TECHNICAL REPORT

Optimizing time resolution and power consumption in a current-mode circuit for SiPMs

R. Manera,[a,*] **S. Gómez,**[b,a] **A. Mariscal-Castilla,**[a] **J. Mauricio,**[a] **R. Ballabriga**[c] **and D. Gascón**[a]

[a]*Institute of Cosmos Sciences (ICCUB), Universitat de Barcelona,*
Martí i Franquès 1, 08028 Barcelona, Spain

[b]*Electronics Departament, Serra Hunter, Universitat Politècnica de Catalunya,*
Eduard Maristany 16, 08019 Barcelona, Spain

[c]*Conseil Européen pour la Recherche Nucléaire (CERN),*
Espl. des Particules 1, 1211 Meyrin, Switzerland

E-mail: rafelmanera@icc.ub.edu

ABSTRACT: Several applications that employ SiPMs require high time precision readout electronics. This work presents a study for the optimization of timing resolution of readout electronics for SiPMs focused on the effect of sensor area, transistor scaling and power consumption on electronic jitter. The design of the most critical stages are presented, specially the front-end input stage in current-mode. The performance of three different technologies (180, 130 and 65 nm) are studied. 65 nm is the best option to obtain good timing resolution with less power consumption. Dividing the sensor into smaller segments improves the Single Photon Electronics Jitter (SPEJ), but does not translate into a better Coincidence Time Resolution (CTR) when keeping the power per unit area constant, performing analog summation or employing an averaging algorithm of the time stamps for small LSO:Ce:%0.2Ca scintillator crystal.

---

*Corresponding author.

Contents

1 Introduction

Silicon Photomultipliers (SiPMs) are unrivalled solid state detectors for high time precision measurements in low light environments, but for large channel areas (≥ 1 mm^2/ch) their time resolution reaches at best 100 ps for analog implementations [1] and 200 ps [2] for digital implementations. Analog SiPMs offer the best performances in terms of Photon Detection Efficiency (PDE), Dark Count Rate (DCR) and optical crosstalk because they are fabricated in processes optimized for the detection of light. However, due to the parallel connection of many Single Photon Avalanche Diodes (SPADs), they present two intrinsic limitations: (1) the large parasitic capacitance degrades the Single Photon Time Resolution (SPTR) and (2) it is not possible to time stamp multiple photons closely spaced in time. On the other hand, Digital SiPMs do not present these limitations, but they require active electronics per SPAD, which limits the Fill Factor (FF) and thus degrading the PDE. Moreover, the requirement of a Time-to-Digital Converter (TDC) [3] per SPAD imposes a large power budget.

Precise time resolution is required for several applications. Time-of-flight Positron Emission Tomography (ToF-PET) demands efficient photodetectors coupled to fast readout electronics in order to determine the position of the anihilation of the positron [4]. In high energy physics experiments,

scintillator-based detectors are employed for the identification of elementary particles. The time resolution allows improving vertex identification, improved transverse momentum resolution, and reducing pile up. Other applications include Light Detection and Ranging (LIDAR) [5] or ToF Mass Spectrometry [6].

The electronic jitter is a good indicator of the time resolution of a photo-detection system. Equation (1.1) defines how the electronic jitter (σ_t), measured as the standard deviation of the time-of-arrival, is affected by the slew rate at a certain threshold (SR_{th}) and the integrated electronic output noise (σ_n) for a system where a leading-edge discriminator is used to determine the photon Time-of-Arrival (ToA).

$$\sigma_t = \frac{\sigma_n}{SR_{th}} \tag{1.1}$$

The employment of smaller SiPMs helps to improve the electronics jitter by virtue of their intrinsic lower capacitance, resulting in an output pulse with a larger slew rate. In addition, electronic noise decreases as the detector capacitance does when series noise dominates [7]. However, decreasing the sensor area, requires the employment of more readout channels and if the channel power consumption is maintained, it means that the power invested per unit area will increase which can not be afforded in some applications. Recently, the possibility of segmenting a SiPM in smaller sensors and then add the signals in the readout electronics has been proposed, but without considering power consumption [8].

This work is focused on the design of the most critical stages of dedicated readout electronics for SiPMs to optimize the time resolution depending on the power invested. More specifically, this study targets the front-end input stage and the fast timing current discriminator, which are the most critical elements to provide the timestamp of the detected photons with low jitter. The main goal is to find the best trade-off between power consumption and time resolution. Starting point is HRFlexToT [1], whose time resolution performance is used as benchmark comparison. From that, three different CMOS technologies are studied: XFAB 180 nm, TSMC 130 nm and 65 nm with the objective of optimizing the trade-off between timing performance and power consumption. The effect of different area SiPMs and different power consumption on timing performance is explored for each technology.

2 Architecture

In this section, the main circuits of this study are presented. Those blocks are the input stage and the Fast Current Comparator (FCC). Auxiliary Operational Transconductance Amplifiers (OTA) are not designed and just modelled as ideal amplifiers, since their contribution to power consumption and to jitter degradation is not dominant.

As a general remark on the design of the circuits, transistors dimensions are adjusted proportionally to the technology node reduction, always maintaining the W/L ratio. Observe that the current flowing through a transistor is proportional to the relation between its width and length (i.e. aspect ratio) [9], referred as W/L. By reducing the W/L of the transistors, the current flowing through them is reduced proportionally (and thus the power consumption) but the operating point, i.e., the voltages that set the proper mode of operation of the transistors, are maintained.

2.1 Input stage

The input stage schematic is based on the HRFlexToT design [1] and it is shown in figure 1. It consists of a current-mode circuit with approximately unity gain. The circuit is based on a current mirror

formed by M_5 and M_{10}. This approach allows generating multiple copies of the input signal. A Low Frequency Feedback (LFF) formed by M_4 and the OTA controls the input node DC voltage regulating M_4 gate. In addition, the OTA also controls M_9 gate, thus improving the current copy accuracy by forcing the drain voltages of M_5 and M_{10} to be as equal as possible. A High Frequency Feedback (HFF) formed by M_4, M_5 and M_6 maintains the input impedance low at higher frequencies (~ 100 MHz). Transistor M_4 is critical in terms of stability and noise, being the main contributor to series noise, which is the dominant noise source when using sensors with a capacitance higher than ~ 1 pF [7]. Transistor M_4 is operating in weak inversion, and it can be assumed that it behaves as a bipolar transistor [10]. The input referred voltage noise for a bipolar transistor is illustrated in equation (2.1), showing the importance of maximizing the transconductance (g_m) of transistor M_4.

$$e_n = \sqrt{\frac{2kT}{g_m}} \tag{2.1}$$

Where k is the Boltzmann constant and T is the temperature of the transistor. For this reason, minimum length transistors are used [9]. The width of the transistor is limited by the parasitic capacitance, which affects the HFF bandwidth.
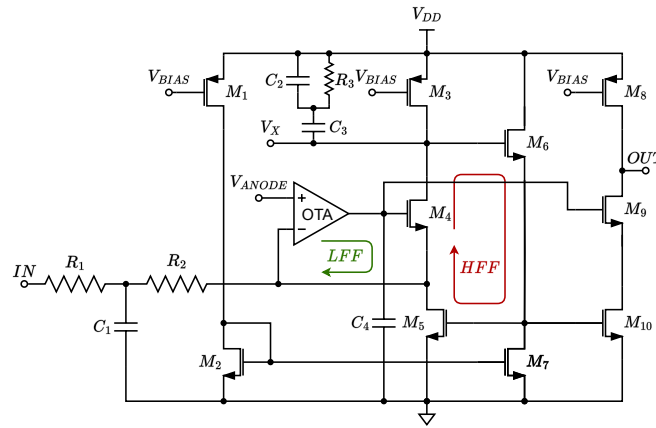

Figure 1. Input stage schematic.

An input 'T' impedance is used to assure stability for low capacitance sensors and for high inductance interconnections. It is formed by R_1, R_2 and C_1. Capacitor C_1 provides a minimum compensation capacitance. Resistors R_1 and R_2 provide damping to minimize the quality factor of a possible resonance of the sensor capacitance with the sensor and the Application Specific Integrated Circuit (ASIC) parasitic interconnection inductance and helps in linearizing the input impedance of the system for large input current pulses.

A linear dynamic range is an important specification in several applications. For instance, in PET, a linear response of the circuit is required to properly identify the number of detected photons equivalent to the incident energy of the 511 keV gamma event. Hence, a key constraint in the design of the input stage is assuring a high dynamic range, to achieve a linear and stable operation for large peak currents. The voltage at the input node is directly connected to the anode of the SiPM and this voltage should be adjustable in order to compensate for non uniformities of the breakdown voltage. This adjustment is done by the LFF network. A DC voltage level shifter M_6 allows increasing input

(anode) voltage range of adjustment, keeping M_4 in active saturation region. To maximize the anode voltage range of adjustment, condition (2.2) must be satisfied.

$$V_X \geq V_{\text{anode}_{\text{MAX}}} + V_{DS\text{sat}}(M4) \tag{2.2}$$

At the same time, the bias current source transistor M_3 must be kept in active saturation region for the whole dynamic range. This condition is illustrated in (2.3).

$$V_X \leq V_{DD} - |V_{DS\text{sat}}(M3)| \tag{2.3}$$

2.2 Cascodes in 65 nm input stage

Simulations performed with the input stage in 65 nm technology showed a significant decrease in the HFF loop gain compared to the other nodes (180 and 130 nm). Since the purpose of the HFF loop is to maintain a low input impedance at high frequencies, a reduction in its gain leads to an increase in the input impedance in its frequency range of operation. The sensor capacitance and the input impedance form a current divider, illustrated in the simplified detector model in figure 2 (Left). Where I_{SENS} represents the output pulse of the sensor, C_{SENS} the sensor capacitance and Z_{IN} the input impedance of the readout electronics. Therefore, for current sensing, is important that the impedance of Z_{IN} is smaller than the impedance of C_{SENS}. In addition, from figure 2 (Left), the current sensed by the electronics (i.e. the current flowing through Z_{IN}) can be computed as

$$I_{Z_{\text{IN}}}(s) = I_{\text{SENS}}(s) \frac{1}{s \cdot \tau_{\text{IN}} + 1} \tag{2.4}$$

where $\tau_{\text{IN}} = Z_{\text{IN}} \cdot C_{\text{SENS}}$ represents the time constant given by the sensor capacitance and the input impedance. The signal shape depends on the relationship between time constants from the sensor τ_{SENS} and τ_{IN}. To preserve the pulse shape, τ_{IN} must be much lower than τ_{SENS}, otherwise, the sensor current is integrated in C_{SENS} and the voltage across Z_{IN} is proportional to the detector charge divided by C_{SENS}. If the pulse shape is not preserved, the slew rate of the signal is decreased, which leads to a higher electronic jitter. The adopted solution to increase the gain of the HFF is adding a cascode transistor to M_4, leading to a lower input impedance. The specific circuit for the input stage in 65 nm technology is shown in figure 2 (Right). For the other technologies, the cascode transistor was not needed.
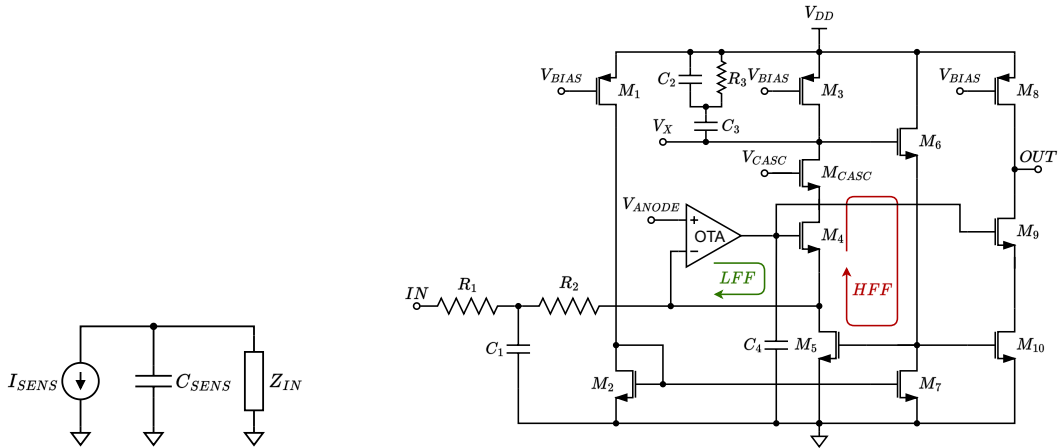

Figure 2. Simplified detector model (Left) and input stage schematic with cascode transistor (Right).

2.3 Fast leading edge current comparator (FCC)

The objective of the comparator is to provide a measurement of the ToA of the events. A leading-edge comparator is here employed, which provides a Time-over-Threshold (ToT) response encoding the arrival time in the rising edge of a binary pulse. The schematic design of the FCC is shown in figure 3. The basic design is very similar to the one used in [11]. Of course, minimal size transistors are scaled down from 180 nm to 130 and 65 nm respectively to exploit the smaller gate delay of the different technologies.
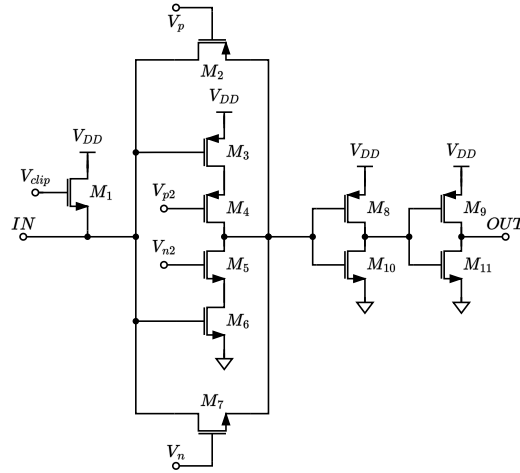

Figure 3. Fast Current Comparator schematic.

The FCC consists of a zero-crossing current-mode comparator. The threshold of the comparator is set by injecting a DC current into the input node, shifting down the signal baseline and moving the desired crossing point to 0. For small currents around threshold, the feedback transistors M_2 and M_7 are in sub-threshold region, so the input impedance of the comparator has a capacitive behavior (i.e. very large in DC) which leads to a very high resolution. When the input current is positive, the current flows through M_7 and the output of the comparator is 0. Conversely, if the input current is negative the current flows through M_2, and the output is 1.

2.4 Readout architectures for segmented sensors

Different strategies can be applied for the readout of segmented sensors. It is important that apart from the input stage and the comparator, two other circuits are needed. First, the summation block, which performs an analog summation of the signal read out from different sensors. Second, the TDC which is in charge of converting the signal from the comparator into a digital data stream that can be processed by the data acquisition system. The design of these blocks is beyond the scope of this study and they are considered ideal, but they need to be accounted in the architecture.

Given a specific number of analog channels, two main strategies of combining the signals of each input stage could be applied (see figure 4). The first option (summation readout) consists of performing an analog summation of all the output signals read out by the input stage before the leading-edge comparator. Time information can be converted into digital by reading the output of the comparator with a TDC. The second option (individual readout) considers an input stage (without analog summation) and a comparator to read each sensor segment. The ToA is digitized by employing a TDC for each

individual channel, and thus several timestamps are obtained depending on the segmentation factor. A single timestamp can be obtained by combining the timestamps using different methods, as it is normally applied in PET applications with monolithic scintillator crystals [1, 12]. In multi-photon systems, the signal is spread between different channels and several timestamps are obtained. In this work, a simple averaging algorithm is used, i.e., the timestamp used to compute the CTR is obtained by averaging the ToA from each readout channel. Although, more sophisticated strategies to combine the timestamps could lead to better results, and may be studied in the future [12, 13]. Block diagrams for the two different approaches (for the case of 4 readout channels) are shown in figure 4.
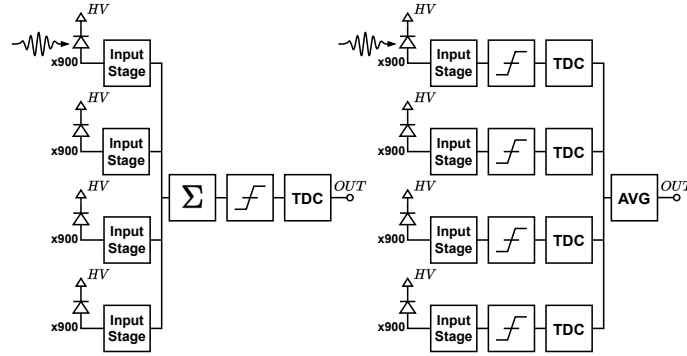

Figure 4. Analog vs digital summation block diagram for the case of 4 readout channels.

3 Methodology

Simulation results are focused on the timing performance of the circuits without considering the RC parasitic of the circuit (layout). The parameters that have been studied to observe their effect on the time resolution of the system are the following:

- *Input Stage Power Consumption*: changing the input stage bias current changes its performance (noise/slew rate) and its power consumption proportionally.

- *Sensor Segmentation*: changing the area of the sensor by connecting a different number of SPADs to each input stage. Hamamatsu Photonics S13360-3050CS SiPM is taken as the reference non-segmented sensor with an area of $3 \times 3\,\text{mm}^2$ and 3600 SPADs.

- *Input Stage Transistors Dimension Scaling*: changing the W/L ratio of the transistors changes the current flowing through the transistor proportionally [9]. The W/L ratio remains constant throughout the three technologies.

3.1 Input stage power consumption

The power consumption of the input stage designed in 180 nm technology is considered as the reference point. Moreover, the same circuit is implemented in the two other technologies (130 and 65 nm), but scaling down the transistors according to the technology node. All those values are considered as the nominal reference points. The possibility of increasing the power consumption by a factor 2 and 4 is also considered. In this work, the power consumption is expressed normalized to the reference non-segmented sensor.

3.2 Sensor segmentation

Table 1 shows the different segmentation factors employed normalized to the reference non-segmented sensor. The segmentation factor represents the number of portions that the sensor is divided into. These SiPMs with a different number of SPADs are connected to the input stage.

Table 1. Sensor segmentation cases under study.

Segmentation Factor	0.5	1	2	4	8	16	32
Microcells (SPADs) per Input Stage	7200	3600	1800	900	450	225	112

The SiPM electrical model [8, 14] used to simulate the firing of a microcell (i.e. the arrival of a photon to the sensor) is shown in figure 5, where N represents the number of cells connected to one input stage. Different values of N are applied to emulate the different sensors specified in table 1 and the value of C_g is scaled inversely proportional to the segmentation factor, i.e., the smaller the sensor, the smaller the parasitic capacitance.
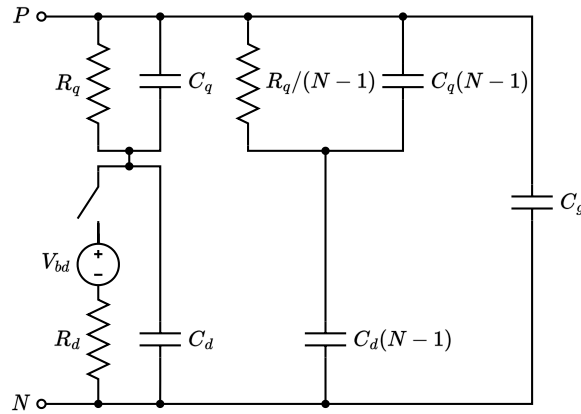

Figure 5. SiPM electrical model. Reprinted from [14], Copyright (2007), with permission from Elsevier.

3.3 Input stage transistors dimension scaling

Regarding the area of the transistors in the input stage, three different scaling strategies have been explored.

- *No Scaling*: the area of the transistors is constant regardless of the segmentation factor.

- *All Scaled*: the area of the transistors is scaled inversely proportional to the segmentation factor.

- *Half Scaled*: the area of the transistors is scaled inversely proportional to the segmentation factor, except for the DC level shifter and its biasing current source (M_1, M_2, M_6 and M_7), which are not scaled. This strategy is included as a possible trade-off between power consumption and time performance, since M_6 plays an important role in the stability of the HFF.

As it has been explained above, scaling the transistors means changing its W/L ratio, which changes the current flowing through the transistor. Changing the current affects the power consumption of the circuit, but also its frequency response. With more current consumption, the circuit presents a higher gain-bandwidth product, which leads to a higher slew rate. Applying the *No Scaling* strategy, the input stage circuit remains unchanged for the different segmentation factors, which means that the input stage has the same power consumption. Although, the power consumption per unit area increases proportionally to the segmentation factor since more input stage circuits are needed to read out the same detection area. On the other hand, the power consumption of the input stage scales inversely proportional to the segmentation factor when applying the *All Scaled* strategy, thus maintaining the power consumption per unit area constant. Lastly, in the *Half Scaled* strategy, only the transistors that form the DC level shifter are not scaled due to its effect on the frequency response.

3.4 Figure of merit

To easily evaluate the trade-off between timing performance and power consumption of the input stage, we defined a Figure of Merit (FOM) that takes into consideration the Single Photon Electronic Jitter (SPEJ) and the power consumption. The equation to compute the FOM is illustrated in (3.1). Higher FOM values indicate either low power consumption or low SPEJ results. The highest FOM value indicates the optimal trade-off between time resolution and power consumption. Note that the highest FOM, does not necessarily indicate the best time performance (lowest SPEJ), but the most efficient configuration to obtain a certain SPEJ. Setting a SPEJ specification and looking for the configuration that shows the highest FOM, ensures that the required timing performance is achieved with the lowest power consumption.

$$\text{FOM} = \frac{1}{\text{SPEJ} \cdot \text{Power} \cdot 10^{12}} \quad [(\text{ns} \cdot \text{mW})^{-1}] \tag{3.1}$$

3.5 Electrical simulations

Electrical simulations shown in this study are performed using the software suite Virtuoso from Cadence [15]. Simulations are performed using the tests benches shown in figure 6, where the values of the components are:

- $HV = 56.4\,\text{V}$

- $R_{\text{BIAS}} = 10\,\text{k}\Omega$

- $C_{\text{HV}} = 1\,\mu\text{F}$

- $C_{\text{PAR}} = 4.5\,\text{pF}$

- $L_{\text{BOND}} = 1.2\,\text{nH}$

- $C_{\text{PAD}} = 1.5\,\text{pF}$

- $R_S = 3\,\Omega$

- $I_{\text{SOURCE}} = $ Exponential current source with peak value from 10 to 40 mA

- $C_{\text{SENS}} = 300\,\text{pF}$

Figure 6 (Left) shows how the SiPM model (figure 5) is connected to the High Voltage (HV) power supply and to the input stage. The connection between the anode of the SiPM and the input stage is modelled trying to mimic the connection of the sensor to an ASIC soldered to a Printed Circuit Board (PCB). C_{PAR} corresponds to the parasitic capacitance from the traces of the PCB, L_{BOND} models the parasitic inductance from the bonding of the ASIC packaging, C_{PAD} corresponds to the parasitic capacitance from the pad of the package and R_S is modelling the resistance of this connection. Figure 6 (Right) shows the test bench for linearity simulations. An exponential current source (I_{SOURCE}) is employed to emulate the behaviour of the SiPM and easily sweep its peak amplitude. C_{SENS} corresponds to the SiPM capacitance to maintain the stability and frequency response of the input stage.
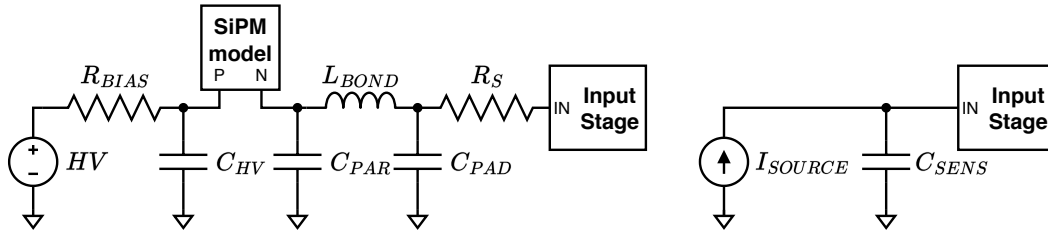

Figure 6. Electrical Test Bench schematic for simulations using the SiPM model (Left) and for linearity simulations (Right).

SPEJ results are obtained performing a transient noise analysis in Virtuoso with 100 iterations to ensure sufficiently good statistics. The SPEJ is obtained as the standard deviation (sigma) of the ToA of a single photon signal. The threshold to obtain the ToA is set at 50% of the peak amplitude of the single photon signal. A single sensor cell is fired at a given time, but the noise generated by the electronics is added to the signal, thus generating an uncertainty in the ToA.

3.6 Coincidence time resolution simulations

In radiation detectors, time degradation (jitter) comes from the three elements of the detection chain: the scintillator crystal, the sensor and the electronics. Gate, which uses GEANT4 as simulation engine [16] is in charge of simulating the time dispersion occurring inside the crystal. Two detectors employing an LSO:Ce:0.2%Ca of $2 \times 2 \times 3$ mm^3 scintillator crystal and Hamamatsu S13360-3050PE SiPM were simulated in a coincidence set-up, where a ^{22}Na source was placed between both, as illustrated in figure 7. Only photoelectric 511 keV events were considered. Observe that a short crystal was employed to highlight the effect of the electronics and thus minimize the effect of the gamma interaction and optical photons travel spread [17]. LSO:Ce:0.2%Ca parameters (decay constants and light yield) were obtained from [18]. A Photo Detection Efficiency (PDE) of 59% was considered for the sensor. The interface material used to glue the crystal and the SiPM was Cargille Meltmount [19] with an index of refraction of 1.582. Additionally, the scintillator crystal was covered with polytetrafluoroethylene (PTFE). Both interfaces were simulated using the Davis LUT model [20].

The simulation generates two antiparallel gamma photons of 511 keV that arrive at the scintillators in coincidence. Then, these gamma photons create optical photons through photoelectric interaction. These optical photons are produced according to a distribution that depends on the timing characteristics of the scintillators. Afterwards, the interaction of the optical photons with the scintillators is simulated until they reach the SiPMs. Therefore, the time distribution of the optical photons arriving at the
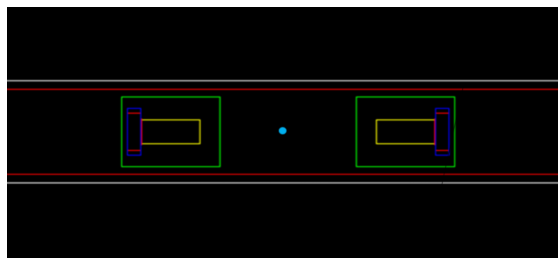
Figure 7. Image of the CTR setup from Gate. It shows the ^{22}Na source (cyan), the detector module (green) and inside, the scintillator (yellow), the SiPM detector (blue) and the optical glue (red).

SiPM will be a convolution between their emission distribution and their transport distribution [21]. Once the Gate simulation is completed, the ToA of the different optical photons impinging each SiPM are saved in a file.

Apart from the contribution of the crystal, the sensor, in this case a SiPM, also contributes to the time degradation. This part can not be simulated due to the lack of information of the sensor, and it is modelled as a random timestamp generated with a Gaussian distribution obtained through experimental measurements. This jitter contribution was added to the ToA of every detected photon and corresponds to the intrinsic Single Photon Time Resolution (SPTR) of the SiPM. This Gaussian distribution had a zero mean value and a standard deviation of 58 ps, which was equal to the one measured for the S13360-3050PE SiPM [18]. Lastly, the SiPM area is scaled according to the segmentation factor and the optical photons are classified according to the segments.

The contribution from the electronics is obtained using an electrical simulator, in this work, Cadence as previously detailed. The output of the Gate simulation with the addition of the SPTR from the sensor is used in the electrical test bench to replicate the distribution of the optical photons detected by the photo-sensor and generated by gamma events on the SiPM electrical model. The input stage circuit is used to evaluate its contribution in terms of noise and slew rate limitation. Figure 8 shows an output signal of the input stage when the Gate results are used in combination with the SiPM model.
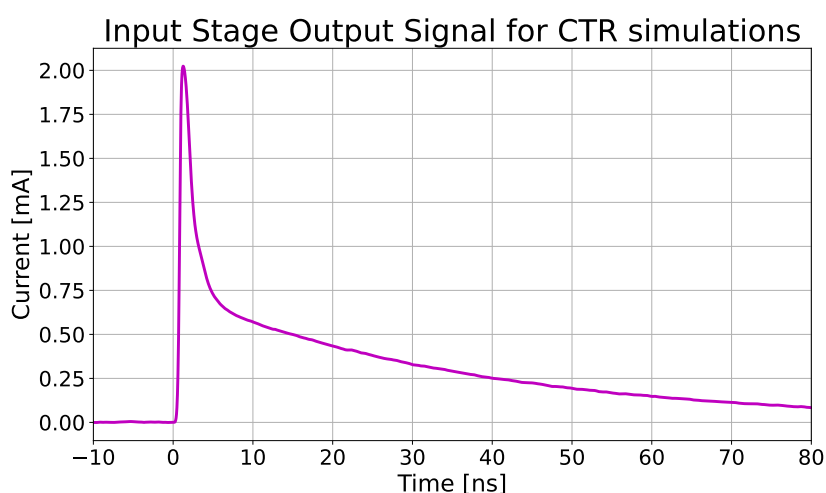

Figure 8. Input stage output signal employing the output of the Gate simulation and the SiPM model.

Note that this signal corresponds to one specific gamma event and each event will result in a different signal. The timestamp of each simulated event is obtained from the discrimination of the output signal of the input stage with a threshold set at 50% of the peak amplitude of a signle photon signal [18]. These timestamps are used to compute the Coincidence Time Resolution (CTR) of the system.

4 Simulation results

4.1 Input stage characteristics

In this section, the main characteristics of the input stage are studied. A linear dynamic range is important for those multi-photon applications (like PET, or High-Energy Physics calorimeters) that require a precise identification of the energy (i.e., the number of photons) of the events [1, 12]. The dynamic range is evaluated by performing a sweep of the peak amplitude of an exponential current source (see figure 6 (Right)) from 10 to 40 mA. This range includes the region of interest for applications where a considerable amount of photons hit the sensor with small difference in ToA.

The main characteristics of the input stage are studied trough simulations depending on the source current amplitude injected. As it can be observed from figure 9, the circuits developed using the 180 and 65 nm nodes start to saturate (enters a non-linear regime) at ~ 25 mA whereas 130 nm shows the best input range, starting to saturate at ~ 30 mA. The loss of linearity is attributed to the increase in the input impedance and the decrease in the HFF loop gain at these current levels, as illustrated in the figure 10. Observe that the input current flows directly through transistor M_5 (figure 1). As this current increases, M_5 gate-source voltage also increases, making the voltage at the node V_X higher. As V_X node voltage rises, M_3 drain-source voltage decreases, up to a point where it enters into the ohmic region and decreases the current flowing through it. The current flowing through M_4 also decreases, which makes its g_m smaller, increasing the input impedance.
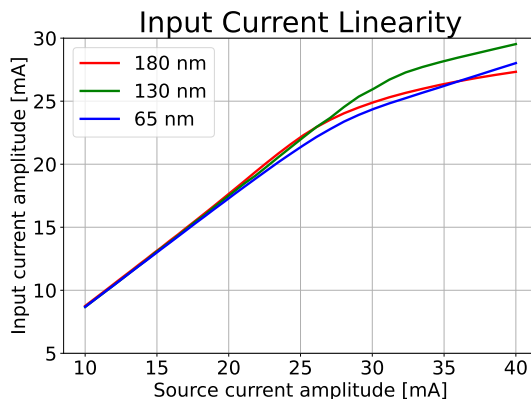

Figure 9. Input current linearity.

Another important characteristic is to keep the input impedance low to optimize its timing performance [7]. A low input impedance is needed in a current-mode readout to maximize the peak current and thus improve the slew rate of the input signal, although some impedance is needed to compensate the resonant circuit formed by the parasitic inductance of the interconnects and the sensor capacitance. The input impedance is measured by dividing the input voltage amplitude by

the input current amplitude when considering the transient response of the input stage, as shown in figure 10 (Left). A low input impedance is ensured for the full dynamic range of each input stage.

The dynamic range of a critical sub-circuit of the input stage, which is the HFF, is shown in figure 10 (Right). This HFF is related to the input impedance, since its function is to maintain it low at the signal frequency range (hundreds of MHz). The gain of the HFF loop is kept constant along the dynamic range of interest, and thus ensuring the proper behaviour of the HFF for the three technologies. In conclusion, these results show that dynamic range is not a limiting factor in any of the three technologies, since the behaviour of the three technologies is very similar up to 25 mA, which is enough for most applications.
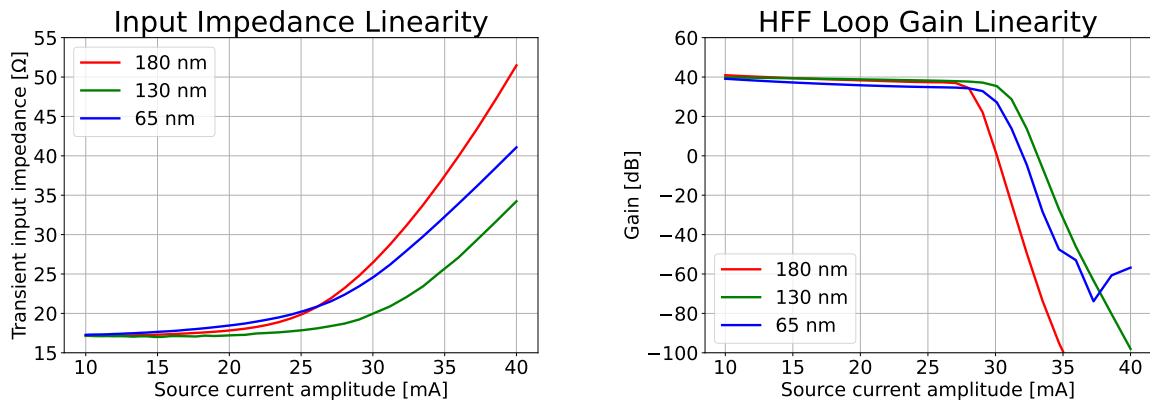

Figure 10. Transient Input Impedance (Left) and HFF Loop Gain (Right) linearity.

4.2 Input stage timing performance

The timing resolution of the input stage is obtained measuring the SPEJ of its output signal. These results are obtained using an ideal threshold at 50% of the peak amplitude and calculating the standard deviation (sigma) of the crossing times of the rising edge with this threshold. Figure 11 shows the SPEJ for the segmentation factors specified in table 1 and different power consumption for the three technologies.

Observe that without scaling the transistors (No Scaling red curves) for nominal and quadruple power consumption, SPEJ always improves with segmentation, but the power consumption per unit area increases proportionally to the segmentation factor. For the cases of All scaled, the lowest SPEJ is obtained with segmentation factors between 2 and 4, while for Half Scaled, a local minimum is found for a segmentation factor between 4 and 8. With this two scaling strategies, 65 nm technology shows the lowest SPEJ for the case of nominal power for segmentation factors lower than 8, as shown in figure 11 (Left). For segmentation factors higher than 8, SPEJ is lower in 130 nm technology. For the case of quadrupled power consumption (figure 11 (Right)), 65 nm is always the best technology. Independently of the scaling strategy, SPEJ always improves when the input stage power consumption is increased.

The FOM for all studied cases is shown in figure 12. It can be observed that 65 nm technology shows the best FOM, for most of the configurations, due to its lower power consumption and lower SPEJ. The highest FOM is achieved employing the Half Scaling strategy for segmentation factor 2 with 65 nm technology and nominal power. Lastly, better FOMs are achieved when configuring
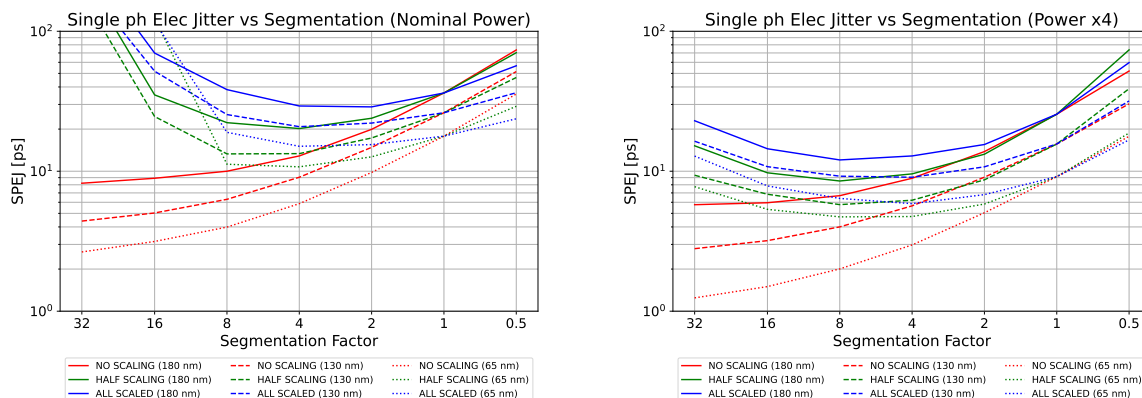
Figure 11. SPEJ in function of the segmentation factor for Nominal Power Consumption (Left) and Quadruple Power Consumption (Right).

the input stage with nominal power independently of the technology, as long as the segmentation factor is 8 or lower. For segmentation factors higher than 8, a higher FOM is obtained with quadruple power if a scaling strategy is employed.
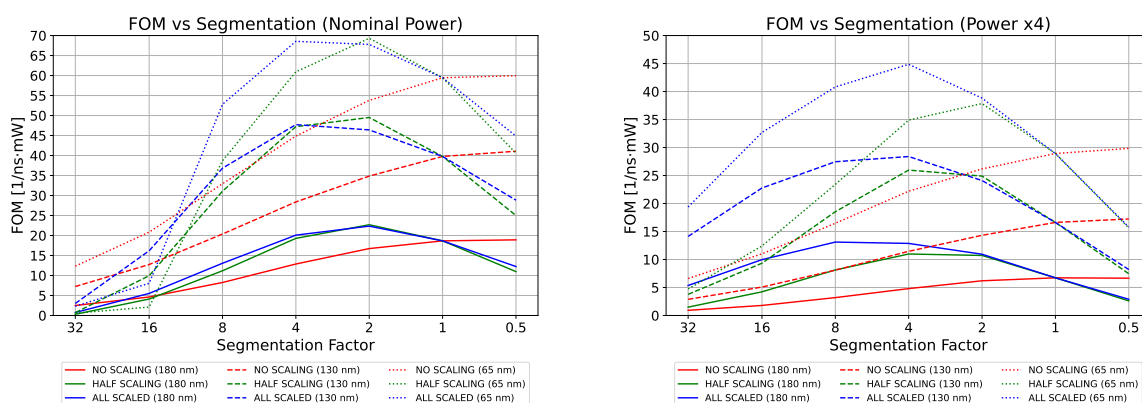

Figure 12. FOM in function of the segmentation factor for Nominal Power Consumption (Left) and Quadruple Power Consumption (Right).

After the evaluation of the time resolution of the input stage with an ideal threshold, the following step is to add the FCC to the system. Table 2 shows the cases where the best FOM is obtained for each technology, with and without considering the FCC. Simulations of the FCC show that its jitter is negligible (< 1 ps). The difference in jitter between using the ideal and the real comparator is due to the difference between the impedance of the real comparator and the impedance from the load connected to the input stage output when using an ideal threshold. The contribution of the comparator itself to the jitter is not significant. This implies that the design of the FCC is critical because it is the output load of the input stage and can degrade the signal slew rate. The best FOM when considering the FCC is achieved for both 130 and 65 nm technologies, achieving a slightly better FOM with 130 nm.

Timing resolution is also studied from a different point of view. In this case, table 3 shows the best scenario for each technology to reach a SPEJ at the level of 10 ps with the lowest power consumption. It is common for the three technologies that the most efficient configuration to obtain a

Table 2. Comparison of best FOM with ideal and the FCC circuit.

Tech	180 nm		130 nm		65 nm	
Scaling	Half Scaled		Half Scaled		Half Scaled	
Power step	Nominal Power		Nominal Power		Nominal Power	
Segmentation Factor	2		2		2	
Comparator	NO	YES	NO	YES	NO	YES
FOM [1/ns·mW]	22.69	17.47	49.51	39.52	69.31	36.99
SPEJ [ps]	23.91	31.06	17.28	21.65	12.67	23.79
Power [mW]	1.843		1.169		1.138	

SPEJ of 10 ps is to apply a segmentation factor 4. It is important to highlight, that in this case the 65 nm technology shows the best efficiency in achieving a SPEJ at the level of 10 ps compared to the other two technologies. Observe that the power consumption of the input stage to obtain a SPEJ lower than 10 ps, should be increased compared to the results shown in table 2 and thus lower FOMs are achieved. More specifically, to achieve this level of time resolution, the power efficiency is lower, i.e., the amount of power needed to reduce the jitter contribution is larger. Despite the FOM is lower compared to table 2, the best FOM is also achieved by the same 65 nm technology, showing that 65 nm node is the most power efficient technology to decrease the time jitter. Lastly, although only the results for nominal and quadruple power consumption are shown, other levels of power consumption have been studied, such as double the power consumption, as illustrated in table 3.

Table 3. Best 10 ps SPEJ with ideal and real comparator.

Tech	180 nm		130 nm		65 nm	
Scaling	Half Scaled		Half Scaled		All Scaled	
Power step	Power x4		Power x2		Power x2	
Segmentation Factor	4		4		4	
Comparator	NO	YES	NO	YES	NO	YES
FOM [1/ns·mW]	10.98	8.959	35.97	26.45	55.68	39.14
SPEJ [ps]	9.575	11.74	8.896	12.10	9.393	13.39
Power [mW]	9.512		3.125		1.912	

4.3 Summation timing performance

The next step is to evaluate the time resolution of performing the summation of the analog output signals from different input stages in order to produce a single output as the non-segmented sensor generates. The aim of this section is to study the impact of a summation circuit and its bandwidth in the time resolution for different readout architectures developed in 65 nm. Only 65 nm technology is employed, since previous results have shown that it has the highest FOM. For example, the impact of the analog and digital summation with a segmentation factor 4 is evaluated using the block diagram illustrated in figure 4, but without considering the contribution to the jitter of the TDC. One of the most important characteristics of the summation block is its bandwidth. Optimizing the bandwidth of the summation is important, since it must be sufficiently low to filter part of the high-frequency

noise, but large enough to ensure a high slew rate, and thus minimizing electronic jitter. The effect of the bandwidth on the summation is studied by performing simulations with an ideal summation block with a passive first-order RC Low-Pass Filter to emulate the bandwidth limitation of a real circuit. Note that this summation circuit does not include real components and is not adding electronic noise, meaning that its only contribution to jitter is in terms of bandwidth limitation. The goal of usign an ideal summation circuit is to isolate the contribution to jitter of adding the signals from the contribution of the bandwidth limitation, because if an ideal summation does not improve jitter, a real circuit will yield worse timing resolution.

Figure 13 shows the SPEJ results for different readout architectures. These architectures include segmentation factors 1 and 4 (3600 and 900 SPADs) with nominal and quadrupled power consumptions. Individual readout of the channels (IND) and analog summation (SUM) are implemented for segmentation factor 4 considering the same power per unit area (All Scaled strategy) and therefore the power budget is distributed between the different channels. A range of bandwidths from 100 MHz up to 2 GHz and the case where no bandwidth limitation is applied are taken into account for each architecture.
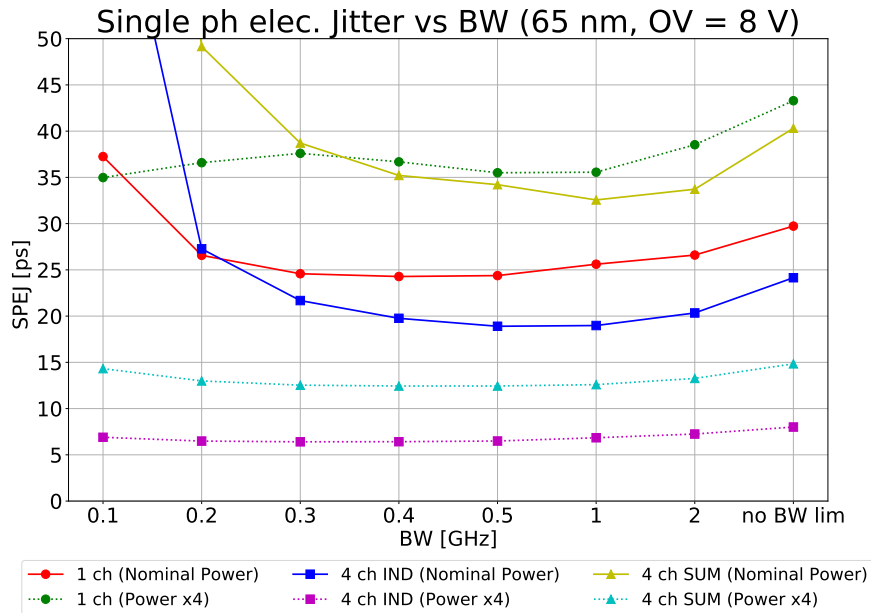

Figure 13. SPEJ vs bandwidth limitation.

Observe in figure 13 that SPEJ improves when the sensor is divided into four segments while maintaining the power consumption per unit area and performing an individual readout. This is evident when comparing the cases of '1 ch (Nominal Power)' with '4 ch IND (Nominal Power)'. When the power consumption is increased 4 times, both segmentation strategies (individual readout and summation) achieve better results, being the individual readout ('4 ch IND (Power x4)') the best option.

It is important to highlight two aspects in the simulations shown in figure 13. First, observe that increasing the power consumption when reading a single channel ('1 ch (Power x4)') deteriorates the jitter when compared to the case of '1 ch (Nominal Power)'. In this case, increasing the power consumption only augments the noise degradation since the increase in slew rate does not provide any improvement because the limiting factor is the slew rate of the input signal. Second, the summation is not effective in single photon applications when comparing for the same power consumption individual

readout (IND) and analog summation (SUM). In this case, summation only adds noise from channels without signal and therefore degrading the overall jitter.

The optimal bandwidth varies depending on each case under study, but in general a bandwidth larger than 1 GHz only degrades the SPEJ. A larger bandwidth only leads to more noise since the slew rate is already limited by the input signal. In particular, for '1 ch (Nominal Power)', the optimal bandwidth can be found between 300 and 500 MHz, but for '4 ch IND (Nominal Power)', the optimal bandwidth is between 500 MHz and 1 GHz.

While SPEJ is a good indicator of the general time performance of photo sensor and readout electronics, for multi-photon systems, such as ToF-PET, CTR is a better indicator of the system time performance. In a ToF-PET system, more than a single photon contributes to the time performance, since all photons arriving before the threshold at the comparator affects the generated time stamps. The slew rate of the resulting signal at the input of the comparator increases as the number of detected photons does. Since the slew rate has a huge impact in the resulting time performance, the optimal bandwidth and power consumption may be different compared to the case of a single photon.
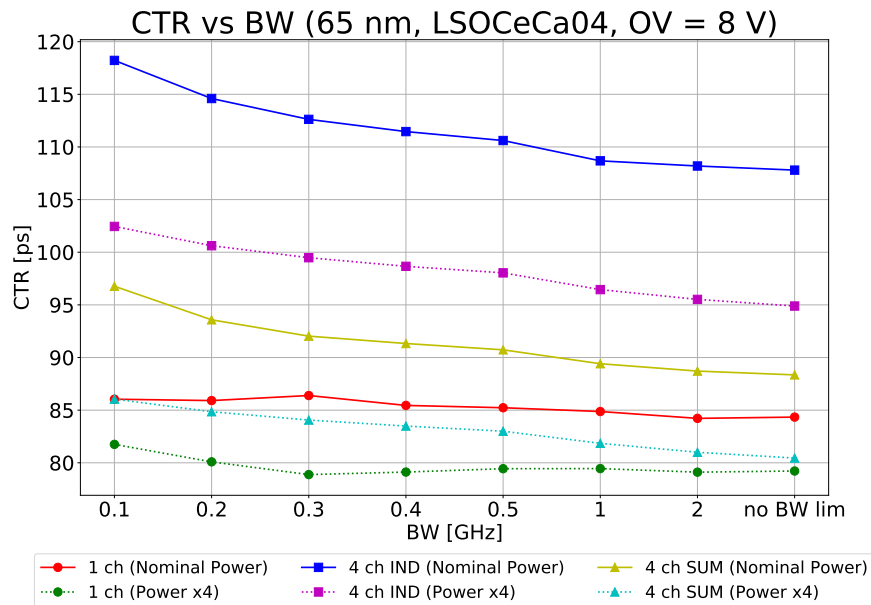

Figure 14. CTR vs bandwidth limitation.

Figure 14 shows the results of CTR simulations for the same cases presented before. In the individual readout cases (4 ch IND), the CTR is obtained by means of averaging of the time stamps, as detailed in section 2.4. According to the simulations, dividing the sensor into 4 segments, digitize each timestamp and compute the CTR using the averaging technique shows the worst CTR. The analog summation of the signals (4 ch SUM) strategy, which generates a single time stamp, shows better results than averaging, due to the increase in the slew rate of the signal before the comparator. Although, for a given power consumption, the best results are obtained using 1 channel, i.e., without applying segmentation. Moreover, the CTR improves when the power is increased by a factor 4 for all the configurations, since in this case, the input signal is not the limiting factor and the improvement in the slew rate is larger than the noise degradation introduced by investing more power.

Lastly, CTR measurements improve in most of the cases when increasing the signal bandwidth because it helps to preserve the larger slew rate of the input signal compared to a single photon application. Although, a signal bandwidth larger than 1 GHz does not introduce a significant improvement in CTR and complicates the design of the circuit. Therefore, this indicates that for the case studied here, an optimal bandwidth is around 1 GHz.

5 Discussion

This study has been centered in the evaluation of the SPEJ for different segmentation factors of a SiPM and different power consumption of a current-mode input stage. 65 nm technology shows the higher FOM values, meaning that it is the most efficient technology to achieve low electronic jitter. Moreover, the possible integration of a TDC in the ASIC to digitize the information, which will benefit from downscaled transistors, makes 65 nm CMOS technology the best option for developing a new ASIC for fast timing applications.

Results show that reducing the sensor area improves SPEJ, as long as the power consumption invested per readout channel remains the same, although this approach increases the power consumption per unit of detection area. This result was expected [7, 8]. However, this strategy can lead to an unmanageable heat dissipation, even employing a cooling mechanism, due to an excessive power consumption. On the other hand, if the power consumption per unit of area is kept constant, i.e., the power budget is split between the different readout channels, an optimal segmentation factor is observed between 4 and 8. The reduction in power consumption of the electronics is given by a reduction in the current flowing through the transistors, which leads to a reduction on the slew rate of the input stage. Since a smaller SiPM provides a higher slew rate and lower electronic noise due to its smaller intrinsic capacitance, it is necessary that the electronics does not degrade the slew rate in order to exploit its potential benefit in terms of time resolution.

A side effect of segmentation is that increasing the number of channels of an ASIC makes the layout more complex. Moreover, segmentation adds complexity at the PCB level since more connections need to be made between the sensor and the electronics. In addition, having a larger number of sensor segments implies that the signal paths for the different readout channels need to be equalized (at ASIC and PCB level), otherwise it can lead to time skew variations that would degrade the time resolution. Additionally, for a given detection area, each segment of the sensor needs to be separated by a gap (dead area) which means a lower fill factor and therefore lower PDE. This is important in ToF-PET applications, where PDE plays an important role in the time resolution [18, 22].

The implementation of a digital SiPM is found when taking the segmentation concept to the limit, which consists on the readout of individual SPADs with integrated electronics. A prototype of a single CMOS SPAD is presented in [23]. The area available for the electronics, require circuits optimized for the characteristics of a single SPAD and simpler than the ones presented in this work. Although no numbers regarding power consumption are detailed in [23], heat dissipation could be an issue for detection areas similar to the ones studied in this work.

Electronic noise directly affects the time resolution of a system with a leading-edge discriminator. Filtering the high-frequency noise by using an adequate bandwidth is useful to reduce the SPEJ. However, reducing the bandwidth of the system excessively degrades the slew rate of the signal. Note that the optimal bandwidth directly depends on the slew rate of the signal at the comparator input, and therefore the selected bandwidth depends on the specific sensor and the amount of photons aimed

to be detected. For signals with larger slew rate, the bandwidth limitation will reduce the slew rate of the signal leading to a worse time resolution. For instance, for a multi-photon system, the effect of filtering the bandwidth is less significant as the slew rate of the signal increases, as can be seen in figure 14 by the improvement of the CTR when using the individual readout instead of dividing the sensor in 4 segments (photons are spread between channels).

In this study, when applying segmentation and individual readout of the timestamps, the CTR results are obtained by means of an averaging algorithm. One aspect, not considered here, is that segmenting a sensor implies that the pixelated crystals behaves as monolithic crystals and therefore more sophisticated reconstruction algorithms could benefit from the information provided by higher segmentation factor. For instance, Convolutional Neural Networks (CNNs) [12, 13] or other data processing techniques that employ the timestamps, energy measurement and Depth of Interaction (DOI) information can improve the CTR. However, the individual readout of a segmented sensor has the drawback that requires a TDC to digitize each time stamp and therefore the power budget per unit area dedicated to the input stage is smaller compared to the case where only one TDC is needed (individual without segmentation and summation).

6 Conclusions

In terms of dynamic range, the three technologies studied present very similar results, 180 and 65 nm technologies starting to saturate at ~ 25 mA while 130 nm shows the best input range, saturating at ~ 30 mA. In terms of input stage time performance, Half Scaling strategy and increasing power consumption are the most efficient techniques to achieve low SPEJ. Simulation results show that the optimal segmentation factor to decrease SPEJ is between 2 and 4. 65 nm technology shows the highest FOM values, which means that it is the most efficient technology to achieve low electronic jitter. Moreover, the possible integration of a TDC in the ASIC, which will benefit from smaller transistors, makes 65 nm CMOS technology the best option for developing a new ASIC for fast timing applications. Smaller transistors lead to the improvement in the electronic jitter contribution of the TDC, which is important to keep it below 10 ps.

The results of this study show that dividing the sensor into smaller portions improves SPEJ, but it is important to not underestimate the drawbacks of the added complexity in the ASIC layout and in the PCB construction and the loss in detection area fill factor.

When segmenting a photosensor, at some point in the signal processing path, some kind of combination of the signals from each segment needs to be performed. It is important to take into account that analog summation of signals increases jitter due to the summation of noise from idle channels. The bandwidth limitation of the summation must be adjusted to filter noise but not decrease slew rate. This bandwidth limitation is very important in terms of SPEJ, but in a multi-photon system it should be as large as possible due to the larger slew rate. SiPM segmentation improves SPEJ as long as an individual channel readout (no analog summation) is performed, but the same results are not observed in CTR simulations under the conditions studied in this work (averaging algorithm and LSO scintillator crystals), where a lot of photons arrive at the sensor with a small time difference. Increasing the power consumption of the electronics and not segmenting the SiPM is the best option to achieve low CTR for the setup studied in this work, although it is not always possible to increase power consumption in a PET system due to heat dissipation. If segmentation is to be exploited, alternative methods to combine the time stamps in a post-processing stage must be explored.

References

[1] D. Sanchez et al., *HRFlexToT: A High Dynamic Range ASIC for Time-of-Flight Positron Emission Tomography*, *IEEE Trans. Radiat. Plasma Med. Sci.* **6** (2022) 51.

[2] Z. Liu et al., *In-depth study of single photon time resolution for the Philips digital silicon photomultiplier*, 2016 *JINST* **11** P06006.

[3] J. Mauricio et al., *MATRIX16: A 16-Channel Low-Power TDC ASIC with 8 ps Time Resolution*, *Electronics* **10** (2021) 1816.

[4] P. Lecoq, *Pushing the Limits in Time-of-Flight PET Imaging*, *IEEE Trans. Radiat. Plasma Med. Sci.* **1** (2017) 473.

[5] P. Padmanabhan, C. Zhang and E. Charbon, *Modeling and Analysis of a Direct Time-of-Flight Sensor Architecture for LiDAR Applications*, *Sensors* **19** (2019) 5464.

[6] J.N. Bull, J.W.L. Lee, S.H. Gardiner and C. Vallance, *Account: An Introduction to Velocity-Map Imaging Mass Spectrometry (VMImMS)*, *Eur. J. Mass Spectrom.* **20** (2014) 117.

[7] J.M. Fernandez-Tenllado et al., *Optimal design of single-photon sensor front-end electronics for fast-timing applications*, in the proceedings of the *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, Manchester, U.K. (2019) [`DOI:10.1109/nss/mic42101.2019.9059805`].

[8] D. Sanchez et al., *Multimodal Simulation of Large Area Silicon Photomultipliers for Time Resolution Optimization*, in the proceedings of the *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, Manchester, U.K. (2019) [`DOI:10.1109/nss/mic42101.2019.9059686`].

[9] P.R. Gray, P.J. Hurst, S.H. Lewis and R.G. Meyer, *Analysis and Design of Intergrated Circuits*, John Wiley & Sons Inc. (2009).

[10] S.M. Sze and K.K. Ng, *Physics of Semiconductor Devices*, John Wiley & Sons Inc. (2006) [`DOI:10.1002/0470068329`].

[11] R. del Rio-Fernandez, G. Linan-Cembrano, R. Dominguez-Castro and A. Rodriguez-Vazquez, *A mismatch-insensitive high-accuracy high-speed continuous-time current comparator in low voltage CMOS*, in the proceedings of the *1997 2nd IEEE-CAS Region 8 Workshop on Analog and Mixed IC Design*, Baveno, Italy (1997) [`DOI:10.1109/amicd.1997.637202`].

[12] A. Gonzalez-Montoro et al., *Evolution of PET Detectors and Event Positioning Algorithms Using Monolithic Scintillation Crystals*, *IEEE Trans. Radiat. Plasma Med. Sci.* **5** (2021) 282.

[13] P. Carra et al., *A neural network-based algorithm for simultaneous event positioning and timestamping in monolithic scintillators*, *Phys. Med. Biol.* **67** (2022) 135001.

[14] F. Corsi et al., *Modelling a silicon photomultiplier (SiPM) as a signal source for optimum front-end design*, *Nucl. Instrum. Meth. A* **572** (2007) 416.

[15] Cadence Design Systems, Inc., San Jose, U.S.A. .

[16] S. Jan et al., *GATE: A Simulation toolkit for PET and SPECT*, *Phys. Med. Biol.* **49** (2004) 4543 [`physics/0408109`].

[17] S. Gundacker et al., *Time resolution deterioration with increasing crystal length in a TOF-PET system*, *Nucl. Instrum. Meth. A* **737** (2014) 92.

[18] S. Gundacker et al., *Experimental time resolution limits of modern SiPMs and TOF-PET detectors exploring different scintillators and Cherenkov emission*, *Phys. Med. Biol.* **65** (2020) 025001.

[19] Cargille Labs, Cedar Grove, U.S.A. .

[20] M. Stockhoff et al., *Advanced optical simulation of scintillation detectors in GATE V8.0: first implementation of a reflectance model based on measured data*, *Phys. Med. Biol.* **62** (2017) L1.

[21] D.R. Schaart, *Physics and technology of time-of-flight PET detectors*, *Phys. Med. Biol.* **66** (2021) 09TR01.

[22] F. Acerbi and S. Gundacker, *Understanding and simulating SiPMs*, *Nucl. Instrum. Meth. A* **926** (2019) 16.

[23] F. Gramuglia et al., *A Low-Noise CMOS SPAD Pixel With 12.1 Ps SPTR and 3 Ns Dead Time*, *IEEE J. Sel. Top. Quant. Electron.* **28** (2022) 3800809.