{BD} = 3V$

A typical bias current for the input stage is 2.5 mA. We see that for an important NSB ($f_{NSB} = 1600$ MHz), 20 % more current will be sunk by the input transistor, which is not negligible.

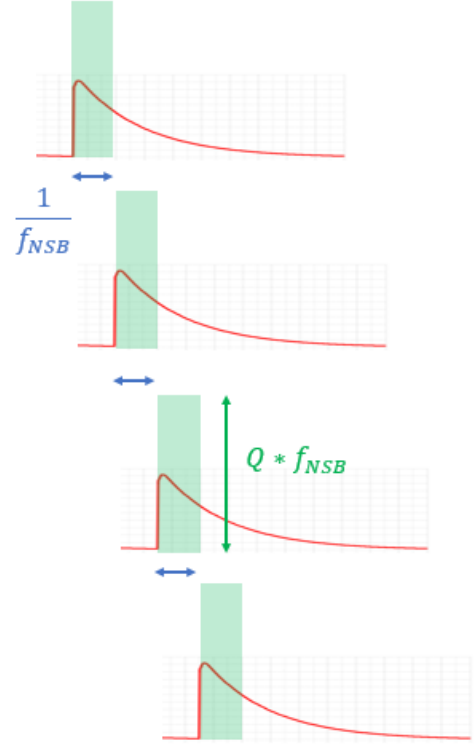


Figure 27: NSB signal with its DC component

5.3. Description of the input-stage

I chose the Regulated-drain current mirror topology described in the first section with one major modification: the gate voltage of N2 is not constant anymore but boosted by a second amplifier. I made that choice because of the following noise considerations:

- The voltage source biasing the gate of N2 adds noise and does not add any gain: it just degrades the SNR.
- Adding a Boosting amplifier of voltage gain A_{boost} also adds noise but divides the sensor contribution by A_{boost} in the SNR expression (please refer to the section [5.3.6](#)).

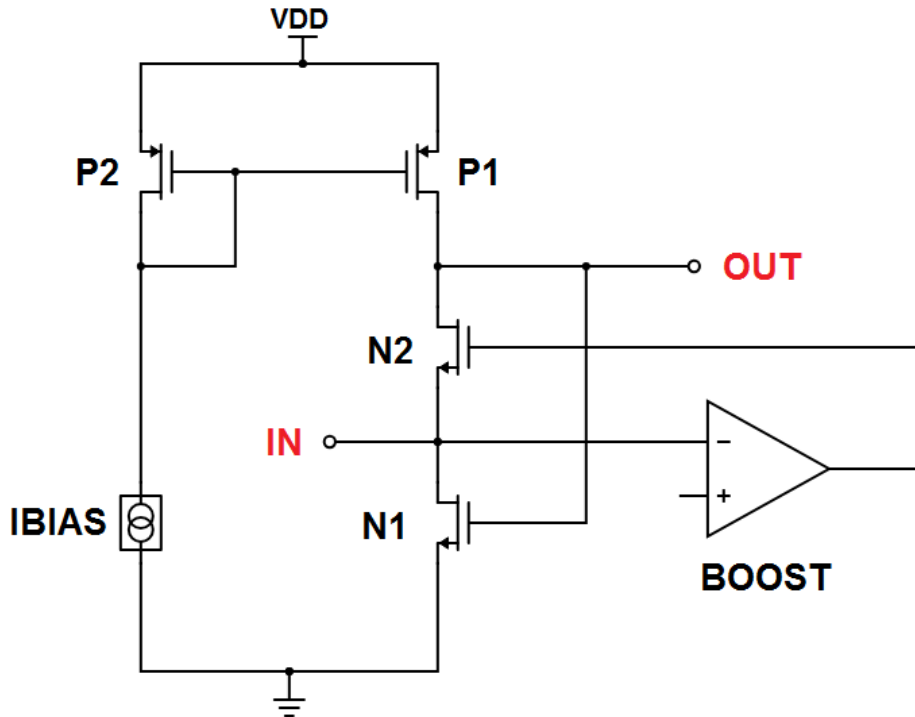


Figure 28: Chosen input stage: Regulated-drain current-mirror with a boosted-common-gate amplifier

The boost amplifier is implemented with a simple common-source amplifier to have a negative voltage gain:

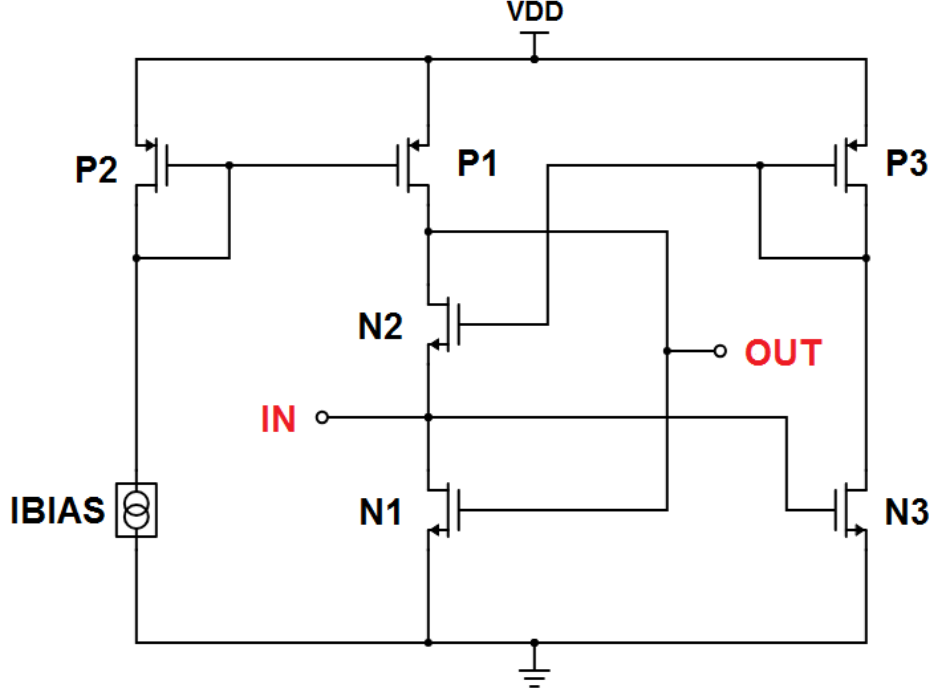


Figure 29: Transistor-level schematic of the input-stage

In the next sections, the equations include the main poles and zeros from the circuits to have the most precise transfer functions in order to optimise the signals shape.

5.3.1. Trans-impedance gain

The first stage is a current amplifier providing a voltage output. By considering its output capacitance C_{out} , mainly set by N1 gate to source capacitance and by the next stage input capacitance, its trans-impedance gain is:

$$Z_m = \frac{V_{out}}{I_{in}} = \frac{1}{gm_1} \cdot \frac{1}{1 + \frac{C_{out}}{gm_1}s} = \frac{1}{gm_1} \cdot \frac{1}{1 + \tau_p s} \quad (13)$$

5.3.2. Voltage gain

We have:

$$\begin{aligned} \frac{V_{out}}{V_{in}} &= A_v = gm_2 Z_{out} \cdot A_{boost} \\ &= \frac{gm_2 R_{out}}{1 + R_{out}C_{out}s} \cdot \frac{gm_{N3}}{gm_{P3}} \cdot \frac{1}{1 + \frac{C_{outB}}{gm_{P3}}s} \\ &= \frac{gm_2 R_{out}}{1 + \tau_{out}s} \cdot \frac{gm_{N3}}{gm_{P3}} \cdot \frac{1}{1 + \tau_B s} \end{aligned} \quad (14)$$

With:

- C_{outB} the boost amplifier output capacitance mainly set by the P3 and N2 gate to source capacitances.
- R_{out} the stage output resistance, set by r_{0N2}/r_{0P1}

5.3.3. Input Impedance

If we divide (13) by (14) we get the stage input impedance:

$$Z_{in} = R_{in}(1 + \tau_{out}s)(1 + \tau_Bs) \frac{1}{1 + \tau_ps} \quad (15)$$

With:

$$R_{in} = \frac{gm_{P3}}{gm_1 gm_2 R_{out} gm_{N3}}$$

The precedent relation confirms the principal goal of a regulated-drain circuit: to divide $1/gm_1$ by a voltage gain. In this case, the voltage gain is the gain of the boosted-common-gate amplifier which is at low frequency: $gm_2 R_{out} gm_{N3}/gm_{P3}$.

5.3.4. Pulse shape optimisation

The section 4 showed a way of cancelling the intrinsic zero R_qC_q with a front-end pole in order to have a pure exponential signal at the output of the first stage. The following paragraph presents how to do it with the chosen input stage.

$Z_{in} = Z_L$ is the stage input impedance and the load of the SiPM. We need now to take into account the sensor capacitance C_{eq} . By using the equations (8) and (14), we get:

$$V_{out} = \frac{1}{gm_1} \cdot \frac{I_{eq}}{1 + (R_{in}C_{eq} + \tau_p)s + R_{in}C_{eq}(\tau_{out} + \tau_B)s^2 + R_{in}C_{eq}\tau_{out}\tau_Bs^3}$$

Because of the boost amplifier, we now have a 3rd-order denominator and extracting its poles is a bit more difficult. However, we can easily find the poles by assuming the following equality:

$$\tau_p \approx \tau_B \iff \frac{C_{out}}{gm_1} \approx \frac{C_{outB}}{gm_{P3}}$$

The capacitance C_{outB} is mainly composed of P3's and N2's gate to source capacitances and is non-negligeable because of the important sizing $W \cdot L$ of N2. By tuning C_{out} , we can fulfill the precedent equality.

Under the precedent assumption, we can rewrite the transfer function:

$$V_{out} = \frac{1}{gm_1} \cdot \frac{I_{eq}}{(1 + \tau_Bs)(1 + s 2m/\omega_0 + s^2/\omega_0^2)} \quad (16)$$

With:

$$\omega_0 = 1/\sqrt{R_{in}C_{eq}\tau_{out}} \quad m = \frac{1}{2}\sqrt{R_{in}C_{eq}/\tau_{out}}$$

To optimise the output pulse shape, we need to cancel the intrinsic zero by having one of the pole located around 30 MHz. The pole $1/\tau_B$ is located at a higher frequency, so we have to set ω_0 around $2\pi \cdot 30$ Mrad/s. This has been done by increasing the output capacitance C_{out} by adding external capacitors. The rest of the parameters such as R_{in} and gm_1 have been set by considering simultaneously the gain and SNR specifications.

The next plot shows the bode-plot of the input stage for 4 different values of C_{out} :

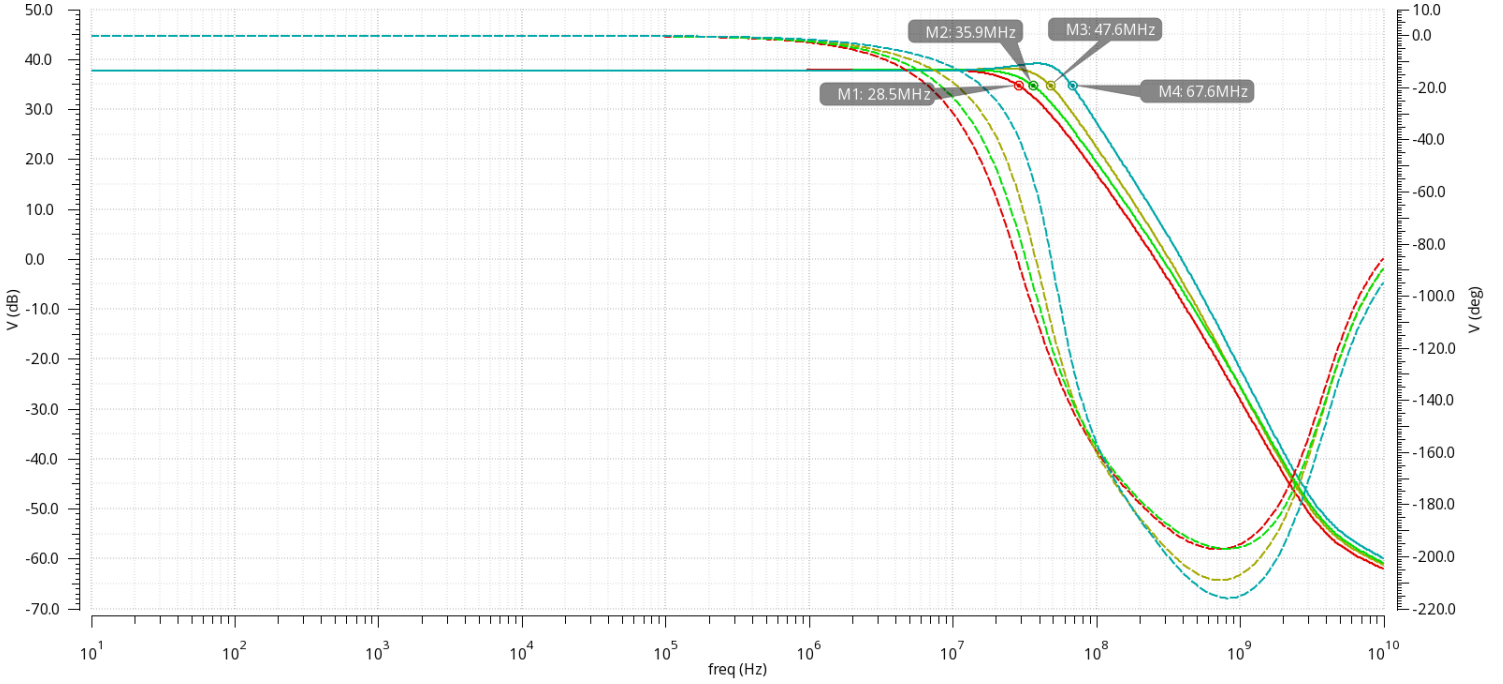


Figure 30: Post-layout Bode-plot of the first stage. The dashed lines are the phases.

The red configuration shows the best output pulse shape because the cut-off frequency $f_{-3dB} = 28.5$ MHz is set before the intrinsic zero $1/(2\pi R_q C_q) = 31.8$ MHz in order to cancel its effect. For the 3 other configurations, the intrinsic zero is not well-cancelled but the speed increases.

5.3.5. Implementation of 4 different configurations

As presented just before, by adding capacitance to C_{out} , we can decrease the speed and have an optimised output pulse shape. I decided to implement 4 different cases with two bits to command two switches:

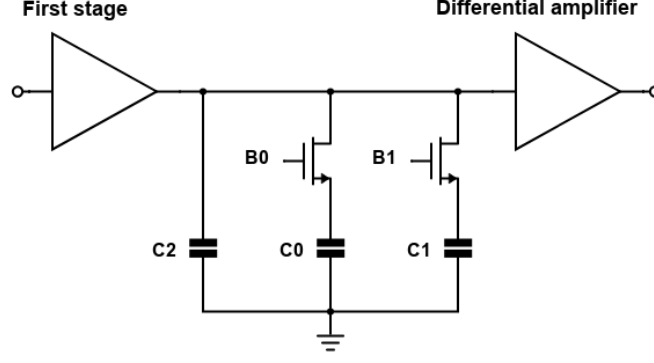


Figure 31: 4 different input-stage configurations

B0\B1	0	1
0	67.6 MHz	47.6 MHz
1	35.9 MHz	28.5 MHz

Table 7: input-stage cut-off frequency set by 2 bits

5.3.6. Signal to Noise Ratio

The goal of this hand-derivation is to identify which parameters are the most important noise contributors and thus to maximise the SNR. I decided to derive the SNR at the input of the stage; therefore it is defined as:

$$SNR = \frac{I_{pp}}{\sigma_N} [A/A]$$

with I_{pp} the peak-to-peak amplitude of the input current pulse, $\sigma_N = \sqrt{\int_0^{\Delta\omega} dI_{Neq}^2}$ the standard deviation of the input-referred noise and $\Delta\omega$ the equivalent bandwidth. Derivations detailed in the appendix B show that:

$$dI_{Neq}^2 = dI_{N1}^2 + dI_{P1}^2 + (dI_{N2}^2 + dI_{P1}^2) \left(\frac{C_{eq} \omega}{gm_{N2} A_{boost}} \right)^2 + (dI_{N3}^2 + dI_{P3}^2) \left(\frac{C_{eq} \omega}{gm_{N3}} \right)^2$$

By only considering the thermal noise and that all the transistors are in strong inversion:

$$dI_{Neq}^2 = 4kT \frac{2}{3} \left[gm_{N1} + gm_{P1} + (gm_{N2} + gm_{P1}) \left(\frac{C_{eq} \omega}{gm_{N2} A_{boost}} \right)^2 + (gm_{N3} + gm_{P3}) \left(\frac{C_{eq} \omega}{gm_{N3}} \right)^2 \right] df$$

We see here that the sensor capacitance C_{eq} gives an important contribution to the noise. We can now identify some ways of minimising the thermal noise generated by the circuit:

- Set all the load-transistors in strong inversion to minimise their transconductances (gm_{P1} and gm_{P3}).
- Set N2 and N3 in weak inversion to maximise their transconductances (gm_{N2} and gm_{N3}).
- The two precedent points imply the maximisation of $A_{boost} = gm_{N3}/gm_{P3}$ which also decreases the noise.

Moreover, one "simple" way of increasing the SNR is to increase the sensor bias V_K because we saw that for a single detected photoelectron: $I_{pp} = (V_K - V_{BD})/R_q$.

5.3.7. Biasing and sizing of the devices

The biasing current of the first branch is 2.5 mA and is copied from a 500- μ A reference. The boost amplifier (second branch) is automatically biased at 2 mA.

	N1	N2	N3	P1	P3	C0	C1	C2	NSW
$W[\mu m]$	300	1000	1000	500	500	50	100	50	200
$L[\mu m]$	2	0.5	0.5	1	0.5	30	30	30	0.18

Table 8: Sizing of the input-stage

The capacitors are implemented with Zero-VT NMOS transistors because they are more compact than Metal-Insulator-Metal capacitors. Moreover, as they are already in inversion for $V_{GS} = 0$ V, the oxide capacitance is constant for the entire dynamic. These capacitors are connected to the output node via NMOS switches called NSW.

5.4. Description of the output stage - Zero/Pole/Cancellation

The output stage has two main functions:

- Implementing a Zero/Pole/Cancellation
- Presenting a differential output voltage

I found a very compact implementation in the book [11]. It looks like a differential pair but the input transistors have independent sources. Between these two nodes, an impedance can be inserted to create a zero in the transfer-function.

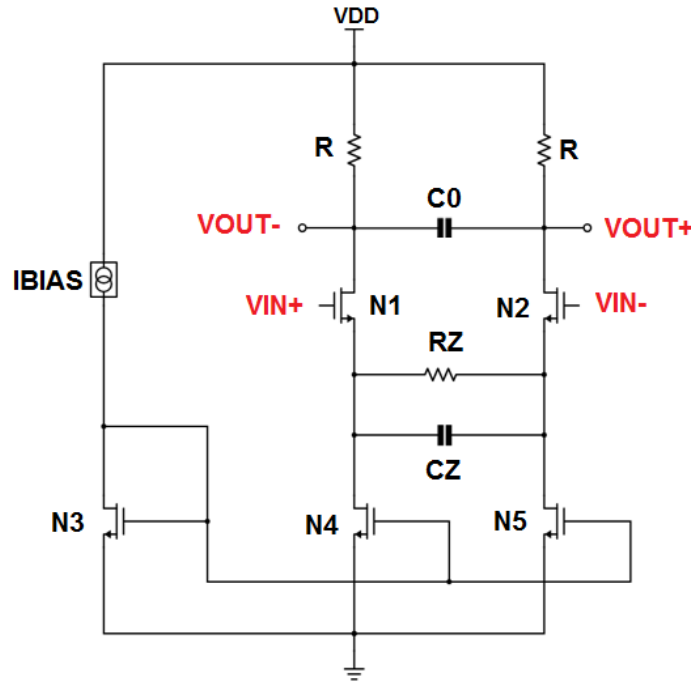


Figure 32: Schematic of the output stage implementing a Zero

I added the output capacitor C_0 to create another pole and thus to reduce the integrated output noise. The load capacitance C_L also contributes to this pole, so C_0 has to be adjusted depending on the input capacitance of the next stage.

5.4.1. Transfer function

Using the small-signals representation, we can derive the transfer function of this circuit:

$$\frac{V_{out+} - V_{out-}}{V_{in+}} = G(s) = \frac{2R}{R_Z} \cdot \frac{1 + R_Z C_Z s}{(1 + R(2C_0 + C_L)s)(1 + s C_Z / g_m)} \quad (17)$$

With:

- g_m the transconductance of the two input transistors N1 and N2.
- $R_Z C_Z$ the time constant defining the inserted zero.

The time constant $R_Z C_Z$ is set such that the zero cancels the sensor intrinsic pole:

$$R_Z C_Z = R_q(C_d + C_q)$$

As shown in (17), the low-frequency gain is defined by a ratio of resistances. In order to assure a reliable value, the 3 resistances have been implemented with the same resistor technology and laid-out in a common-centroid configuration to enhance the matching. This is why the resistances R have not been implemented using diode-connected transistors.

5.4.2. Noise analysis

Details of the derivation are given in the appendix B. The voltage input-referred power spectral density of noise is given by:

$$dV_{N_{eq}}^2 = \left| \frac{R_Z}{1 + R_Z C_Z s} \right|^2 \left(\frac{dI_{N1}^2 + dI_{N2}^2}{2} + \frac{2dI_{NR}^2 + dI_{NRZ}^2 + dI_{N4}^2 + dI_{N5}^2}{4} \right)$$

By considering only the thermal noise generated by the devices and that all the transistors are in strong inversion:

$$dV_{N_{eq}}^2 = 4kT \frac{R_Z^2}{1 + (R_Z C_Z \omega)^2} \left(\frac{gm_{N1} + gm_{N2}}{3} + \frac{1}{2R} + \frac{1}{4R_Z} + \frac{gm_{N4} + gm_{N5}}{6} \right) df$$

We can see that to minimise the noise generated by this stage we need high biasing currents. Indeed, for a fixed DC voltage drop ΔV across R , increasing the biasing current decreases R . Moreover, for a fixed low-frequency gain $2R/R_Z$ we also need to decrease R_Z . The transconductances can scale linearly with the biasing current (weak inversion), but the overall noise expression is still inversely proportional to the biasing current I :

$$dV_{N_{eq}}^2 \sim \left(\frac{\Delta V}{I} \right)^2 \frac{I}{U_T} df$$

The drawbacks of increasing I are of course the power consumption but also the big capacitance C_Z to keep $R_Z C_Z$ constant.

5.4.3. Biasing and sizing of the devices

Each branch of the circuit is biased with 1 mA copied from a 500- μ A reference.

	N1	N2	N4	N5	R	RZ	CZ	C0
$W[\mu m]$	640	640	200	200	20	20	120	20
$L[\mu m]$	0.25	0.25	1	1	39	39	50	12

Table 9: Sizing of the output stage

The resistors are implemented with P-doped poly-resistors and the capacitors with Zero-VT NMOS transistors. I chose the maximum resistors width to minimise self-heating and the impact of process variations.

5.5. Biasing circuits

To bias the circuits, we need at least one current reference. Moreover, we need a voltage reference for the differential amplifier (signal on V_{in+} pin and reference on V_{in-} pin). I used Beta-multiplier circuits for the two references for good power supply noise rejection and to have a similar temperature dependence as explained in the next paragraphs.

5.5.1. Current reference

I used a Beta-multiplier circuit: $(W/L)_{N1} = k (W/L)_{N2}$, implementing a 500- μ A current reference for the entire front-end:

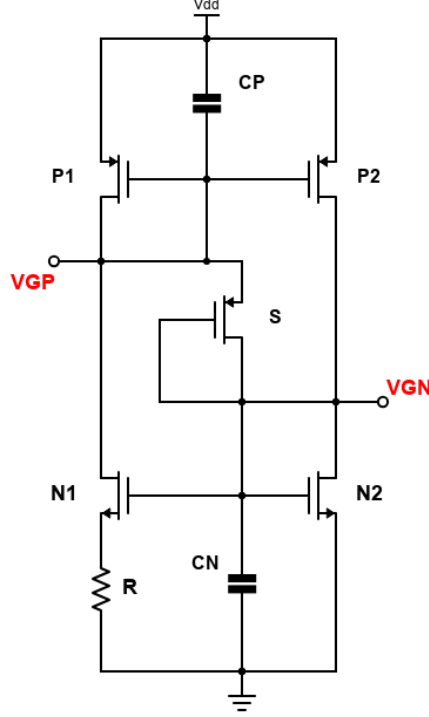


Figure 33: Schematic of the Beta-multiplier - current reference

The current equations are:

$$I_{ref} = \frac{2}{n \beta_{N2} R^2} (1 - 1/\sqrt{k})^2 \quad \text{or} \quad I_{ref} = 0$$

To avoid the case $I_{ref} = 0$, a start-up transistor S is inserted to inject current through P1 and N2 after turning-on the power supply (transient). In steady-state, S does not conduct because we have the condition:

$$V_{GP} - V_{GN} < V_T(S)$$

For $(W/L)_{N2} = 100/1$, $k = 4$ and $I_{ref} = 500 \mu\text{A}$, we get $R = 192 \Omega$. The resistor is implemented with a N-doped poly-resistor. Big capacitors (Zero-VT NMOS) are connected to the reference voltage nodes V_{GN} and V_{GP} to filter-out noise.

$k = 4$	N1	N2	P1	P2	S	R	CN	CP
$W [\mu m]$	400	100	100	100	100	20	300	300
$L [\mu m]$	1	1	1	1	1	33.6	30	30

Table 10: Sizing of the current reference circuit

5.5.2. Voltage reference for the differential amplifier

In order to have a similar temperature dependence as the current reference circuit, I chose a PMOS-Beta-multiplier for the voltage reference.

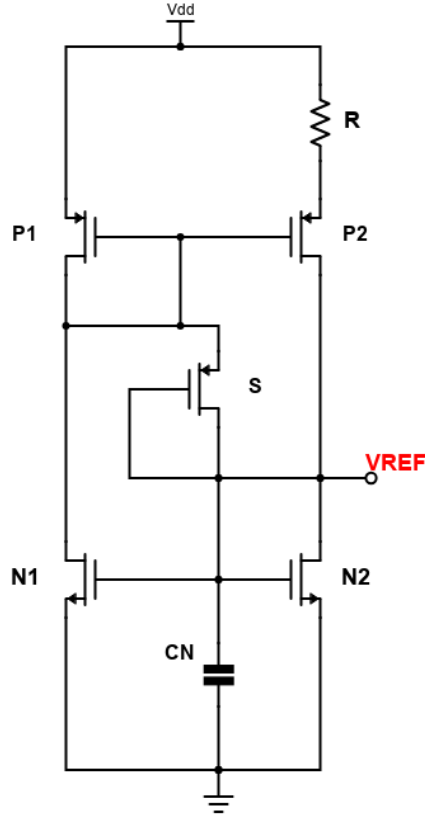


Figure 34: Schematic of the Beta-multiplier - voltage reference

The transistors are in strong inversion so we can write:

$$V_{REF} = \sqrt{\frac{2I_2}{\beta_{N2}}} + V_{TN2} \quad (18)$$

The transistors sizes and the resistance value have been chosen such that:

- $V_{REF} = V_{out}(input\ stage)$

- $\frac{dV_{REF}}{dT} \approx \frac{dV_{out}(input\ stage)}{dT}$

The last relation is detailed in the next section.

$k = 4$	N1	N2	P1	P2	S	R	CN
$W [\mu m]$	50	50	100	400	100	4	300
$L [\mu m]$	8	8	8	8	1	72.8	30

Table 11: Sizing of the voltage reference circuit

5.5.3. Temperature dependence of the biasing circuits

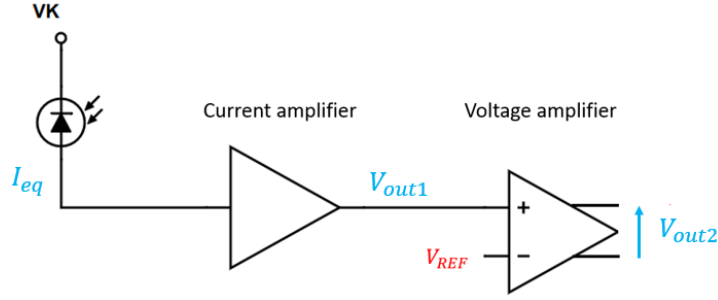


Figure 35: Front-end overview

Ideally, we would like that the DC output of the front-end does not vary with temperature to not have any NSB error (NSB: section 5.2.3):

$$\forall T \quad V_{OUT2} = 0 \quad \Longleftrightarrow \quad \frac{dV_+}{dT} = \frac{dV_{REF}}{dT} \quad \& \quad V_+(T_0) = V_{REF}(T_0) \quad (19)$$

V_+ is the output voltage of the input stage and is the gate voltage of N1 (please refer to the [input stage schematic](#)). Let's call I_1 the input stage current flowing through N1 for which in strong inversion:

$$V_+ = \sqrt{\frac{2I_1}{\beta_{N1}}} + V_{TN1}$$

By taking V_+ and V_{REF} (18) derivatives:

$$\begin{aligned} \frac{dV_+}{dT} &= (TCI_1 - TC\beta_{N1}) \left(\frac{I_1}{2\beta_{N1}} \right)^{\frac{1}{2}} + \frac{dV_{TN1}}{dT} \\ \frac{dV_{REF}}{dT} &= (TCI_2 - TC\beta_{N2}) \left(\frac{I_2}{2\beta_{N2}} \right)^{\frac{1}{2}} + \frac{dV_{TN2}}{dT} \end{aligned}$$

With the temperature coefficient $TCX := \frac{1}{X} \frac{dX}{dT}$.

I_1 is the current copied from the current reference circuit and I_2 is the current flowing in the voltage reference circuit. Moreover, We have:

$$\frac{dV_{TN1}}{dT} \approx \frac{dV_{TN2}}{dT} \quad \& \quad TC\beta_{N1} \approx TC\beta_{N2} \quad \& \quad \left(\frac{I_1}{2\beta_{N1}} \right)^{\frac{1}{2}} < \left(\frac{I_2}{2\beta_{N2}} \right)^{\frac{1}{2}}$$

Finally, the equality (19) has been obtained by using a PMOS-Beta-multiplier for the voltage reference circuit.

Indeed, I found by simulation that the mobility temperature coefficient differs between holes and electrons:

$$|TC\mu_p| < |TC\mu_n| \quad \Rightarrow \quad TCI_2 < TCI_1$$

The following plot shows the evolution of V_{REF} , V_+ and the anode voltage with temperature:

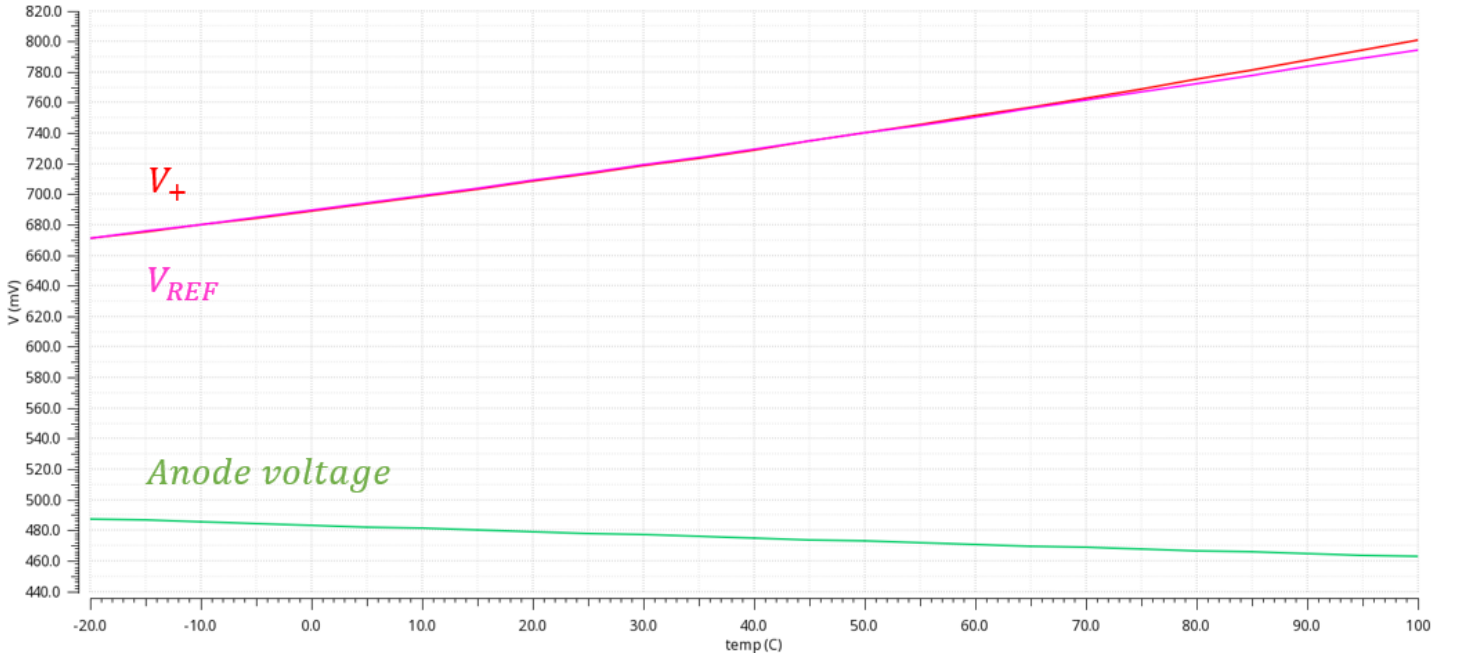


Figure 36: Front-end temperature dependence - post-layout simulation

Because the final design consumes only 14 mW, the junction temperature in the semiconductor will be close to the ambient temperature. Let's take the example of a typical thermal resistance of a package between the junction and the air $R_{\theta-JA} = 100 \text{ }^\circ\text{C}/\text{W}$ (typical 8-pins package): then the temperature difference between the semiconductor and the atmosphere is $0.014 \cdot 100 = 1.4 \text{ }^\circ\text{C}$.

We can see on this plot that $V_{OUT} = V_+ - V_{REF} \approx 0$ for a wide ambient temperature

range. Therefore there will be a negligible error on the NSB computation regarding temperature variations.

6. Physical implementation of the front-end and post-layout simulations

6.1. Layouts of the circuits

For every layout, I took care of respecting the following rules:

- Important currents are involved $\Rightarrow$ large metal widths to avoid aluminium electro-migration ($1\text{ mA} \Rightarrow W > 2\text{ }\mu\text{m}$).
- The devices that need to be matched have been implemented using common-centroid structures + dummies transistors on the sides.
- The transistors fingers length is limited to $1\text{ }\mu\text{m}$ to decrease the poly-gate resistance and thus its noise contribution.
- The largest areas for the resistors have been chosen to minimise the die-to-die process variations.

With all the blocks together, the full layout area of the front-end is $330 \cdot 450\text{ }\mu\text{m}^2$.

6.1.1. Layout of the input stage

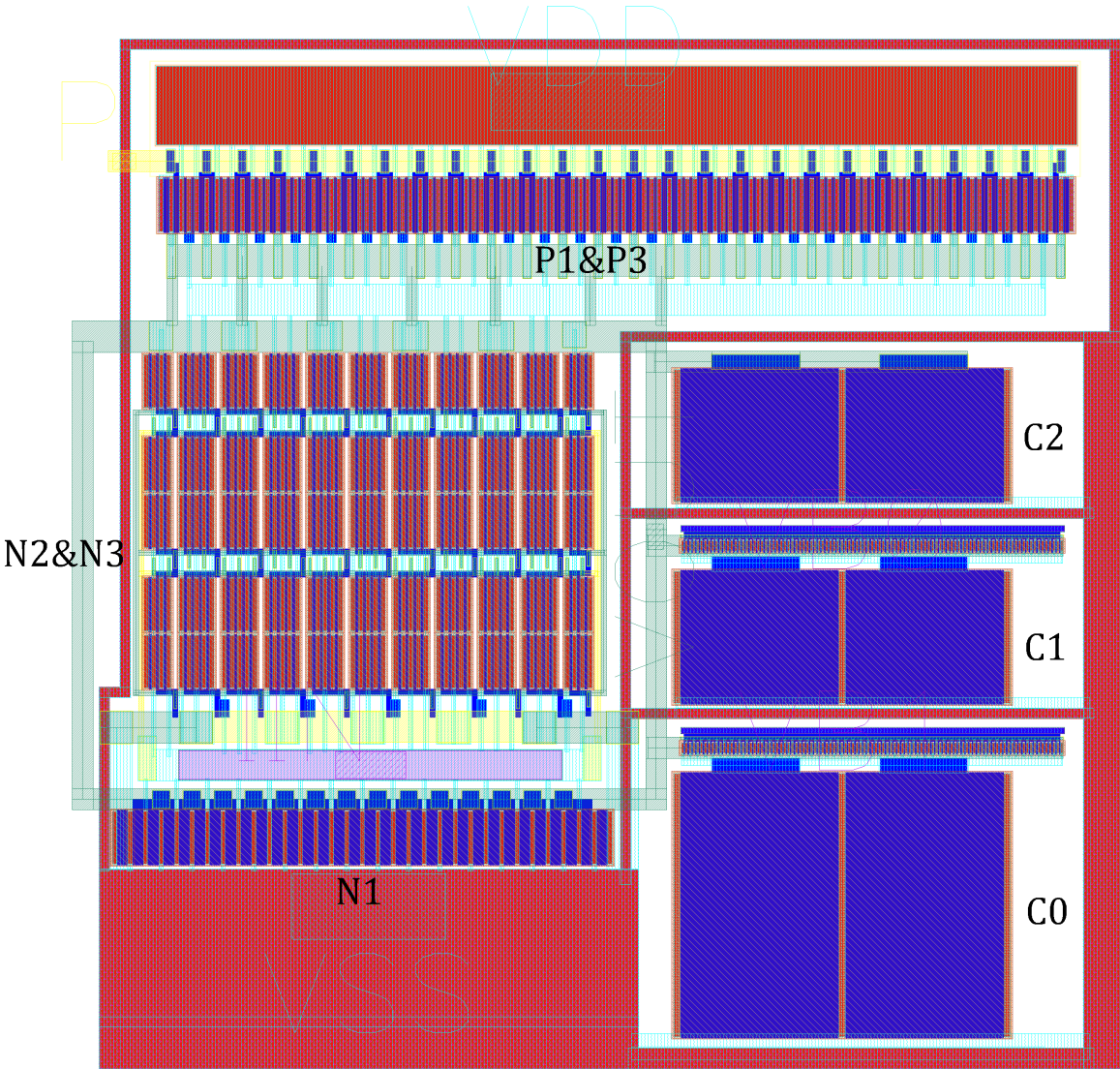


Figure 37: Layout of the input stage

6.1.2. Layout of the differential amplifier

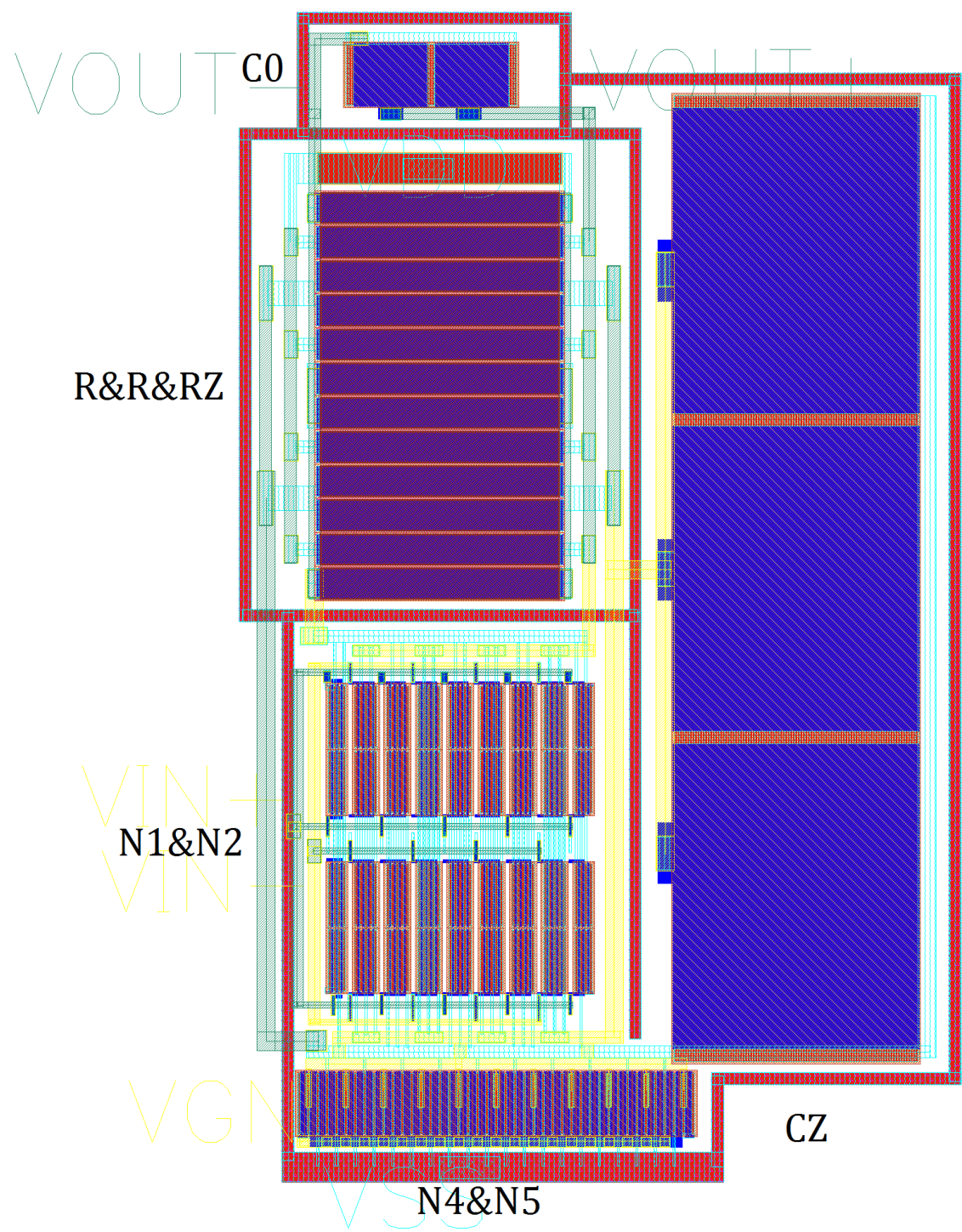


Figure 38: Layout of the differential amplifier

6.1.3. Layout of the biasing circuits

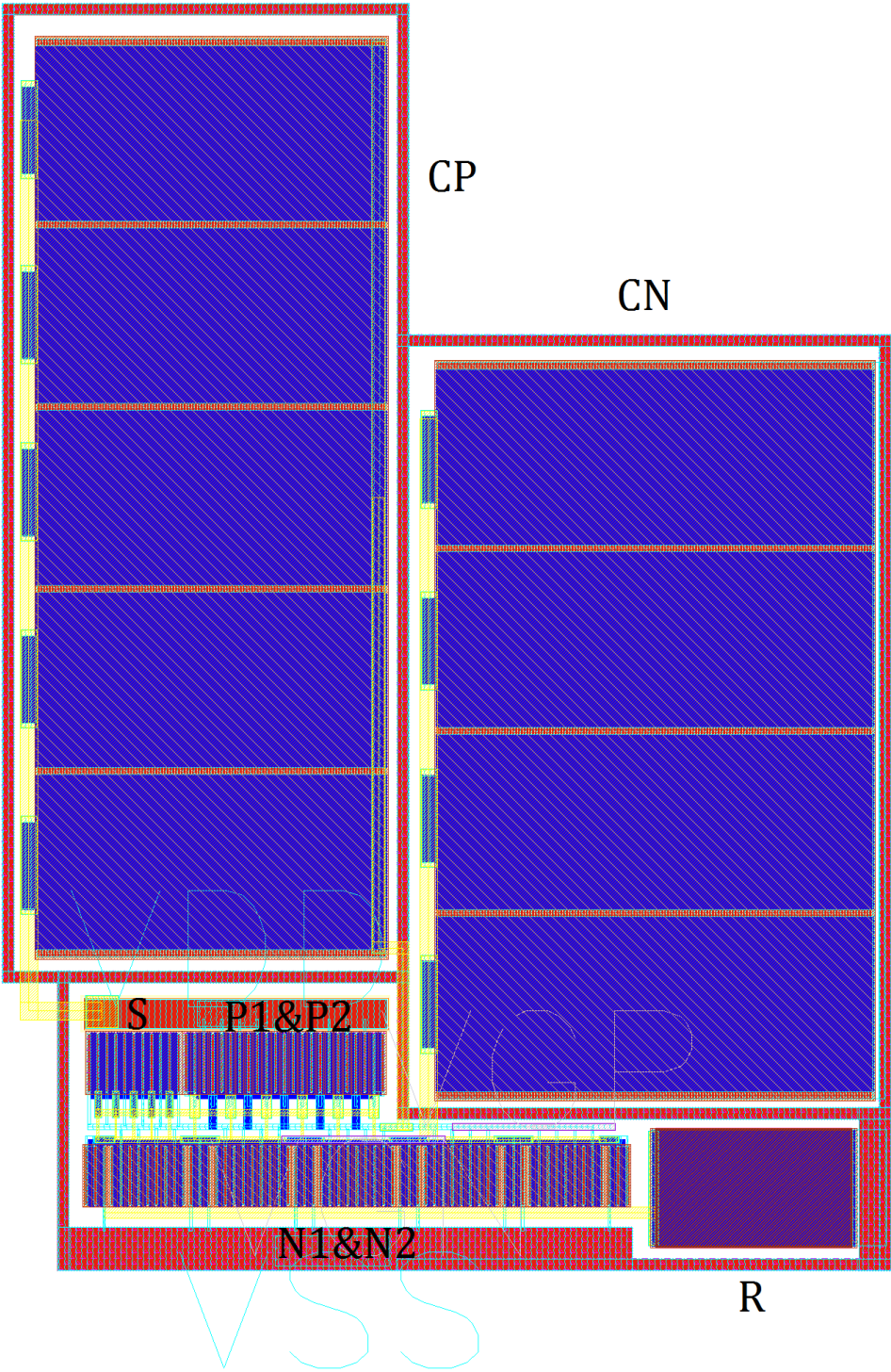


Figure 39: Layout of the current reference

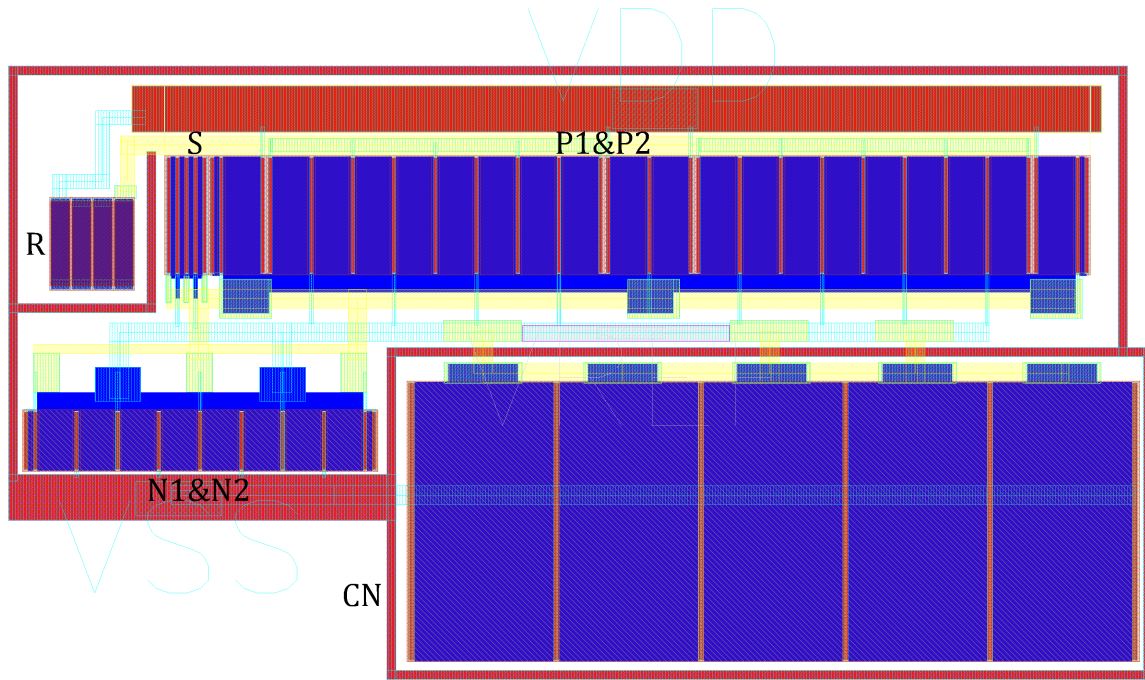


Figure 40: Layout of the voltage reference

6.2. Post-layout results

The following results are obtained using the post-layout RC-extracted netlist.

6.2.1. Front-end signals

The next plot shows the output voltage for one p.e and for the 4 different versions that we can select with two bits. The pulse shape optimisation has been made for the low-gain case (blue curve).

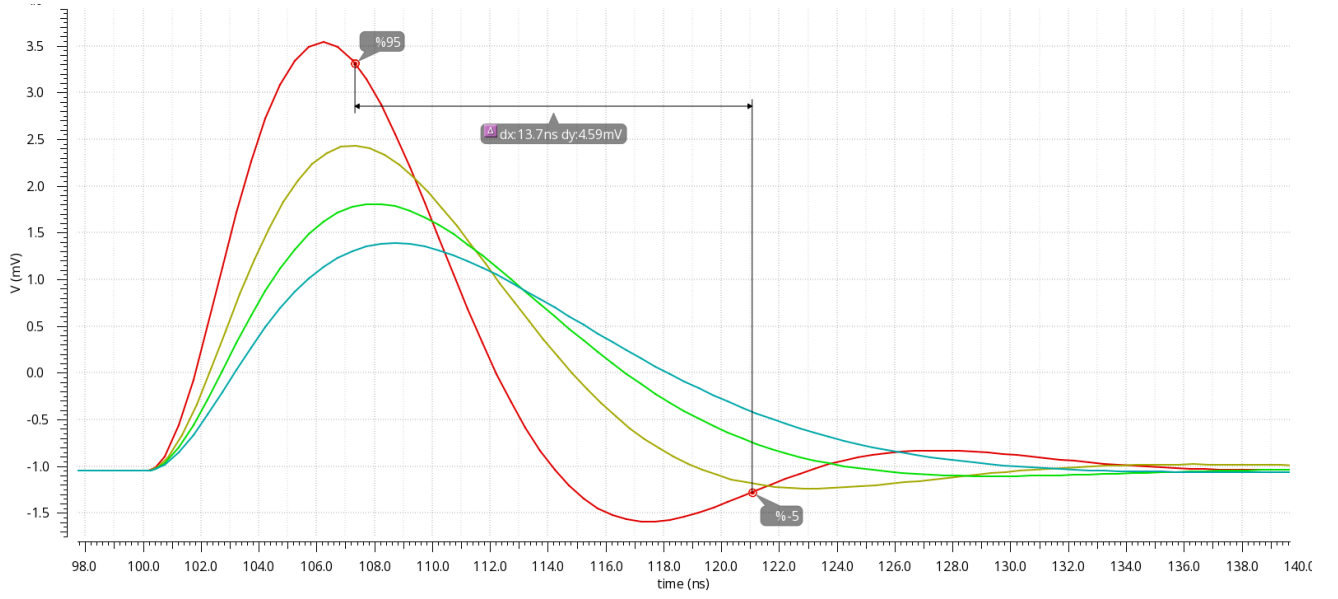


Figure 41: Output voltage for 1 photoelectron and for the 4 versions

If we sweep the number of detected p.e at the same time (from 1 to 1000), we obtain the following waveforms:

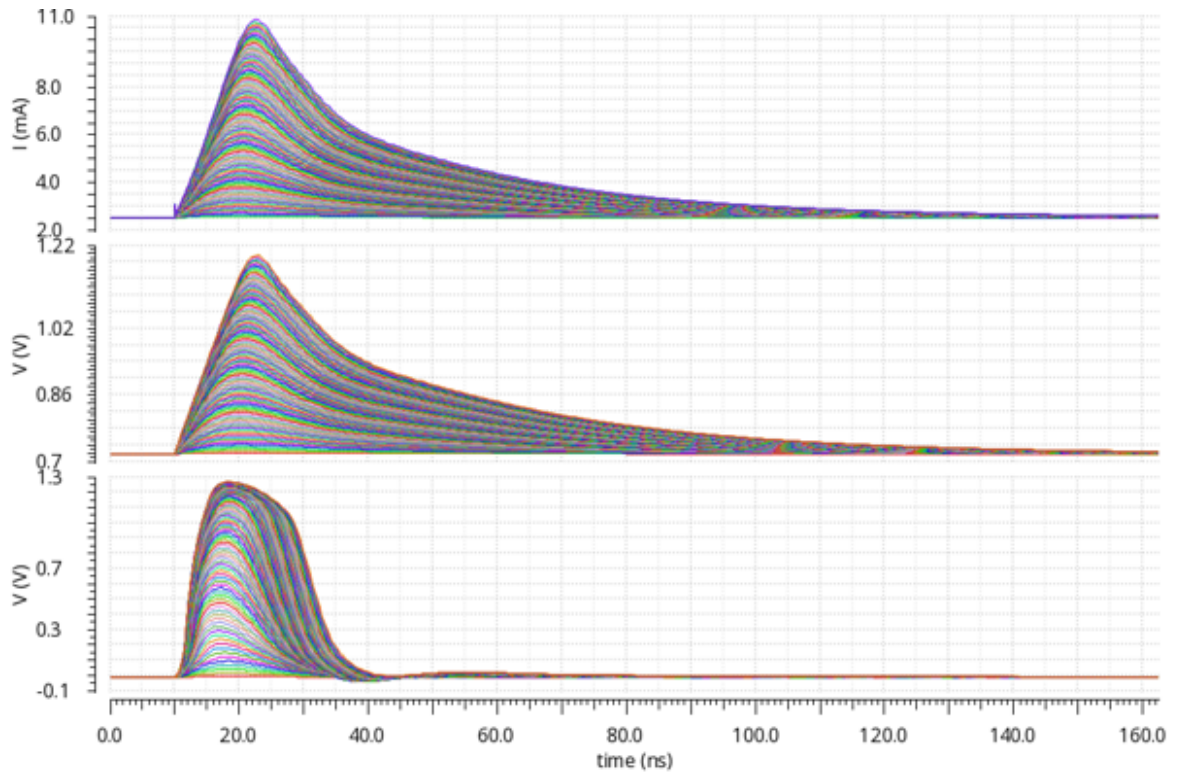


Figure 42: Input current (top), output voltage of the first stage (middle) and front-end output voltage (bottom) - Low gain case (V4)

6.2.2. Gain, FWHM and SNR

The gain, the Full Width at Half Maximum of the output pulse and the SNR are computed for one detected photoelectron.

To analyse the pixel-to-pixel variations, Monte Carlo simulations for each one of the 4 versions have been made using 50 samples. I used two types of simulations:

- Process variations: for each new MC sample, a random variation of every technology-dependent parameters is applied (same parameters for all the devices within a die).
- Local Mismatch simulation & Process variations: In addition to the process variations, other random variations for every parameters are applied between all the devices in the die.

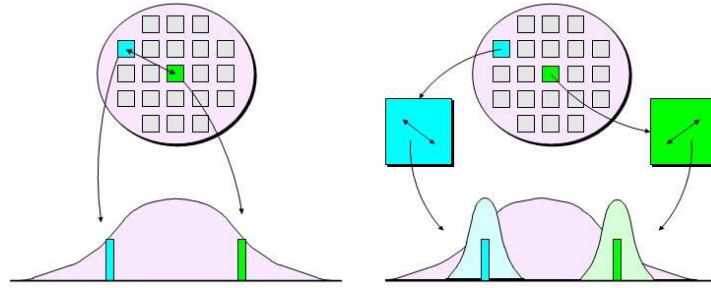


Figure 43: Representation of MC process variations (left) and MC process & mismatch variations (right). Credit: <https://www.eetimes.com/overcoming-32-28-nm-ic-implementation-challenges/>

The local mismatch simulation does not take into account how the circuit has been laid-out and applies variations between every transistor fingers and devices. It corresponds then to a "worst layout" scenario. In the following results, the combination of Process & Mismatch MC variations were applied. The most sensible parameters regarding the performances variations of the front-end are the resistances which have a relative standard deviation around 8% in this technology. Indeed, resistors are used in both biasing circuits (described in the section 5.5) and determine the amount of current biasing each stage.

Unfortunately, I could not perform Monte Carlo variations for the resistors but only corner simulations. Therefore, I simulated the transistors Process & Mismatch variations for the three resistor corner cases: minimum resistance, nominal resistance and maximum resistance.

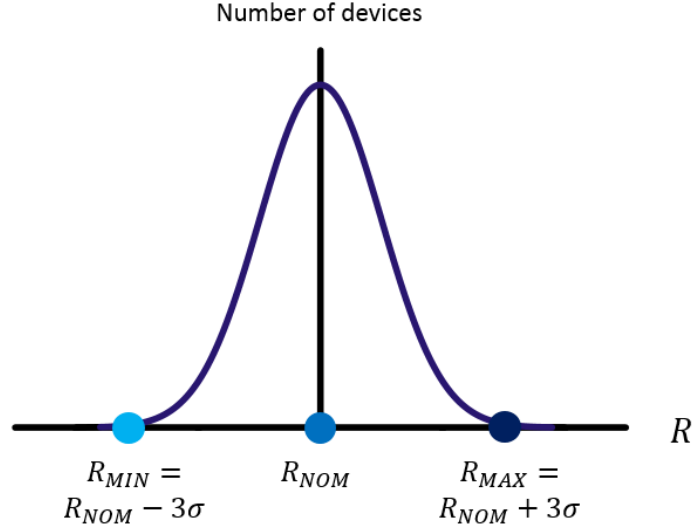


Figure 44: Resistance corner values given by the foundry. They are defined as a $\pm 3\sigma$ deviation from the nominal value

The next plot shows the gain versus the FWHM for the 4 versions and for the three resistance corner cases:

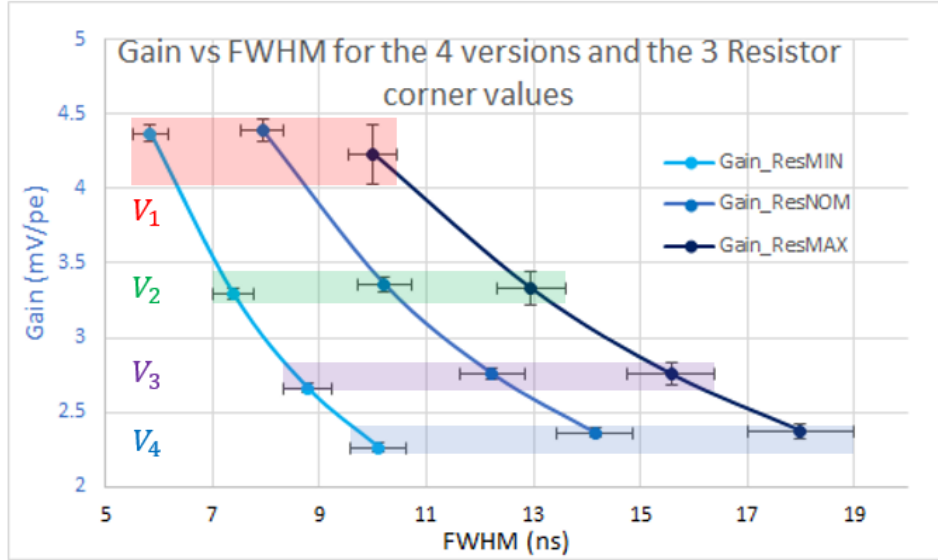


Figure 45: Simulation results of MC process & mismatch variations for the three resistance corner cases and for the 4 versions

The error bars correspond to 3 standard deviations of the resulting transistors Process & Mismatch simulations. The colored rectangles represent for each version the intervals

where at least 99.7 % of the front-ends lie in. We see that the resistance variations have an important impact on the pulse FWHM. Indeed, the biasing currents directly determine the front-end bandwidth. However, the gain variations are quite low for each one of the 4 versions.

After noticing that the results for the minimum & maximum resistance corners are equally distributed around the nominal resistance corner result, I approximated the results distribution with a Gaussian distribution to estimate the standard deviation.

σ/μ (%)	V1	V2	V3	V4
FWHM	8.7	9.1	9.3	9.3
Gain	0.15	0.69	1.2	1.4
SNR	5.0	4.4	4.0	3.8

Table 12: Relative standard deviation estimation - Transistors & Resistors Process & Mismatch variations

Finally, the gain, SNR versus FWHM for the 4 versions is presented in the next plot. The error bars correspond to one standard deviation.

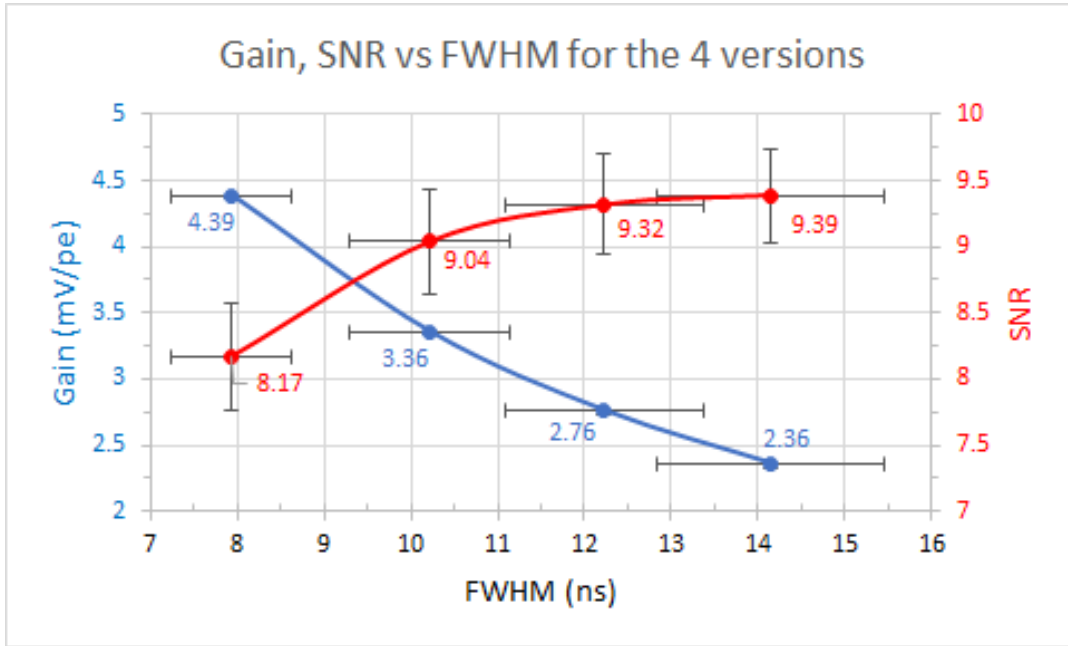


Figure 46: Simulation results of MC process & mismatch variations for the 4 versions - Standard deviation estimated thanks to the 3 resistance corner cases

6.2.3. Linearity and Dynamic range

The following plot shows the peak-to-peak amplitude of the output pulse versus the number of detected photoelectrons for the 4 versions, associated with the linear deviation. The errors bars correspond to the maximum possible errors calculated with the corner resistances cases R_{max} and R_{min} . Indeed, as shown before, the pixel-to-pixel variations due to the transistors process & mismatch variations are negligible but are rather dominated by the resistance variations.

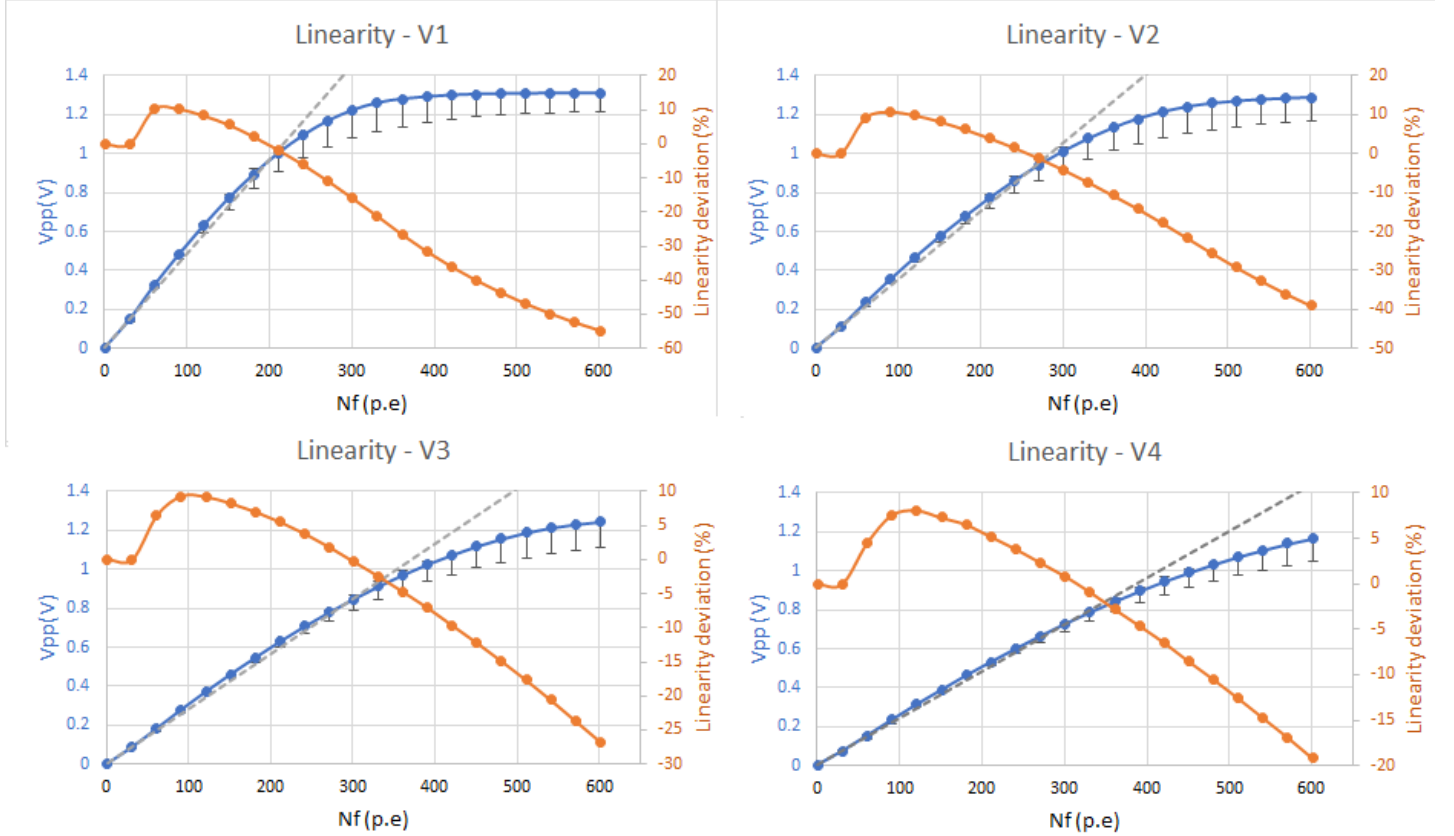


Figure 47: Front-end linearity for the 4 configurations - The error bars correspond to the maximum error

The best linearity the front-end can achieve is in the Low Gain mode (V4): if we consider a deviation from linearity smaller than 8%, then the dynamic range is 460 p.e. However, if we consider a smaller deviation from linearity, the front-end is quite non linear right after the second point of the plot. This non linearity is mainly caused by the Zero/Pole/Cancellation function in the second stage as explained in the section 7.

Nevertheless, the non-linearity is not detrimental for our application. Indeed, in their precedent design for the SST-1M camera [7], the scientists used the charge information

by integrating the output pulse. The output pulse integral is shown in the next plot with its one-standard deviation error bars. Because the die-to-die FWHM variations are quite important ([Relative standard deviation](#) around 9 %), those variations impact the integral variations.

The curve can easily be fitted with polynomial functions. For example, in the next plot, two independent fitting functions are used for the [0;300] p.e and the [300;2000] p.e ranges with respectively a 3rd order and a 2nd order polynomial function. The quantity $(data - fit)/\sigma$ is also shown in grey to compare the fitting residuals with the die-to-die variations and its modulus is always smaller than one.

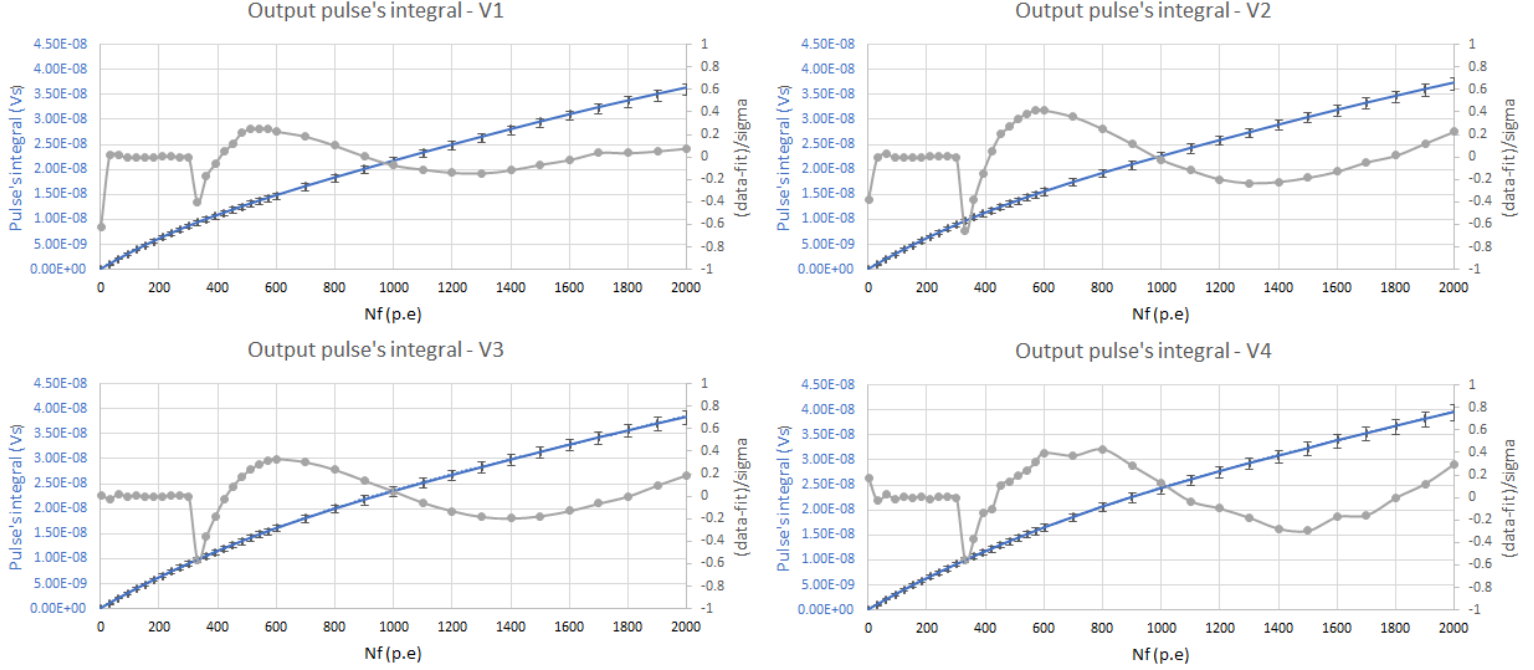


Figure 48: Output pulse integral for the 4 versions - The error bars correspond to one standard deviation

Beyond 2000 photoelectrons, the charge is quasi linearly-dependent.

6.2.4. Gain and FWHM variations with temperature

As shown in the paragraph about the DC output shift (section 5.5.3), the temperature has almost no influence on the DC output voltage in the $[-20; 100]^\circ\text{C}$ range and therefore on the NSB information. Let's now look at the gain and FWHM variations with temperature:

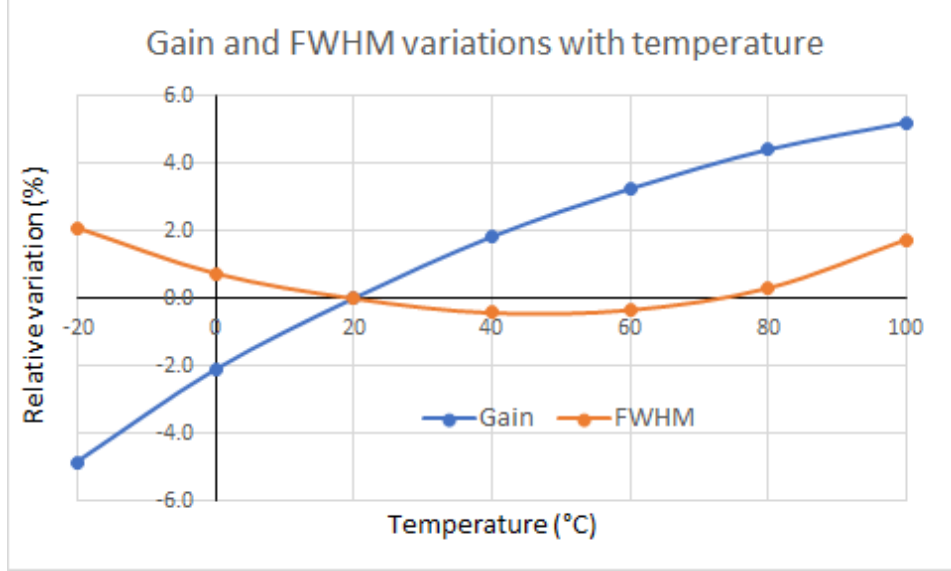


Figure 49: Gain and FWHM variations with temperature - Low gain case (V4)

6.2.5. Gain and FWHM variations with I_{NSB}

Any DC current from the sensor will be sunk by the transistor N1 from the input stage. This augmentation of the bias current I_{bias} of N1 will increase its transconductance gm_1 . N1 is in strong inversion and by taking the derivative dgm_1/dI_{bias} :

$$\Delta gm_1 \approx \sqrt{\frac{\beta_1}{2I_{bias}}} I_{NSB}$$

Moreover, two poles are proportional to gm_1 (16), so the main NSB effect will be an increase of the bandwidth, i.e a decrease of the FWHM.

In addition, we could expect a decrease of the gain because it is inversely proportional to gm_1 (16) but because the bandwidth increases, the output signal holds more power at high frequency and finally the gain is quite stable regarding I_{NSB} . These results are summarised in the following plot for the Low gain version:

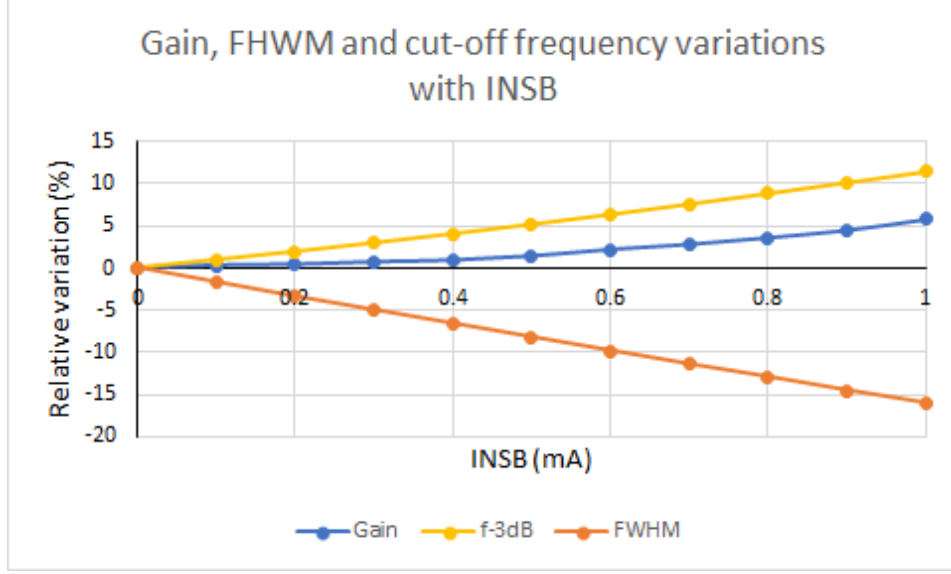


Figure 50: Gain and FWHM variations with I_{NSB} - Low gain case (V4)

We saw in the NSB section 5.2.3 that we expect a maximum NSB current of 0.5 mA (1.6 GHz). For that value, the FWHM sees a variation of -8%, which is not negligible and could induce an important error on the charge extraction. However, because the front-end is DC-coupled, we could apply a correction knowing I_{NSB} and the linear FWHM variation with I_{NSB} .

The next plot shows the front-end gain in the frequency domain for $0 \leq I_{NSB} \leq 1$ mA:

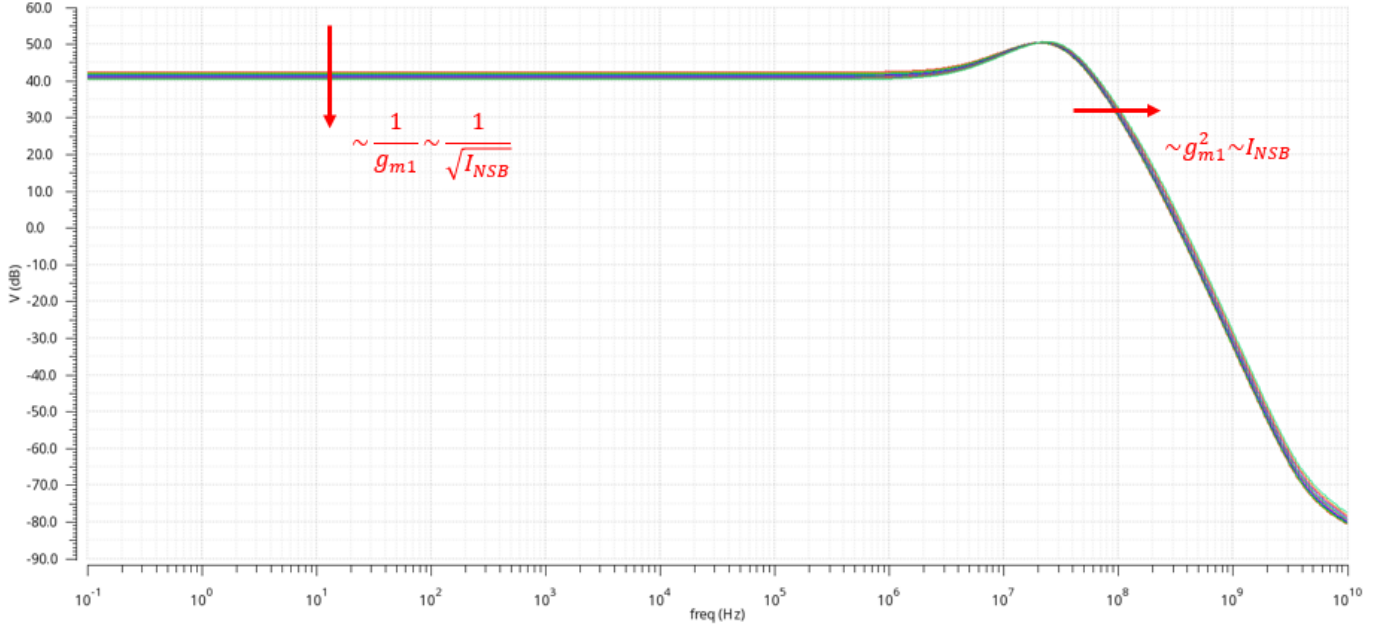


Figure 51: Front-end gain in frequency domain for $0 \leq I_{NSB} \leq 1$ mA

7. Limitations and possible improvements

7.1. Linearity

As seen before in the section 6.2.3 about linearity, the peak-to-peak amplitude deviates from linearity up to 8% around 100 p.e in the low gain version (V4). This non-linearity essentially comes from the Zero/Pole/Cancellation function, aimed to speed-up the front-end response. The following plot shows on the left the front-end with ZPC and on the right the same front-end with $C_Z = 0$ and thus without ZPC.

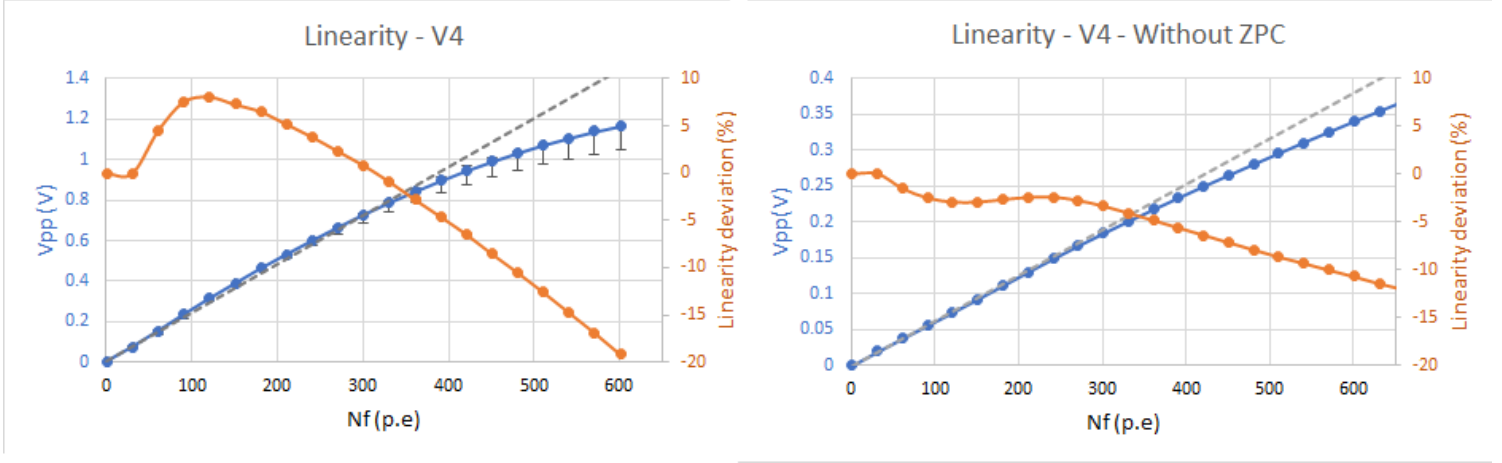


Figure 52: Front-end linearity with ZPC (left) and without ZPC (right) - V4

We clearly see that the ZPC is responsible for the systematic linearity deviation at the beginning of the dynamic, with a peak around 100 p.e. More investigations could be done concerning the ZPC implementation, in order to improve the linearity while keeping a high speed response.

7.2. New biasing circuits to decrease the FWHM and Charge variations

We saw in the post-layout results (table 12) that the FWHM die-to-die variations reach 9.3% (thus inducing variations of the pulse integral, i.e the charge), due to the important resistors variations in this technology. One way of reducing those variations is to use resistor-less biasing circuits. I studied the circuit from [13] and used pairs of transistors in weak inversion to generate a voltage difference between the two NMOS gates of a Beta-Multiplier instead of using a resistor.

7.2.1. Resistor-less Beta-Multiplier - current reference

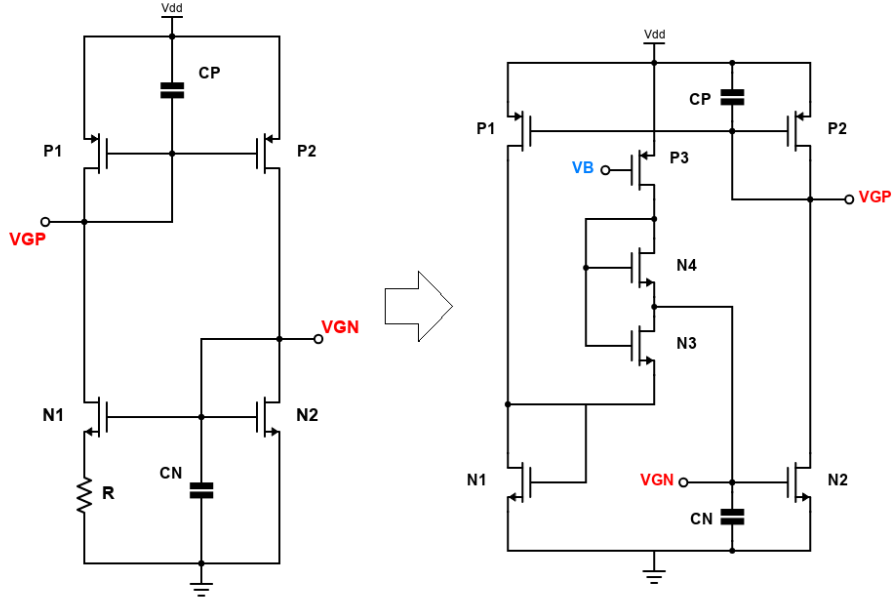


Figure 53: Upgrade of a classic Beta-Multiplier (left) to a Resistor-less Beta-multiplier (right) [13] - Startup circuits are not shown

The difference between N1 and N2 overdrive voltages $V_{GS2} - V_{GS1}$ determines the current flowing in any Beta-Multiplier circuit. The classical circuit uses a resistor to create this voltage reference but it is also possible to use a pair of transistors in weak inversion (N3 and N4 in the right schematic). Let's write the expression of this voltage difference V_d . N3 and N4 are in weak inversion and in saturation (bulk effect neglected: $n = 1$), biased by I_B and with $S = (W/L)_{N4}/(W/L)_{N3} > 1$:

$$I_B = I_S \exp\left(\frac{V_{GS3} - V_T}{U_T}\right) = S \cdot I_S \exp\left(\frac{V_{GS4} - V_T}{U_T}\right)$$

$$\Rightarrow V_d = V_{GS3} - V_{GS4} = U_T \cdot \ln(S)$$

In theory, V_d does not depend on the biasing current I_B of N3 and N4. So it does not matter if the circuit implementing the reference I_B has important variations (I used another classic Beta-Multiplier to generate a multiple of $I_B \approx 50 \text{ nA}$).

I implemented this upgrade by keeping the initial transistors dimensions and with $V_d = 100 \text{ mV}$ ($\approx R \cdot I_{REF}$).

	N1	N2	N3	N4	P1	P2	P3	CN	CP
$W [\mu m]$	400	100	10	590	100	100	0.5	300	300
$L [\mu m]$	1	1	0.5	0.5	1	1	1	30	30

Table 13: Resistor-less Beta-Multiplier current reference sizing

7.2.2. Resistor-less Beta-Multiplier - voltage reference V_{REF}

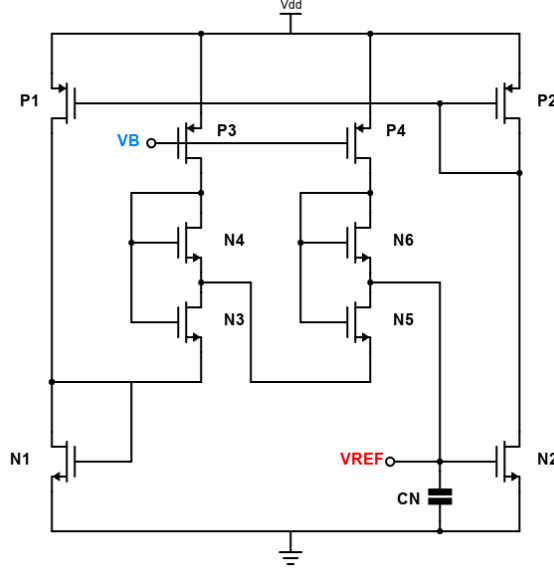


Figure 54: Resistor-less Beta-multiplier used as a voltage reference - Startup circuit is not shown

As for the "classic" version with a resistor, the goal of this circuit is to generate a stable 0.715 V reference for the differential amplifier. It is possible to use several pairs of transistors in W.I in series to increase V_d , as I did here with two cells.

	N1	N2	N3,N5	N4,N6	P1	P2	P3	P4	CN
$W [\mu m]$	40	10	22.5	800	100	100	0.5	0.5	300
$L [\mu m]$	3	3	0.5	0.5	1	1	1	1	30

Table 14: Resistor-less Beta-Multiplier voltage reference sizing

7.2.3. Temperature dependence of the two new biasing circuits

As we saw in the section 5.5.3, the output DC shift of the first stage and the reference voltage shift with temperature must be equals in order to not have any NSB errors with temperature. The advantage of the Resistor-less Beta-Multiplier circuit is that

the voltage V_d implemented by the pair of transistors in WI is a PTAT voltage source (positive temperature coefficient). It can decrease the effect of the carriers mobility variations with temperature (negative temperature coefficient). Therefore, the overall DC shift with temperature is reduced, as shown in the next plot. The red curve is the output voltage of the first stage and the pink is the reference voltage needed in the differential amplifier (second stage).

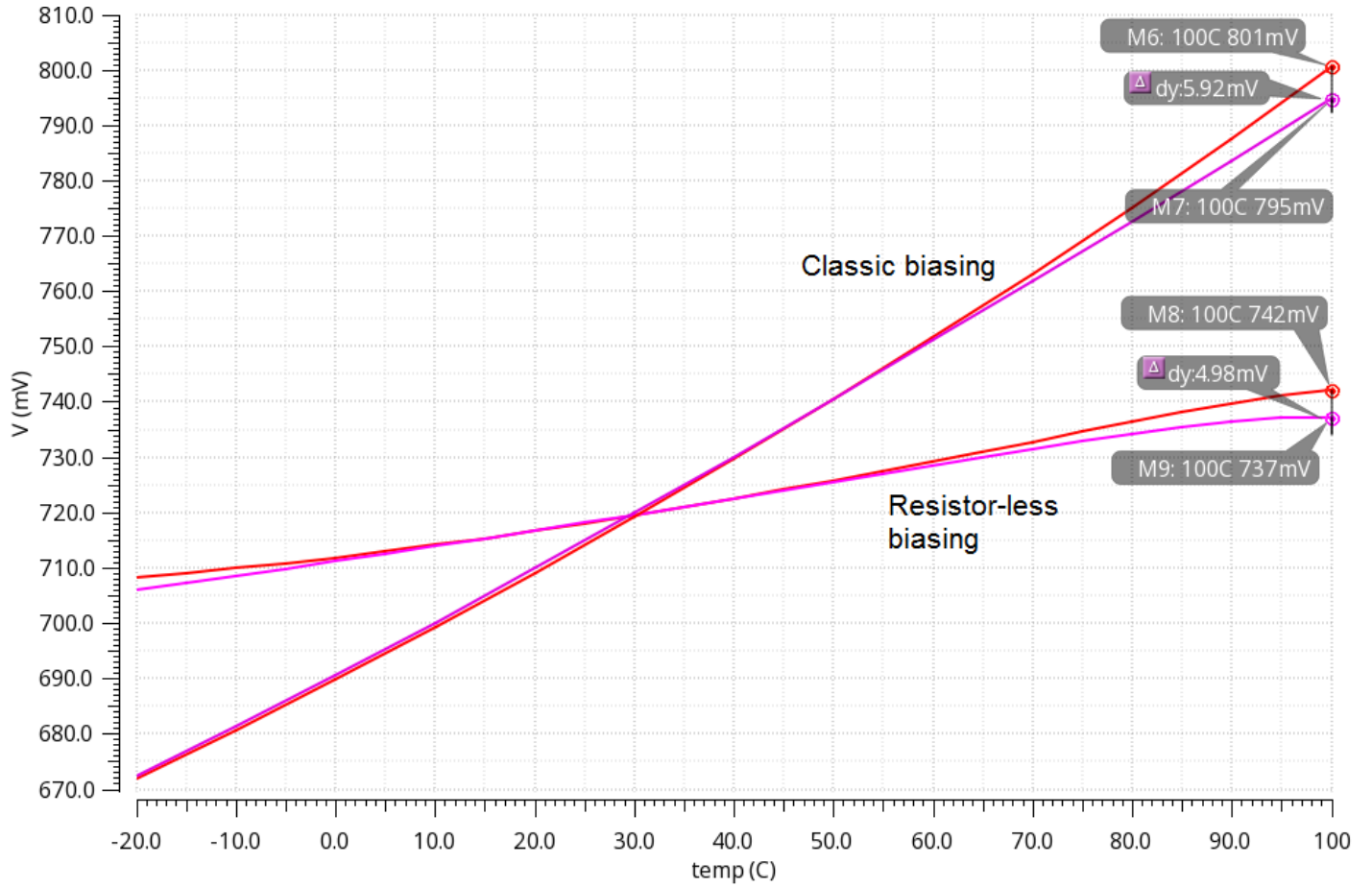


Figure 55: DC voltages shift with temperature. Classic and Resistor-less biasing are compared

At 100 °C and for the classic biasing case, the 6 mV DC error induces approximately an NSB error of 160 MHz.

7.2.4. FWHM, Gain and SNR variations with the Resistor-less biasing circuits

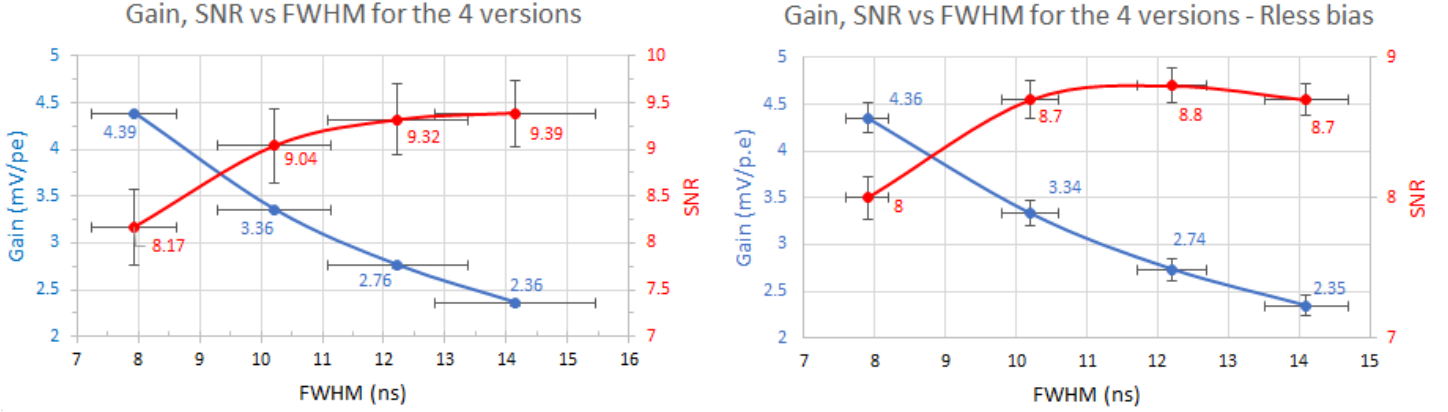


Figure 56: Gain & SNR vs FWHM - classic biasing (left) and resistor-less biasing (right).
Error bar = one standard deviation

We see that using resistor-less biasing circuits reduces the FWHM variations thanks to the smaller variations of the biasing currents (FWHM RSD of 9% reduced to 4%). In this latter case, the resulting variations are due to the MOS transistors process & mismatch variations. The gain sees more variations due to the remaining resistor R_Z variations which implements the zero in the second stage. Finally, the SNR is a bit worse in the second case because of the flicker noise produced by N3 and N6 in the new biasing circuits. This problem can be solved by choosing bigger transistors.

8. Conclusions

8.1. Pros and Cons of the proposed front-end

Advantages:

- Low power (14 mW/pixel=104 W for the entire camera)
- High SNR (up to 9.4 @ 1p.e)
- DC-coupled (NSB information)
- Robust to temperature variations

Drawbacks:

- Bad linearity (up to 8% deviation in the [0;450 pe] dynamic in Low gain mode). However, it is not detrimental for our application since we are interested by the charge information.
- Important die-to-die FWHM variations (RSD=9%). Nevertheless, by using Resistor-less biasing circuits, the FWHM RSD falls to 4%.

8.2. Comparison with the state-of-the-art front-ends

	Power/pixel [mW]	Total power (7420 pixels) [kW]	Gain [mV/pe]	SNR @ 1 p.e	FWHM [ns]	Dynamic Range [p.e]
SST-1M front-end	400	2.97	2.4	10 (@ 1 GHz)	12	< 600
INFN front-end	800	5.94	2.0	4.3 (@ 1 GHz)	2.5	< 400
MUSIC ASIC	100	0.742	[0.63 ; 1.25]	[2.25 ; 2.9] (@ 20 MHz)	[12 ; 20]	[500;1000]
This work	14	0.104	[2.4 ; 4.4]	[8.2 ; 9.4] (@ 10 GHz)	[7.9 ; 14.1]	<450

Table 15: Performances comparison between the front-ends candidates for the LST camera update and this work

My first design priority was to have a low power consumption which is achieved by using the minimum number of circuits dedicated to the LCT5 sensor. The front-end processes the 4 channels of the sensor and consumes 7 times less power than the ASIC MUSIC.

Next, I have focused on a high-speed response with a sufficient SNR as required in the specification list (section 1.4). However, I could not reach both SNR and speed specifications. It is indeed a difficult trade-off because a small speed enhancement induces inevitably a larger bandwidth and thus more integrated thermal noise. However, these two characteristics have been improved regarding MUSIC and are still comparable with the ones from the SST-1M front-end.

Moreover, I tried to achieve a good linear response for low-light levels (before saturation) but the Zero/Pole/cancellation circuit degrades it, highlighting another trade-off between linearity and speed. Nevertheless, the front-end peak-to-peak response is easily fitted with a polynomial function, as well as the pulse integral which gives the charge information when the front-end saturates (this method has been successfully used in the SST-1M camera [6, 7]).

Finally, the Monte Carlo simulations show that the die-to-die variations are reasonable (FWHM RSD=4% and Gain RSD=4.5% using Resistor-less biasing circuits), leading to low pixel-to-pixel variations and therefore a good camera precision.

The next steps would be to produce a test chip and compare the measures with simulations to validate the proposed design. The test chip could include several front-end configurations with different circuits parameters such as for example, various load capacitance for the second stage to select the fastest one while ensuring a single photon detection capability.

A. Derivation of the SiPM equivalent circuit

This section presents the derivations using the Norton/Thevenin theorem, in Laplace's domain, to reduce the full Corsi model into an equivalent current source in parallel with its output impedance. Let's first recall the parameters of the Corsi model:

$$\begin{aligned} R_{q,N_f} &= \frac{R_q}{N_f} & R_{q,N_p} &= \frac{R_q}{(N - N_f)} \\ C_{d,N_f} &= C_d N_f & C_{d,N_p} &= C_d (N - N_f) \\ C_{q,N_f} &= C_q N_f & C_{q,N_p} &= C_q (N - N_f) \end{aligned}$$

C_G is the total parasitic grid capacitance:

$$C_G = C_{m,N_f} + C_{m,N_p}$$

The first step is presented below:

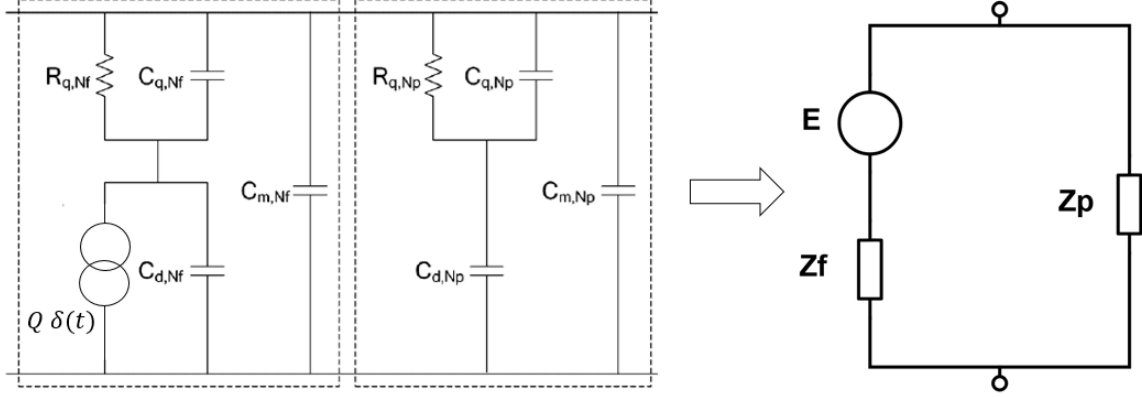


Figure 57: 1st step in the computation of the SiPM equivalent circuit. Left-sided schematic adapted from [10]

$$\begin{aligned} E &= \frac{Q}{C_d N_f s} \\ Z_f &= \frac{1}{C_d N_f s} + \frac{R_q / N_f}{1 + R_q C_q s} \\ Z_p &= \frac{1 + R_q (C_d + C_q) s}{s \left(C_G (1 + R_q (C_d + C_q) s) + C_d (N - N_f) (1 + R_q C_q s) \right)} \end{aligned}$$

The final step is presented in the next figure:

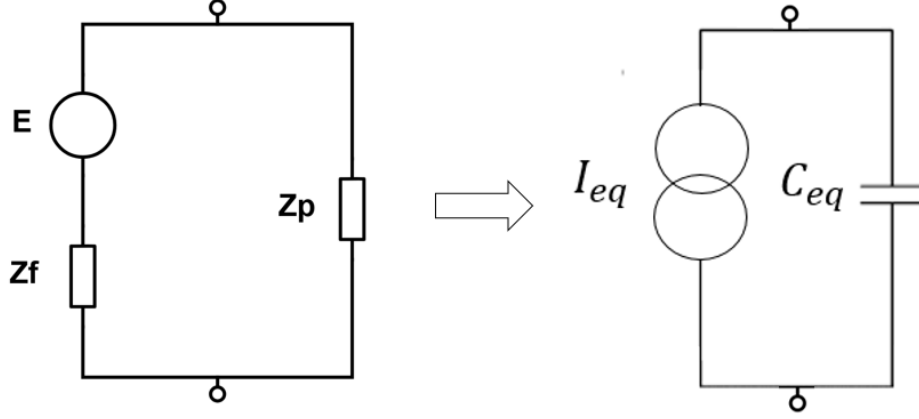


Figure 58: Last step in the computation of the SiPM equivalent circuit

$$I_{eq} = \frac{E}{Z_f} = Q \cdot \frac{1 + R_q C_q s}{1 + R_q (C_d + C_q) s}$$

The output impedance is $Z_f // Z_p$. For Fourier analysis ($s = j\omega$) in the frequency domain, the output impedance can be simplified at low frequencies and at high frequencies to a single capacitance C_{eq} :

$$\begin{aligned} \omega &<< \frac{1}{R_q(C_d + C_q)} & \Rightarrow & C_{eq} = C_G + NC_d \\ \omega &>> \frac{1}{R_q C_q} & \Rightarrow & C_{eq} = C_G + N \frac{C_d C_q}{C_d + C_q} \end{aligned}$$

B. Derivations - noise analysis

B.1. Input-stage

In this section, I derive the expression of the input-referred noise power spectral density for the input stage dI_{Neq}^2 . I chose to derive expressions using currents rather than voltages because we know the peak amplitude of the input current pulse (1), so it is easier to compute the SNR. The equivalent circuit is the following:


$$Z_m = \frac{1}{gm_{N1}} \cdot \frac{1}{1 + s R_{in} C_{eq}}$$

69

$$R_{in} = \frac{1}{A_{boost} gm_{N2} R_{out} gm_{N1}}$$

B.1.2. Voltage output-referred power spectral density of noise

Because we model the noise with small-signals, we can use the superposition theorem and compute each contribution of the current noise sources to the output voltage. By taking the modulus squared, we have a relation with the noise power spectral densities:

$$dV_{Nout}^2 = |Z_m|^2 \left[dI_{N1}^2 + dI_{N2}^2 \left| \frac{C_{eq} s}{gm_{N2} A_{boost}} \right|^2 + dI_{P1}^2 \left| 1 + \frac{C_{eq} s}{gm_{N2} A_{boost}} \right|^2 + (dI_{N3}^2 + dI_{P3}^2) \left| \frac{C_{eq} s}{gm_{N3}} \right|^2 \right]$$

B.1.3. Current input-referred power spectral density of noise

By dividing the output noise voltage by the transimpedance gain we get the equivalent input current noise:

$$dI_{Neq}^2 = dI_{N1}^2 + dI_{P1}^2 + (dI_{N2}^2 + dI_{P1}^2) \left(\frac{C_{eq} \omega}{gm_{N2} A_{boost}} \right)^2 + (dI_{N3}^2 + dI_{P3}^2) \left(\frac{C_{eq} \omega}{gm_{N3}} \right)^2$$

By considering only the thermal noise generated by the transistors, all in strong inversion:

$$dI_{Neq}^2 = 4kT \frac{2}{3} \left[gm_{N1} + gm_{P1} + (gm_{N2} + gm_{P1}) \left(\frac{C_{eq} \omega}{gm_{N2} A_{boost}} \right)^2 + (gm_{N3} + gm_{P3}) \left(\frac{C_{eq} \omega}{gm_{N3}} \right)^2 \right] df$$

B.2. Differential amplifier - ZPC

As the circuit is a voltage amplifier, it is more natural to consider the voltage input-referred noise power spectral density dV_{Neq}^2 rather than the current one. The equivalent circuit is then the following:

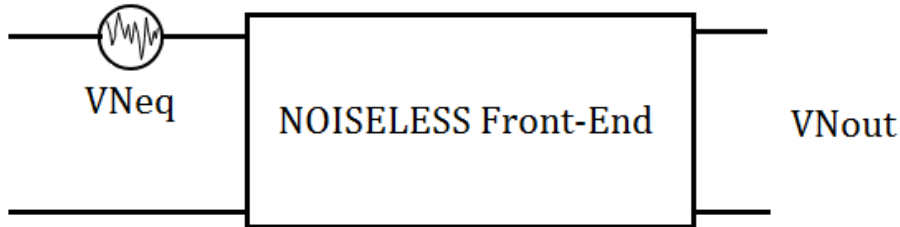


Figure 61: Noisy output stage: equivalent schematic

This time, we need to short-circuit the input to compute the output noise V_{Nout} . Then, we have to divide it by the voltage gain G to bring the noise at the input.

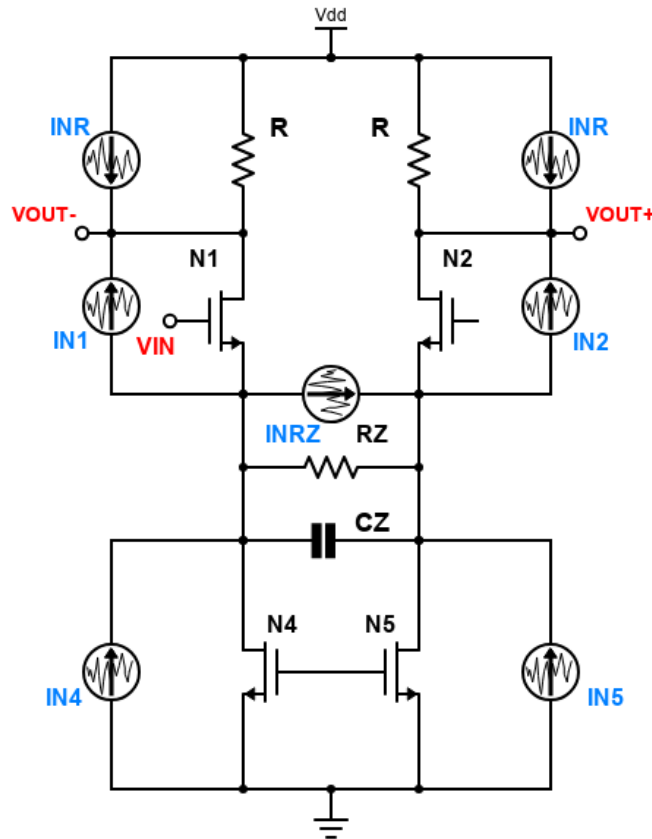


Figure 62: Output stage schematic with the noise sources

B.2.1. Voltage gain G

It is the same expression as (17) but for simplicity, I neglected the 2 poles created by $R//C_L$ and $(1/g_m)//C_Z$. Indeed, the goal is to have a simple expression to know the most important parameters to decrease the noise rather than knowing the exact dynamic of the system.

$$\frac{V_{out+} - V_{out-}}{V_{in+}} = G(s) \approx \frac{2R}{R_Z} \cdot (1 + R_Z C_Z s)$$

B.2.2. Output-referred power spectral density of noise

Let's use again the superposition theorem to compute the total noise at the output:

$$dV_{Nout}^2 = 2R^2(dI_{N1}^2 + dI_{N2}^2) + R^2(2dI_{NR}^2 + dI_{NRZ}^2 + dI_{N4}^2 + dI_{N5}^2)$$

B.2.3. Input-referred power spectral density of noise

By dividing dV_{Nout}^2 by $|G|^2$ we get:

$$dV_{Neq}^2 = \left| \frac{R_Z}{1 + R_Z C_Z s} \right|^2 \left(\frac{dI_{N1}^2 + dI_{N2}^2}{2} + \frac{2dI_{NR}^2 + dI_{NRZ}^2 + dI_{N4}^2 + dI_{N5}^2}{4} \right)$$

If we consider only the thermal noise generated by the devices and that all the transistors are in strong inversion:

$$dV_{Neq}^2 = 4kT \frac{R_Z^2}{1 + (R_Z C_Z \omega)^2} \left(\frac{gm_{N1} + gm_{N2}}{3} + \frac{1}{2R} + \frac{1}{4R_Z} + \frac{gm_{N4} + gm_{N5}}{6} \right) df$$

References

- [1] Cherenkov Telescope Array Observatory. <https://www.cta-observatory.org/>. 2020.
- [2] B.S. Acharya et al. “Introducing the CTA concept”. In: *Astroparticle Physics* 43 (2013), pp. 3–18.
- [3] Slawomir Piatek, Hamamatsu Corporation & New Jersey Institute of Technology. <https://hub.hamamatsu.com/us/en/webinar/low-light-detection-pmt-sipm/index.html>. 2016.
- [4] Editorial Committee Hamamatsu Photonics K.K. *Photomultiplier Tubes - Basics and Applications*. Hamamatsu Photonics K.K., Electron Tube Division, 2017.
- [5] K. Kobayashi A. Ghassemi K. Sato. *MPPC*. Hamamatsu Photonics K.K., Solid State Division, 2017.
- [6] Matthieu Heller et al. “An innovative silicon photomultiplier digitizing camera for gamma-ray astronomy”. In: *The European Physical Journal C* 77 (July 2016). DOI: [10.1140/epjc/s10052-017-4609-z](https://doi.org/10.1140/epjc/s10052-017-4609-z).
- [7] Juan Antonio Aguilar Sánchez et al. “The front-end electronics and slow control of large area SiPM for the SST-1M camera developed for the CTA experiment”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 830 (May 2016). DOI: [10.1016/j.nima.2016.05.086](https://doi.org/10.1016/j.nima.2016.05.086).
- [8] Sibila Gómez et al. “MUSIC: An 8 channel readout ASIC for SiPM arrays”. In: Apr. 2016, 98990G. DOI: [10.1117/12.2231095](https://doi.org/10.1117/12.2231095).
- [9] F. Corsi et al. “Electrical Characterization of Silicon Photo-Multiplier Detectors for Optimal Front-End Design”. In: *2006 IEEE Nuclear Science Symposium Conference Record*. Vol. 2. 2006, pp. 1276–1280.
- [10] D. Marano et al. “Silicon Photomultipliers Electrical Model Extensive Analytical Analysis”. In: *IEEE Transactions on Nuclear Science* 61.1 (2014), pp. 23–34.
- [11] Willy M. C. Sansen. *Analog Design Essentials*. pages 386 and 406. Springer, 2006. ISBN: 10 0-387-25747-0.
- [12] Angelo Rivetti. *CMOS: Front-End Electronics for Radiation Sensors*. page 207. CRC Press, 2015. ISBN: 13: 978-1-4665-6311-7.
- [13] W. M. Sansen, F. Op’t Eynde, and M. Steyaert. “A CMOS temperature-compensated current reference”. In: *IEEE Journal of Solid-State Circuits* 23.3 (1988), pp. 821–824.