

SiGe Time Resolving Pixel Detectors for High Energy Physics and Medical Imaging

Présentée le 6 mai 2022

Faculté des sciences et techniques de l'ingénieur
Laboratoire d'architecture quantique
Programme doctoral en microsystemes et microélectronique

pour l'obtention du grade de Docteur ès Sciences

par

Fulvio MARTINELLI

Acceptée sur proposition du jury

Prof. C. Enz, président du jury
Prof. E. Charbon, Prof. M. Nesi, directeurs de thèse
Prof. I. Peric, rapporteur
Dr A. Rivetti, rapporteur
Prof. A. Rubbia, rapporteur

Acknowledgements

It has always been difficult for me to express gratitude. To be honest, this might have been the hardest section of the thesis. I know, it may sound not very original, but I was able to write this manuscript also (and especially) thanks to all those people that helped me during these difficult, hectic, crazy and wonderful years that I will never forget for the rest of my life.

First of all, I have to thank prof. Giuseppe Iacobucci, prof. Edoardo Charbon and prof. Marzio Nessi who gave me the possibility (and the honor) to start this journey. Thank you for all the trust you put in me. I have also to extend my gratitude to Pierpaolo Valerio and Lorenzo Paolozzi, mentors and friends who allowed me to become the engineer I am right now, and to all my current and former colleagues that contributed to create the best working environment I could possibly imagine: Antonio Picardi, Chiara Magliocca, Vincenzo Silvestro, Jordi Sabater, Théo Moretti, Matteo Milanesio, Rafaella Kotitsa, Roberto Cardella, Luca Iodice, Mario Cardoso, Emanuele Ripiccini, Daiki Hayakawa, Isabella Sanna, Jihad Saidi, Mateus Vicente Barreto Pinto, Yana Gurinskaya, Andrea Pizarro and Noshin Tarannum.

Obviously, I have to infinitely thank my girlfriend Derya, my parents Michele and Luisa, my brother Daniele, my grandparents Giuseppina, Carmine, Fulvio and Rosa, my dear friend A. Landolfi and the rest of my family who always supported me even during the most difficult moments of these years.

It is necessary for me to thank my former flatmates and amazing friends Antonio Cristiano, Fabio Luchetti and Camille Gourouvadou for the wonderful time we spent living together.

I need to thank all the friends that were with me during these years and that made my road to the PhD an incredible journey like Adnan Kurtulus, Gennaro Termo, Leandro Intelisano, Francesco Giordano, Giulia Simoni, Samuele Altruda, Dario D'Andrea, Sebastiano Costanzo, Sharon Salumu, Michela Pirozzi, İlgin Kuyumcu, Tayfun Yıldız, Bianca Borsarini, Andrea Bacconi, Giulia Bigliani, Luca Pizzimento, Francesco Piro, Eugenio Senes, Salvatore Race, Dimitra Gkimousi, Roland Sipos, Luca Gardi, Luca Atzori, Antonella Catte, Alessio Amodio, Giovanni Cavallero, Silvia Gambetta, Francesco Gramuglia. A special thanks goes to my friends from Naples Alessandro Barrella, Antonio Murino, Vincenzo Fortino, Antonio Mancini, Maria Paparone, Ilaria De Benedictis, Giovanni Adamo, Armando Casino, Andrea Pennone, Manuel Noviello, Giuseppe Piro, Roberto Valente and Rossella Russo that were able to support me despite the large distance between us.

I hope I did not forget anyone. Thank you all.

Geneva, March 10, 2022

Abstract

In the last years, sub-nanosecond time-resolved particle detectors have been object of research by many companies and institutes since they represent an efficient tool to improve the performance of detecting systems for various applications such as High Energy Physics (HEP) and medical imaging. The work summarized in this thesis focuses on the design, implementation and testing of high-performance electronics in SiGe BiCMOS technology for the development of timing systems such as γ -photons detectors for PET scanners or Long-Lived Particles (LLPs) detectors for HEP applications. Design solutions and architectures are presented and analysed and their impact on the performance of timing detectors are emphasized. The main contributions of the thesis also include the development of a non-linearity model for the analysis of the impact of mismatches in free-running ring-oscillator based Time-to-Digital Converters (TDCs) and the implementation of various Genetic Algorithms (GAs) for circuit tuning. A comparison of different types of GAs is presented and the impact of their properties on the performance of the circuits to optimize is analysed.

This work describes the design solutions adopted for the implementation of two pre-production ASICs for the upgrade of detecting system of the FASER experiment at CERN. One of them was produced to perform an extensive study on different level of integration of the front-end system inside the pixel area. The second represents a smaller version of the final full-reticle ASIC for the FASER experiment. This work reports and highlights the main design challenges related to the implementation of a monolithic $23.2 \times 15.3 \text{ mm}^2$ detector.

Part of the thesis is focused on the TT-PET project. The latter aims to develop a monolithic pixel detector characterized by a $\sim 30 \text{ ps}$ time resolution to be integrated in a small-animal PET scanner. The design of a compact TDC for the TT-PET project is presented. Its architecture is based on a multi-path free-running Ring-Oscillator (RO) featuring a PLL-less event-by-event calibration system. This system is characterized by a $\sim 33 \text{ ps}$ Least Significant Bit (LSB) and its compact area, simplicity and power consumption makes this solution particularly suitable for systems in which the integration of many converters is required. Moreover, the aforementioned non-linearity model allowed demonstrating the source of the performance superiority (in terms of linearity) of the presented design. The PLL-less synchronization system was also integrated in the $< 200 \text{ ps}$ LSB TDC designed for the FASER experiment ASICs.

Despite most of the thesis work was focused on the development of monolithic systems, an IC for hybrid pixel detector is presented. The ASIC was designed to explore the limits of

Abstract

the electronic systems integrated in other systems without the constraint of the monolithic architecture and for the testing of different kind of external sensors (wire-bond or flip-chip assembled). For this reason, the design process was mainly focused on making the testing of the electronics and external circuits as flexible as possible.

Keywords: Timing, detectors, TDC, front-end, readout, GA, SiGe, BiCMOS, PET, HEP.

Résumé

Ces dernières années, les détecteurs de particules à résolution temporelle sub-nanoseconde ont fait l'objet de recherches de la part de nombreuses entreprises et instituts car ils représentent un outil efficace pour améliorer les performances des systèmes de détection pour diverses applications telles que la physique des hautes énergies (HEP) et l'imagerie médicale. Le travail résumé dans cette thèse se concentre sur la conception, l'implémentation et le test d'une électronique de haute performance en technologie SiGe BiCMOS pour le développement de systèmes de synchronisation tels que les détecteurs de photons γ pour les scanners PET ou détecteurs de Long-Lived Particles (LLPs) pour les applications HEP. Les solutions et les architectures de conception sont présentées et analysées et leur impact sur les performances des détecteurs de synchronisation est souligné. Les principales contributions de la thèse comprennent également le développement d'un modèle de non-linéarité pour l'analyse de l'impact des déséquilibres dans les Time-to-Digital Converters (TDCs) basés sur des ring-oscillators à fonctionnement libre et la mise en œuvre de différents algorithmes génétiques (GAs) pour le réglage des circuits. Une comparaison de différents types de GAs est présentée et l'impact de leurs propriétés sur les performances des circuits à optimiser est analysé.

Ce travail décrit les solutions de conception adoptées pour la mise en œuvre de deux ASICs de pré-production pour la mise à niveau du système de détection de l'expérience FASER au CERN. L'un d'eux a été produit pour réaliser une étude approfondie sur les différents niveaux d'intégration du système frontal dans la zone des pixels. Le second représente une version plus petite de l'ASIC final à particules entières pour l'expérience FASER. Cette thèse rapporte et met en évidence les principaux défis de conception liés à la mise en œuvre d'un détecteur monolithique de $23,2 \times 15,3 \text{ mm}^2$.

Une partie de la thèse est consacrée au projet TT-PET. Ce dernier vise à développer un détecteur monolithique de pixels caractérisé par une résolution temporelle de $\sim 30 \text{ ps}$ à intégrer dans un scanner PET pour petits animaux. La conception d'un TDC compact pour le projet TT-PET est présentée. Son architecture est basée sur un Ring-Oscillator (RO) multi-trajets fonctionnant en free-running et doté d'un système de calibration événement par événement sans PLL. Ce système est caractérisé par un LSB (Least Significant Bit) de $\sim 33 \text{ ps}$ et sa surface compacte, sa simplicité et sa consommation d'énergie rendent cette solution particulièrement adaptée aux systèmes dans lesquels l'intégration de nombreux convertisseurs est nécessaire.

Résumé

En outre, le modèle de non-linéarité susmentionné a permis de démontrer la source de la supériorité des performances (en termes de linéarité) de la conception présentée. Le système de synchronisation sans PLL a également été intégré dans le TDC LSB < 200 ps conçu pour les ASICs de l'expérience FASER.

Bien que la plupart des travaux de thèse se soient concentrés sur le développement de systèmes monolithiques, un circuit intégré pour détecteur de pixels hybrides est présenté. L'ASIC a été conçu pour explorer les limites des systèmes électroniques intégrés dans d'autres systèmes sans la contrainte de l'architecture monolithique et pour tester différents types de capteurs externes (wire-bond ou flip-chip assemblé). Pour cette raison, le processus de conception a été principalement axé sur la flexibilité des tests de l'électronique et des circuits externes.

Mots-clés : Timing, detectors, TDC, front-end, readout, GA, SiGe, BiCMOS, PET, HEP.

Contents

Acknowledgements	i
Abstract (English/Français)	iii
List of Figures	x
List of Tables	xvi
Glossary and Acronyms	xvi
1 Introduction	1
1.1 Time resolved detector applications	2
1.1.1 Medical imaging and TT-PET project	2
1.1.2 High Energy Physics and FASER experiment	7
1.1.3 Others	12
1.2 Thesis goals and contributions	12
1.3 Organization of the thesis	13
2 Timing detector basics	14
2.1 Transducers	14
2.1.1 Silicon sensors for ionizing radiation	14
2.1.2 Scintillators and PMTs	17
2.1.3 Other materials	19
2.2 Electronics	19
2.2.1 Front-end	20
2.2.2 Time digitization system	24
2.2.3 Readout logic	29
3 ASICs for FASER detector	31
3.1 General architecture and specifications	32
3.2 Front-end system	34
3.2.1 Pre-amplifier and discriminator design	35
3.2.2 Flavours adopted for this FASER prototype-chip	39
3.2.3 Cross-talk compensation and layout	42
3.2.4 Measurements and comparisons	43

Contents

3.2.5	Future developments and alternative solutions	50
3.3	Analog memory and MUX	55
3.4	Digital logic	59
3.4.1	Super-column logic	59
3.4.2	Periphery logic	62
4	Time-to-Digital Converters	65
4.1	Integration in timing detectors	65
4.1.1	PLL-less synchronization system and calibration	67
4.1.2	Bubble correction	69
4.2	Multi-path TDC for TT-PET project	70
4.2.1	Architecture	71
4.2.2	Non-linearity model	73
4.2.3	Layout	77
4.2.4	Post-layout simulations	78
4.2.5	Test-chip measurements and state-of-the-art comparison	81
4.3	Low-power TDC for FASER experiment	86
4.3.1	Architecture and integration in pre-shower detector chip	86
4.3.2	Measurements	89
4.4	Possible improvements and new features	91
5	Readout IC for Hybrid Pixel Detectors	93
5.1	Motivation	93
5.2	System description and architecture	94
5.2.1	Front-end	96
5.2.2	Digital logic	101
5.2.3	TDC	103
5.2.4	Power gating system and biasing	104
5.3	Future developments	105
6	Genetic Algorithms for Circuit Optimization	106
6.1	Introduction to Genetic Algorithms	106
6.2	Design framework for D-Latch optimization	109
6.3	Chosen algorithms and implementation	111
6.3.1	GA with linear aggregation	111
6.3.2	Second implementation of a linear aggregation-based GA	113
6.3.3	Third implementation of a linear aggregation-based GA	115
6.3.4	Pareto-based evolutionary algorithm	117
6.4	Discussion and comparisons	118
	Conclusions	120
A	Ionization and Bethe-Bloch formula	123

	Contents
B The Shockley-Ramo theorem	125
C Chip Gallery	127
Bibliography	128
Curriculum Vitae	137

List of Figures

1.1	Representation of true (a), random (b) and scatter coincidence due to Compton effect (c) in a generic PET scanner. The annihilation points are depicted in red while in (c) the blue cross indicates a photon scattering event.	4
1.2	True coincidence in a non-TOF (a) and TOF (b) PET scanner.	5
1.3	Comparison of the reconstruction of a Derenzo phantom without (left) and with (right) time of flight (a ~ 30 ps coincidence time resolution was considered). Source: [12].	5
1.4	Small animals scanner geometry (a) and cross-section of two consecutive detection layers (b) of the TT-PET project. The picture (a) was generated with Geant4. Source: [12].	6
1.5	Reconstruction of 78 vertices in CMS detector system. Source: [31].	9
1.6	Position of the FASER experiment with respect to the LHC (a) and 3D model of the FASER detector (b). The LHC representation in (a) is not in scale. Source of (b): [43].	10
1.7	Two fermion signal (top) and two photon signal (center and bottom). The latter can only be sensed with the new version of the pre-shower detector (bottom). On the right, new pre-shower architecture depicted in GEANT4 simulation. . .	11
2.1	Schematic of the doping profile and electric field (E) distribution in a generic silicon sensor without (a) and with (b) gain layer.	15
2.2	Bias configuration of a set of SPADs to implement an Analog SiPM featuring quenching resistors R_q	16
2.3	Scintillator coupled with a PMT. Source: [54].	17
2.4	Schematic of a generic front-end system for the readout of a timing detector. Source: [21].	20
2.5	Working principle of a discriminator with hysteresis. Left: input-output characteristics (non-inverting in this example). Centre: no-hysteresis discriminator. Right: hysteresis discriminator. A noisy input can make the circuit switch several times as the signal crosses the threshold V_{th} . On the other hand, a discriminator with hysteresis shows a different equivalent threshold depending on the state of the output ($V_{th-lh} > V_{th-hl}$).	21

2.6	Schematic of a charge amplifier (a) and simplified input-output characteristics over frequency (b). C_{in} (depicted in red) is the equivalent input capacitance and can be calculated with Miller's theorem.	22
2.7	behavior of the ENC of a charge amplifier as function of the shaping time τ_m for various active devices. Source: [66].	22
2.8	RO-based TDC general schematic.	25
2.9	Schematic of an interpolating VCO composed of NOR cells. CT is used as a control signal to modulate the oscillation frequency. Source: [79].	27
2.10	Vernier line TDC. Schematic (on the left) and working principle (on the right). .	28
2.11	Simplified block diagram of the compression logic of the FASER prototype chip.	29
3.1	Architecture and pixel distribution of the final version of the FASER pre-shower detector chip.	32
3.2	FASER chip substrate cross-section representation (a) and GEANT4 simulation of a typical event in the detector (b) [91].	33
3.3	Super-pixels and pixel size (left) and a photograph of the prototype ASIC (right); the ASIC total area is $1.7 \times 2.6 \text{ mm}^2$	35
3.4	Pre-amplifier architecture. The stage and resistance values highlighted in red refer to the inverting solution that characterizes only one of the flavours adopted for the test chip.	36
3.5	Configuration of the first pre-amplifier stage integrated in the first prototype (left) and the pre-production chip (right) for the FASER pre-shower.	37
3.6	Discriminator schematic and connection to the pre-amplifier.	39
3.7	Distribution in the pixel matrix of the front-end variants of this test chip.	40
3.8	Layout of 2×2 pixels (left) and zoom (right) on the shielding line to reduce the cross-talk between the output of the discriminator and the adjacent pixel wells.	41
3.9	Discriminator output and pixels signal with (right) and without (left) cross-talk protection lines for an input charge of 0.5 fC	41
3.10	Discriminators output behavior with (orange) and without (blue) self-induced noise compensation lines.	42
3.11	Example of self-induced noise compensation line. The metal2 line is connected to the negative discriminator output to increase the coupling with the pixel well and avoid self oscillation induced by the effect of the positive output.	42
3.12	Mapping of the pixels in the eight groups used for the threshold calibration. . .	44
3.13	DAC code (a), DAC average LSBs (b) and front-end threshold (c) distribution in the prototype chip. The x-axis indicates the pixel number. The pixel are labeled from bottom to top of Figure 3.7 and split in even and odd on the right and the left part of the chip. The results in (a) were obtained setting the global threshold at 1.1 V	45
3.14	(a) Threshold scan representation for gain evaluation and (b) event rate as function of the threshold for different front-end configurations. The x-axis of (b) is referred to the baseline.	47

List of Figures

3.15	Gain of a selection of pixels that have been measured with the ^{109}Cd source for the five front-end configurations.	48
3.16	Event rate as a function of the threshold obtained with a ^{55}Fe radioactive source.	49
3.17	Peak measuring system schematic and working principle.	51
3.18	Laser trigger, reset (active low) and slow discriminator output signals for the tests of the peak-measuring system.	52
3.19	Event rate as function of trigger delay for various laser intensities. The results with the peak-measuring system activated were obtained choosing $bias_{delay} = 17\mu\text{A}$ and $bias_{discharge} = 40\text{ nA}$	53
3.20	TOT (a) and jitter over the average TOT variation (b) as function of laser attenuation.	54
3.21	Average TOT as function of the discharge current. This characteristics was calculated setting the laser attenuation at 68%.	55
3.22	Analog memory schematic.	55
3.23	Simple (a) and proposed (b) design solutions for the implementation of the analog multiplexer. In this picture a 3-bit example is reported.	56
3.24	Impact of the charge sharing in the simple (a) and proposed (b) solutions for the analog multiplexer. The simulations have been performed on the 256-to-1 (8-bit) schematic integrated in the pre-production chip. In the plot (b), the drop of the switch output before the rising edge of the gating signal is due to a reset signal used to discharge the output capacitance of the MUX before reading another pixel. More details in Section 3.4.1.	57
3.25	Charge sharing problem caused by address switching during polling (left) and proposed solution (right).	58
3.26	Super-column logic (dark gray), periphery and TDC block diagram (not in scale).	59
3.27	Pixel well and test-pulse generator schematic.	60
3.28	Simplified time diagram of the polling of the pixels measured charge.	61
3.29	Amount of data to readout in the frame-based solution adopted in FASER pre-production ASIC and in a packet-based solution as function of the amount of pixels stimulated during the events (a square distribution is assumed).	63
4.1	Possible disposition of a 4×4 pixel matrix connected to different TDCs through fast-OR blocks. The squares represent generic structures composed of pixels (active area) and front-end system (preamplifier and discriminator).	66
4.2	Block diagram of the system for the event-by-event calibration.	67
4.3	Reference clock signal CLK (up) and gating signals G_j (down).	68
4.4	Bubble correction algorithm implementation.	70
4.5	Delay cell (a) and buffer (b) of the proposed RO.	71
4.6	Architecture of the proposed RO. I and If represent the input of the cells connected to the direct and feedforward paths respectively. OB indicates the buffer outputs.	72

4.7	Connections of the delay cells of the RO. Inverting the inputs of one of the stages allows the oscillation of the system.	72
4.8	Schematic of the latches used to sample the state of the RO.	72
4.9	An example of a 5 stage multi-path RO with two types of feedforward connections (red dotted line: proposed solution). A mismatch on the delay of the first buffer Δ_0 is assumed for this analysis.	74
4.10	RMS (top) and maximum of the absolute value (bottom) of DNL as function of η of both of the solutions depicted in Figure 4.9 (calculated with Equation 4.6 for the usual connection case, with Equation 4.14 for the proposed solution scenario and exploiting the edge time distribution of Equation 4.9 for the more detailed model).	75
4.11	RMS (top) and maximum of the absolute value (bottom) of DNL as function of the cell delay (calculated with Equation 4.6 for the usual connection case, with Equation 4.14 for the proposed solution scenario and exploiting the edge time distribution of Equation 4.9 for the more detailed model).	76
4.12	Picture of the test chip of the proposed TDC (total area: $0.9 \times 0.9 \text{ mm}^2$) (a) and layout of the RO with cells disposition (b).	77
4.13	LSB and power consumption of the TDC for typical, Fast/Fast (F/F), Fast/Slow (F/S), Slow/Fast (S/F) and Slow/Slow (S/S) corners and for V_{DD} equal to 1.4 V and 1.6 V.	78
4.14	Maximum values and RMS distributions of DNL and INL calculated over various Monte Carlo simulations. In this case, the supply $V_{DD} = 1.6 \text{ V}$	79
4.15	Measured output distribution (before and after correction) of the TDC for $V_{DD} = 1.6 \text{ V}$ and for all the latch stages connected to the RO.	80
4.16	Probability density function of the quantization error for each latch stage ($V_{DD} = 1.6 \text{ V}$).	81
4.17	Block diagram of the measurement system to evaluate the SSP of the converter.	82
4.18	Output distribution of the data obtained with the measurement system depicted in Figure 4.17 for $V_{DD} = 1.4 \text{ V}$	83
4.19	Zoom of the power spectrum of the divider output for $V_{DD} = 1.6 \text{ V}$ around the fundamental component of the signal.	83
4.20	Area and LSB of the presented TDC compared to the works of Table 4.1 and the ones reported in [103–115]. The size of the dots on the plot is proportional to the power consumption of the analyzed TDCs (logarithmic scale).	84
4.21	Block diagram of the TDC designed for the demonstrator chip.	87
4.22	Integration of the TDC inside the readout logic a $\sim 96 \%$ cell density.	88
4.23	Integration of the TDC in the new pre-production ASIC for FASER pre-shower (left) and block-diagram (right).	89
4.24	Average output words obtained with different clock frequencies as function of the time difference (top) and words progression with linear fit (bottom) for supply voltage $V_{DD} = 1.4 \text{ V}$	90

List of Figures

5.1	Block diagram of the readout chip (top) and layout (bottom).	94
5.2	Shape and size (in scale) of the pads of the matrix and their passivation aperture.	94
5.3	Front-end configurations for each mini-block in the pad matrix.	95
5.4	Schematic of the pre-amplifiers of low and high Q_{in} front-end configurations. Differences are highlighted in red.	97
5.5	Schematic of the TT-PET pre-amplifier configuration of the readout chip.	97
5.6	TOT dynamic range for the low Q_{in} front-end configuration for various values of delay bias current I_{delay} (top) and for the high Q_{in} and TT-PET variants (bottom).	98
5.7	Standard deviation of the TOA as function of the input charge for the low Q_{in} (a), high Q_{in} (b) and TT-PET (c) front-end configurations.	100
5.8	Simplified block diagram of the Readout Control system. The PG word sent to the encoding blocks in the first part of the readout are stored in a separate register since they are used to identify the first pixel that sensed a hit in the readout time frame.	102
5.9	Time diagram example of the output data and PG words during readout.	102
5.10	Chip and GPIO board communication for the periphery logic based on SPI standard.	103
5.11	Switches used for the power gating of the readout chip matrix.	104
6.1	Representation of the Pareto-optimal front of an evolutionary algorithm with $M=2$ optimization parameters F_1 and F_2	108
6.2	Block diagram of a generic GA for the optimization of D-latches.	109
6.3	Setup time calculation for D-latches.	110
6.4	behavior of the best solution over generations in terms fitness, area, setup time and power consumption with the first algorithm based on G3 model.	112
6.5	behavior of the best solution over generations in terms fitness, area, setup time and power consumption with the second algorithm based on G3 model. This time the fitness function directly includes P_C and A	114
6.6	behavior of the best solution over generations in terms fitness, area, setup time and power consumption with the third algorithm based on G3 model with a different mutation operator.	115
6.7	Comparison between the first and last generation of the Pareto-based evolutionary algorithm. For the sake of simplicity, only power consumption and setup time have been considered in the plot.	116
A.1	Average stopping power (energy loss) in various materials (from top: liquid hydrogen, helium, carbon, aluminium, iron, tin and lead). Source: [130].	123
B.1	Representation of the problem solved by the Shockley-Ramo theorem.	125
C.1	FASER first prototype ASIC.	127
C.2	Feedforward TDC test-chip.	127
C.3	Layout of the three FASER pre-production ASIC variants.	127

C.4 Readout ASIC layout.	127
----------------------------------	-----

List of Tables

1.1	Comparison of the main characteristics and specifications of FCC and HL-LHC. Sources: [36] (FCC), [32, 38] (HL-LHC).	9
2.1	Photoelectric absorption (σ_{abs}), scattering (σ_{scat}) and total (σ_{tot}) cross section of lead (Pb) and silicon (Si) referred to 511 keV γ -photons. Source: [55]	18
2.2	Pros and cons of the discussed TDC architectures.	29
3.1	RMS threshold dispersion $\sigma_{V_{th}}$ for each front-end configurations integrated in the chip.	46
3.2	Noise contribution (σ_v), charge gain (G_c), ENC of one channel for each front-end configurations integrated in the chip. The error associated to G_c does not represent the channel-to-channel dispersion but is the uncertainty on the gain measurement of the analyzed channel.	50
4.1	Multi-path simulations and measurements results. A comparison with other works is also reported.	85
4.2	Performance of the TDC integrated in the first prototype of the FASER pre-shower detector.	90
5.1	Expected performance of the three front-end configurations integrated in the readout chip. The gain and the ENC of the pre-amplifier have been calculated as the average gain in the input charge range of interest. However, since in the high Q_{in} configuration the pre-amplifier saturates for $Q_{in} > 3$ fC, the gain reported in the table was extracted from simulations performed with lower values of input charge. The gain and the ENC reported for the TT-PET configurations are the values for 1 fC and 5 fC Q_{in} respectively. For $Q_{in} > 5$ fC the front-end saturates.	99
6.1	Comparison of the presented GAs	119

Glossary and Acronyms

Band gap Distance in terms of energy between the valence and the conduction band of a certain material.

BiCMOS Technology for ASIC production in which bipolar transistors and CMOS devices can be integrated in the same die.

DNL Differential Non-Linearity. In a converter (ADC, DAC or TDC), it is a function that describes the difference between the step widths of the input-output characteristics of the converter and the ones of an ideal version of it.

ENC Equivalent Noise Charge. For a certain charge amplifier, it is defined as the input charge that has to be injected in a noiseless version of the amplifier for which the corresponding output voltage has the same power of the noise of the original system when no input is set.

FASER ForwArd Search ExpeRiment. Experiment at CERN for the research on dark matter through the detection of long-lived particles. It extends the physics program of other experiments at LHC such as ATLAS and CMS.

Flip chip Technique to interconnect integrated circuits with PCBs, their packages or other ASICs through the usage of solder bumps. It generally allows a reduction of the connection size compared to wire bonding.

FWHM Full Width at Half Maximum. Width of a function $y = f(x)$ calculated as difference between the values of x such that the corresponding y are equal to half of the maximum of f . For a gaussian distribution, the FWHM is ~ 2.355 times its standard deviation.

HBT Heterojunction Bipolar Transistor. Bipolar transistor characterized by different materials for its emitter, base and collector regions.

HEP High Energy Physics. Branch of physics that studies particles and their interactions, often through the usage of particle accelerators and colliders.

INL Integral Non-Linearity. In a converter (ADC, DAC or TDC), it is a function that describes the difference between the input-output characteristics of the converter and the ideal one. It can be calculated as the integral of the DNL.

Glossary and Acronyms

LFSR Linear Feedback Shift Register. A shift register in which the input bit is a linear function of the state of each element of the system.

LHC Large Hadron Collider. A 27 km circular particle accelerator at CERN.

LOR Line of Response. In PET-based imaging it represents the region inside the scanner in which pairs of detectors can locate annihilation points.

Luminosity In a particle collider, it is defined as the ratio between the number of events per second and the cross section of the collision events under exam.

Medical imaging Diagnostic technique based on generating a set of images of a patient under exam.

PET Positron Emission Tomography. A medical imaging technique used for metabolic process observation and based on the detection of 511 keV photons produced by the annihilation of positrons in the body of the patient under exam.

Phase noise Random fluctuations of the phase of a signal, often represented in the frequency domain.

RO Ring Oscillator. Circuit composed by an odd number N of inverting gates which oscillate at a certain frequency that depends on N and the delay of the cells.

Setup time For latches or flip-flop, it represents the time interval before the active clock edge of the device in which the input signal has to be stable in order to guarantee the correct working operation of the latch or the flip-flop and to avoid metastability.

SNR Signal-to-Noise Ratio. A parameter used to compare the intensity of a signal with the noise that affects it.

SPI Serial Peripheral Interface. Interface used for the synchronous serial communication among ASICs or microcontrollers. It is based on a master-slave architecture.

SSP Single Shot Precision. For a TDC, it is defined as the standard deviation of the output distribution produced by the converter when it is repeatedly stimulated with the same input time interval.

TDC Time-to-Digital Converter. System able to measure time intervals. The result of the measurement is provided as a digital representation of the period of time to evaluate.

TOF Time of Flight. Measure of the time that a particle (or, more generally, an object) takes to cross a certain medium or material.

TT-PET project Thin TOF-PET project. Collaboration among CERN, the Department of Nuclear Physics (DPNC) at University of Geneva, University of Rome "Tor Vergata", University of Bern, Hôpital Cantonale de Geneve and University of Stanford for the development of a small-animal PET scanner with a time resolution of 30 ps RMS.

Wire bonding Technique to interconnect integrated circuits with PCBs, their packages or other ASICs through the usage of metal wires.

1 Introduction

Since the beginning of the 20th century, many progresses have been made in the field of particle physics due to the development of efficient systems for the detection of various types of particles. Over the years, several detectors played a crucial role for fundamental scientific discoveries e.g., radioactivity (Becquerel in 1896), atomic nuclei (Rutherford in 1911), electron-positron pairs (Blackett in 1948), pions (Powell in 1950) [1].

Notable and more recent results have been achieved with the experiments at the Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN) in Geneva as the discovery of the Higgs boson in 2012 [2]. As it will be clearer in the course of this chapter, in the last few decades, particle detectors have been exploited also for other applications, such as medical imaging [1].

An efficient way to improve the performance of particle detectors is to design systems with time-measurement capabilities. Time resolved detectors are useful to enhance the discrimination of different kind of particles and to better reconstruct the phenomena under analysis: the development of a high time resolution system is crucial (and necessary for some application) to improve the performance of detectors integrated in systems like particle accelerators or tomography scanners.

This chapter aims to highlight the importance and the role of time-resolved particle detectors describing the most important applications in which this kind of systems can be adopted. Section 1.1 mainly focuses on medical imaging and High Energy Physics but also gives a brief introduction on other applications in which precise time measuring systems are required.

Section 1.2 reports the objectives and the contributions of the work performed during the doctoral program, that will be developed in the next chapters of the thesis. Finally, Section 1.3 describes how the thesis is organized.

1.1 Time resolved detector applications

1.1.1 Medical imaging and TT-PET project

With the term "medical imaging" it is possible to indicate all the techniques exploited to generate a set of images of a patient for diagnostic purposes [3] [4]. The first and oldest medical imaging technique is radiography. This procedure is based on exposing the body of the patient under exam to a source of X-rays. A detector is then used to analyze the interactions of the latter with the body (e.g., absorption, scattering, reflection) [3]. Other examples of medical imaging techniques are Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). CT is based on the same principle of a traditional radiography system: the interactions of a X-ray source with the body of the patient is exploited to analyze the properties of its tissues. However, this time the X-ray source is rotating around the body and the collection of all the data provided by the detectors allows the generation, through a computer, of a high-contrast image of the patient [3]. For what concerns the MRI, this technique exploits the properties of the protons inside the body of a patient to align to highly intense magnetic fields and radio-frequency signals. As reported in details in various textbooks as [3] [5], MRI scanners are able to discriminate various types of tissues on the basis of how they react to electromagnetic stimulus.

The aforementioned imaging techniques do not require detector systems able to sense particles or radiations with a high time accuracy even if various researches are exploring the possibility of using the timing information to improve the performance of various medical imaging scanners like the ones used for CT [6]. However, timing detectors are extensively used for the enhancement of Positron Emission Tomography (PET) scanners. Part of this thesis was focused on the design of high performance electronic systems for the PET scanner detector of the TT-PET project in collaboration with University of Geneva and CERN.

Positron Emission Tomography

Positron Emission Tomography (PET) is a medical imaging technique that is used for metabolic process observation [7]. Unlike X-ray tomography or MRI, PET does not provide a structural image of the body under exam but it allows to analyze the physiological functions and biological properties of organs and tissues. This technique is exploiting the emission of positrons for the reconstruction of the patient image. A radiopharmaceutical, composed by a radioisotope and a ligand, is injected in the body [7]. The ligand is usually a glucose and it is used to interact with the body by being metabolized and captured by tissues. One of the most common radio pharmaceuticals used for PET applications is fludeoxyglucose ^{18}F (often indicated with FDG) which contains the radioactive isotope fluorine-18. In [7], a list of radionuclides that can be used for PET is reported. The main properties that radionuclides for PET are required to have is related to their ability to achieve β^+ decay:



Equation 1.1 indicates that when a radioisotope ${}^A_Z X$ β^+ decays in ${}^A_{Z-1} Y$ an electron neutrino ν_e and a positron e^+ are generated. After approximately 10^{-10} s, a process of annihilation of the positron with an electron starts, generating two back-to-back γ -photons with an energy of 511 keV each i.e., the angles between the photons is almost 180° . The detection of these couples of photons generated in the body can be used for the reconstruction and the calculation of the annihilation point of the positrons. Indeed, the radiopharmaceuticals tend to accumulate in organs or tissues of interest like malignant cancers [8], generating a statistically relevant excess of photon emissions. For this reason, one of the most important PET application is clinical oncology. However, PET scanners can be used also in neurology: in the case of pathologies such as Alzheimer's disease, the brain metabolism is lower or, more in general, different compared to the one of a healthy patient and a PET scanner is useful to monitor the use of glucose in the brain and its activity [9].

The structure of a PET scanner surrounds the patient and it is composed by a set of detectors that are able to sense the above-mentioned γ -photons. The localization of the annihilation points is performed with the reconstruction of the so-called Lines of Response (LOR): if two photons are detected in a certain time frame, indicated with coincidence window, it is possible to assume that they were generated from the same annihilation point. This situation is indicated with the name of "true coincidence". The annihilation points will lay on the LOR that connects the regions of the detectors in which the photons were sensed. The computation of more LOR allows reconstructing the position of the statistically relevant annihilation points in the body. In Figure 1.1a a representation of a true coincidence is depicted. However, the actual reconstruction of the position of the points of interest is more complicated. The radioisotopes injected in the patient diffuses in all of the body and this determines the emission of positrons (and so of photons) everywhere with an higher concentration in the organs or tissues in which the metabolism is faster. For this reason the sensors will detect a higher rate of events (generally millions per second) but only a small fraction of them represents the signal of interest. The scanner is able to generate the correct final image using a reconstruction algorithm that will be able to reveal only the points in which the excess of decays is statistically relevant. Moreover, true coincidences represent only one of the events that the scanner is able to detect. Indeed, from two separate annihilation points a couple of unrelated photons can be sensed by the scanner in the same coincidence window. This kind of event, indicated with "random coincidence" and depicted in Figure 1.1b, leads to an incorrect evaluation of the LOR (dashed line in the figure). Furthermore, one or both of the two γ -photons generated from a certain annihilation point can be deflected due to Compton effect as shown in Figure 1.1c [7]. This event is indicated as "scatter coincidence". It is important to underscore that both scatter and random coincidences are reducing the quality of the reconstruction and the signal-to-noise (SNR) of the image since, as demonstrated in [10]

$$SNR \propto \sqrt{\frac{T^2}{T + S + R}}, \quad (1.2)$$

where T , S and R are true, scatter and random coincidence rates respectively.

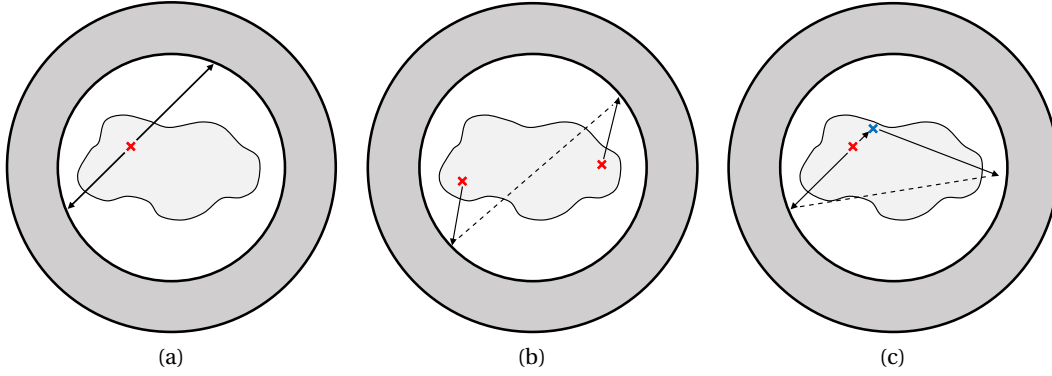


Figure 1.1: Representation of true (a), random (b) and scatter coincidence due to Compton effect (c) in a generic PET scanner. The annihilation points are depicted in red while in (c) the blue cross indicates a photon scattering event.

Importance of timing in PET scanner detectors

As explained in [8], it is possible to enhance the quality of the reconstructed image by measuring the Time of Flight (TOF) of the γ -photons produced in a PET scanner. Figure 1.2a depicts a representation of a true coincidence in a conventional non-TOF scanner. If two photons are detected, the annihilation point can be anywhere on the LOR and only more coincidences will allow the system to determine its exact position. A TOF PET scanner, as shown in Figure 1.2b, is also able to calculate the difference between the times in which the back-to-back photons hit the detectors (often indicated with TOF difference [11]). In this way, a more precise evaluation of the position of the annihilation point is possible because the latter will be statistically confined in a reduced portion of the LOR [8]. It can be demonstrated that the improvements in terms of SNR is given by the following relation [11]:

$$SNR_{TOF} \approx SNR_{non-TOF} \sqrt{\frac{D}{c\sigma_t}} \quad (1.3)$$

where SNR_{TOF} and $SNR_{non-TOF}$ are the SNR with and without TOF information respectively, D is the scanner diameter and σ_t is the time resolution of the detectors (i.e., the uncertainty in the time evaluation). Figure 1.3, reported in [12], highlights the impact of TOF information to distinguish and detect small annihilation points (down to 0.5 mm, as described in the paper) in a Derenzo phantom. The improvement in the image reconstruction given by Equation 1.3 is useful to facilitate the detection of smaller regions of interest and to reduce the dose of potentially harmful radionuclides to be injected in the patient under exam: indeed, in non-TOF PET scanners it is possible to partially compensate the reduced reconstruction capability increasing the annihilation events to be detected.

The work reported in [13] describes the development of the CPS Hi-Rez PET scanner, characterized by a 1.2 ns Full-Width-Half-Maximum (FWHM) time resolution, leading to an SNR increase of 50 % compared to a non-TOF PET scanner. The work also addresses the important

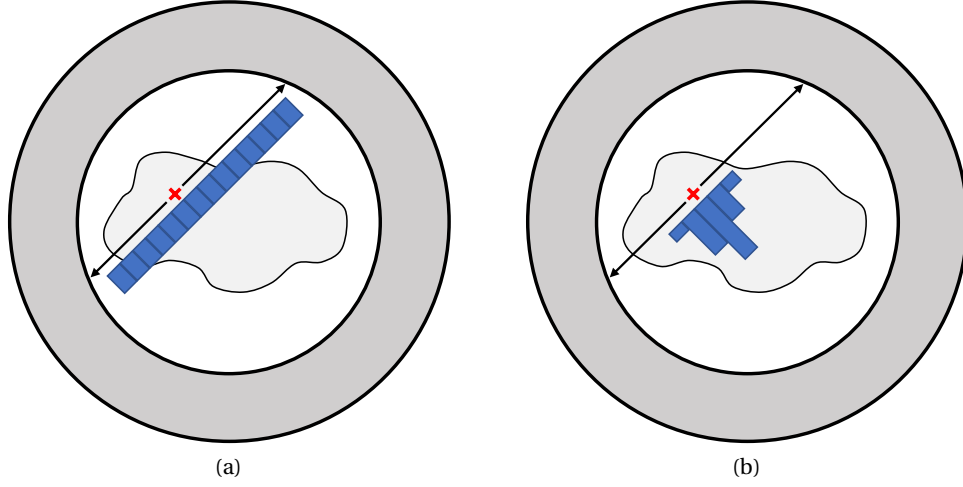


Figure 1.2: True coincidence in a non-TOF (a) and TOF (b) PET scanner.

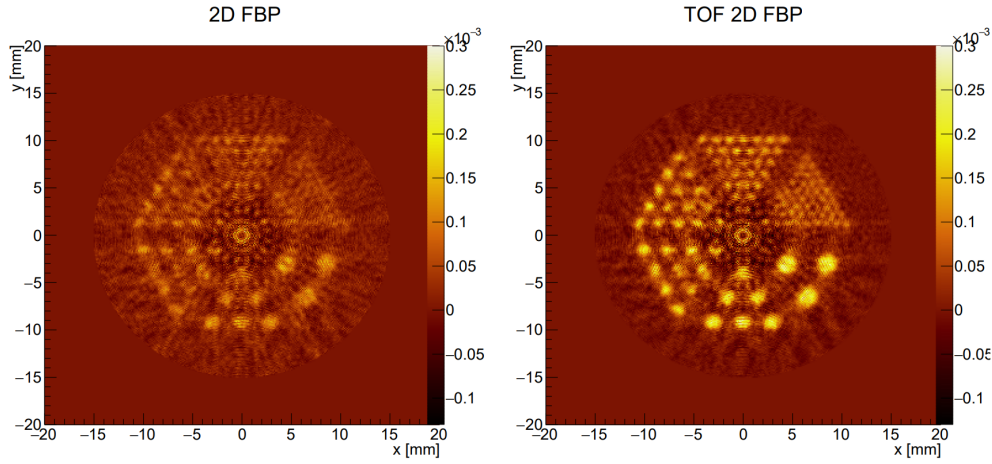


Figure 1.3: Comparison of the reconstruction of a Derenzo phantom without (left) and with (right) time of flight (a ~ 30 ps coincidence time resolution was considered). Source: [12].

roles of scintillator crystals (that will be briefly described in Section 2.1) for the development of high-time resolution PET scanners. In 2011, the Biograph mCT TOF PET/CT scanner from Siemens Medical Solutions was shown in [14]. This system is able to measure the TOF of the γ -photons with a 527.5 ps (FWHM) resolution. Other works [15, 16] describe detectors for PET scanners with a FWHM resolution smaller than 100 ps, underlining the importance of an efficient coupling of the transducer (scintillator) with the rest of the system (Section 2.1) and the crucial role of the electronics in the overall performance of the scanner (more details in Section 2.2).

Most of the currently available commercial PET scanner are characterized by a time resolution in the order of few hundreds of picoseconds. However, as stated in various works like [17–19], one of the most ambitious goal for PET researchers is the design of a 10 ps time resolution

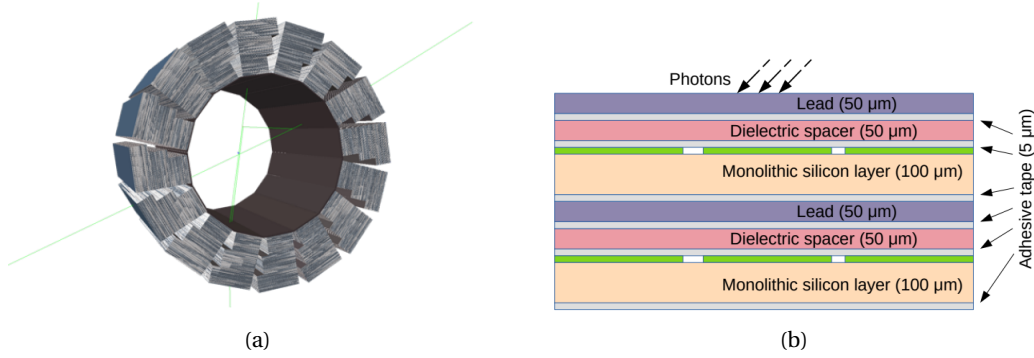


Figure 1.4: Small animals scanner geometry (a) and cross-section of two consecutive detection layers (b) of the TT-PET project. The picture (a) was generated with Geant4. Source: [12].

scanner. In this case, the spatial uncertainty on the LOR would be approximately 1.5 mm and the system will not need to perform any kind of reconstruction algorithm. Indeed, this accuracy would be similar or smaller than minimum achievable spatial resolution of a PET scanner and the TOF information would directly give an indication of the volume element (often indicated as voxel) of the space within scanner in which the annihilation of the positrons occurred. As explained in [7], the theoretical limiting factor in the spatial PET resolution is related to the mean free path of the positron between the point in which it is emitted and the one in which annihilation occurs, often indicated with positron range. This value oscillates between fractions of a millimeter and few millimeter. Moreover, it is possible to demonstrate that the positron range is also dependent on the radionuclide [7]. Because of these reasons, the development of a ~ 10 ps PET scanner would represent a crucial milestone for the medical imaging community.

The TT-PET project

Part of the work performed for this thesis was focused on the design of various circuits and systems for the detector of the Thin TOF-PET (TT-PET) project [8, 12, 20–24].

The TT-PET project is a collaboration among CERN, the Department of Nuclear Physics (DPNC) at University of Geneva, University of Rome "Tor Vergata", University of Bern, Hôpital Cantonale de Geneve and University of Stanford. Its main goal is the development of a small-animal PET scanner with a time resolution of 30 ps. In order to fulfil this specification, a 130 nm Bi-CMOS technology was used for the detector ASIC.

The structure of the scanner is depicted in Figure 1.4a. This architecture is meant to be integrated in a MRI scanner for the analysis of small animals. The structure is composed of 16 units denominated cells or towers. Each of them is characterized by 60 detection layers with a length of 50 mm and a width of 7 mm, 9 mm or 11 mm. The different values of widths of these layers and the resulting shape of the towers were chosen to maximize the detection efficiency of the scanner. The detector systems in each layer are monolithic i.e., the active area (transducer) and the electronics that senses the signal are integrated on the same die.

As indicated in [25], this solution has its main advantage in a lower production cost than the hybrid alternative (in which the active area is on another chip as it will be shown in Chapter 5). Conventional PET scanners are usually featuring a set of scintillator crystals coupled with a Silicon Photo-Multipliers (SiPM) to convert the 511 keV photons into electrical signals for the front-end system. However, scintillators may represent a limiting component for the resolution of the scanner (see Section 2.1 for more details). For this reason, the TT-PET detector aims to implement the γ -photons conversion avoiding this type of devices. In Figure 1.4b the cross section of the detector is depicted. The scintillator-SiPM structure is replaced by a system composed by a layer of high atomic number material (e.g. lead with $Z=82$) and a silicon sensor (Section 2.1 will present more details about this kind of conversion). The total thickness of the detection layer is 200 μm , and it is divided in 100 μm of silicon sensor layer, 50 μm of lead and 50 μm of flexible PCB, used principally for the extraction of the signals. A demonstrator chip for the TT-PET detector is described in [24]. The system was tested on a beam line of Minimum Ionizing Particles (MIPs, see Section 2.1) and achieved a time resolution of approximately 110 ps. This result is consistent with the target 30 ps resolution due to the MIPS operated for the tests (approximately a third of the measured resolution is expected with the charges produced by the 511 keV photons). More details about the circuits solutions implemented for the chip related to the TT-PET project are discussed in Section 2.2.

1.1.2 High Energy Physics and FASER experiment

Time-resolved particle detectors can be used in many High Energy Physics (HEP) experiments such as the ones exploiting particle accelerators. Since the 1950s, physicists have benefited from the great possibilities offered by accelerators for the study and the discovery of new particles [1]. For example, the Large Hadron Collider (LHC) at CERN was used for the discovery of the Higgs boson [2] through a system capable of producing proton-proton collision with a 40 MHz bunch crossing frequency [26, 27]. Indeed, as its name indicates, the LHC is a collider, a particular type of circular accelerators in which particle beams traveling in different directions are interacting through high energy collisions. The products of these interactions are then analyzed with various detectors at specific stations along the accelerator. One of the main properties of a this kind of machine is the luminosity \mathcal{L} . This parameter can be calculated as [1, 27]

$$\mathcal{L} = \frac{N}{\sigma} \quad (1.4)$$

where N is the number of event per second generated into the accelerator and σ is the cross section of the collision event under exam. As explained in [28, 29], many experiments during the last years have been moving towards the direction of high luminosity colliders. The LHC is characterized by a $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ luminosity which gave the possibility to a large community of scientists to improve the research of a new physics beyond the standard model. However, an upgrade of the collider, indicated with Hi-Lumi LHC (HL-LHC), is planned for the 2020s and the expected peak luminosity is approximately 5 times larger than the previous one [30]. A brief analysis on the importance of high-resolution timing detectors for HEP applications is

reported below.

Timing detector requirements for HEP experiments

The increase of the luminosity of an accelerator ([28] indicates that this trend will be followed by many machines besides LHC) requires the development of a detecting system that is able to work at the higher event rates. As described in [26], detectors characterized by excellent timing performance are fundamental for high luminosity collider-related experiments for various reasons, including the above-mentioned event rate and the pile-up suppression.

During the operation of an accelerator, only a significantly small portion of the total amount of the produced events ($<0.001\%$ for ATLAS experiment at LHC as reported in [26]) will be actually sensed and registered by the detectors of the machine. If the luminosity increases, the number of events will also rise and discarding the non-significant events will be a fundamental procedure for the correct operation of the accelerator especially in high luminosity cases. Thus, using time resolved detectors with a resolution of various hundreds of picoseconds is necessary for a more efficient discrimination of the significant events [26]. However, more demanding specifications are required to deal with the problem of pile-up. In a collider, pile-up indicates the reduction of detection efficiency caused by multiple hits in a short timeframe. High luminosity accelerators are characterized by large rates of interactions generated in different moments (in the same or neighbouring bunch-crossing) but close in terms of space, indicated as pile-up events, that cause a reduction of the ability of the detection system to efficiently reconstruct the interaction points. Therefore, timing detectors can be useful to improve the pile-up suppression resulting in a better tracking accuracy and a reduction of the background noise. For the above-mentioned HL-LHC, [32] reports that the proposed High-Granularity Timing Detector (HGTD) is expected to be characterized by a 30 ps time resolution that will improve the pile-up suppression of a factor ~ 6 .

Figure 1.5 shows the reconstruction of 78 events of a single bunch crossing in the detector of the CMS experiment at LHC. In the picture, it is possible to see that various events can approximately overlap in space (especially in the central part). As underlined in [33], the high luminosity of HL-LHC will increase the number of events per bunch crossing up to 200 making the exploitation of the timing information necessary for the reconstruction of the collision points. The work presented in [33] also addresses the importance of measuring the timing characteristics of each hit in order to better reconstruct the tracks associated to the various events.

Trackers with timing capabilities are indicated as 4D trackers [33]. The HGTD, reported in [32], represents an example of state-of-the-art timing detector for 4D tracking with an expected 30 ps time resolution and a large set of high demanding performance parameters to sustain the specification of HL-LHC. Other example of notable detector at CERN is the Gigatracker [34] used for the NA62 experiment [35] at the Super Proton Synchrotron (SPS) characterized by a 65 ps time resolution on tracks.

As anticipated before, high luminosity colliders will be used by the scientific community for the exploration of the hidden sectors of the standard model or beyond [28] like the Future

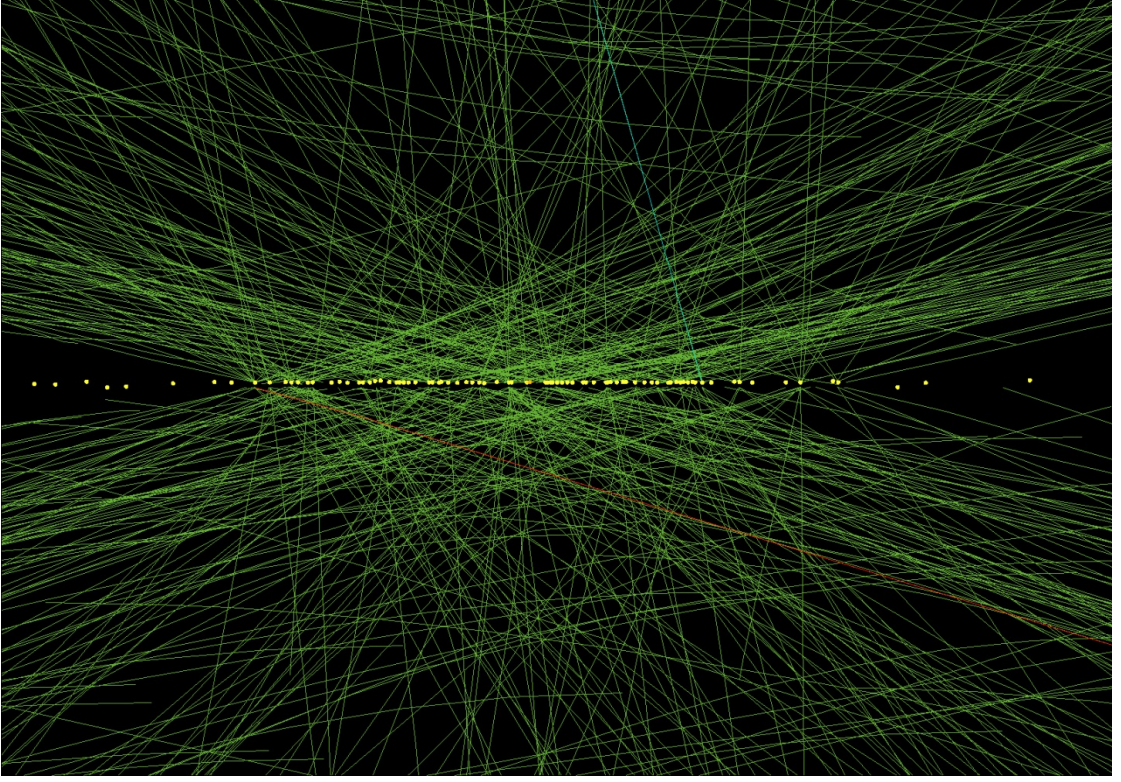


Figure 1.5: Reconstruction of 78 vertices in CMS detector system. Source: [31].

Parameter	Unit	FCC	HL-LHC
Circumference	km	97.75	26.7
Collision energy	TeV	100	14
Bunch spacing	ns	25	25
Peak luminosity	$10^{34}\text{cm}^{-2}\text{s}^{-1}$	30	5
Events per bunch crossing	-	1000	132

Table 1.1: Comparison of the main characteristics and specifications of FCC and HL-LHC. Sources: [36] (FCC), [32, 38] (HL-LHC).

Circular Collider (FCC) at CERN [36, 37], a proposed 100 km circumference circular collider that will represent an evolution of the above mentioned LHC. Table 1.1 shows a comparison of the main characteristics of the HL-LHC and the FCC. It is possible to highlight an increase of a factor 6 in terms of peak luminosity with a consequent rise of the number of events per bunch crossing and pile-up. Therefore, the detectors that will be installed in the trackers of the FCC will be required to be characterized by a few picosecond level time resolution or less in order to guarantee an acceptable pile-up suppression.

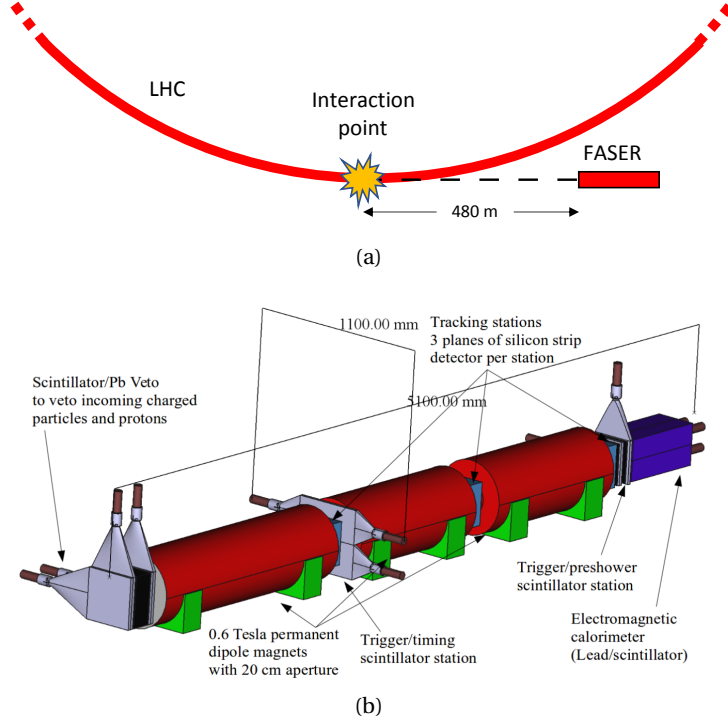


Figure 1.6: Position of the FASER experiment with respect to the LHC (a) and 3D model of the FASER detector (b). The LHC representation in (a) is not in scale. Source of (b): [43].

The FASER experiment

In Chapter 3, the design of various circuits integrated in a detector chip for the upgrade of the ForwArD Search Experiment (FASER) at CERN will be presented. Part of this work was published in

- [39] F. Martinelli et al. “Measurements and analysis of different front-end configurations for monolithic SiGe BiCMOS pixel detectors for HEP applications”. In: *Journal of Instrumentation* 16.12 (Dec. 2021), P12038. DOI: 10.1088/1748-0221/16/12/p12038.

The main goal of the FASER experiment [40–43] is the research on dark matter through the detection of long-lived particles (LLP), in order to extend the physics programs of other experiments at the LHC such as ATLAS and CMS. Indeed, the discovery of potential LLPs like dark photons and axion-like particles (ALP) may represent an important step for the study of the hidden sector [44]. For the achievement of this goal, the FASER experiment will exploit a large flux of hadrons at 0 degree angles with respect to the beam line of the LHC that experiments like ATLAS or CMS are not well-suited to detect [43]. The FASER detector was installed 480 m away from the ATLAS interaction point as depicted in Figure 1.6a.

The complete detector system, shown in Figure 1.6b and 1.7, is composed by the following elements:

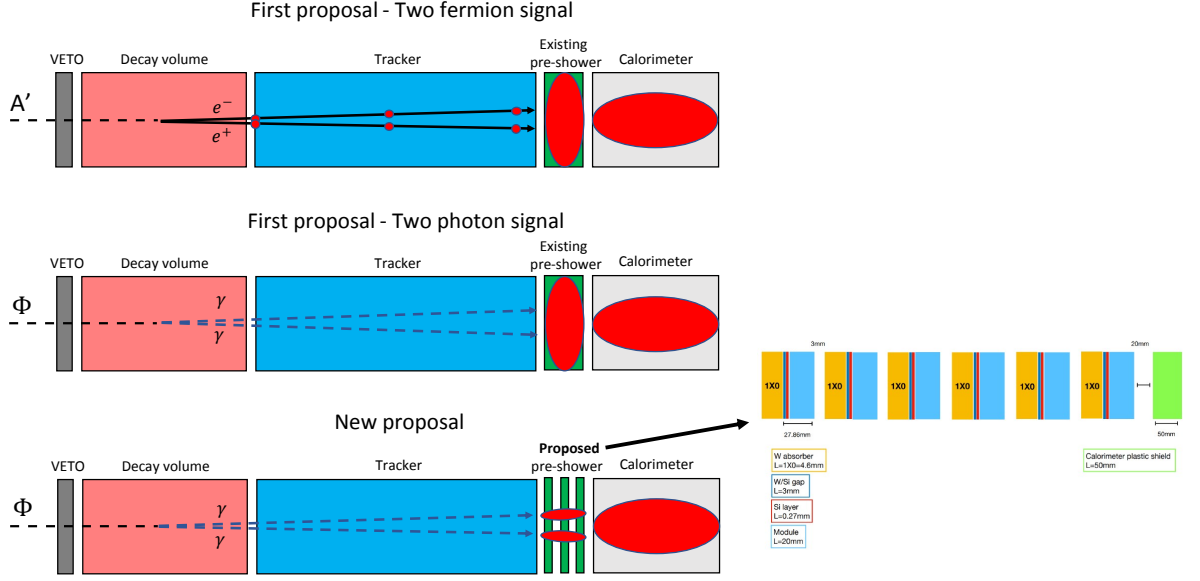


Figure 1.7: Two fermion signal (top) and two photon signal (center and bottom). The latter can only be sensed with the new version of the pre-shower detector (bottom). On the right, new pre-shower architecture depicted in GEANT4 simulation.

- a veto station used to filter those particles that are not generated into the decay volume. It composed by two layers of plastic scintillators separated by lead.
- a decay volume made of 1.5 m dipole magnet.
- a spectrometer (indicated as tracker in Figure. 1.7) composed by three silicon strip detector stations and two 1 m long dipole magnets.
- a scintillator-based preshower.
- a calorimeter.

This detector was designed to sense pairs of charged particles generated by the decay of LLPs [43] (two fermion signals in Figure 1.7) but it is unable to detect and discriminate pairs of photons (two photon signals of Figure 1.7 at the bottom) that may represent an indication of the existence of ALPs. Indeed, those photons are invisible to the tracker, their trajectory can not be affected by the magnetic field and the current pre-shower system was not designed to properly distinguish them: because of their energy (100 GeV - 4 TeV) their separation (due to only kinematics) is in the few hundreds of micrometers range. For this reason, a new high-granularity pre-shower system was proposed to replace the old scintillator-based structure in order to enhance the research of LLPs. The main idea is to adopt various timing detectors interleaved with photon-conversion layers of tungsten as depicted in the right portion of Figure 1.7. The timing information, with a target time resolution of 100 ps, is useful to extract the direction of the incoming particles and the position of the interaction points.

Chapter 3 describes the architecture of the FASER detector ASIC while in Chapter 4 a focus on the TDC for the first prototype of the chip of the new pre-shower system is reported.

1.1.3 Others

HEP and medical imaging represent some of the most important applications in which time resolved particle detectors can be used [28]. However, interesting detector solutions have been developed for other fields of interest. For instance, the Medipix Collaboration at CERN [45] has been designing ASICs for imaging applications since the 1997 (with the first Medipix chip). In 2005, this collaboration started the development of the Timepix chip [46] initially intended to be used for the readout of gas detectors. Timepix chips (like the others produced by the Medipix collaboration) belong to the family of hybrid pixel detectors (in which the active area is not integrated in the same chip of the readout electronics). As explained in [46], these ASICs have been used for space dosimetry applications: detectors featuring Timepix readout electronics were used on both the International Space Station (ISS) and on the Orion test vehicle developed by the National Aeronautics and Space Administration (NASA). Timing detectors can also be used for time-of-flight ranging [47], X-ray and γ -ray imaging [46] and TOF mass spectroscopy [48].

The applications cited in this subsection have been only mentioned to highlight the importance of time resolved timing detectors and the fields in which their characteristics can be used. However, they are not part of the scope of the presented thesis.

1.2 Thesis goals and contributions

This thesis aims to describe the design process of several ASICs that can be integrated in a heterogeneous set of timing systems such as γ -photons detectors for PET scanners (TT-PET project) or LLP detectors for HEP applications (FASER experiment). The work described in the thesis includes the design of:

- various configurations of front-end systems integrated in a first prototype chip for the FASER experiment (Chapter 3)
- a test chip of a Time-to-Digital Converter (TDC) for the TT-PET project ASIC (Chapter 4)
- the TDC for a first prototype chip for the FASER experiment (Chapter 4)
- the digital logic and TDC for a second FASER pre-production chip (Chapter 3). Some of the front-end configurations of the first prototype chip (first point of this list) were also integrated in the second pre-production ASIC.
- a readout ASIC for hybrid pixel detectors (Chapter 5).

The thesis provides a description of the designed ASICs attempting to highlight the most

critical points in the development of these systems and addressing the main challenges. Moreover, the main contributions of this thesis also include:

- the development of a non-linearity model for the analysis of the impact of mismatches in free-running ring-oscillator based TDCs (Chapter 4). This model was built for the study of the TDC architecture designed for the TT-PET project ASIC and demonstrated that a particular configuration of the output buffers can improve the performance of the converter in terms of linearity. The work is also meant to highlight the advantages of having an analytical model that can guide the designer architectural choices. The development of state-of-the-art circuits often requires the optimization of many aspects of such systems: this process can benefit from an analytical model helping the designer to move in the space of the parameters to optimize.
- the development of a framework for circuit tuning and optimization based on Genetic Algorithms (GAs) (Chapter 6). GAs represent a type of procedure that can be used to minimize (or maximize) a set of parameters. These algorithms can be used for a large variety of complex optimization problems including, as it will be explained in this thesis, circuit design. Indeed, they can be particularly useful for the sizing of the devices of a circuit once a certain architecture has been chosen. This work will also highlight the importance of the choice of the function (fitness) to be minimized for the optimization of the main performance parameters of the system to design.

The work described in this thesis followed all the steps of the design flow of ASICs for timing detectors including: design of analog systems; digital circuits implementation; mixed-signal optimization; ASIC testing and data analysis. The ICs presented in this manuscript were designed using the Cadence Virtuoso Platform and all the simulations were performed using Spectre.

1.3 Organization of the thesis

The thesis is organized as follows. Chapter 2 provides a brief overview on timing detectors describing how different types of transducers interact with the particles that need to be sensed and explaining the importance of the electronics in the development of these detecting systems. Chapter 3 describes two pre-production ASICs designed for the enhancement of the FASER pre-shower detector. Chapter 4 is focused on the development of two TDCs for the TT-PET project (including the above-mentioned non-linearity model) and for the FASER experiment. Chapter 5 describes a readout ASIC for hybrid pixel detectors that aims to be used for the testing of external sensors. Chapter 6, shows the development of the GAs used for circuits tuning and optimization. After the conclusions, a brief introduction to the Bethe-Bloch formula and Schockley-Ramo theorem are given in Appendix A and B respectively. A gallery of the submitted ASICs is reported in Appendix C .

2 Timing detector basics

The detection of particles can only be implemented through their interaction with matter e.g., ionisation in the case of charged particles or photon scattering [1]. In many detectors, various types of transducers are used to generate, through their interactions with particles, electrical signals that can be read by an electric circuit. A brief overview on the most common transducers used in HEP or medical imaging applications and the way they can sense the particles is reported in Section 2.1.

As anticipated, this thesis is mainly focused on the design of electronic systems used for timing detectors. The role of the electronics and its importance in the development of high-performance detectors will be described in Section 2.2. More in detail, this section will give a description of the state-of-the-art of the front-end (Section 2.2.1), digitization (Section 2.2.2) and readout (Section 2.2.3) systems underlining their role and the aspects to be improved for the enhancement of the detectors in which they are integrated.

2.1 Transducers

2.1.1 Silicon sensors for ionizing radiation

One of the simplest detecting structures able to exploit the above-mentioned ionization process is the PIN diode depicted in Figure 2.1a. The system is composed of a semiconductor divided in three main regions with a $p(+)/n-/n(+)$ doping profile. The $n-$ zone is the thickest and, due to its reduced doping concentration, is indicated as intrinsic region. The PIN diode needs to be reverse-biased with a voltage that allows the intrinsic region to be completely depleted, as the electric field distribution depicted in Figure 2.1a indicates. When a particle able to interact with the semiconductor crosses the diode, the ionization process leads to the generation of a certain amount of electron-hole (e^-h^+) pairs that will induce the electrical signal to be read by the readout electronics. More in detail, the production of e^-h^+ pairs happens when an electron in the valence band of the semiconductor, due to the interaction with the incoming particle, acquires a sufficient amount of energy to move into the conduction band leading to the generation of a hole in the valence band. It is possible to demonstrate

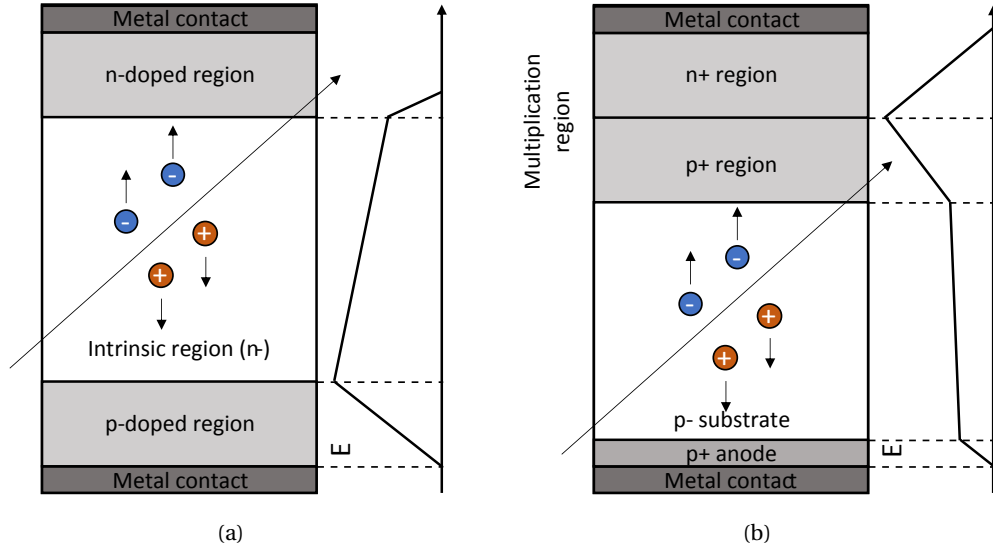


Figure 2.1: Schematic of the doping profile and electric field (E) distribution in a generic silicon sensor without (a) and with (b) gain layer.

that this solution is able to guarantee good timing performance, with time resolution down to several tens of ps [1]. This sensor architecture is the one adopted by the TT-PET project for its monolith pixel detectors (see Section 1.1.1).

A possible improvement of this architecture is represented by the structure in Figure 2.1b. This solution is usually indicated as Low Gain Avalanche Diode (LGAD) [49] and was first developed for the implementation of semiconductor photodetectors like Avalanche Photo-Diode (APD). LGADs are characterized by the implantation of an additional highly-doped p+ region separating the p- (intrinsic) and n+ regions. The presence of the p+/n+ junction leads to an increase of the electric field. When the ionization occurs, the holes are absorbed by the metal connected to the p+ anode while the electrons move to the high-field region. The bias potential is such that these electrons are generating a second ionization, amplifying the signal read by the electronics [26]. A proper modulation of the amplification gain G is useful to obtain an optimal SNR. As explained in [1], the component of the output signal associated to the charge multiplication in the high-field zone is affected by statistical fluctuations increasing the shot noise:

$$d\langle i^2 \rangle \propto \langle I_0 G^2 F \rangle df, \quad (2.1)$$

where $d\langle i^2 \rangle$ is the power spectral density of the shot noise, I_0 is the bulk contribution to the leakage current and F is a factor that describes the above-mentioned fluctuations. Depending on the characteristics of the detector it is possible to find the optimal value of G for the maximization of the SNR.

Silicon sensors with an integrated gain layer can either be designed to work as LGAD, showing a usual gain in the order of 10 [49], or to work in Geiger mode [1], setting their bias higher than their breakdown voltage. In this way the sensor will be characterized by a gain $G \approx 10^6$. This

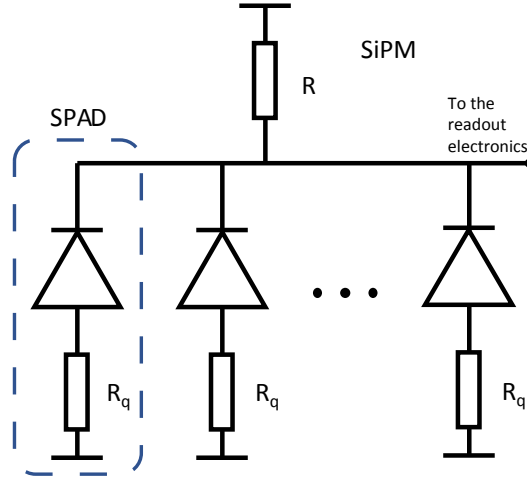


Figure 2.2: Bias configuration of a set of SPADs to implement an Analog SiPM featuring quenching resistors R_q .

solution is often adopted for the implementation of photodetectors able to activate also when a single photon is passing through the sensitive area. For this reason, this kind of devices are named Single-Photon Avalanche Diodes (SPADs). The activation of the Geiger regime requires the system to be connected to a quenching device to discharge the SPAD when it detects an ionizing photon. The quenching system can be either a simple resistor R_q (passive quenching) or a transistor (active quenching). SPADs are characterized by sub-nanosecond output rise times (due to the fast multiplication process) and discharge times in the order of 100 ns [50]. Because of the large gain, the output of a SPAD rapidly saturates when it detects a photon. For this reason, SPADs can be considered as counting devices that are able to generate a digital pulse when they sense an event of interest. However, in some applications, the detection of more than one photon at the same time can be required. A solution to this problem can be represented by the integration of a set of SPADs in a matrix-like structure, usually named Silicon PhotoMultipliers (SiPMs) [1, 51]. Figure 2.2 depicts the bias configuration of a generic SiPM: the SPADs with their respective quenching resistors R_q are connected together to a single bias resistor R . In this way, the output of the matrix will be proportional to the number of SPADs that sensed a photon. Despite the good time resolutions achievable with this kind of solutions ($\lesssim 100$ ps), one of the main performance limiting factors of SiPM is the dark pulse rate [1]. Avalanche processes can occur also when no photon is crossing the sensitive area of any SPAD in the matrix. The main cause of dark pulses is related to the multiplication of carriers due to the traps located in the bandgap¹ and the high gain of the SPADs [51]. Moreover, during a process of ionization due to an incoming photon, it is possible that a group of charges remain trapped in the bandgap and then released only after the pulse of the event extinguished, causing another avalanche process named afterpulse [51]. However, as explained in [1], the dark pulses are usually characterized by a lower amplitude than the real events in which a large group of photons are crossing the SPADs at the same time. Hence, a proper threshold on

¹Traps in the bandgap can be generated by many causes including crystallographic defects and impurities [52].

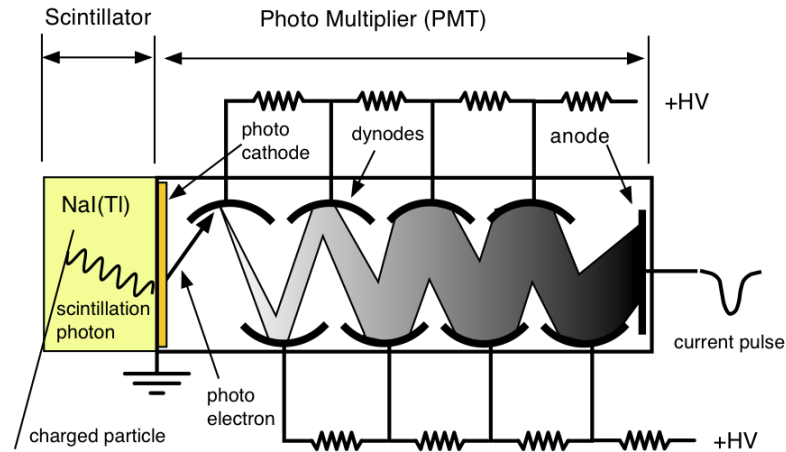


Figure 2.3: Scintillator coupled with a PMT. Source: [54].

the output of the detector can significantly reduce the impact of the dark pulses.

The system depicted in Figure 2.2 represents an Analog SiPM since the output of the SPADs are shorted to send a single signal to the readout electronics. An alternative solution is represented by Digital SiPMs [53], in which each SPAD is connected to its own electronics able to send fast triggers and timing information associate to the events of the corresponding cell to the common readout. This architecture allows quenching each SPAD individually and selectively disabling a cell of the SiPM matrix if its dark count rate is too high [1]. However, it is characterized by larger complexity and power consumption compared to the Analog SiPM solution.

More details about the generation of the electrical signal of a silicon detector for ionizing radiation in Appendix B.

2.1.2 Scintillators and PMTs

As explained in Section 1.1.1, the working principle of a PET scanner is based on the detection of 511 keV γ -photons produced by the annihilation of e^+e^- pairs. For this reason, the development of an efficient PET scanner requires detectors capable of sensing photons in the above-mentioned energy range through photoelectric effect i.e., the absorption of photons with a consequent generation of free carriers. The interaction characteristics of a particle with a certain material is usually modelled with the concept of cross section (see Appendix A for more details). The latter can be defined as the area of the particles of the target material seen by the incoming interactive particle. A larger cross section indicates a higher interaction probability. In [1], the behavior of the cross section of silicon as function of the energy of incoming photons is described, showing a drop of several order of magnitude above 100 eV. For this reason, silicon sensors usually are not used alone in applications in which γ -photons are required to be sensed. Many commercial PET scanner, for instance, exploit the properties of scintillators to efficiently interact with photons in the γ energy range and convert them

Material	σ_{abs} [cm ² /g]	σ_{scat} [cm ² /g]	σ_{tot} [cm ² /g]
Pb	$7.84 \cdot 10^{-2}$	$7.775 \cdot 10^{-2}$	$1.562 \cdot 10^{-1}$
Si	$1.74 \cdot 10^{-4}$	$8.649 \cdot 10^{-2}$	$8.666 \cdot 10^{-1}$

Table 2.1: Photoelectric absorption (σ_{abs}), scattering (σ_{scat}) and total (σ_{tot}) cross section of lead (Pb) and silicon (Si) referred to 511 keV γ -photons. Source: [55]

into visible or UV light. An exhaustive analysis and description of the scintillators and their characteristics is out of the scopes of this work. A detailed analysis can be found in textbooks as [56] [1] or papers as [57].

As described in [56], scintillators can be divided in two main groups: organic and inorganic. The latter are also denominated scintillator crystals and are largely used in medical imaging and HEP applications. The scintillation process in inorganic crystals can be summarized as follows. When a photon (usually in the x-ray or γ range for the above-mentioned applications) crosses a scintillator its interaction with the material will move one or more electrons from the valence to the conduction band. The scintillation process occurs when the material is featuring a bandgap characterized by luminescence centers [1]. An electron in conduction band can move to one of these centers due to phonon scattering (energy loss). At this point, when the electron transitions from a high energy center to a smaller energy one a scintillation photon is then generated. The absorption of the latter is avoided because of the smaller energy distance among the luminescence centers in the bandgap compared to the one absorbed by the electrons of the first ionization process. Moreover, the emitted light will be characterized by a wavelength in the visible or UV range. This phenomenon is called Stoke shift [56].

In PET scanners, various scintillator crystals (such as Bismute Germanium Oxide (BGO) or Lutetium-Yttrium Oxyorthosilicate (LYSO) whose characteristics are reported in [7]) are coupled with SiPMs. An alternative is represented by a scintillator-PhotoMultiplier Tube (PMT) system as the one shown in Figure 2.3. PMTs are composed by a photocathode, a structure able to absorb an incoming photon and emit an electron. The latter is then transferred to a system composed by several dynodes that are increasing the total amount of electrons through multiplications and then sent to an anode from which the output signal will be read. PMTs are usually characterized by large gains in the order of 10^6 and represent good solution for low-intensity input signals [1]. A recent implementation of this structure is proposed in [58] featuring a light-guide for a better coupling between scintillator and the PMT.

One of the main limiting factors for the time resolution of a scintillator-based detector is the variance associated to the interaction depth of the photons inside the crystal. A possible solution to overcome this issue is presented in [15], where the readout system is coupled along the side of the crystal to reduce the uncertainties associated to the interaction point of the incoming photons and, at the same time, guarantee an efficiency comparable to the ones of usual architectures. The TT-PET project proposed an alternative solution, based on avoiding the use of scintillators and SiPMs: the γ -photon conversion is implemented exploiting layers of high atomic number Z material like lead interleaved with monolithic silicon detectors. As reported in Table 2.1, lead is characterized by a significantly higher photoelectric absorption

cross section σ_{abs} for 511 keV γ -photons. Indeed, it can be demonstrated that σ_{abs} is increasing as function of the atomic number being $\propto Z^4$ and $\propto Z^5$ for low and high photon energy respectively [59]. This implementation is one of the reasons behind the aim of the project to develop a 30 ps time resolution PET scanner (see Section 1.1.1).

2.1.3 Other materials

Over the years, silicon has been used for a large number of applications because of its unique properties. However, many researches have pursued to develop alternative materials that are able to overcome the limits of silicon. Few examples are reported below.

Gallium arsenide (GaAs) is an interesting semiconductor especially for photonic applications [60]. It is characterized by a direct bandgap of 1.43 eV that makes it suitable for the detection of photons: an electron can move from the valence to the conduction band directly absorbing a photon with energy similar to the bandgap². Moreover, GaAs is a compound made of two atoms with a relatively high atomic number ($Z = 31$ for As and $Z = 33$ for Ga) that, as previously explained, increases the cross section of the material and makes it useful also for high-energy radiation detection. Also, the GaAs bandgap energy is higher than the silicon one: for this reason, the bulk resistivity is larger and its depletion occurs at lower voltages. However, production of GaAs is more complex and not as efficient as that of silicon.

Another interesting semiconductor compound is cadmium zinc telluride ($\text{Cd}_{1-x}\text{Zn}_x\text{Te}$). The main properties of this material are related to the high atomic number of its atoms ($Z = 48$ for Cd, $Z = 30$ for Zn and $Z = 52$ for Te) and the large bandgap that can be modulated modifying the amount of Zn in the compound (from 1.44 eV if $x=0$ to 2.2 eV when $x=1$). For all these reasons, $\text{Cd}_{1-x}\text{Zn}_x\text{Te}$ is a valid candidate for x-ray and γ -ray detectors [61].

In the last few years, the interest in diamond for HEP experiments has increased [62]. This interest is due to the diamond high bond energy (almost twice the one of silicon) and the low amount of free charges, that makes this material a good insulator. For this reason, diamond detectors potentially represent the most suitable solution for the development of high rad-hard systems (see Section 1.1.2). However, the performance of recent implementations proves that there is still room for improvement (e.g., [62] explains that one of the main limiting factor of timing detectors is the difficulty in the generation of large signals to be read by the front-end electronics).

2.2 Electronics

The development of a picosecond level time resolution particle detector requires the design of an ultra-fast low-noise electronics. The time resolution σ_t of a timing detector is affected by

²In an indirect semiconductor like silicon, the absorption of a photon with an energy similar to the bandgap (e.g., ~ 1.12 eV for silicon) is very unlikely. Indeed, this process could happen only if the electron is able to absorb both a sufficient energy to cross the bandgap and a certain amount of crystal momentum. Low energy photons are characterized by a negligible momentum, being significant only above the x-ray range [52].

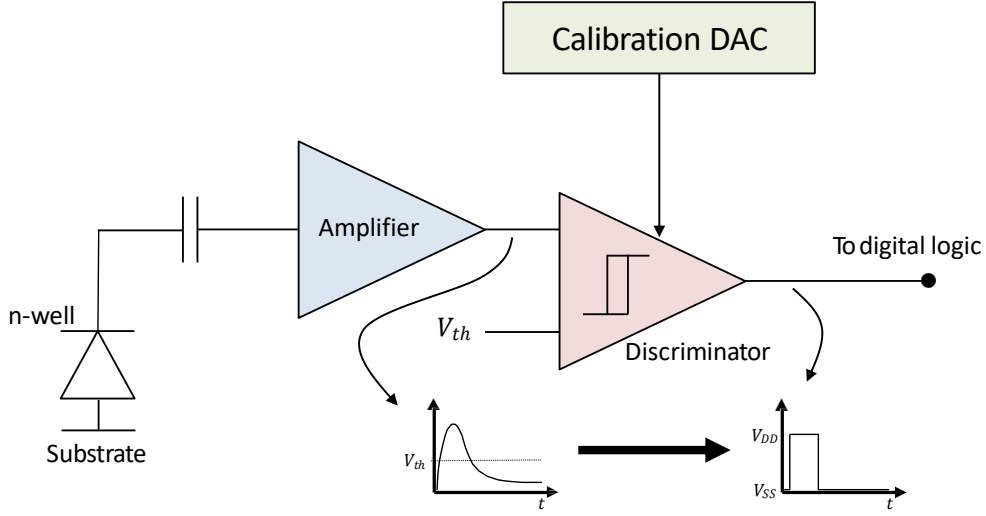


Figure 2.4: Schematic of a generic front-end system for the readout of a timing detector. Source: [21].

several factors. Assuming their statistical independence, it is possible to express σ_t as [22, 28]

$$\sigma_t = \sqrt{\sigma_{elec}^2 + \sigma_{TDC}^2 + \dots}, \quad (2.2)$$

in which σ_{elec}^2 and σ_{TDC}^2 represent the contributions of the front-end electronics (Section 2.2.1) and TDCs (Section 2.2.2) respectively. The dots have been added to highlight the fact that many other effects can worsen the equivalent time resolution of the system (e.g., pixel cross-talk, clock distribution [28]) and their nature and intensity depend on the characteristics and the design of the chosen detector. In this section, a first overview of the electronics for timing detector is reported, focusing on the main and most critical blocks whose optimization is crucial if high-performance are aimed to be achieved.

2.2.1 Front-end

In a detector, the main role of the front-end electronics is to acquire and amplify the signals produced by the sensors. In this way the digital logic is able to read those signals and extract from them all the information needed for the event reconstruction. The block diagram of a generic front-end system for timing applications is reported in Figure 2.4. The sensor is commonly modeled as a diode connected to an amplifier, commonly through an AC coupling capacitor, although a DC connection is also possible [63] [64]). The amplifier is needed to make the signal large enough to be read by the digitization blocks and it plays a crucial role for the performance of the detector [26]. The output of the amplifier is digitized with a discriminator e.g., an open-loop OPERational AMPlifier (OPAMP) that compares the input with a fixed threshold voltage V_{th} . The discriminator is often designed with a hysteresis input-output characteristics. In this way, it is possible to prevent the discriminator to switch several

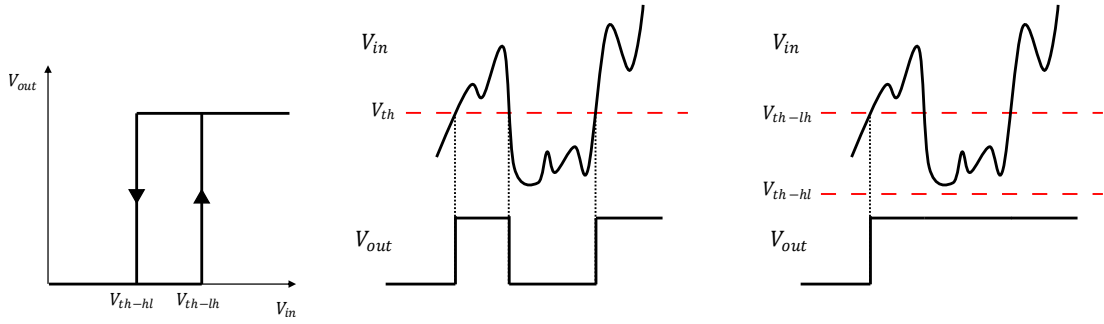


Figure 2.5: Working principle of a discriminator with hysteresis. Left: input-output characteristics (non-inverting in this example). Centre: no-hysteresis discriminator. Right: hysteresis discriminator. A noisy input can make the circuit switch several times as the signal crosses the threshold V_{th} . On the other hand, a discriminator with hysteresis shows a different equivalent threshold depending on the state of the output ($V_{th-lh} > V_{th-hl}$).

times during the input transition due to noise or cross-talk (the working principle is explained in Figure 2.5). A possible implementation is based on the so-called Schmitt's Trigger [65]. The Digital-to-Analog Converter (DAC) in Figure 2.4 is used to calibrate the discriminator and to compensate for pixel-to-pixel variability.

In time-resolved detectors, the output of the discriminator is read by a time-digitization system (Section 2.2.2) that will extract the Time-of-Arrival (ToA) from this signal, a timestamp given to each event. In many cases, it is useful to perform an evaluation of the input charge Q_{in} injected in the front-end for each event to improve the quality of the reconstruction. Q_{in} can be calculated measuring the period of time in which the discriminator output is over the threshold, often indicated as Time-Over-Threshold (TOT). This period can be used for the evaluation of Q_{in} . The measurement of the TOT can be also used to perform the correction of the time-walk [1]: the slope of the output of the pre-amplifier may depend on the value of Q_{in} leading to a variation of the time in which the signal crosses V_{th} , i.e. the ToA of the corresponding event. This effect can be compensated for by measuring Q_{in} through the TOT. One of the most critical blocks for the performance of a detector is the amplifier (often indicated with pre-amplifier). Various configurations can be adopted for the design of this circuit depending on the characteristics of the sensor and the target SNR. Indeed, [1, 26] explain that, in order to improve the SNR, one of the best solutions is to implement a large-bandwidth amplifier that is able to follow and amplify the input current (or voltage) produced by the sensor. On the other hand, lower SNR and/or fast input signals scenarios require architectures with a limited bandwidth to reduce the impact of the noise on the output. One of the most common approaches is based on the implementation of a charge amplifier, modeled in Figure 2.6a. As the name suggests, charge amplifiers are circuits whose output voltage V_{out} is proportional to the input charge Q_{in} . Thus, the amplifier acts as an integrator of the current produced by the sensor and injected into the circuit. The integration characteristics are obtained implementing a system with a reduced bandwidth that, as anticipated, is a useful solution to obtain better timing performance in noisy environments. Assuming that the

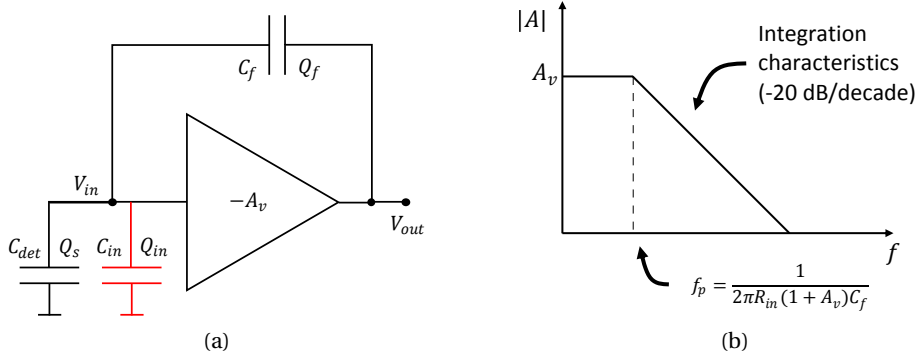


Figure 2.6: Schematic of a charge amplifier (a) and simplified input-output characteristics over frequency (b). C_{in} (depicted in red) is the equivalent input capacitance and can be calculated with Miller's theorem.

voltage amplifier with gain $-A_v$ of Figure 2.6a is characterized by a large input impedance compared to the one shown by the feedback capacitance C_f , the latter will absorb all the current produced by the sensor and store the entire input charge $Q_{in} = Q_f$, indicating with Q_f the charge in C_f . Under this assumption and supposing that $A_v \gg 1$, the charge gain of the

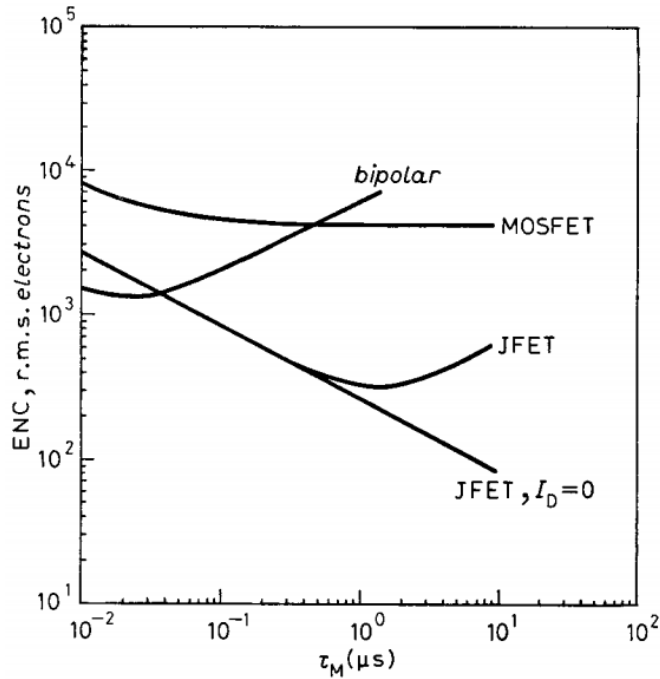


Figure 2.7: behavior of the ENC of a charge amplifier as function of the shaping time τ_m for various active devices. Source: [66].

system will be [26]

$$A_q = \frac{\partial V_{out}}{\partial Q_{in}} = \frac{-A_v}{(1 + A_v)C_f} \approx -\frac{1}{C_f}. \quad (2.3)$$

Equation 2.3 can be easily calculated remembering that, according to the Miller's theorem [67], the equivalent input capacitance C_{in} (depicted in red in Figure 2.3), can be expressed as

$$C_{in} = (1 + A_v)C_f. \quad (2.4)$$

Considering that $V_{out} = -A_v V_{in}$ and $Q_{in} = C_{in} V_{in}$, from Equation 2.4, it is possible to obtain Equation 2.3. However, as reported in [26], the detector capacitance C_{det} acts as a charge divider, making Q_{in} only a portion of the total sensor charge Q_s . For this reason, the real gain of the amplifier A_q needs to be corrected

$$A_q = \frac{C_{in}}{C_{in} + C_{det}} \frac{\partial V_{out}}{\partial Q_{in}} \approx -\frac{1}{C_f + (C_{det}/A_v)}. \quad (2.5)$$

Even in this case, a large voltage gain ($A_v \gg C_{det}/C_f$) will lead to the condition $A_q \approx -1/C_f$, in which the charge gain of the amplifier is only dependent on the feedback capacitance. All this results are obtained assuming that the input impedance of the voltage amplifier is infinite. While in a MOSFET-based architecture this approximation can be considered valid in many scenarios (due to an input gate resistance in the $M\Omega$ order of magnitude), this is not necessarily true for bipolar transistors. Indicating with R_{in} the input resistance of the voltage amplifier, the above analysis is valid only if [26]

$$R_{in} \gg \frac{1}{2\pi f C_{in}} = \frac{1}{2\pi f (1 + A_v) C_f} \rightarrow f \gg \frac{1}{2\pi R_{in} (1 + A_v) C_f}. \quad (2.6)$$

This equation indicates that the integration condition (that makes the system a charge amplifier) is valid only if the impedance associated to the C_f is small enough compared the R_{in} , making the input current flow entirely inside the feedback. Moreover, the second part of Equation 2.6 highlights that the system acts like a charge amplifier only at high frequencies compared to the first pole of the amplifier, in a region of the input-output characteristics with a -20 dB/decade slope as depicted in Figure 2.6b. The reduction of the bandwidth is also useful to decrease other noise contributions associated to high frequency sources cross-talk from the discriminator output and other digital signals.

As explained at the beginning of this section, the time resolution of a detector σ_t is proportional to the contribution related to the front-end electronics σ_{elec} . The latter can be expressed as [26]

$$\sigma_{elec} = \frac{\sigma_v}{dV/dt}, \quad (2.7)$$

where σ_v is the output noise of the amplifier. Equation 2.7 emphasizes that a better resolution is obtained with fast and high-gain amplifier able to maximize the slope dV/dt . In the case of a charge amplifier and approximating $dV/dt \approx V_{out}/t_{rise}$, it is possible to resolve σ_{elec} as

function of SNR

$$\sigma_{elec} \approx \frac{t_{rise}}{SNR} = \frac{t_{rise} ENC}{Q}, \quad (2.8)$$

since $SNR = V_{out}/\sigma_v$. The equation is obtained expressing the SNR as function of the charge $SNR = Q/Q_{noise}$ and referring the noise contribution Q_{noise} to the Equivalent Noise Charge (ENC). This parameter is defined as the input charge of a noiseless version of the amplifier for which the corresponding output voltage has the same power (RMS value) of the noise of the original system when no input is set [68]. A detailed analysis on charge amplifiers and their performance, reported in [66], highlights the behavior of the ENC as function of the shaping time τ_m and

$$ENC(\tau_m) = [a(C_{det} + C_{in})^2 \frac{1}{\tau_m} + b(C_{det} + C_{in})^2 + c\tau_m]^{1/2}. \quad (2.9)$$

In this equation, a , b and c include the characteristic noise sources of the active device used for the implementation of the amplifier. In particular, the first term is related to the series white noise component, the second to $1/f$ noise and the third to parallel white noise source. In Figure 2.7, the behavior of the ENC described by Equation 2.9 for charge amplifiers with different active devices is reported. For short shaping times (hundreds of nanoseconds), bipolar transistors represent the best solution since these devices are usually characterized by small series noise components (first term of Equation 2.9) compared to MOSFETs [26]. Furthermore, as explained in [26], for BJTs, the series noise contribution decreases by increasing the current gain β and small base spreading resistance (in order to limit both shot and thermal noise). For these reasons, Silicon-Germanium Heterojunction Bipolar Transistors (SiGe HBTs) are particularly suitable for this application. Indeed, as explained in [69], the introduction of Germanium in the base of a BJT bends of the band structure, introducing a drift field in the base that increases the current gain³. For this reason, the base of SiGe HBTs can also be highly doped without compromising the performance in terms of β and consequently reducing the equivalent base resistance. This technology represents the solution adopted for the development of the front-end system for TT-PET project and FASER experiment prototype chips (more details in Chapter 3). The readout chip produced and described in Chapter 5 also features SiGe HBTs for the implementation of pre-amplifiers and discriminators.

2.2.2 Time digitization system

A portion of the following analysis, describing the impact of Time-to-Digital Converters (TDCs) on the timing performance of a detector and possible implementation solutions, was presented in

- [71] F. Martinelli et al. “A massively scalable Time-to-Digital Converter with a PLL-free calibration system in a commercial 130 nm process”. In: *Journal of Instrumentation* 16.11 (Nov. 2021), P11023. DOI: 10.1088/1748-0221/16/11/p11023.

³It can be demonstrated that the current gain increases as $\beta \propto \exp \frac{E_{G-Si} - E_{G-SiGe}}{kT}$, where E_{G-Si} and E_{G-SiGe} are the bandgap energies of Si and SiGe respectively, T is the temperature and k is the Boltzmann constant [69]. SiGe bandgap can vary from ~ 0.66 eV (Ge) to ~ 1.12 eV (Si) depending on the concentration of germanium [70].

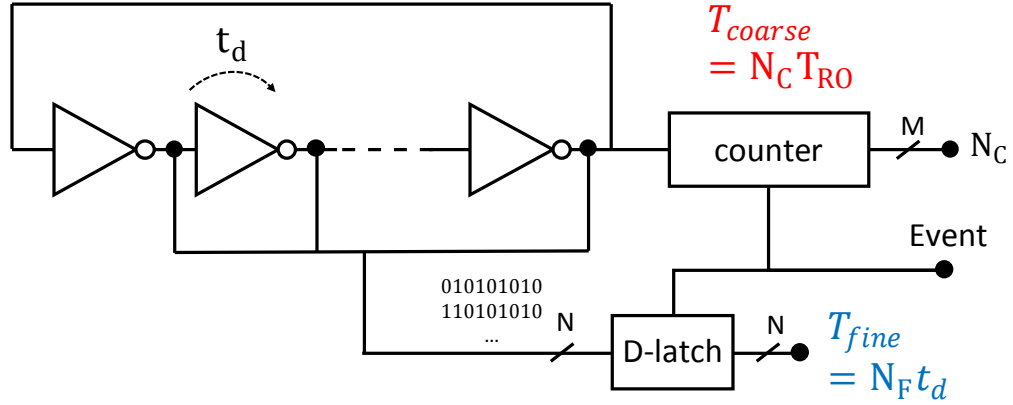


Figure 2.8: RO-based TDC general schematic.

When a particle hits a pixel (or, more generally, the active area of a detector) and the front-end system generates a corresponding discriminated signal, the timing information related to this event has to be measured by a TDC. As reported in Equation 2.2, the time resolution of a detector σ_t is directly dependent on the contribute provided by the TDC because it represents the accuracy of the time measurement. A particle detector characterized by a picosecond level σ_t requires a system that is able to measure time intervals with a precision in the same order of magnitude. Indeed, even assuming an ideal TDC, its measurements outcomes are affected by a quantization error ϵ_q . In this case, the latter can be modeled as random variable with a uniform distribution inside each time bin. In this work, the average time bin will be referred to as Least Significant Bit (LSB) T_{LSB} . Its standard deviation σ_q is equal to the contribution of the TDC on the final time resolution of the detector [22] (Equation 2.2) and can be expressed as [72]

$$\sigma_{TDC} = \sigma_q = \frac{T_{LSB}}{\sqrt{12}}. \quad (2.10)$$

σ_{TDC} also represents the product of the quantization noise [72]. Indeed, in the uniform distribution case, the probability density function (pdf) of $\epsilon_q = \epsilon_q(t)$ is equal to $1/T_{LSB}$ if $0 < t < T_{LSB}$ and 0 elsewhere. The average value of ϵ_q is given by

$$\mu_q = \int_0^{T_{LSB}} \frac{1}{T_{LSB}} t dt = \frac{T_{LSB}}{2} \quad (2.11)$$

while its Root Mean Square (RMS) value rms_q is

$$rms_q = \int_0^{T_{LSB}} \frac{1}{T_{LSB}} t^2 dt = \frac{T_{LSB}^2}{3}. \quad (2.12)$$

σ_q can be calculated as

$$\sigma_q^2 = rms_q - \mu_q^2. \quad (2.13)$$

Replacing Equation 2.11 and 2.12 in 2.13, it is possible to obtain Equation 2.10. σ_q is often used to indicate the resolution of a converter [73] because it represents the uncertainty that

affects the measurement. However, for the rest of the thesis, the resolution of a TDC will refer to the LSB. In Section 4.2, the evaluation of σ_q for a non-ideal TDC is described.

Common TDC architectures

Different approaches can be adopted to design a TDC [72]. One of the traditional and most common solutions is exploiting Ring Oscillators (ROs) [74–76]. A RO is a circuit composed, in its simplest implementation, by a chain of N inverters connected as shown in Figure 2.8. N needs to be an odd number in order to satisfy the Barkhausen criterion, a necessary condition to make the system oscillate [67, 77, 78]. Indicating with t_d the transmission delay of the inverters, assuming that they are equal per each stage of the RO and that the delays for the 0-to-1 and 1-to-0 transitions are identical, a simple large signal analysis of the system in Figure 2.8 allows computing the oscillation period T_{RO} as

$$T_{RO} = 2Nt_d. \quad (2.14)$$

The working principle of a RO-based TDC is summarized in Figure 2.8. Indicating with T the time interval to be measured, a counter connected to one of the output of the RO will be used to evaluate the coarse component of the measurement

$$T_{coarse} = N_C T_{RO}, \quad (2.15)$$

where N_C is the number of oscillator cycles within T . However, in order to increase the precision of the measurement, it is possible to sample the state of the RO i.e., the value provided by the output node of each inverter, at the beginning and at the end of T obtaining the fine component proportional to the cell delay t_d

$$T_{fine} = N_F t_d. \quad (2.16)$$

Combining Equation 2.15 and 2.16, it is possible to express the time interval under measurement T as

$$T = N_C T_{RO} + N_F t_d + \epsilon_q, \quad (2.17)$$

where ϵ_q is the quantization error mentioned above. Equation 2.16 and 2.14 highlights that a RO-based TDC is characterized by a LSB equal to the delay t_d and that increasing the oscillation frequency turns into an improvement of the accuracy of the converter. A possible approach to reduce the oscillation period of a RO is exploiting multi-path architectures. In these implementations, each node of the oscillator is also used to pre-charge or discharge further nodes in order to increase the frequency of the system. An example of multi-path architecture is reported in [79] and in Figure 2.9: this architecture is composed of pseudo-NMOS NOR delay cells whose output (p , q , r , s and t) are connected to one of the inputs of the cells placed two stages further in the oscillator. As explained in [79], this architecture features a RMS jitter value of 1.25 ps and a maximum frequency of 4.6 GHz in 180 nm CMOS

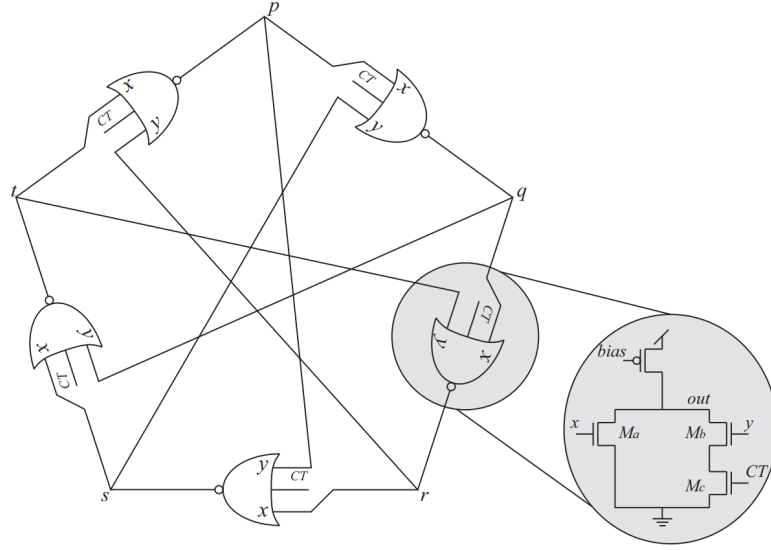


Figure 2.9: Schematic of an interpolating VCO composed of NOR cells. CT is used as a control signal to modulate the oscillation frequency. Source: [79].

technology. This solution can be used for the implementation of TDCs or Phase-Locked Loops (PLLs). A similar approach was followed for the design of the time measurement system of the analog SiPM-based chip Blumino at EPFL [80] and for the TDC described in Section 4.2. Another solution exploits cyclic interpolation of a switched-frequency RO to design a TDC with a 4.2 ps resolution and a maximum measurable time interval of 375 μ s [81].

The previous analysis highlighted that the resolution of a RO-based TDC is limited by the technology since it is dependent on the performance of the inverters composing the oscillator. This limitation can be overcome adopting architectures based on Vernier lines [72] like the one proposed in [82]. In Figure 2.10, the schematic of a generic Vernier line TDC is reported. Two delay lines with different propagation times $t_{d1} > t_{d2}$ are connected to a set of D-flip-flop as in the left part of the figure. The slower line (t_{d1}) is activated at the first edge of the time interval to measure T (START signal) while the faster (t_{d2}) propagates the edge corresponding to the end of T (STOP signal) and its nodes are connected to the sampling inputs of the flip-flops. As shown in the right part of Figure 2.10, the sampling of the nodes of the slower line will generate an output code D characterized by N ones, where N is the number of stages needed by the faster line to be in phase with the slower one. As shown in the picture, it is possible to express T as

$$Nt_{d1} = T + Nt_{d2} + \epsilon_q \longrightarrow T = N(t_{d1} - t_{d2}) + \epsilon_q, \quad (2.18)$$

where, also in this case, ϵ_q is the quantization error. From Equation 2.18, it is possible to conclude that this architecture is characterized by an LSB equal to $T_{LSB} = t_{d1} - t_{d2}$. One of the main problem of this solution is related to the measurement range. Indeed, for a given T_{LSB} the maximum measurable time is equal to $T_{max} = N_{tot} T_{LSB}$ where N_{tot} is the total number of stages of the delay lines. For this reason, if a larger range is required, the TDC needs to feature more delay cells resulting in a consequent increase of the power consumption.

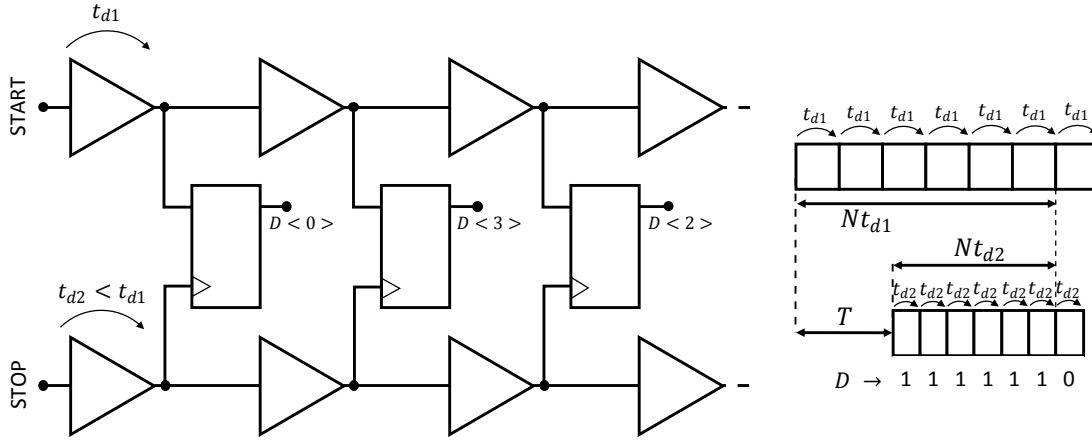


Figure 2.10: Vernier line TDC. Schematic (on the left) and working principle (on the right).

Various architectures can be implemented to improve this trade-off like cyclic [83] or 2D Vernier lines [84]. The former features two ROs with different oscillation frequencies and a counter to extend to total range. On the other hand, 2D Vernier lines exploit a set of sampling blocks connecting the nodes of two delay lines in different combinations. This solution allows obtaining N quantization levels using only \sqrt{N} stages.

Other configurations are possible. For instance, [73] proposes a Time-to-Amplitude Converter (TAC) architecture. The working principle is based on charging a capacitor with a constant current during the interval to measure T . In this way the output voltage will be linearly proportional to T . A digital counter is added to increase the measurable time range. This solution is characterized by a simple architecture but requires a precise biasing of the circuits for the generation of the ramp signals if high resolutions want to be achieved. In [85], a Σ - Δ configuration is proposed. This architecture, featuring various time subtractors, exploits previous measurements implementing a quantization noise shaping. The system is suitable for Light-Detection And Ranging (LIDAR) and rad-hard applications [85]. Σ - Δ configurations may achieve picosecond level resolutions but their main drawback is related to the complexity of their architecture. Moreover, the noise shaping is effective when repeated measurement are performed in a limited amount of time. A detailed analysis of various TDC configurations can be found in [72, 85]. The pros and cons of the discussed TDC architectures are briefly summarized in Table 2.2.

In Chapter 4, two TDC architectures, designed and tested for TT-PET and FASER experiment test chips, will be described and analyzed. As explained before, in HEP and medical imaging applications, the integration of a large number of TDCs with high time resolution (e.g., ≤ 100 ps) is often required to improve the image or track reconstruction. For this reason, a simple, compact, easily scalable, low-power design is crucial for this kind of applications. In Section 4.1, possible TDC integration strategies in a pixel detector system will be discussed and the importance of a design characterized by a simple architecture and good time performance will be highlighted.

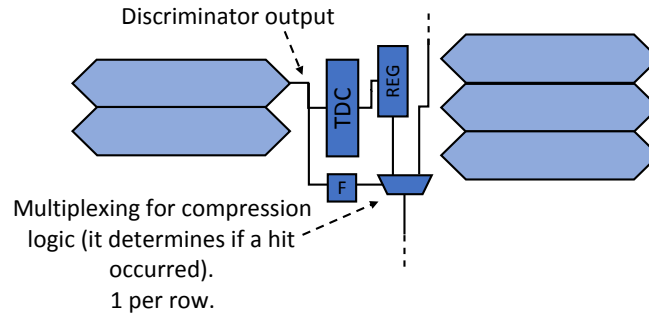


Figure 2.11: Simplified block diagram of the compression logic of the FASER prototype chip.

2.2.3 Readout logic

Each time an event is detected by a pixel, the detector needs to extract from that event the information required for the hit reconstruction. The number of bits to acquire and store can be very high depending on the characteristics of the system (e.g., the length of the addresses used to identify the pixels in the matrix, the number of bits provided by the TDC and so on) and the event rate. For this reason, various detecting systems require data compression to reduce the total amount of data to handle during the readout. One of the possible readout technique is the packet-based approach with compression, an event-driven solution in which the number of data to sent is linearly proportional to the number of hits in the chip and ideally zero if no event occurred. This architecture was implemented for the Timepix3 chip [86]. An alternative solution is represented by the frame-based compression [87]. In this solution, every pixel of the matrix is read but the number of bits to sent is smaller if no event occurred on the pixel. This data compression approach is similar to the solution implemented for the FASER prototype chip depicted in Figure 2.11: a flag is used to indicate whether an event occurred on at least one of the pixels of a row or not. In this way, it is possible to skip the row even if a certain amount of data (i.e., the bits of the flag) needs to be sent. The frame-based technique can be implemented with different levels of compression. One of the possible approaches is

TDC architecture	Pros	Cons
RO-based	Simplicity of the design	Calibration or a PLL is often required
Vernier line	Resolution below the single delay cell	Extended measurement range may be expensive in terms of area
TAC	Simple architecture and potentially good resolution	Precise biasing of the ramp generators is required
Σ - Δ	Picosecond level resolution and rad-hardness can be achieved	Complex design that can be suitable only for specific applications

Table 2.2: Pros and cons of the discussed TDC architectures.

Chapter 2. Timing detector basics

based on dividing the matrix in group of pixels indicated as super-pixels. A set of flags indicate if at least one hit occurred in the super-pixel. In this way, it is possible to skip the superpixel in case no pixels belonging to it is hit. This solution is similar to the one adopted for the second FASER pre-production chip (Section 3.4.2).

3 ASICs for FASER detector

As anticipated in Chapter 2, monolithic architectures represent an interesting solution for the development of particle detectors due to the reduced production costs and budget material. The integration of the sensitive area and the electronics in the same die also gives the possibility to reduce the pixel capacitances (no external bonding is required) which helps with the optimization of the time performance. However, despite the above-mentioned advantages, the design of this type of architecture is also characterized by many challenges which make monolithic pixel detectors not as diffused as hybrid ones in HEP applications [88]. Indeed, hybrid detectors allows a separated and easier optimization of the sensors and the electronics which also made this solution more attractive for the development of the trackers used in the LHC at CERN. Despite the above-mentioned reasons, the possibility of using monolithic pixel detectors has been explored over the last years: [88] discusses that higher input charge-to-capacitance ratios (Q/C) in monolithic solutions can lead to a significant reduction of the power consumption (which in HEP application is a fundamental requirement) but highlights that an efficient integration of the electronics and the pixel in the same die is particularly challenging performance-wise. FASER (Section 1.1.2) will be one of the first HEP experiments featuring monolithic pixel detectors in SiGe BiCMOS technology.

This chapter focuses on the design solutions adopted for the implementation of two pre-production ASICs for the upgrade of the pre-shower detector of the FASER experiment. The challenges related to the development of monolithic architectures are highlighted and design solutions for a functional integration of electronics in pixel area are provided. The main specifications and an overview of the detector chip are discussed in Section 3.1 while Section 3.3 and 3.4 describe the input charge analog memory system and the digital logic respectively. In Section 3.2, the front-end system is presented and the measurement results of the first test-chip prototype are reported. These results were published in the following paper:

- [39] F. Martinelli et al. “Measurements and analysis of different front-end configurations for monolithic SiGe BiCMOS pixel detectors for HEP applications”. In: *Journal of Instrumentation* 16.12 (Dec. 2021), P12038. DOI: 10.1088/1748-0221/16/12/p12038.

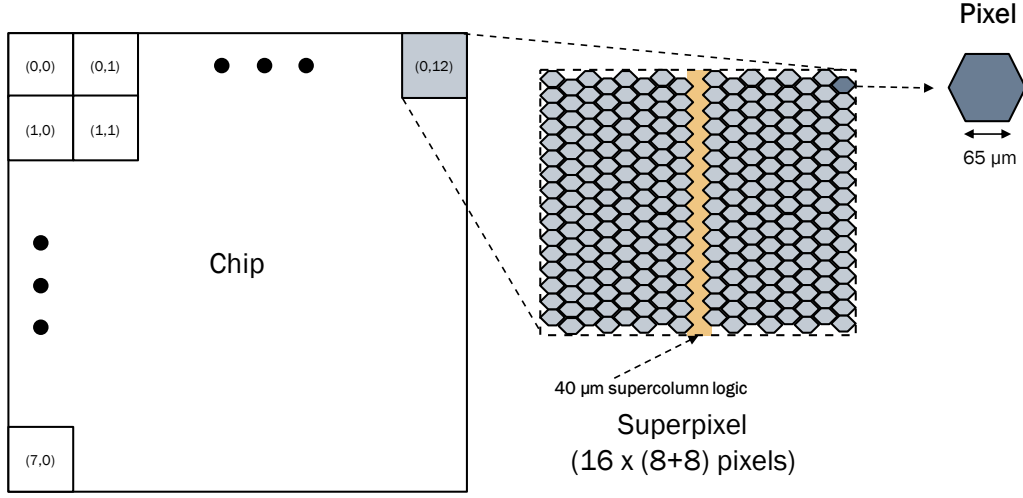


Figure 3.1: Architecture and pixel distribution of the final version of the FASER pre-shower detector chip.

3.1 General architecture and specifications

As anticipated in Section 1.1.2, the enhancement of the research on dark matter with the FASER experiment requires the adoption of a high-granularity pre-shower detector able to sense and correctly distinguish pairs of photons. The latter can be characterized by spatial separations in the order of few hundreds of μm due to their expected energy (100 GeV - 4 TeV range). For this reason, the final ASIC of the pre-shower (that is going to be submitted for production in 2022) will feature a matrix of pixels with a 100 μm pitch in both the x and y direction approximately.

Also for this project, a monolithic architecture has been chosen thus the pixels are integrated with the rest of the electronics in the same die. Figure 3.1 displays the block diagram of the final version of the FASER chip. The structure is supposed to be composed of 13 super-pixel columns, indicated as super-columns. The super-pixels are sub-matrices of 16 by 16 pixels divided in two groups of 126 pixels each by the super-column logic (yellow area in the central part of Figure 3.1). More details on this block are provided in Section 3.4.1. The size of the final chip is $23.2 \times 15.3 \text{ mm}^2$ while the pixels are characterized by a hexagonal shape with a 65 μm side (this length is necessary to satisfy the $\sim 100 \text{ }\mu\text{m}$ pitch requirement). The hexagonal shape of the pixels has been chosen to increase their break-down voltage: a sharper p-n junction can reach the critical electric field E_C ¹ with smaller voltages [89, 90]. Therefore, the adoption of hexagonal shapes is useful to increase the break-down voltage of the p-n junction of the pixel with respect to a rectangular structure. In this way, it is possible to bias the pixels with higher voltages potentially leading to better timing performance [26] (a discussion on the topic is also reported in Section 2.1.1 and Appendix B).

¹ E_C represents the minimum value of electric field in a p-n junction that leads to the onset of the multiplication regime. Its value is also dependent on the doping profile of the junction. More details in [89].

A 3D representation (not in scale) of the ASIC substrate cross-section is displayed in Figure 3.2a. As done for the TT-PET project (Section 1.1.1), a monolithic architecture has been chosen to design a detector ASIC in 130 nm SiGe BiCMOS technology provided by IHP Microelectronics. This foundry also gave the design group the possibility to integrate a custom high-resistivity ($\rho=350 \Omega\text{cm}$) 50 μm EPI-layer which can be completely depleted with a substrate bias voltage of $-HV < -120 \text{ V}$. Moreover, from the data in [92], it is possible to calculate that this high-resistivity portion of the substrate allows the production of a 0.5 - 1.0 fC input charge when the sensor is stimulated by MIPs. For this reason, this value of charge is set as the minimum of the target input dynamic range. Various simulation analysis performed with GEANT4 platform (see Figure 3.2b) highlighted that photon signals can produce input charges up to 64 fC. Moreover, Figure 3.2b also shows that couples of photons can be characterized by an asymmetrical charge release in the above mentioned 0.5 - 64 fC input dynamic range. The FASER detector ASIC needs to satisfy a set of specifications that can be found in [91]. It is important to underline a required power consumption of less than 7 $\mu\text{W}/\text{pixel}$ and a readout time of $\lesssim 30 \mu\text{s}$. The latter is a crucial specification that must be satisfied to not compromise the charge information stored in the analog memories of the pixels. More details will be provided in Section 3.3.

The final ASIC of the FASER pre-shower detector is supposed to feature 13 super-columns and will be submitted for production in 2022. However, three different pre-production chips have been designed and submitted: one of them has 4 super-columns while the other two only 3. For all of the variants, the size and the characteristics in the vertical direction (referring to the Figure 3.1) are the same of the final chip. In the 4 super-column base variant, the pixel pre-amplifiers are connected to MOS-based discriminators while in one of the two 3 super-column versions, this circuit is implemented with a HBT-based input differential pair (more details will be shown in Section 3.2). Both of the versions have been integrated in two different ASICs for the testing and comparison of their performance in terms of speed, noise and input offset. In many front-end system, the input offset is corrected using DACs

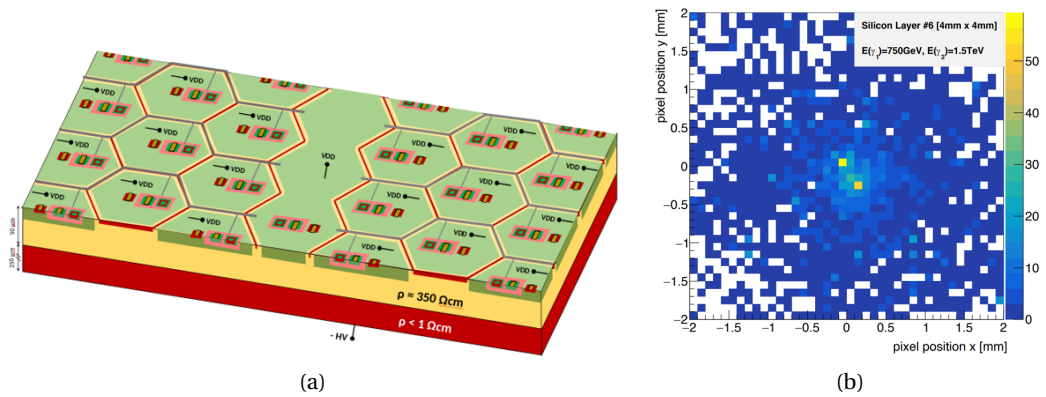


Figure 3.2: FASER chip substrate cross-section representation (a) and GEANT4 simulation of a typical event in the detector (b) [91].

for the biasing of local threshold. Indeed, in order to improve the timing performance of the pre-amplifier, a discriminator solution with small area devices for the differential input pair is often adopted [87]. However, as explained in [93], small area transistors are more affected by mismatch effects² hence local DACs are required for pixel calibration [87]. For the FASER ASIC, the high pixel density and the requirement of a high detection efficiency (>97%) did not allow the integration of a local DAC per pixel. Therefore, the adopted solution is based on dividing the matrix in 8 regions and use a single 8-bit DAC for each of them to set their local threshold. This solution represents a compromise to deal with the mismatches of the front-end systems in large ASICs. The BJT version of the discriminator should further improve the characteristics of the chip in terms of mismatch, making the above-mentioned solution more efficient. Indeed, a simulation analysis showed that the input offset of BJT discriminators is approximately 3 times smaller than the one shown by the MOS transistor-based version. In BiCMOS technological nodes, bipolar transistors are usually vertical implantation devices in which the doping profiles are sharper and more precise if compared to horizontal MOS transistors. This explains the reduced mismatch of HBT-based amplifiers.

The third variant of the pre-production chip is featuring Linear Feedback Shift Registers (LFSR) as counters connected to the discriminator for the measurement of the input charge: the latter is calculated by counting the number of a reference clock cycles within the Time-Over-Threshold (TOT) i.e., the period of time in which the discriminator is activated. This chip variant is particularly inefficient in terms of dead area since the integration of the counters required a three times thicker super-column logic (Section 3.4.1). However, this chip has been designed as a safe variant and to compare its performance and efficiency with the analog memory version of the front-end.

3.2 Front-end system

This section presents a monolithic pixel detector test-chip developed to study different design solutions to be used for the high-precision pre-shower upgrade of the FASER experiment at CERN. The prototype chip presented here, also indicated as Prototype-0, was designed to study the integration of the front-end electronics (or part of it) in the sensor area, a solution explored to minimise the detector inactive area, that is a potential limiting factor of monolithic architectures. A photograph of the prototype chip is shown in Figure 3.3 (right). The pixel matrix of the ASIC is composed by two superpixels, as seen in Figure 3.3 (left), each featuring 16×4 pixels with a $200 \times 50 \mu\text{m}^2$ area³, a Time-to-Digital Converter (TDC) and a digital

²A couple of nominally identical MOS transistors, because of mismatches given by doping fluctuations, will show different values of threshold voltages. Their difference is characterized by a standard deviation that follows Pelgrom's relation $\sigma_{\Delta(V_{th})} = A_{V_{th}} / \sqrt{WL}$, where $A_{V_{th}}$ is a parameter depending on the MOS characteristics (such as doping concentration and oxide thickness) and WL is the transistors' area. For this reason, smaller devices are more affected by mismatch effects. More details in [93].

³The pixel area and shape implemented in this prototype were those considered for the FASER high-precision pre-shower at the time of the submission of this ASIC. Successive full-simulation studies have shown that the optimal pixel size is hexagonal with $65 \mu\text{m}$ side (corresponding approximately to a pixel pitch of $100 \mu\text{m}$), which has been adopted for the final ASIC, as described in Section 3.1

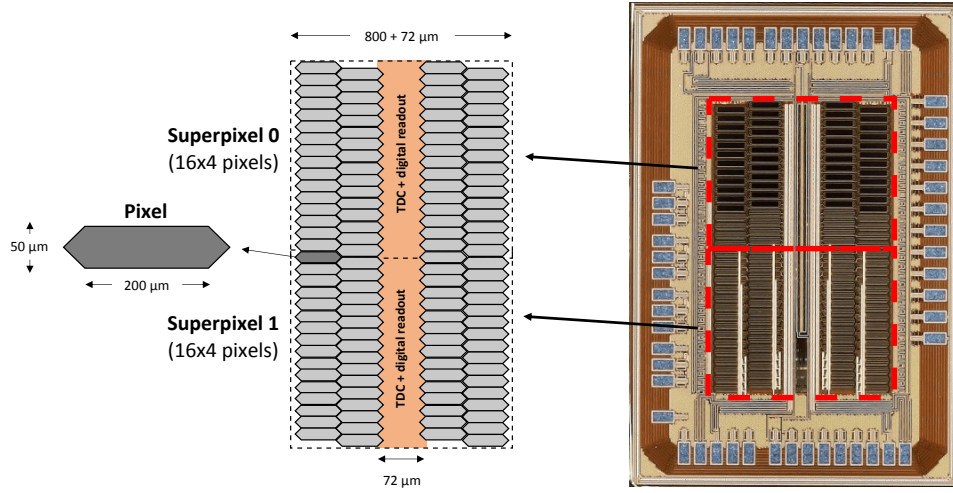


Figure 3.3: Super-pixels and pixel size (left) and a photograph of the prototype ASIC (right); the ASIC total area is $1.7 \times 2.6 \text{ mm}^2$.

readout logic placed in the $72 \mu\text{m}$ thick region highlighted in orange in Figure 3.3.

3.2.1 Pre-amplifier and discriminator design

The sensor implemented in this prototype chip is a PN junction operating in reverse bias. In this mode, the anode is connected to the p-doped substrate of $50 \Omega\text{cm}$ bulk resistivity while the cathode is an n-well with an elongated hexagonal shape, as described before. When a negative voltage $HV = -120 \text{ V}$ is applied, a depletion region of approximately $20\text{-}25 \mu\text{m}$ is created in the substrate. Such depletion region leads to the generation of $1200\text{-}1800$ electron-hole pairs [92] when a minimum-ionizing particle passes through the sensor.

The pre-amplifier is the first stage connected to the sensor (depicted as a diode in reverse bias in Figure 3.4). The pixel n-well is biased at a low voltage through a polycrystalline silicon $443 \text{ k}\Omega$ resistor. The pre-amplifier features a single-ended BJT-based first stage with active load. The latter provides a bias current of few μA and is implemented with a PMOS transistor. Since the connection with the pixel sensor is in AC, an additional bias is needed for the base current of the bipolar transistor. For this purpose the block on the left part of Figure 3.4 is used, in which the MOSFET on the right, used as a feedback impedance, is able to show a resistance of several hundred $\text{k}\Omega$ (simulations highlight a value of approximately $670 \text{ k}\Omega$ for a feedback bias current of 30 nA). The AC coupling capacitor is implemented with several PMOS transistors whose body terminals are connected to the n-doped cathode of the pixel and the gates, shorted to sources and drains, to the base of the BJT.

The first stage of the front-end system behaves as a charge amplifier, capable of producing output voltage signals directly proportional to the charge injected at the input. Indeed, simulations show that the condition of Equation 2.6 is satisfied. The main contribution of the feedback capacitance C_f is given by the base-collector capacitance of the BJT and it is

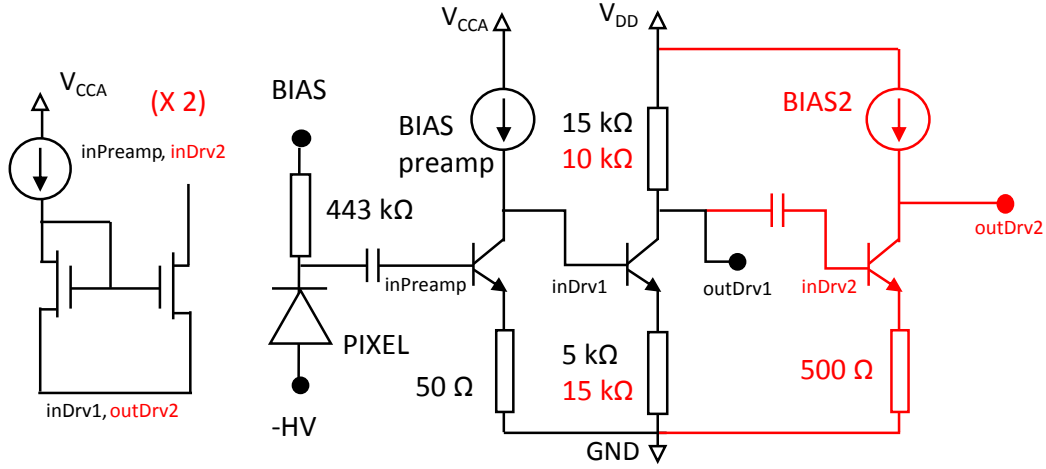


Figure 3.4: Pre-amplifier architecture. The stage and resistance values highlighted in red refer to the inverting solution that characterizes only one of the flavours adopted for the test chip.

approximately $C_f \approx C_{BC} \approx 2\text{fF}$. Moreover, considering an input resistance $R_{in} \lesssim 100\text{ k}\Omega$ and a voltage gain A_v in the order of few tens, the pole described in the second term of Equation 2.6 will lay in the 100 MHz range. Because of the above-mentioned 20-25 μm depletion of the substrate, the input signals of the pre-amplifier will have rise times in the order of few hundreds of picoseconds. Therefore, the spectrum of these signals will be over the first pole of the system and the amplifier will work in integration regime i.e., as a charge amplifier. As explained in 2.2.1, this solution is useful to cope with noisy environments and to improve the timing performance. More in detail, considering the small signal model, the voltage gain of the first stage is

$$A_{v1} = Z_O \frac{\frac{1}{Z_f} - \frac{g_m r_\pi}{r_\pi + (1+\beta)R_E}}{1 + Z_O/Z_f}, \quad (3.1)$$

where Z_O is the output impedance of the stage (mainly given by the impedance of the bias PMOS connected to the collector), R_E is the 50 Ω emitter resistance and Z_f is the feedback impedance of the HBT. The latter can be estimated as the parallel between the C_f impedance and the resistance shown by the feedback block depicted on the left of Figure 3.4. Simulations show that the value of R_E is small enough to justify the following approximation

$$A_{v1} \approx Z_O \frac{\frac{1}{Z_f} - g_m}{1 + Z_O/Z_f}. \quad (3.2)$$

R_E has been integrated to improve the stability of the system, acting as a negative feedback in order to compensate potential spurious feedbacks through the ground line or temperature variations. However, because of its small value, R_E has a negligible effect on the gain of the system.

The way the feedback block is connected to the HBT can have different effects on the characteristics of the pre-amplifier. In the first prototype designed for the FASER pre-shower, the

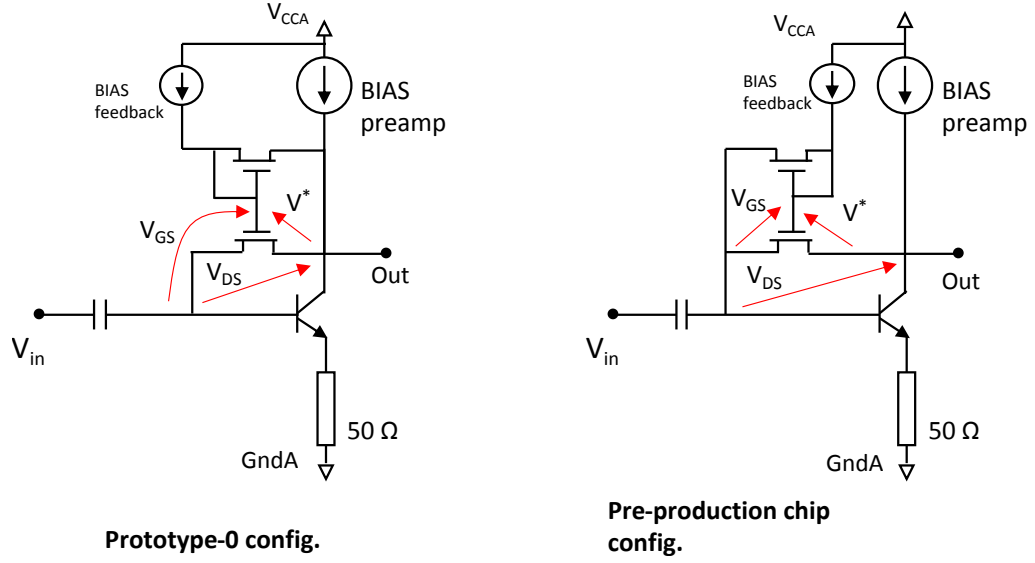


Figure 3.5: Configuration of the first pre-amplifier stage integrated in the first prototype (left) and the pre-production chip (right) for the FASER pre-shower.

circuit can be redrawn as in Figure 3.5 on the left. Using the Kirchhoff's law, it is possible to highlight that

$$V_{GS} = V_{DS} + V^*. \quad (3.3)$$

In this circuit, the voltage V^* is equal to the gate-source voltage of M_1 , $V^* = V_{GS-M1}$. Since M_1 is biased with a current in the order of few tens of nA, this MOSFET works in Weak Inversion. For this reason, and as confirmed by simulations, $V_{GS-M1} = V^*$ will be relatively small compared to V_{GS} and V_{DS} , leading the second MOS of the feedback block towards saturation and, consequently, increasing its output resistance. This solution, that has also been implemented for the ASIC for hybrid pixel detector described in Chapter 5, is useful to increase the gain of the pre-amplifier. In the second ASIC for the FASER pre-shower, the connection of the feedback block is different as reported in Figure 3.5 on the right. This time $V_{GS} = V_{GS-M1}$ that, for the same reason mentioned above, will be significantly below the threshold voltage of the MOSFETs. As highlighted by Equation 3.3, a small V_{GS} leads to a small V_{DS} and, for this reason, this solution will push the second MOS of the feedback towards the linear region, reducing its output impedance and the voltage gain of the pre-amplifier. Two main considerations related to the pre-amplifier configurations are reported below:

- because of the expected rise time of the input signals, the main contribution to the gain of the pre-amplifier is given by the capacitance C_f that will dominate R_f in the feedback impedance Z_f . Hence, in the FASER pre-production ASIC the first pre-amplifier stages are expected to show similar performance in terms of gain compared to the ones of the first prototype.

- full-simulation studies performed after the production of the first prototype highlighted that the expected input charge range spans between 0.5 and 64 fC. The latter is slightly larger than the one expected during the design of the Prototype-0. The second configuration is characterized by a better TOT linearity and range, especially for large charges (>5-10 fC). In this case, the rise of the output signal (due to a negative input) would lead to a situation in which the MOSFETs of the feedback current mirror are better biased (i.e., larger drain-source and gate-source voltages). The discharge of the output signal would therefore be (approximately) linear. In the Prototype-0 configuration, large charges and the consequent rise of the output would turn off M_1 for a certain period before the discharge. Its gate would rise to the positive supply and this would make the output discharge in a less linear fashion. Because of a larger input charge range, the second configuration represents a more optimal solution for the pre-production ASIC and its future iterations.

The gain of a charge amplifier is largely dependent on its feedback capacitance C_f . As explained before, in the solution presented in this section, C_f is mainly given by the parasitic base-collector capacitance of the HBT. An extensive analysis based on Monte Carlo simulations highlighted that process variations and the consequent variability of C_f have a negligible impact on the performance of the pre-amplifier. Indeed, as anticipated in the previous section, HBTs in the SiGe BiCMOS technology are vertical implantation devices. Therefore, the behavior of their parasitics is less subject to mismatches and process gradients compared to MOSFETs.

The second stage of the pre-amplifier is used to increase the voltage gain of the system by a factor $A_{v2} \approx R_C/R_E$ where R_C and R_E represent the collector and emitter resistances. It is also lowering the output impedance of the pre-amplifier, making it more robust to routing. The coupling with the first stage is in DC, hence this stage does not need another bias system for the base current.

The third stage, highlighted in red in Figure 3.4, is an additional gain stage used to produce an output signal (outDrv2) with the same polarity of the input of the pre-amplifier system (inPreamp). This particular configuration is featured only in one of the pre-amplifier variants that have been integrated in the test-chip. The design process of this configuration aimed to implement a structure in which the stability of the amplifier was the main optimization parameter to take into account. Indeed, outDrv2 follows the same behavior of inPreamp inducing a negative feedback that is meant to avoid unwanted oscillations, making this architecture suitable for an integration inside the active area of the pixel. This configuration presents a second stage with a reduced gain compared to the other (10 and 15 k Ω instead of 15 and 5 k Ω as indicated by the red labels of the resistors in Figure 3.4) in order to further reduce the impact of the output (outDrv1) that shows a positive feedback coupling with the pixel. The other pre-amplifier flavours only feature the part of the circuit depicted in black in Figure 3.4. The differences among these configurations will be described in Section 3.2.2. More details about the techniques to improve the stability of the system are reported in Section 3.2.3. Simulations show that the 433 k Ω pixel bias resistor is responsible for $\sim 10\%$ of the total noise

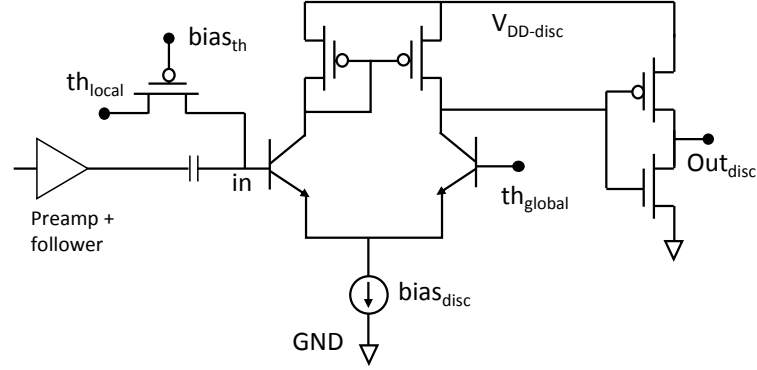


Figure 3.6: Discriminator schematic and connection to the pre-amplifier.

on the output of the pre-amplifier. The other main contributions are provided by the feedback block and the bias PMOS connected to the collector of the HBT of the first stage ($\lesssim 20\%$ each). Moreover, the performance in terms of noise of the front-end systems satisfies the specifications, as confirmed by simulations and by measurement results (Section 3.2.4). The solution based on using a resistor to bias the pixel was adopted for its simplicity. The schematic of the discriminator chosen for the front-end is reported in Figure 3.6. The differential pair is designed also in this case using SiGe HBTs with a PMOS-based active load. The threshold of the circuit is set through the global threshold signal th_{global} , distributed to the front-end of every pixel in the chip, and a local threshold th_{local} . As it will be explained in Section 3.2.3, the design of the discriminator plays an important role in the stability of the front-end, especially when the system is integrated in pixel.

The presented front-end chain contains various AC coupling connections. This solution can lead to potential baseline drifts in the case of high signal rates (if it is not properly handled). However, the FASER pre-shower is expected to be characterized by relatively small (muon) background rates in the order of $\sim 0.5 \text{ Hz/cm}^2$. Therefore, the aforementioned AC coupling connections are not expected to have a negative impact on the performance of the ASIC.

3.2.2 Flavours adopted for this FASER prototype-chip

The aim of this prototype chip was to perform an analysis of several front-end configurations and to choose the best solution for the final full-reticle ASIC of the FASER pre-shower. These configurations are characterized by different degrees of integration in pixel of the electronics. In general, integrating circuits inside the sensitive area can lead to an increase of the detector capacitance and, consequently, to a reduction of the SNR. Therefore, the noise performance of the versions of the front-end with many blocks in pixel were expected to show worse performance in terms of noise. Further details will be provided in Section 3.2.4.

Figure 3.7 shows the floorplan of the front-end variants:

- the first one, reported in purple, features the whole front-end outside the pixel. Because

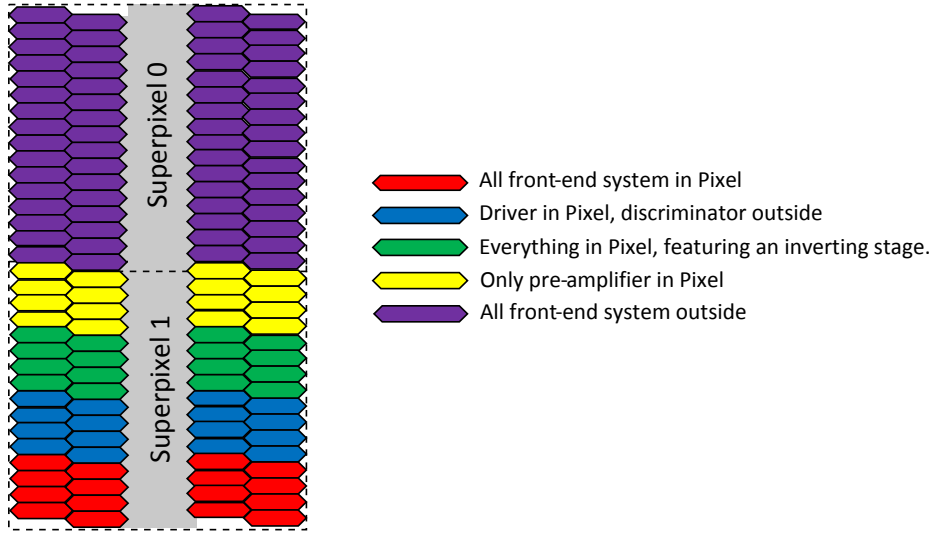


Figure 3.7: Distribution in the pixel matrix of the front-end variants of this test chip.

of the demanding specification of the final FASER ASIC in terms of dead area and pixel density, this version will not be inserted in future iterations of the chip. However, it was integrated in the presented prototype ASIC to compare its performance with the other versions of the front-end: this solution was expected to be the least critical one for the stability and noise thus it can be used as a reference for the evaluation of other configurations.

- the pixel reported in yellow feature a front-end with only the pre-amplifier inside the pixel. The gain of this architecture was expected to be smaller than others since increasing the length of the routing that connects the output of the pre-amplifier with the driver can significantly reduce the band-width and the gain of the system.
- in blue and green the pixel configurations with pre-amplifier and driver integrated in the sensitive area are reported. In particular, the green ones indicate the solution with an additional inverting stage to further improve the stability (reported in red in Figure 3.4).
- finally, the configuration reported in red, has all its blocks, including the discriminator, integrated in the pixel well. As it will be highlighted in Section 3.2.3, particular care was dedicated to the layout of this architecture in order to reduce self-induced oscillations.

During the design process, the last three solutions were expected to be the best candidates to be integrated in the final chip of the FASER detector because they are the least demanding in terms of dead area. Their performance and a more detailed comparison of the configurations are reported in Section 3.2.4.

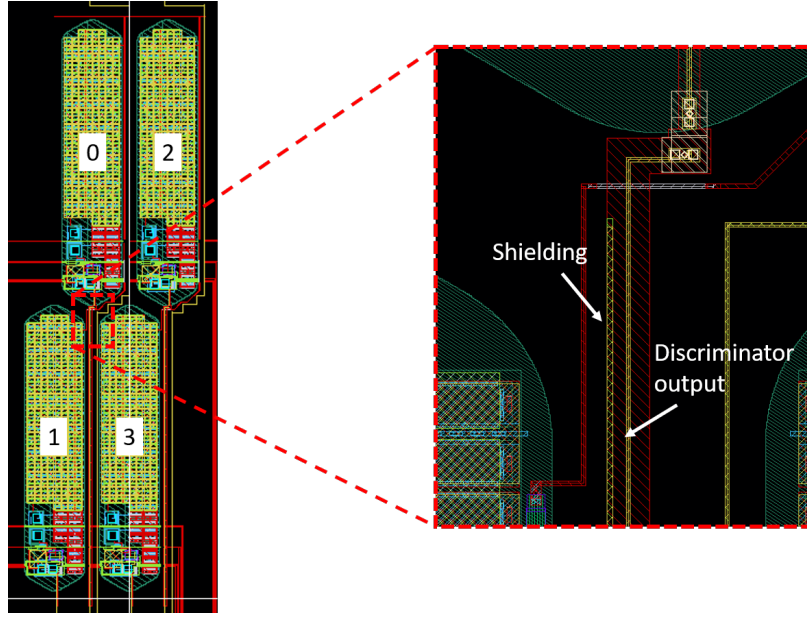


Figure 3.8: Layout of 2×2 pixels (left) and zoom (right) on the shielding line to reduce the cross-talk between the output of the discriminator and the adjacent pixel wells.

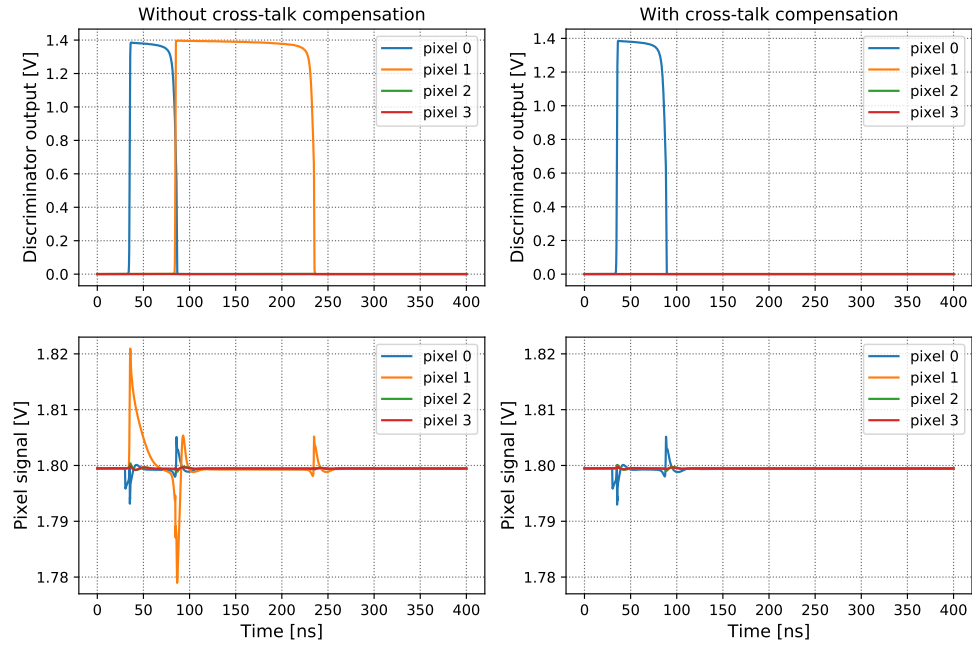


Figure 3.9: Discriminator output and pixels signal with (right) and without (left) cross-talk protection lines for an input charge of 0.5 fC.

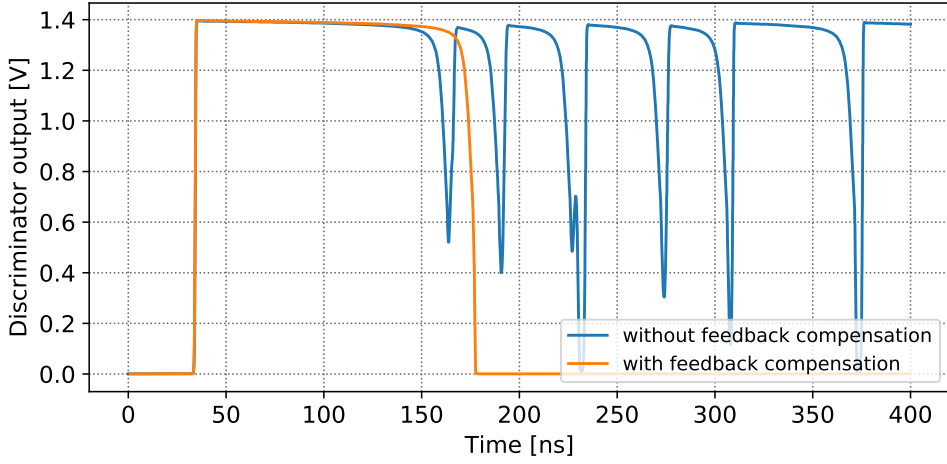


Figure 3.10: Discriminators output behavior with (orange) and without (blue) self-induced noise compensation lines.

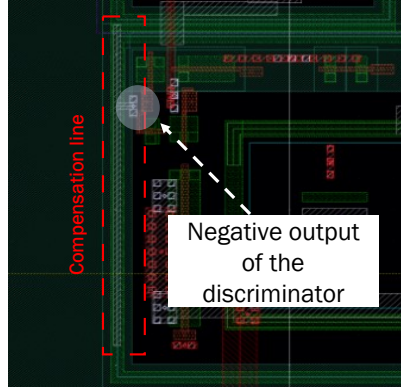


Figure 3.11: Example of self-induced noise compensation line. The metal2 line is connected to the negative discriminator output to increase the coupling with the pixel well and avoid self oscillation induced by the effect of the positive output.

3.2.3 Cross-talk compensation and layout

A crucial part of the design process focused on the optimisation of the layout. An extensive simulation campaign was launched to analyze the performance of the chip. Layouts with 2×2 pixels sub-matrices featuring various front-end configurations were simulated to evaluate the effects of coupling between discriminator output lines and neighbouring pixels, as seen in Figure 3.8 (left). The analysis showed that, if no shielding line is included in the layout, the falling edge of the discriminator output of pixel 0 of Figure 3.8 can induce a spurious hit on another pixel (in this case, pixel 1) as it is reported in the plots of Figure 3.9 (left). The shielding lines avoid this problem, as seen in Figure 3.9 (right), without any significant impact on the discriminator performance.

As mentioned in Section 3.2.2, the front-end configuration that present a complete integration of the discriminator inside the pixel active area may be critical for the stability of the system.

Indeed, as clearly shown in Figure 3.10, the coupling between the discriminator output and the pixel could generate a positive feedback leading to unwanted oscillations of the front-end chain. In the present prototype the problem was solved inverting the polarity of the output of the discriminator and increasing its coupling with the pixel exploiting additional metal lines as shown in Figure 3.11. This compensation technique led to a correct behavior of the front-end (orange curves in Figure 3.10).

During the process of integration of the metal fillers in the ASIC, the regions of the pixels and the front-end systems were excluded. In this way, potential negative effects of the fillers on the performance of the amplifiers are prevented. The same solution was adopted for the pre-production ASIC.

3.2.4 Measurements and comparisons

The ASIC has been tested and qualified with the UNIGE USB3 GPIO system that was initially developed by the engineering team of the Département de Physique Nucléaire et Corpusculaire (DPNC) of the University of Geneva for the Baby-MIND experiment at CERN [94]. The GPIO system includes a readout board that uses an Altera Cyclon V FPGA that can control several detectors at once.

All the measurements reported in this section were performed setting the power consumption of the chip at 144 mW/cm^2 thus within the 150 mW/cm^2 specification of the FASER experiment.

Calibration

As shown in Figure 3.6, the front-ends of this prototype feature a global threshold th_{global} signal connected to the base of the negative input of the discriminators. In addition to this global threshold, each pixel also features a 6-bit Digital-to-Analog Converter (DAC) used to set the local threshold th_{local} . The role of this signal is to bias the positive input of the discriminator and move the threshold of the discriminators with respect to th_{global} . The th_{local} of each front-end can be set independently of each other to compensate for mismatch effects and guarantee that the equivalent threshold is as uniform as possible across the matrix.

For a certain value of th_{global} , the calibration algorithm is based on a scan of the local threshold for each pixel performed by changing the input of the associated DAC. In this way, it is possible to obtain the values of the input codes of the converter that make the discriminator output switch. Indeed, in a certain range of th_{local} the baseline of the input will be so close to its threshold that the noise will activate the discriminator multiple times. The correct value of DAC input for the calibration is chosen such that the noise hit rate is low (in the 0.1 - 0.01 Hz range). The fastest way to perform this operation is to make a simultaneous scan of all pixels. However, the activation of several discriminators at the same time will produce peaks of absorption that can compromise the accuracy of the calibration process. Measurements highlighted that the difference between the threshold value obtained with the calibration of a single pixel and the one given by a simultaneous scan of the whole chip can be up to 22% of

the DAC dynamic range. For this reason, an alternative process was developed in which the pixels were divided in eight groups such that the calibration was performed independently for each group. An efficient distribution of the mapping of these eight groups of pixels is displayed in Figure 3.12: the distance maintained among the pixels of the same group drastically reduces the calibration error to values up to 1 Least Significant Bit (LSB) of the local DACs, i.e. less than 3% of their dynamic range.

Figure 3.13 displays the results of the threshold calibration. Figure 3.13a shows the values of the DAC codes associated to the local threshold of the pixels while Figure 3.13b displays the distribution of the average LSB of the converters in the matrix. The LSB was calculated performing a calibration with four different values of the global threshold (0.95, 1.00, 1.05 and 1.10 V) and evaluating the corresponding calibration code for each pixel. At this point, the average LSB of the i -th pixel LSB_i can be obtained as

$$LSB_i = \frac{\Delta(th_{global})}{\mu(\Delta(O_i))}, \quad (3.4)$$

where $\Delta(th_{global})=50$ mV is the global threshold step implemented for the measurement and $\mu(\Delta(O_i))$ is the average value of the difference between the code obtained for a given global threshold and the previous one. In these two plots, it is possible to highlight an asymmetry between the DAC outputs of the pixels on the left (even pixels) and right (odd pixels) portion of the matrix: this effect can be attributed to a gradient in the fabrication process of the chip and to asymmetries in the converters. The decreasing trend of the average LSB in Figure 3.13b can also be associated to process gradients and to voltage drops on the supply.

Figure 3.13c shows the distribution of the equivalent threshold of all the front-ends in the chip. The threshold $V_{th,i}$ of the i -th pixel is calculated combining the information of the two previous plots as

$$V_{th,i} = V_{DD} - (10 + O_i)LSB_i, \quad (3.5)$$

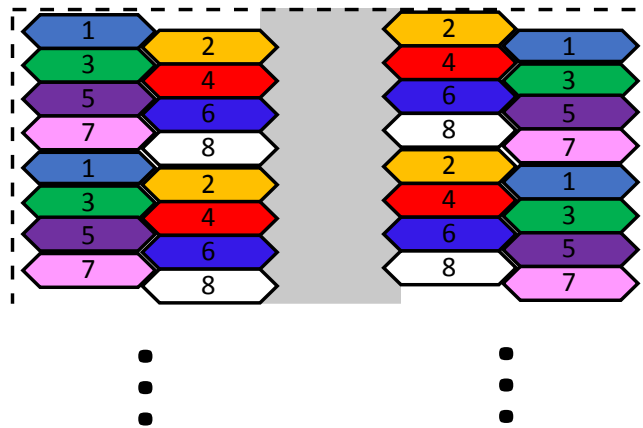


Figure 3.12: Mapping of the pixels in the eight groups used for the threshold calibration.

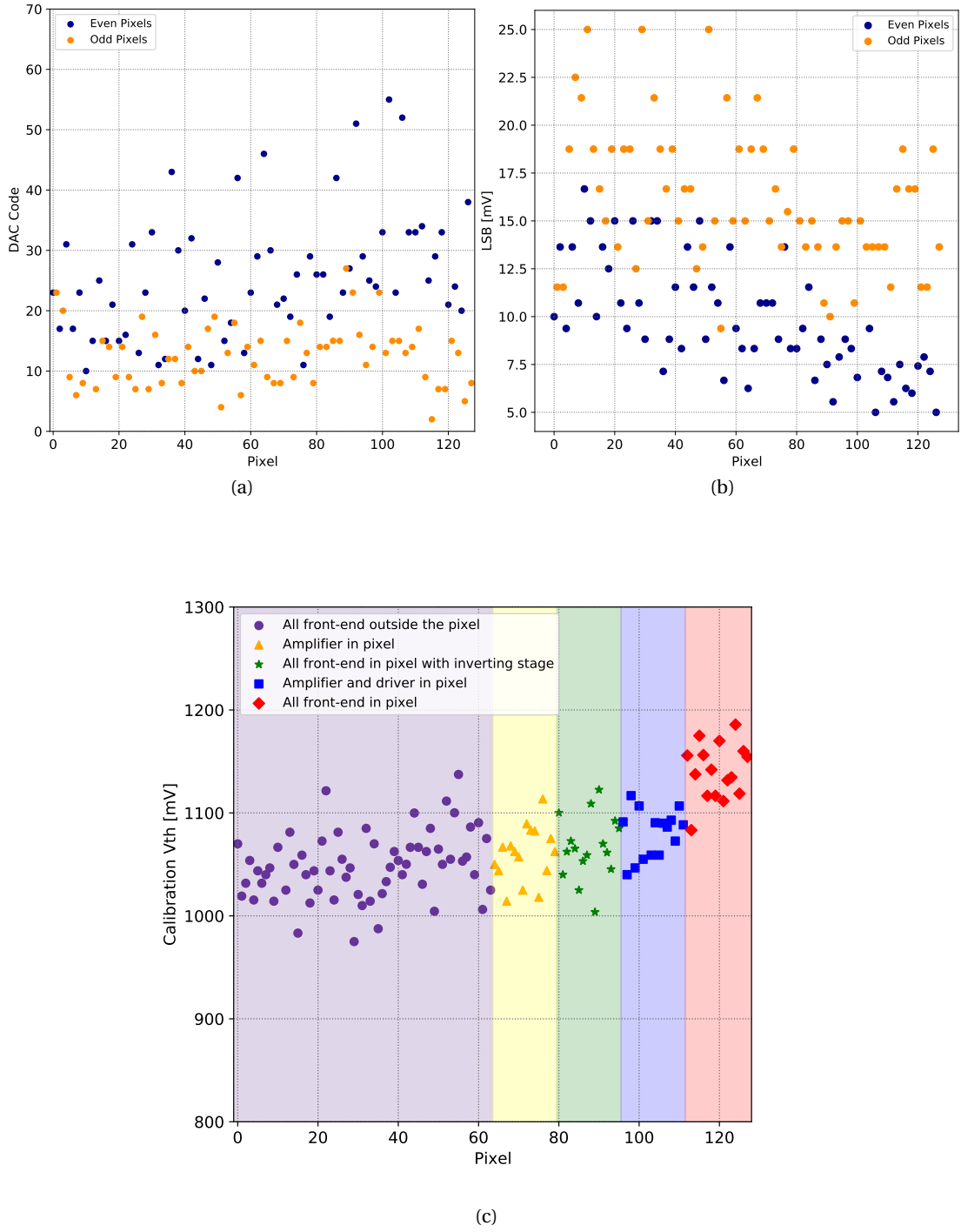


Figure 3.13: DAC code (a), DAC average LSBs (b) and front-end threshold (c) distribution in the prototype chip. The x-axis indicates the pixel number. The pixel are labeled from bottom to top of Figure 3.7 and split in even and odd on the right and the left part of the chip. The results in (a) were obtained setting the global threshold at 1.1 V.

where the factor 10 is associated to a current offset of 10 LSB_i on the output of the converter. The asymmetry between even and odd pixels and the decreasing trend of the LSBs cannot be deduced by the plot of Figure 3.13c because the latter only describes the dispersion of the equivalent threshold given by mismatches of the electronics, i.e. without the contribution of DAC. The plot also highlights a threshold dispersion $\sim 30 \text{ mV}$ (exact values are reported in Table 3.1). In addition, it is possible to notice that the pixels with the whole front-end integrated in the sensitive area are showing a higher threshold than the others. This effect is caused by the fact that, in this configuration, the PMOS transistors of the discriminator share the body with the pixel n-well, thus their threshold voltage is different from the one associated to PMOS integrated in external wells.

The DACs integrated within the test-chip are based on a multiple current mirrors architecture in which the i -th input bit drives a mirror that doubles i times a bias current (LSB). The variation of the LSBs showed in Figure 3.13b is caused by the mismatches of the converters. Because of the size of the DACs and their dispersion (a better matching would be obtained with bigger MOS transistors [93]) this architecture will not be integrated in future prototypes of the FASER experiment.

The final ASIC will feature either R-2R ladder DACs [95] (characterized by a more compact architecture) or a set of converters placed in the periphery of the chip that will calibrate small sub-matrices. This choice is motivated by the small threshold dispersion obtained in the test-chip and also by the significantly larger number of pixels that will be integrated in the final FASER ASIC. Moreover, the gain measurements reported in the following section confirm that the threshold dispersion is small enough to make the system able to meet the experiment requirement on the minimum input charge to discriminate (1 fC) even without using the local DACs.

Tests with ^{113}Cd and ^{55}Fe sources

A ^{109}Cd radioactive source was used to measure the gain of the front-end circuits. The emission spectrum of this radioisotope is reported in [96]. The gain evaluation was performed with a threshold scan as displayed in the representation of Figure 3.14a: using the calibration data, the local threshold was set to an initial value close the baseline and then decreased (or increased, depending on the chosen front-end configuration). For each threshold step, the

Configuration	$\sigma_{V_{th}}$ [mV]
All f.e. outside pixel	32.3
Only pre-amp. in pixel	26.9
All f.e. in pixel, inv. stage	30.8
Pre-amp. and driver in pixel	23.4
All f.e. in pixel	27.1

Table 3.1: RMS threshold dispersion $\sigma_{V_{th}}$ for each front-end configurations integrated in the chip.

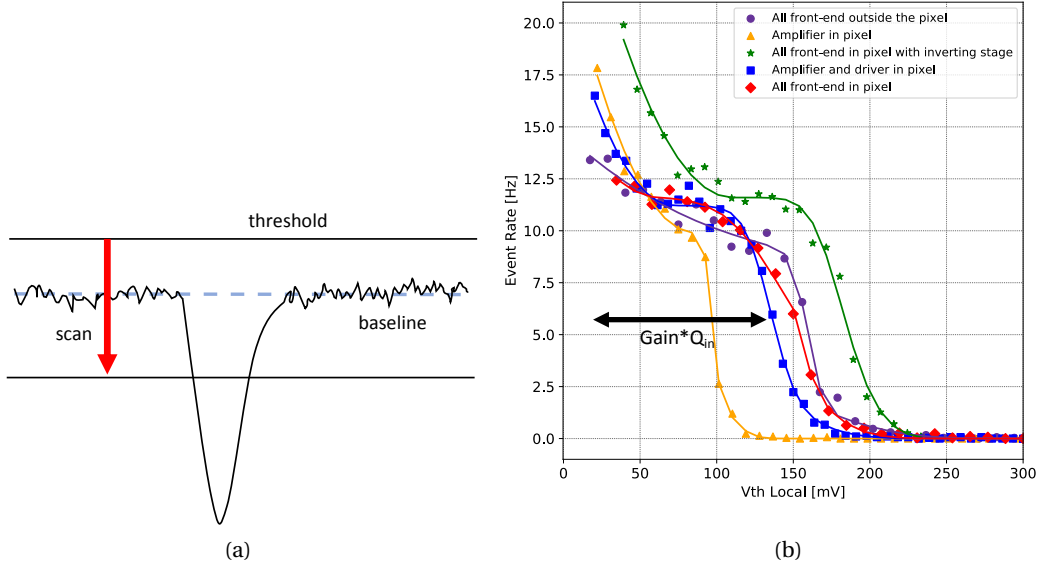


Figure 3.14: (a) Threshold scan representation for gain evaluation and (b) event rate as function of the threshold for different front-end configurations. The x-axis of (b) is referred to the baseline.

events were acquired and the event rate recorded. Figure 3.14b shows the result of one of such measurements. The data were then analyzed using the fitting function

$$F_{Cd}(x) = \begin{cases} N \cdot \text{erfc}\left(\frac{x - \mu}{\sqrt{2}\sigma_1}\right) + 0.28 \cdot N \cdot \text{erfc}\left(\frac{x - 1.13\mu}{\sqrt{2}\sigma_2}\right) + a + bx + cx^2 & x \leq \mu - 2\sigma \\ N \cdot \text{erfc}\left(\frac{x - \mu}{\sqrt{2}\sigma_1}\right) + 0.28 \cdot N \cdot \text{erfc}\left(\frac{x - 1.13\mu}{\sqrt{2}\sigma_2}\right) + d & x > \mu - 2\sigma \end{cases} \quad (3.6a)$$

where N , a , b , c , d , $\sigma_{1,2}$ and μ are the fitting coefficients and $\text{erfc}(x)$ is the complementary error function⁴. The values 1.13 and 0.28 in the function are obtained evaluating the emission spectrum of the source and its peaks [96]. The polynomial is added to be able to fit the first part of the curve, related to the region in which the threshold is close to the baseline. The charge gain in mV/fC is equal to $\mu/0.98$ because the ~ 22 keV photons emitted by the ^{109}Cd source generate a ionization charge of approximately 0.98 fC in the 20-25 μm depletion zone of this sensor. Figure 3.15 shows the value of the gain obtained with the above-mentioned method for some of the pixels for five of the front-end configurations under study.

A ^{55}Fe radioactive source was used to measure the ENC associated to the front-end amplifiers. The ^{55}Fe emission spectrum is characterized by a main peak at an energy of ~ 5.9 keV [97], that produces approximately 1650 e^- in the depletion region of our sensor. This charge is low enough to guarantee a linear response of the amplifiers. In addition, the ^{55}Fe peak is narrower

⁴The function $\text{erfc}(x) = 1 - \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt$ where $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ is the error function.

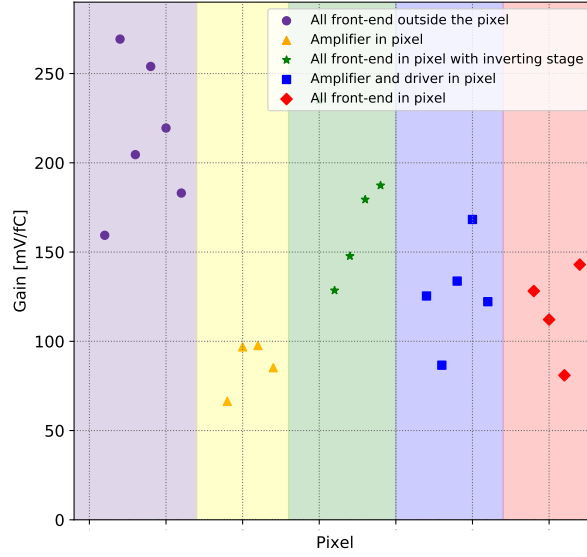


Figure 3.15: Gain of a selection of pixels that have been measured with the ^{109}Cd source for the five front-end configurations.

than the 22 keV peak produced by the ^{109}Cd radioisotope that generates an input charge of approximately $1 \text{ fC} \approx 6240 \text{ e}^-$. Because of the narrower peak and lower energy, the ^{55}Fe source is more suited to analyze the noise contribution of the front-end system and was used for our measurements.

A threshold scan was performed with the ^{55}Fe source and the event-rate data were fitted with the function

$$F_{Fe}(x) = N \cdot \text{erfc}\left(\frac{x - \mu}{\sqrt{2}\sigma_v}\right) + 9.3 \cdot N \cdot \text{erfc}\left(\frac{x - 0.9\mu}{\sqrt{2}\sigma_v}\right) \quad (3.7)$$

to calculate the σ_v component of Equation 2.7 and from it obtaining the equivalent-noise charge as

$$ENC = \frac{\sigma_v}{G_c}, \quad (3.8)$$

where G_c is the charge gain of the amplifier measured with the ^{109}Cd source. Figure 3.16 shows the event rate distribution as a function of the threshold obtained with the ^{55}Fe source for one of the pixels of the matrix.

Performance and comparison

Table 3.2 reports a summary of the measurements performed on one channel for each front-end configurations integrated in the ASIC. The configuration in which the front-end system is completely outside the pixel is the one that shows the second best performance in terms of

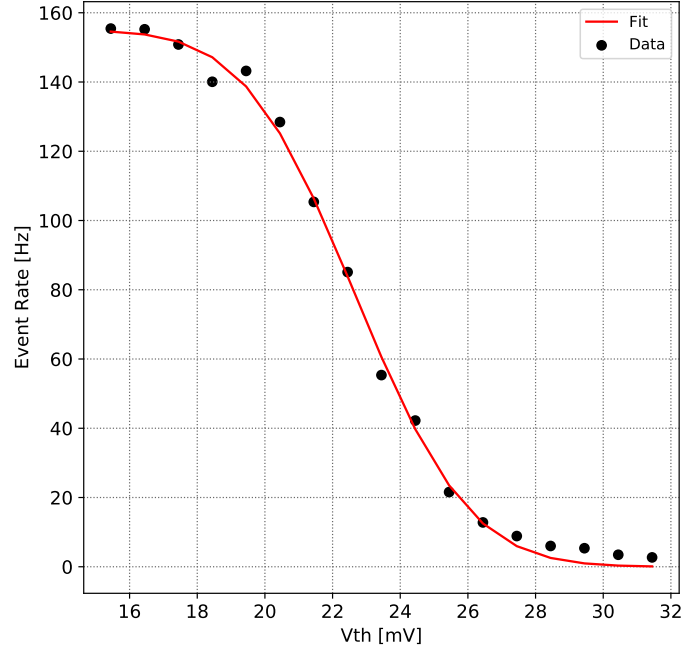


Figure 3.16: Event rate as a function of the threshold obtained with a ^{55}Fe radioactive source.

noise. Integrating the electronics outside the sensitive area is useful to reduce the noise on the pixel induced by the amplifiers. The gain of the architecture is one of the highest among all because, despite the connection with the pre-amplifier input is longer than the other versions (in which at least the first stage of the front-end is inside the pixel), the capacitance of the sensitive area is smaller since no triple-well is needed for the integration of the electronics. However, this solution is the one that requires using the most area outside the pixel and it is vulnerable to the scaling of the pixel density and, as anticipated in Section 3.2.2, it will not be adopted in the final FASER chip.

Integrating only the pre-amplifier inside the pixel leads to a significant reduction of the charge gain. This effect is caused by the need of a longer connection between the pre-amplifier and the driver and consequently of the increase of the output capacitance of the former which results into a reduction of its bandwidth. Moreover, a longer pre-amplifier output line increases the coupling with the pixel in which the circuit is integrated. The first stage of the front-end is inverting the pixel signal leading to a negative feedback with the sensor. Therefore, a more intense coupling (due to the longer lines) is further reducing the charge gain.

The configurations with a third (inverting) stage is characterized by the largest measured gain. The additional block is improving the decoupling between the first driver and the discriminator resulting into an increase of G_c . However, a second driving stage worsens the noise performance of this solution which makes it the worst in terms of ENC among all.

The front-end variants that include every stage inside the sensitive area of the pixel are

Configuration	σ_v [mV]	G_c [mV/fC]	ENC [e ⁻]
All f.e. outside pixel	4.2 ± 0.2	159 ± 1.0	165 ± 9
Only pre-amp. in pixel	2.5 ± 0.1	96.8 ± 0.5	161 ± 9
All f.e. in pixel, inv. stage	6.9 ± 0.5	179 ± 1.0	241 ± 19
Pre-amp. and driver in pixel	3.8 ± 0.2	133.7 ± 0.6	178 ± 9
All f.e. in pixel	5.4 ± 0.4	148 ± 1.0	228 ± 20

Table 3.2: Noise contribution (σ_v), charge gain (G_c), ENC of one channel for each front-end configurations integrated in the chip. The error associated to G_c does not represent the channel-to-channel dispersion but is the uncertainty on the gain measurement of the analyzed channel.

characterized by the second worst ENC performance. Also, the charge gain is not the highest among the designed configurations because, as explained before, the triple-well in which the circuits are integrated increases the sensor capacitance. This explains also the noise performance. Similar performance in terms of gain but better ENC are achieved by the configurations with only pre-amplifier and driver in pixel. In this case the exclusion of the discriminator inside the sensitive pixel area significantly improves the noise level of the front-end and reduces the ENC. The last two solutions represent a good compromise between performance and compactness, a crucial requirement for the design of the final version of the FASER pre-shower chip. For this reason, they will be taken into consideration for the next iterations.

Table 3.1 showed that all the solutions are characterized by a threshold dispersion ~ 30 mV, or smaller, in some configurations. In particular, the variant with the discriminator outside the pixel shows a peak-to-peak dispersion of $6 \cdot \sigma_{V_{th}} = 140.4$ mV and an average gain around 130 mV/fC. This front-end configuration will therefore meet the specification of the high-precision FASER pre-shower to be able to discriminate input charges of $Q_{in} \gtrsim 1$ fC even without using local DACs for pixel-to-pixel calibration. As anticipated, this solution will be investigated for future implementations of the chip.

3.2.5 Future developments and alternative solutions

Other front-end architectures will be investigated for future iterations of the ASICs for the FASER experiment. One of the analyzed design solutions is displayed in Figure 3.17 and it is indicated as peak-measuring system. This architecture has been integrated in various chips and tested for the evaluation of its performance. The main goal of this solution is the maximization of the TOT dynamic range.

The system works as follows: the pre-amplifier is connected to a switch that when no event occurs is closed. The output of the switch Out_{switch} is fed in the input of a follower and discriminator that will compare the analog signal with the threshold Th . The discriminator output is then connected to a tunable delay line designed as a set of current-starved digital inverters. When an event is sensed, the outputs of the delay line, $Out_{disc-pos}$ and $Out_{disc-neg}$, will open the switch. At this point Out_{switch} will be linearly charged (or discharged, depending

on the polarity of the signals) to the baseline. The charge (discharge) process is implemented through a PMOS (in blue in Figure 3.17) or a NMOS (in red in Figure 3.17) transistor that usually carries a current up to few hundreds of nA. This current does not have any significant impact on the waveform of Out_{switch} during the first edge of the signal. Moreover, the delay line and the above-mentioned charge/discharge current can be tuned during chip calibration in order to maximize the dynamic range of the TOT. In the chip, the peak-measuring system can also be deactivated keeping the switch always closed. In this way, the pre-amplifier would directly be connected to the input of the follower even during the trailing edge. This solution is useful to analyze the impact of the peak-measuring system on the front-end comparing its performance with the ones shown by the circuit when the switch is always connected.

Tests with laser source

The tests of the front-end with the peak-measuring system were performed by stimulating the chip with an infrared laser source. The FPGA of the GPIO was programmed to produce a periodic low frequency (~ 70 Hz) signal to be connected to the reset pin of the ASIC with the front-end electronics under test. The GPIO board also produces a signal used as laser trigger characterized by the same period of the reset signal. As shown in the representation of

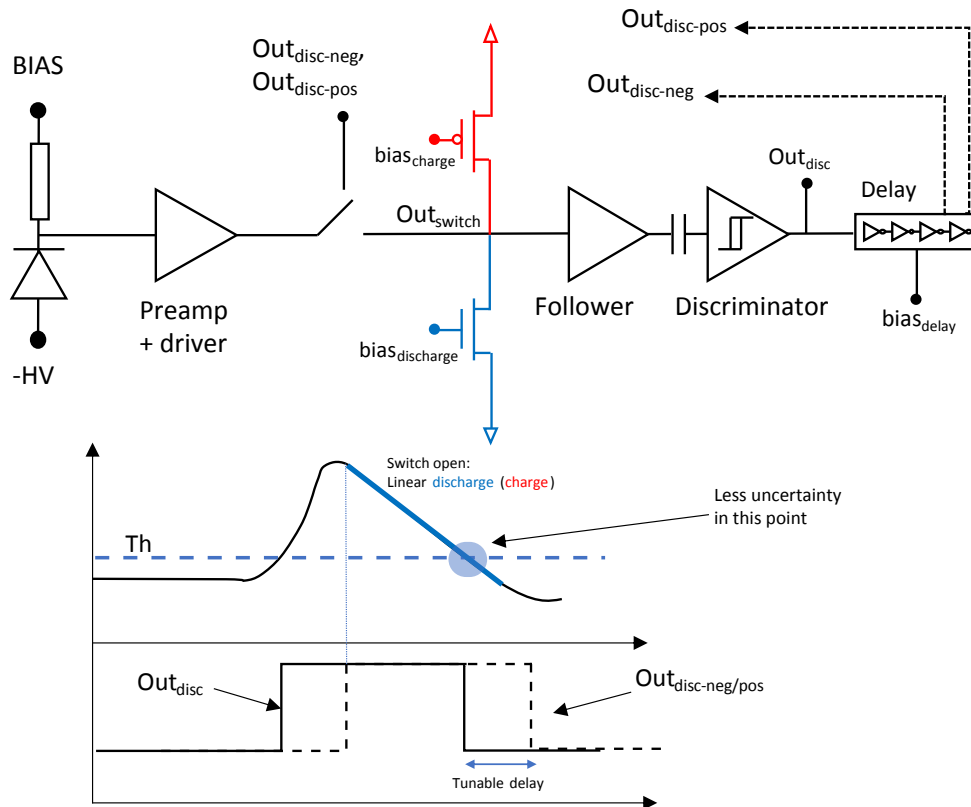


Figure 3.17: Peak measuring system schematic and working principle.

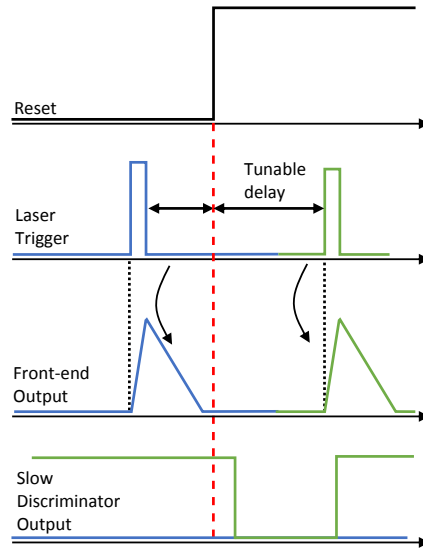


Figure 3.18: Laser trigger, reset (active low) and slow discriminator output signals for the tests of the peak-measuring system.

Figure 3.18, the pulse of the laser trigger is sent around the positive edge of the reset signal with a delay that can be tuned by the FPGA controller (with a minimum step of 20 ns, i.e. the period of the FPGA clock) and an external delay line (with a 500 ps minimum step). If the laser trigger is sent before the positive edge of the reset, which in this case is active-low, the pixels of the chip may be stimulated in a period in which the ASIC is not able to produce any output (blue in Figure 3.18). On the other hand, when the laser trigger is sent after the reset edge the output of the discriminator of the stimulated front-end will be latched and readout, producing the signal indicated as slow discriminator output (SDO). This means that, in the situation depicted in green in Figure 3.18, the SDO turns low after the positive edge of the reset and then switches to high again when the laser stimulates the chip.

The goal of the measurement setup is to analyze the rate of the transitions of the SDO signal as function of the relative temporal position of the laser trigger and the positive edge of the reset signal for the evaluation of the TOT and its jitter. Indeed, starting from the scenario depicted in blue in Figure 3.18 and reducing the delay between the laser trigger and the reset, it is possible to see a gradual increase of the transition rate of the SDO. The slope of the event rate characteristics is directly related to the jitter of the trailing edge of the output of the front-end, i.e. of its TOT. After this phase, reducing further the laser trigger delay leads to a scenario in which the SDO is always high, i.e. a 100% rate is detected: in this case, the reset edge always lays in the period in which the discriminator output is activated. Finally, moving further the trigger towards the reset edge, the SDO will show the behavior reported in green of Figure 3.18. The rate of the transition of the SDO in this phase is directly related to the jitter of the rising edge of the front-end output, i.e. the ToA.

Figure 3.19 displays the event rate distribution of the SDO for various laser intensities (reported in terms of attenuation) showing a comparison of the performance when the peak-measuring

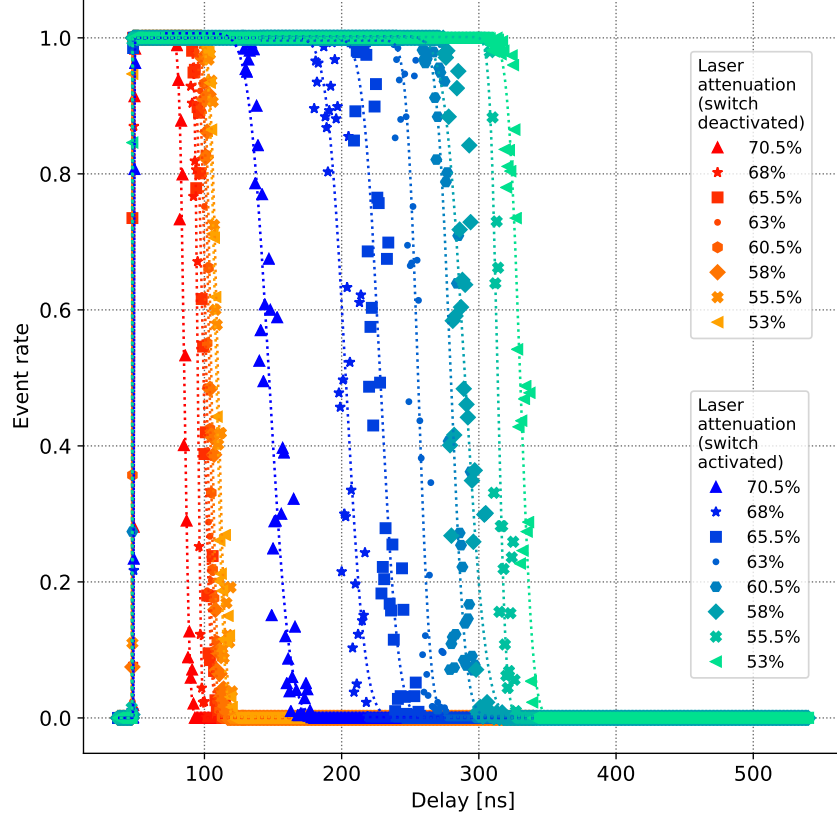


Figure 3.19: Event rate as function of trigger delay for various laser intensities. The results with the peak-measuring system activated were obtained choosing $bias_{delay} = 17\mu\text{A}$ and $bias_{discharge} = 40\text{ nA}$.

system is activated and when it is not. In the system under analysis, a reduction of the laser attenuation of 2.5 % corresponds to an increase of input charge of approximately 1 fC. This value has been calculated stimulating various analog channels in the ASIC under test and comparing the outputs with the ones obtained with a ^{109}Cd source. The TOT and the correspondent jitter have been evaluated by performing a fit of the event rate with error functions $\text{erfc}(x)$ and $\text{erf}(x)$ as done for the gain and noise analysis of the FASER prototype front-end (described in Section 3.2.4).

The measurements confirm a significant increase of the TOT dynamics when the switching system is enabled: as shown in Figure 3.20a, the TOT spans in a 100 ns to 280 ns when the peak-measuring system is activated and in a 38 ns to 63 ns when it is not in an equivalent $\sim 8\text{ fC}$ input charge range. This effect is caused by the above-mentioned linear discharge of the output of the pre-amplifier which leads to an increase of the TOT range but also of the corresponding jitter due to the smaller slope with which the input of the discriminator crosses

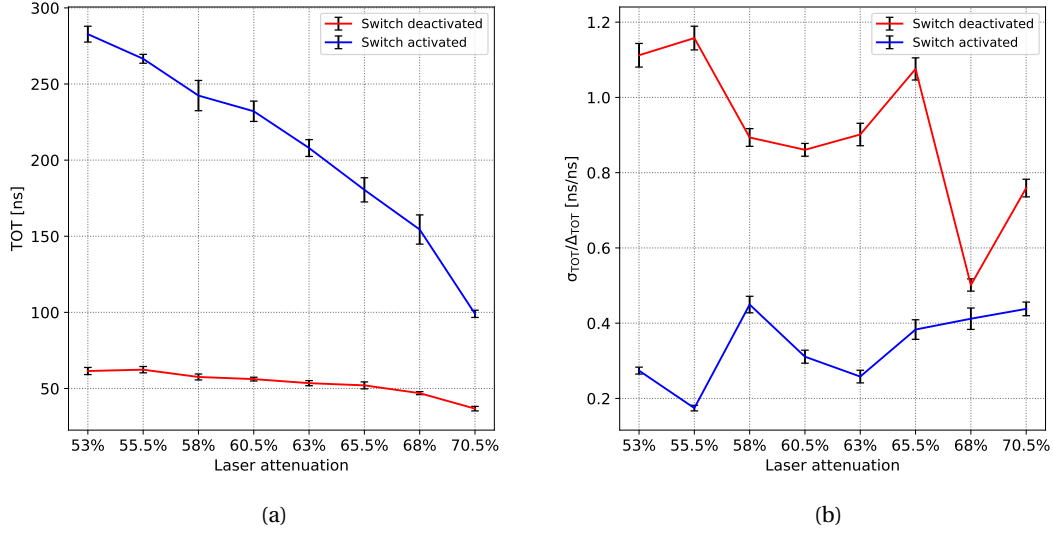


Figure 3.20: TOT (a) and jitter over the average TOT variation (b) as function of laser attenuation.

the threshold of the front-end. Measurements highlight a 3.3 ns average TOT jitter when the peak-measuring system is not enabled and 8.9 ns when it is activated. However, the large TOT dynamics in the second case leads to better performance in terms of relative jitter as reported in Figure 3.20b. This parameter was calculated as the ratio between the absolute jitter and the average TOT characteristics slope. The results shown in Figure 3.20b demonstrate that the relative jitter performance of the architecture when the peak-measuring is enabled can significantly smaller than the ones obtained without its activation.

A fit on the other edge of the event rate characteristics also allowed obtaining an estimation of the jitter of the ToA. The average value of this parameter for the chosen laser intensities is ~ 95 ps. However, potential jitters of the reset signal and the laser trigger generated by the GPIO FPGA and internal noise on the logic for the reset of the ASIC under test make the reported value an upper limit of the ToA jitter that can potentially be significantly smaller.

The value of the TOT for a certain Q_{in} is heavily influenced by the discharge current which is used to regulate the slope of the input of the discriminator during the trailing edge. The distribution of the TOT versus this current is reported in Figure 3.21. The plot shows, as expected, a hyperbolic behavior of the TOT which demonstrates that the slope of the front-end output decreases approximately in a linear fashion with the discharge current. From the figure, it is also possible to highlight that the TOT can be extended up to few microseconds if small currents ($\lesssim 10$ nA) are set.

The measurements reported in this section demonstrate the efficiency of the proposed architecture that will be further developed and implemented in future ASICs.

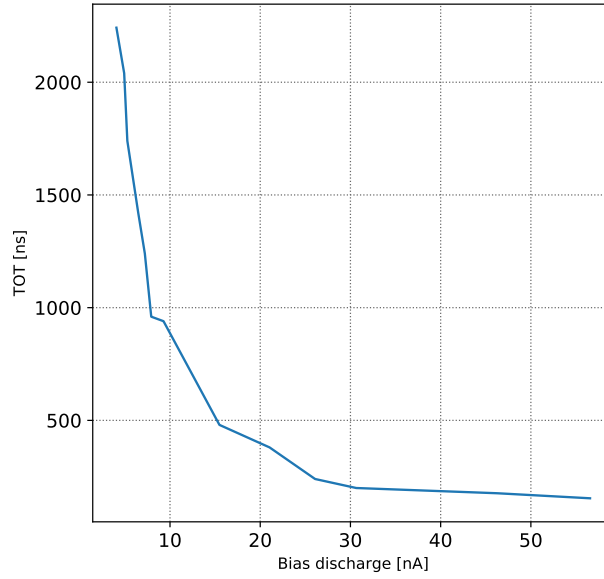


Figure 3.21: Average TOT as function of the discharge current. This characteristics was calculated setting the laser attenuation at 68%.

3.3 Analog memory and MUX

Simulation analysis on the FASER experiment (see Figure 3.2b) highlighted that, in order to better reconstruct the electromagnetic showers, charge information associated to the events sensed in each pixel need to be measured and stored to detect clusters of hits. As explained before, in one of the three variants of the submitted pre-production ASICs the charge is measured by evaluating the TOT with a set of counters. However, this solution features various problems including the large dead area needed for the integration of the required logic. A more compact architecture featuring analog memories has been chosen for

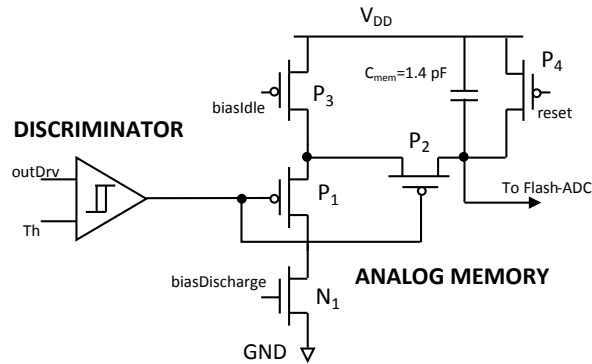


Figure 3.22: Analog memory schematic.

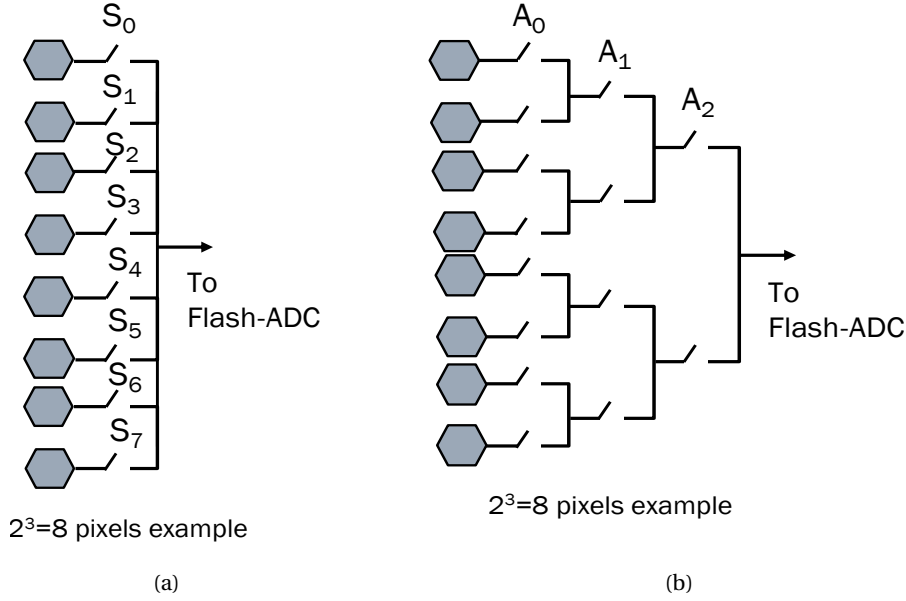


Figure 3.23: Simple (a) and proposed (b) design solutions for the implementation of the analog multiplexer. In this picture a 3-bit example is reported.

the other two versions of the FASER ASIC. As for the counter-based solution, also in this case a charge-to-time conversion is implemented: when an event occurs the discriminator will activate a circuit designed to charge, with a constant current, a Metal-Insulator-Metal (MIM) capacitor. The voltage on this capacitor will be directly proportional to the TOT and, therefore, to the intensity of the sensed event. The schematic of the analog memory control circuit is reported in Figure 3.22. The output of the discriminator is connected to the gate of two PMOS transistors, P_1 and P_2 , used as switches. When an event is sensed, these PMOSTs will activate a path between one of the nodes of the MIM 1.4 pF C_{mem} capacitor and ground, charging it with a current that depends on the one of the NMOS N_1 . The latter is used to control the charging speed of the capacitor and its current is set with a 8-bit global DAC. The PMOS P_4 acts as a reset device which, when activated, increases the voltage of the output node of the analog memory consequently discharging C_{mem} . P_4 is used as a switch and its gate is driven by the super-column logic. The transistor P_3 is required to deal with the leakage currents I_{leak} of the switches: setting its current I_{P_3} such that $I_{leak} \ll I_{P_3} \ll I_{N_1}$, P_3 will prevent the leakage to charge C_{mem} when the system is in idle state without compromising the normal behavior of the system when an event has to be sensed.

The output of the analog memory of a certain pixel is connected to the input of a 4-bit flash-ADC. The latter is shared with the remaining 255 pixels of a super-pixel. During readout, the ADC will convert the input charge data by polling all the elements in a super-pixel through a 256-to-1 analog multiplexer (MUX). A fully functional implementation of this block required to deal with some challenges that could compromise the measurement of the charge stored inside the MIM capacitors. The simplest design for an analog MUX is displayed in Figure 3.23a.

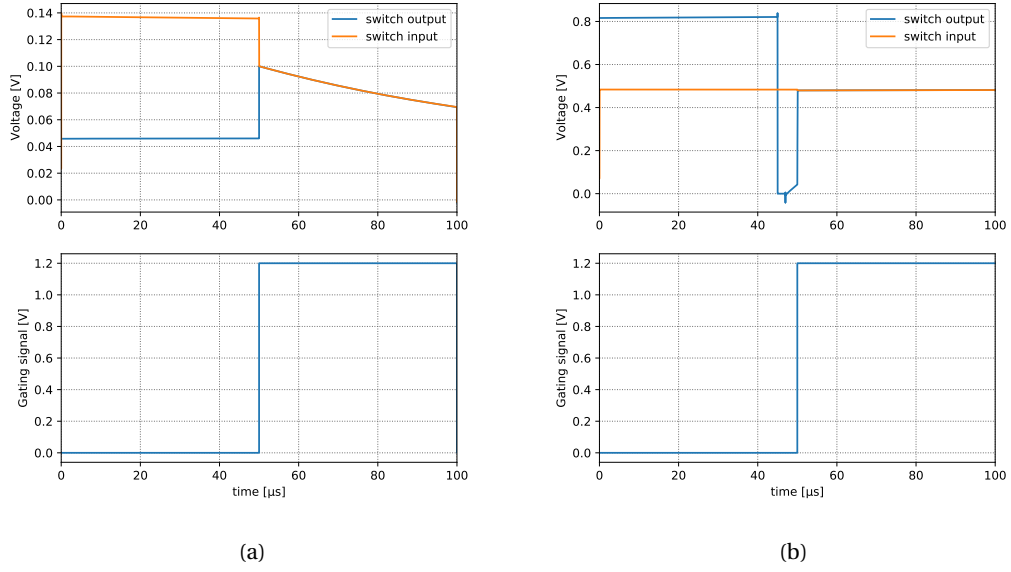


Figure 3.24: Impact of the charge sharing in the simple (a) and proposed (b) solutions for the analog multiplexer. The simulations have been performed on the 256-to-1 (8-bit) schematic integrated in the pre-production chip. In the plot (b), the drop of the switch output before the rising edge of the gating signal is due to a reset signal used to discharge the output capacitance of the MUX before reading another pixel. More details in Section 3.4.1.

In this 3-bit (8 elements) example, each pixel is associated to a single switch, all connected to the same line. Despite its simplicity, this architecture shows several problems including the need of an additional logic able to generate the correct gating signals to poll all the capacitors and a significant charge sharing effect: in the 256 pixels version, due to the large capacitance connected to each switch, half of the charge can be lost during switching. The plot reported in Figure 3.24a shows the impact of this effect on the input signal of the flash ADC with such an architecture. Therefore, the implemented solution is characterized by a design similar to the 3-bit example reported in Figure 3.23b: 8 stages of 2-to-1 analog MUX were connected to the 256 elements of each super-pixel, with a consequent reduction of the capacitance seen at each node. Gating signals are coded in 8-bit words forming the addresses associated to each pixel. Figure 3.24b highlights a significant reduction of the charge sharing effect in the proposed solution. This multi-stage solution also features an additional logic that improves the evaluation of pixel charges. As illustrated in the left part of Figure 3.25, the switching of the LSB of the gating words during polling will lead to the sharing of the charge of pixels with higher addresses (highlighted in red in the figure). In order to avoid this detrimental effect, an additional set of switches has been added as shown in Figure 3.25 on the right. These switches are driven by a combination of the bits of the address in order to disconnect the pixels to the MUX when they do not have to be read. In the 3-bit example of the figure, a functional combination of gating signals S_i (where $i = 0, \dots, 7$) could be the following

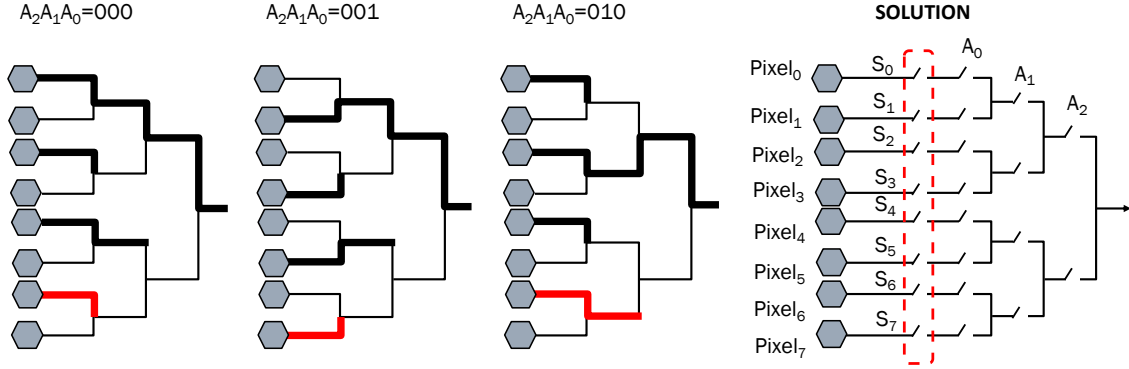


Figure 3.25: Charge sharing problem caused by address switching during polling (left) and proposed solution (right).

- S_{0-1} always active
- S_{2-3} connected to A_1
- S_{4-5} connected to A_2
- S_{6-7} connected to $(A_1 \text{ AND } A_2)$.

In a N -bit case, there will be 2^{N-1} couples of pixels. The couples can be coded in binary words C of $N - 1$ bits. Iterating the process shown before, it is possible to see that the number of AND inputs connected to the switches associated to the i -th couple of pixels is equal to the number of ones in the corresponding word C_i . Therefore, considering a N -bit analog MUX and, therefore, $(N - 1)$ -bit words for the couples of pixels, the total number of additional AND inputs N_{AND} can be calculated with the following relation

$$N_{AND} = \sum_{i=1}^{N-1} \binom{N-1}{i} i. \quad (3.9)$$

Equation 3.9 highlights that N_{AND} can be obtained calculating the number of words with a certain amount of ones i (this explains the binomials) multiplied by i . For $N = 8$, Equation 3.9 provides $N_{AND} = 448$. The inefficient solution of Figure 3.23a requires an additional logic to provide the gating signals to the switches during polling. A simple architecture would feature a 256-bit shift register (as a token ring to activate one switch at the time) or an 8-bit encoder. However, the calculation of N_{AND} shown before proves that both of these solutions would require comparable or more area than the one with 448 AND inputs implemented in the proposed analog MUX.

The architecture presented in this section is still vulnerable to leakage effects that can compromise the charge information stored in the MIM capacitors: the implementation of a readout logic able to read all the pixel in a short time is crucial to prevent this effect. However, simulations show that, because of the leakage, the analog memory loses a charge corresponding to

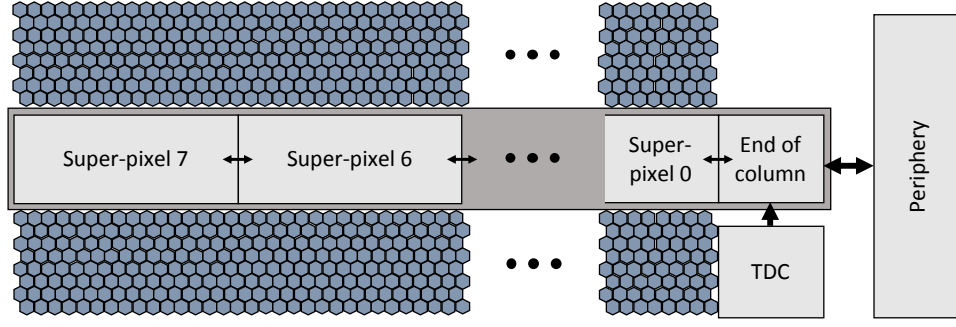


Figure 3.26: Super-column logic (dark gray), periphery and TDC block diagram (not in scale).

approximately 1 LSB of the flash-ADC ($\approx 0.5V_{DD}/2^4$) in 250 μs . The readout time depends on the number of events the pixel matrix senses, however, for the FASER experiment the event rate is supposed to be extremely low (up to few Hz): in this case, the readout time may be in the order of few tens of μs , an acceptable time frame which is not able to compromise the charge evaluation.

3.4 Digital logic

3.4.1 Super-column logic

Since the significant events that the FASER detector aims to sense can be particularly rare (potentially only a few in one year), the specifications of the pre-shower ASIC impose a high detection efficiency to maximize the possibility to sense these events. Therefore, the dead area of the chip i.e., the non-sensitive zones of the ASIC, needed to be minimized. In the 3 and 4 super-column versions with analog memories of the pre-production chips, the non-sensitive regions occupy <6% of the total super-column area. This result could be achieved only by adopting the analog memory architecture and designing a 37.7 μm thick logic in the middle of the super-column. The super-column logic is a digital system that aims to handle the communication between the pixels and the rest of the periphery. Figure 3.26 shows a block diagram of this system and its connections with the rest of the chip. The super-column logic is composed of 8 identical (in terms of functionality) blocks that for the sake of simplicity will be indicated as super-pixels. These blocks directly communicate with the correspondent pixels and include the logic for the enabling of the input paths of the analog-MUX explained in Section 3.3. They also feature a set of shift-registers to set the masking bits of the pixels. The shift-registers of the super-pixels are connected in series and fed by the end-of-column logic. The masking bits are also exploited to drive the selection signals of the 2-to-1 multiplexers used for the generation of the test-pulses: in all the pre-production chip variants, there is the possibility to stimulate the pixels with an internally generated current through the circuit depicted in Figure 3.27. Activating the PMOS with $EN_{testpulse}$ and switching the selection signal of the MUX $Sel_{testpulse}$, the system is able to feed the pixel with a current proportional to the value provided by a DAC and C_1 , a 55 fF MIM capacitor. The DAC has a 8-bit dynamic

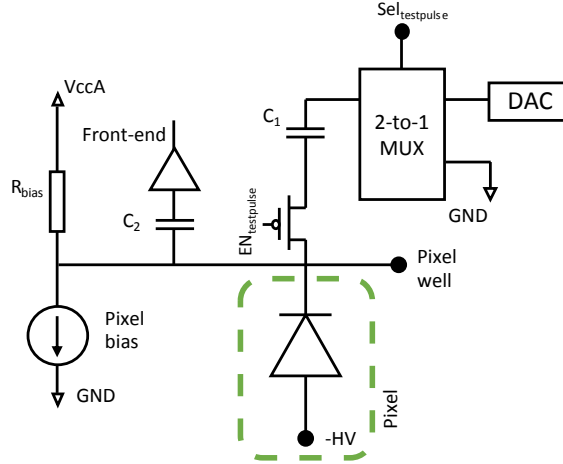


Figure 3.27: Pixel well and test-pulse generator schematic.

and its output voltage is set through a 8 k Ω resistor. The $Sel_{testpulse}$ signals are provided to the pixels as outputs of the super-pixel modules.

As anticipated, the masking bits are provided to the super-pixels by the end-of-column block. This logic also handles the readout process and allows the communication between the digital periphery and the corresponding super-column. During readout, the end-of-column performs the polling process to sequentially read the charge stored in the pixels that sensed a hit through the flash-ADC. A finite state machine (FSM) is implemented to send the proper pixel addresses to the super-pixels which will drive the corresponding analog-MUX to redirect the charges in the analog memories to the ADC. A simplified time diagram of the polling process is reported in Figure 3.28. When switching between codes, glitches and non-uniform delays may activate unwanted addresses and connect the output of the analog memory of different pixels, compromising the charge evaluation process. For this reason, the FSM of the end-of-column block provides a MUX-enable signal to avoid this kind of situation. The reset ADC is set to charge the input of the ADC to the positive supply before changing the pixel address: the analog-MUX has a certain capacitance associated to its active path that needs to be discharged before reading the next pixel. In this way it is possible to avoid that the evaluation of a certain charge is influenced by the result associated to the previous pixel. The end-of-column logic also aims to acquire the data provided by the TDC (a detailed description of this system is reported in Section 4.3) and send them to the digital periphery for the readout.

Notes on physical implementation

As mentioned above, the super-column logic is 37.7 μm thick and spans almost the entire chip with a length of approximately 1.4 cm. The aspect-ratio of this system made its physical implementation particularly challenging. Indeed, the density of the design (slightly over 40%) and the necessity to connect the end-of-column logic with the rest of the super-pixels significantly increased the effort for the synthesis of the clock tree and the routing phase: a

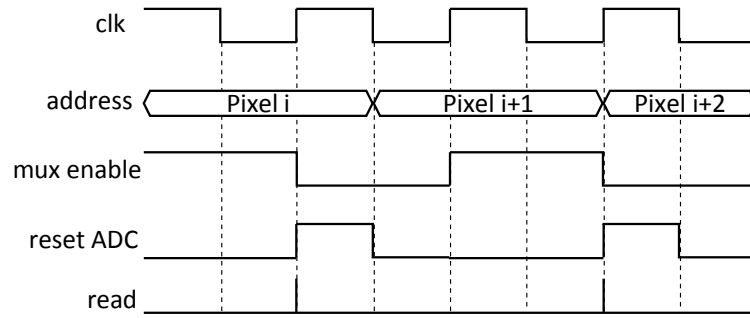


Figure 3.28: Simplified time diagram of the polling of the pixels measured charge.

first routing attempt lead to more than 1.5 millions design rule check (DRC) errors due to the limited thickness of the architecture (in $37.7\ \mu\text{m}$ approximately up to 91 parallel metal lines are routable). For this reason, the end-of-column and the super-pixels where designed to directly communicate only with adjacent blocks as illustrated in Figure 3.26. Moreover, since the logic is particularly repetitive for each super-pixel, the digital implementation tool was able to close the timing constraints and complete the routing only by forcing the position of super-pixel modules in their corresponding regions inside the matrix. This solution was implemented during the placing phase exploiting the following TCL code⁵ for Innovus Implementation System

```

1 for {set i 0} {$i < 8} {incr i} {
2     createRegion column/\superpixel[[expr {7-$i}]] [expr {$i*1787.2}] 0 [
3         expr {1787.2+$i*1787.2}] 37.7
4 }

```

The command in the for-cycle is used to define the regions that must contain all the elements associated to the i -th super-pixel. However, other blocks can be instantiated inside these regions in order to allow the implementation tool to perform potential optimization (e.g. clock tree and buffering). The coordinates in the code are referring to the size of the super-pixels and they are reported in microns. The solution presented above, highlights that the implementation of digital systems with strong constraints in terms of area (a typical challenge for the design of large timing detector systems) may be solved reducing the freedom of the tools (in order to set a sort of guideline) by organizing the geometry of the design according to its characteristics (in this case, the periodicity of the structure).

The super-column logic and the periphery can work in two different operating modes: programming, in which the chip ignores external stimuli and the digital logic defines the settings of the chip (e.g. biases), and readout, in which the ASIC is enabled to acquire external data and to transfer them outside. For this reason, two sets of constraints were defined to implement a multi-mode multi-corner (MMMC) synthesis. The main difference between the two operating modes is related to the clock signals: the programming process is synchronized by a 5 MHz clock while the readout is performed with a 200 MHz signal. The MMMC synthesis is necessary

⁵More information on tool command language (TCL) in [98].

to reduce the constraints and preventing the implementation tool to over-optimize portions of the logic that are not designed to work at high frequencies.

3.4.2 Periphery logic

The digital periphery of the FASER pre-shower ASIC aims to handle the communications among super-columns, TDCs and internal biases and the chip with external systems. It includes

- a readout control block. It features various FSMs to collect the ADC and TDC data. The communication exploits an SPI bus. This block is agnostic to the number of super-columns making the design of the periphery easily reusable for a larger version of the FASER ASIC with minimal changes.
- a controller to regulate, during programming, the biases of the chip. The latter has no analog inputs since all the biases are provided through internal bandgap references and DACs.
- column and row encoders to identify the pixels that sensed an event.

The digital periphery was designed to work with a 200 MHz clock. For this reason, during readout, the ASIC is able to provide data at 200 Mbs. The readout is characterized by a frame-based architecture which sends data associated to the pixels of a certain super-column only when at least one hit occurred on the latter. The output data associated to each super-column have the following structure:

- 8-bit header. 2 bits are used as super-column ID.
- 1-bit hit flag. It indicates if there is at least one hit on the super-column.
- in case of one or more hits on the super-column, a 5-bit word is readout for each pixel that sensed an event. In this word, the first bit is used as a flag (to indicate that the corresponding pixel was activated) while the remaining 4 bits represent the output of the flash-ADC, i.e. the pixel input charge. On the other hand, if the pixel was not activated, only the flag bit is sent (to indicate that no event occurred).
- 20-bit word for TDC calibration. 14 bits are used to store the state of the TDC at the edge of a reference signal period (fine component of the measurement) while the remaining 6 bits represent the output of the counter for the course component of the measurement of the reference signal. More detail on the calibration of the TDC will be provided in Chapter 4.
- 1-bit flag for each TDC channel. If the flag indicates that the channel was stimulated, a additional (7+11)-bit words are sent (indicating the fine and the coarse component

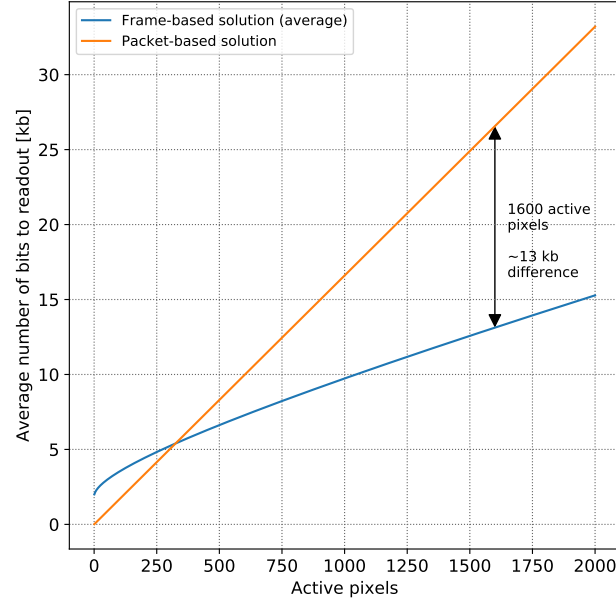


Figure 3.29: Amount of data to readout in the frame-based solution adopted in FASER pre-production ASIC and in a packet-based solution as function of the amount of pixels stimulated during the events (a square distribution is assumed).

of the ToA measurement). If no-event on a TDC channel is sensed, only the flag is sent. More details on the TDC are reported in Chapter 4.

As explained before, if no hit occurs on a super-column, only the associated 8-bit header and the 1-bit flag are readout. This solution was chosen because of the expected pixel occupancy for each significant event in the FASER experiment. Figure 3.2 highlights that approximately 1600 pixels in a concentrated region are activated for each typical event in the FASER detecting system. A rough estimation of the amount of data produced by the chip in this case can be computed assuming that the typical event is able to activate a 40×40 sub-matrix of the ASIC. In this situation, considering the size of the super-pixels, only three or four super-columns can be stimulated producing ~ 14 kb and ~ 12 kb respectively. In Figure 3.29, the average amount of data to readout with the frame-based architecture of the FASER pre-production chip as function of the number of activated pixel is reported. The distribution was obtained calculating the data produced when three or four super-columns are stimulated. These values are then combined on the basis of the probability of these scenarios⁶. This distribution is compared with the one obtained in an equivalent packet-based readout architecture: in this

⁶The average amount of data $N = N(p)$ is function of the number of active pixels p . For each value of p , a sub-matrix of $\sqrt{p} \times \sqrt{p}$ elements is assumed to be hit. The value of N can be calculated as $N(p) = \sum_i N_i(p) P_i(p)$, where $N_i(p)$ are the data produced when i super-columns are stimulated and $P_i(p)$ is the probability to produce hits on i super-columns.

alternative solution, only the data associated to the pixel that sensed a hit are sent during readout. In order to identify these pixels (2048 per super-column), an (11-bit) address has to be sent together with the output of the ADCs. As shown in Figure 3.29, the packet-based solution is advantageous if $\lesssim 400$ pixels are stimulated but for higher occupancies it produces a significantly larger amount of readout data: with 1600 active pixels, ~ 26 kb are generated. Moreover, the design of a packet-based architecture would require a bigger super-column logic (because it would have to generate the address of the pixels) that would make this solution unfeasible for the strict dead-area specifications mentioned before.

4 Time-to-Digital Converters

This Chapter focuses on the description and the analysis of various RO-based TDCs, designed and implemented during the doctoral activities. As explained in Chapter 2, high-performance timing detectors require TDCs capable of measuring time intervals with a precision compatible with the time resolution set by the specifications. For this reason, the design process of this kind of converters plays a crucial role for the definition of the performance of the final system. The TDCs presented in this Chapter have been designed in the context of various experiments and projects and submitted for production. Measurement results and performance of the converters will be reported and described.

In Section 4.1, a brief overview on how TDCs can be efficiently integrated and calibrated in timing detectors is reported. Section 4.2 provides a detailed description of the converter designed for the TT-PET project chip. A non-linearity model for the characterization of the TDCs is also described. This model is useful to compare the proposed solution with more conventional architectures. The converter designed and integrated inside a prototype chip for the FASER experiment at CERN is described in Section 4.3. Possible improvements and new features are briefly discussed in Section 4.4. The architecture and the analysis described in Section 4.2 and the discussions of Section 4.1 are part of the following paper:

- [71] F. Martinelli et al. “A massively scalable Time-to-Digital Converter with a PLL-free calibration system in a commercial 130 nm process”. In: *Journal of Instrumentation* 16.11 (Nov. 2021), P11023. DOI: 10.1088/1748-0221/16/11/p11023.

4.1 Integration in timing detectors

The converters described in this Chapter were designed for the time measuring systems of high-performance monolithic pixel detectors. For this reason, as it will be clarified in the next sections, these circuits needed to be characterized not only by good performance in terms of time resolution but also by a compact and simple structure that could make them suitable for the integration of many measuring channels in pixel chips such as the ones for the TT-PET project and FASER detectors.

Having a precise time measuring system is critical for systems like ToF-PET scanner, as introduced in Chapter 2 and more in detail in [14, 99], to reduce the positional uncertainty of the annihilation points produced in the scanner.

The number of TDCs and the way they are connected to the pixels in a timing detector can have a significant impact on the characteristics of the system in terms of detection efficiency. In an ideal architecture each pixel is connected to its own TDC channel: this situation would allow the system to be able to store the timing information of the received signals in the case in which all pixels are hit at the same time. However, especially for monolithic pixel detectors, this solution is difficult to implement for various reasons including area, complexity of the routing and power consumption. For this reason, multiple pixels can be multiplexed to the same TDC channel, as shown in Figure 4.1. The whole matrix of a detector can be split in several sub-matrices that, in the case of the illustration, are composed by 2 x 2 pixels. Each of them is connected to a certain TDC channel (or to a different converter) together

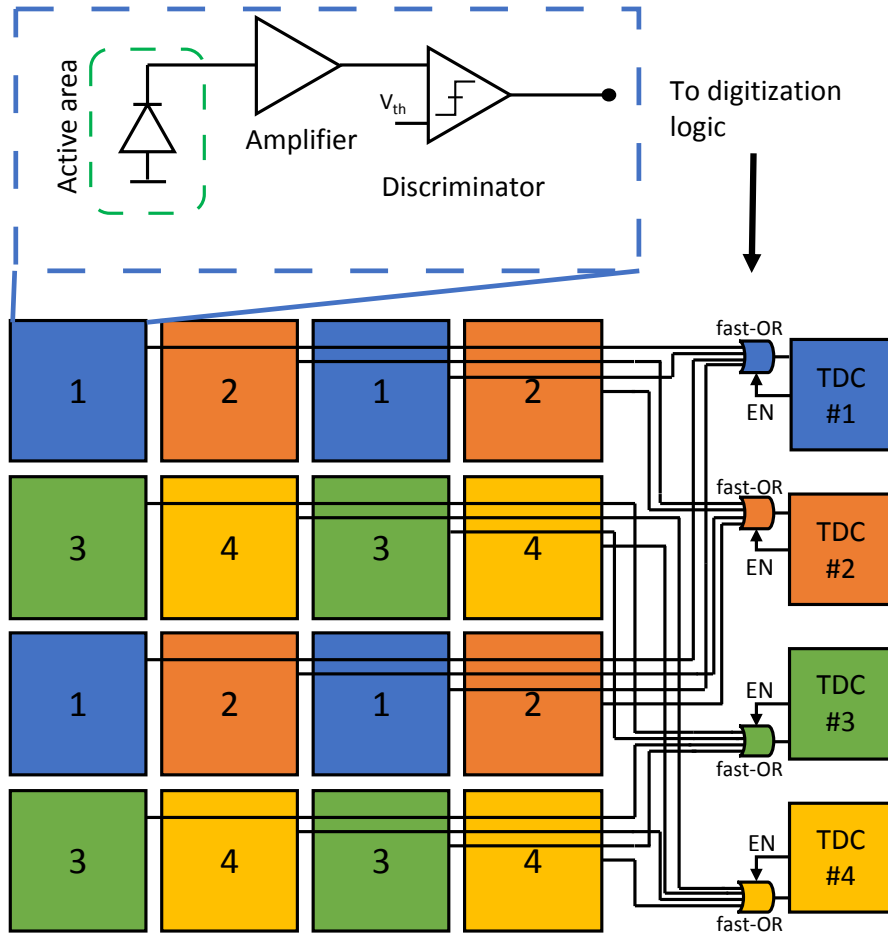


Figure 4.1: Possible disposition of a 4 x 4 pixel matrix connected to different TDCs through fast-OR blocks. The squares represent generic structures composed of pixels (active area) and front-end system (preamplifier and discriminator).

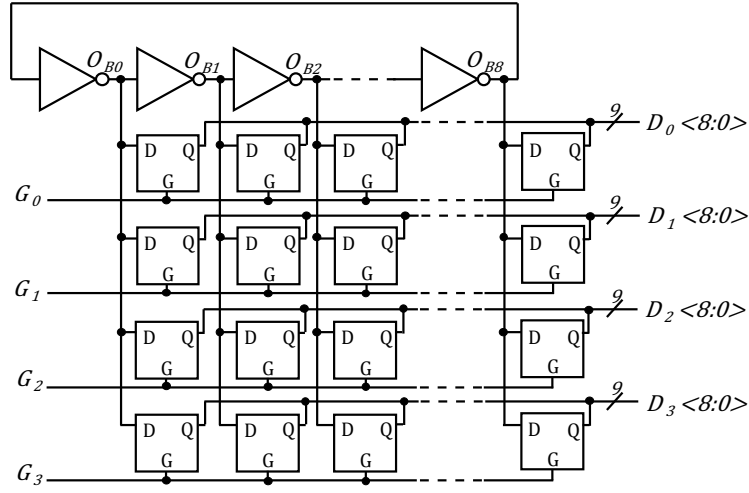


Figure 4.2: Block diagram of the system for the event-by-event calibration.

with the the corresponding pixels of the other sub-matrices inside the system. Even if this solution is still far from the above-mentioned ideal architecture, it allows the correct detection of simultaneous hits on different TDC channels (in Figure 4.1, they are referred to different colors and numbers between 1 to 4). The fast-OR blocks are used for the multiplexing of the signals coming from the pixels to the respective TDC channel. Having sub-matrices of pixels connected to separated converters allows avoiding problems related to high cluster sizes because, in many detectors, the particles that need to be sensed can generate hits in groups of adjacent pixels (depending on their size) [100]. The number of TDCs is chosen on the basis of the cluster size and the event rate, taking into account, as mentioned before, the power consumption and the area of the converter. In a system organized as in Figure 4.1, if multiple events occurs on the same channel in a shorter time frame than the dead time of the TDC, the converter, after the first one, will disable the fast-OR block in order to prevent other hits from interfering with the measurement. A possible improvement of this architecture has been implemented for the readout chip described in Chapter 5: in a multiple hit scenario, the readout system is able to store the codes associated to all the pixels that sensed an event signal as well as the timing information of the first one. For all the above explained reasons and since the TDCs described in Section 4.2 and 4.3 were designed to be integrated in monolithic pixel detectors, the design process of these converters aimed to implement compact, simple, low-power architectures and with good performance in terms of time resolution. Moreover, both of the proposed converter are featuring a PLL-less synchronization system described in Subsection 4.1.1.

4.1.1 PLL-less synchronization system and calibration

The circuit depicted in Figure 4.2 represents a self-calibration system based on the one presented in [101]. As stated before, the TDCs described in Section 4.2 and 4.3 are exploiting its

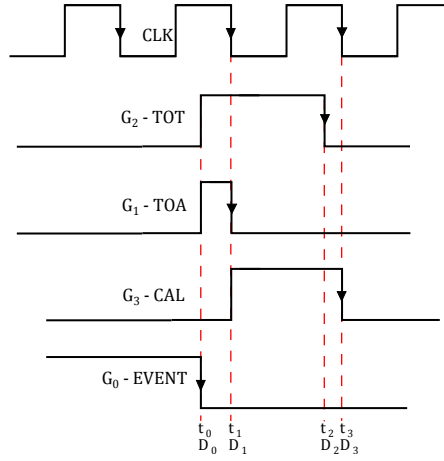


Figure 4.3: Reference clock signal CLK (up) and gating signals G_j (down).

properties due to the simplicity of the architecture.

Considering a RO-based TDC (like the ones that will be later presented), each node of the oscillator O_{Bi} with $i = 0, 1, \dots, N-1$ ($N = 9$ in the example of Figure 4.2) is connected to 4 stages of D-latch. Their outputs $D_j < N-1 : 0 >$ with $j = 0, \dots, 3$ follows the signals produced by the RO when the latches are in transparent mode (in this case when gating signals $G_j = 1$). The falling edge of G_j switches the latches to hold mode and sample the oscillator signals into the D_j outputs. Three counters are connected to three of the four latch stages. The gating signal G_0 is connected to the EVENT line. A falling edge occurs every time there is an event. The latches associated to the EVENT line are not connected to any counter. A piece of logic then generates the remaining gating signals $G_{1,2,3}$ that are used for the measurement of the various fundamental time intervals for the detection system. For instance, in the case of the TDC designed for the TT-PET project chip, $G_{1,2,3}$ are associated to ToA, TOT and the period of a reference clock (indicated with CAL in Figure 4.3) respectively. However, it is worth to highlight that a different number of stages can be adopted depending on the applications and the specifications of the system in which the TDC is integrated. Indeed, the converters designed for the detector of the FASER experiment (see Section 4.3) features only 3 or 2 latch stages per channel.

The counters calculate the number of oscillator cycles N_C in these time intervals distributed as in Figure 4.3, producing coarse measurements of these periods $T_{coarse} = N_C T_{RO}$. The difference between the states of the TDC at the beginning and at the end of ToA, TOT and CAL intervals defines the fine contributions of the measurements $T_{fine} = (D_i - D_j) t_d$ where D_i and D_j are the binary values associated to the outputs¹ of two of the latch stages and t_d is the resolution of TDC (as stated before, it corresponds to the delay of the cells of the RO). From Figure 4.3, considering both of the fine and coarse contributions and resolving the RO period as $T_{RO} = 2N t_d$ (where, as explained in Chapter 2, t_d is the elementary delay of the TDC cell), it

¹The outputs of a RO-based TDC are non-binary. For this reason, it is possible to associate a binary number indicating the state of the RO to each of the possible outcomes of the converter.

is possible to express the ToA, TOT and CAL intervals as

$$T_{ToA} = t_d[N_{C1}2N + (D_1 - D_0)] \quad (4.1)$$

$$T_{TOT} = t_d[N_{C2}2N + (D_2 - D_0)] \quad (4.2)$$

$$T_{CAL} = t_d[N_{C3}2N + (D_3 - D_1)] \quad (4.3)$$

The measurement of T_{CAL} is necessary to estimate the oscillation frequency of the RO. Performing this measurement each time an event occurs is useful to compensate for potential parasitics, device mismatches, voltage drops of the supply, temperature gradients and in general all those factors that may cause a variation of the t_d and a consequent worsening of the accuracy of the converter. Indeed, the value of T_{CAL} is nominally equal to an external clock reference. For this reason, Equation 4.3 can be exploited to calculate the average value of t_d , i.e. the LSB, as function of the clock period every time an event occurs. In this way, the TDC is able to give a output coherent with the time interval to be measured. This approach allows avoiding the use of any PLL-based synchronization system reducing the complexity of the whole architecture, power consumption and noise. Moreover, in a chip with many ROs and only one PLL, all the frequencies would be synchronized to the slowest one. The approach shown above, instead, allows avoiding this situation, since all the ROs will oscillate at their own natural frequency.

The integration of a TDC inside a timing detector system requires a calibration process. For example, if a system like the one described in this Subsection is implemented, the difference among the delays of the ring oscillator and the counters used for the coarse component of the measurement can worsen the accuracy of the converter. In order to compensate this effect, a possible calibration approach consists of sending a periodic known event (synchronous with the reference clock) to the TDC. At this point, a set of offset parameters will be applied to the outputs of the system (given by Equation 4.1, 4.2 and 4.3) in order to minimize the standard deviation of the measured values.

The jitter of the CLK signal of Figure 4.3 directly affects the precision of the measurement. If an LSB in the order of tens of picoseconds is needed, a jitter of a few picoseconds is required. The distribution of a clock with a picosecond level jitter in a large ASIC is a challenging task in terms of area and power consumption. Fortunately, a reference signal can be sent only when a calibration is needed: the clock can be gated for most of the time, sending it only when an event is detected. Another solution is sending the clock at a fixed rate, depending on the expected drift in frequency of the clock source.

4.1.2 Bubble correction

Considering a N -bit RO-based TDC (as the ones of Section 4.2 and 4.3), the total number of legal states in which the oscillator can be sampled is only $2N$ making the dynamics of the converter significantly smaller than the ideal one (2^N). However, because of mismatches and metastability of the latches, it is possible that the sampled word is not included among the $2N$ correct states and it is characterized by a group of more than two consecutive equal bits called

bubble [72].

A simple bubble correction algorithm was implemented for the RO-based TDCs described in this Chapter. Simulation analysis highlighted that the most probable bubbles are the ones in which the output words has four consecutive zeros or ones. Figure 4.4 shows the way the chosen algorithm work. If four consecutive bits are 0 (word on top), assuming that the others are correct, there are only 5 possible states in which the RO can be (bottom). The numbers on the right represent the associated code (arbitrary) and they are ordered in the way the TDC goes through these states (e.g. 2 follows 1). The implemented correction is based on inverting the two middle bits of the incorrect portion of the word (in the full rectangle). The code provided with this procedure will be the one in the middle of the 5 potential correct states. This choice reduces the maximum potential error and it is also the most probable value (proved after a simulation analysis). This algorithm does not represent a general procedure for the correction of potential bubbles but it is specific for the presented implementation. Section 4.2 shows that this simple algorithm allows to reduce illegal state rate to 0.03%. Other implementations can be found in [72].

4.2 Multi-path TDC for TT-PET project

A first TDC was designed to be integrated in the prototype chip for the TT-PET project detector (introduced in Chapter 2). The design process of this architecture was not only focused on the compactness and simplicity of the converter but also on its resolution. The circuit was validated through extensive simulations and analytical modeling of the non-linearity of the system. The TDC was designed in a 130 nm BiCMOS technology, used for the design of many of the circuits and detector ASICs described in this thesis due to the beneficial properties of HBTs introduced in Chapter 2. However, no bipolar transistor was used in this TDC, so the analysis and the architecture can be implemented in a pure CMOS node.

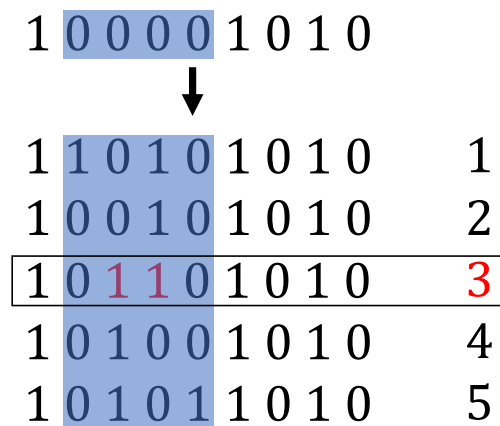


Figure 4.4: Bubble correction algorithm implementation.

4.2.1 Architecture

The presented converter is composed of a free-running RO with 9 pseudo-differential pseudo-NMOS delay cells, depicted in Figure 4.5a. Each of the output pairs of these cells is connected to the inputs of the following cell and to a pseudo-NMOS Differential Cascode Voltage-Switch-Logic (DCVSL) buffer [102], shown in Figure 4.5b. The pseudo-NMOS architecture was chosen to increase the oscillator frequency, as the load connected to each cell does not include the gate capacitances of PMOS transistors.

As introduced in Chapter 2, the oscillation frequency is

$$f_{RO} = \frac{1}{18t_d} \quad (4.4)$$

where t_d is the delay of the single stage that represents the limit in time-resolution, i.e. the LSB, of a TDC with a conventional RO. However, the feedforward design technique (also indicated as multi-path) has been applied to increase the speed of the system and to reduce the delay of the cells of Equation 4.4. Indeed, each delay cell of Figure 4.5a features two differential inputs: one of them is connected to the output of the previous cell while the other to the outputs of the buffer related to the cell placed four stage before in the RO. In this way, each buffer will be used to anticipate the charge or the discharge of the input of a further cell (as shown in Figure 4.6), resulting in a consequent increase of the oscillation frequency. Moreover, the inputs of one of the delay cells must be inverted as displayed in Figure 4.7 in order to have an odd number of inverting stages and make the ring oscillate. Indeed the outputs of each stage are not inverting. For this reason, the cross connection in blue of Figure 4.7 is necessary to satisfy the Barkhausen criterion (as also explained in Section 2.2.2). The choice of having only one inversion in the ring was made to simplify the layout and improve its symmetry. It must be highlighted that the choice of a differential solution, despite the increase of power consumption (and area), is also useful to improve the linearity of the system: simulations show that the Differential Non-Linearity (an important parameter used for the evaluation of the non-linearities of a converter that will be introduced in the next Subsection) of a single-ended solution is $\sim 14\%$ higher than the one of an equivalent differential structure. Moreover, a differential architecture is useful to reject common-mode noise such as noise induced through

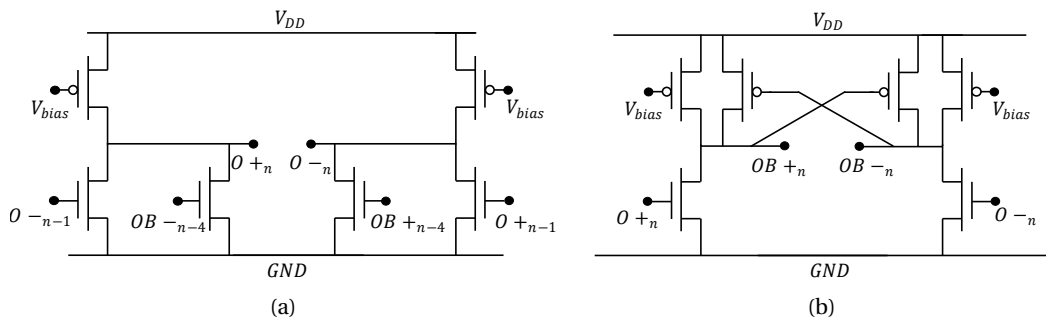


Figure 4.5: Delay cell (a) and buffer (b) of the proposed RO.

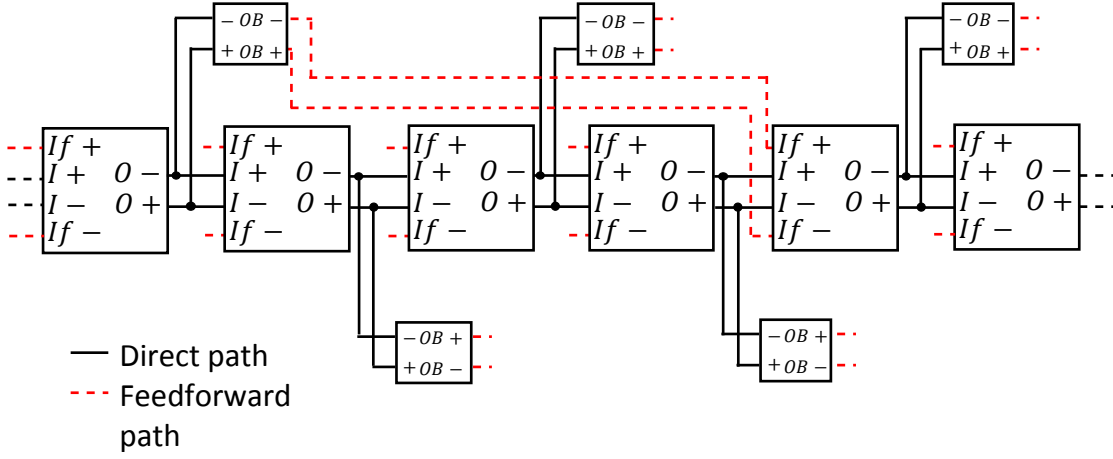


Figure 4.6: Architecture of the proposed RO. I and If represent the input of the cells connected to the direct and feedforward paths respectively. OB indicates the buffer outputs.

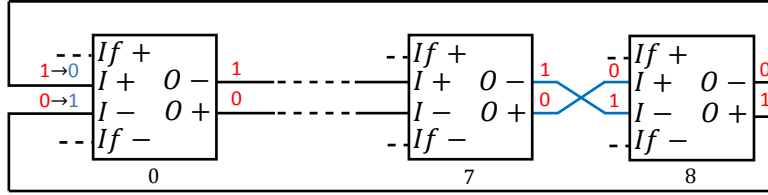


Figure 4.7: Connections of the delay cells of the RO. Inverting the inputs of one of the stages allows the oscillation of the system.

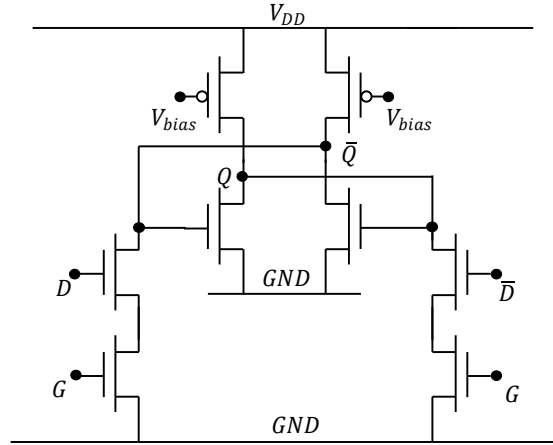


Figure 4.8: Schematic of the latches used to sample the state of the RO.

the power supplies. As introduced before, the RO was designed to be part of a TDC featuring a PLL-less calibration system explained in Section 4.1.1. From the block diagram depicted in Figure 4.2, it is possible to see that the D-latches used for the sampling of the RO need to be able to follow its oscillation frequency f_{RO} . In order to increase the speed of these blocks,

full-custom differential pseudo-NMOS D-latches were designed. Their schematics is shown in Figure 4.8.

The role of the buffers is to decouple the output nodes of the RO and the loads of the circuit, i.e. the latch stages used to sample the state of the oscillator. However, in the proposed solution, these blocks also provide inputs to later delay cells in order to increase the linearity and to reduce the effect of mismatches among the buffers by exploiting the feedback loops of the oscillator. A non-linearity model was developed to explain this improvement and it is described in the following pages.

4.2.2 Non-linearity model

The non-linearity model developed for the TDC of the TT-PET project is meant to analyze the impact of the mismatch in the propagation delay of a single output buffer on the linearity of the converter in both of the connection configurations depicted in Figure 4.9. For this purpose, the simple 5 stage multi-path RO illustrated in the figure is considered (the following results are general and can be applied also for structures with a larger number of stages). The blue dashed line represents the conventional multi-path architecture in which the feedforward is provided directly by the outputs of the delay cells while with the red dotted connections, the buffers are included in the feedforwards paths. The parameters t_{di} with $i = 0, 1, \dots, 4$ are the delay of the inverters of the oscillator and the (non-inverting) buffers show a nominal delay given by Δ . In order to analyze the linearity of the system, it is possible to exploit the Differential Non-Linearity (DNL) defined as

$$DNL(i) = \frac{t_{di} - t_d}{t_d} \quad (4.5)$$

where i is the code of the converter and t_d is the ideal delay which, as stated before, corresponds to the ideal LSB. Considering the first case (dashed line connection) with ideal delays $t_{di} = t_d \forall i$ and assuming that, because of mismatches, the delay of the first buffer is $\Delta_0 \neq \Delta$, the DNL will be

$$DNL(i) = \begin{cases} \frac{t_d + (\Delta_0 - \Delta) - t_d}{t_d} = \frac{\Delta_0 - \Delta}{t_d} & i = 0 \\ 0 & i \neq 0 \end{cases} \quad (4.6a)$$

$$(4.6b)$$

since the Δ_0 will only affect the value of DNL related to the first cell. More in detail, the mismatch $\Delta_0 \neq \Delta$ may possibly generate a bubble in the output code (see Section 4.1.2). In the proposed example, it is possible to evaluate the DNL associated to the RO using Equation 4.6 only by assuming that an efficient bubble correction algorithm has been implemented. The same assumption will be used for the rest of the Section.

The characterization of the behavior of the RO requires the introduction of a parameter that links the effect of the feedforward connections with the speed of the system. The value of t_d is function of the difference between the arrival times of the inputs of each cell δ . Expanding

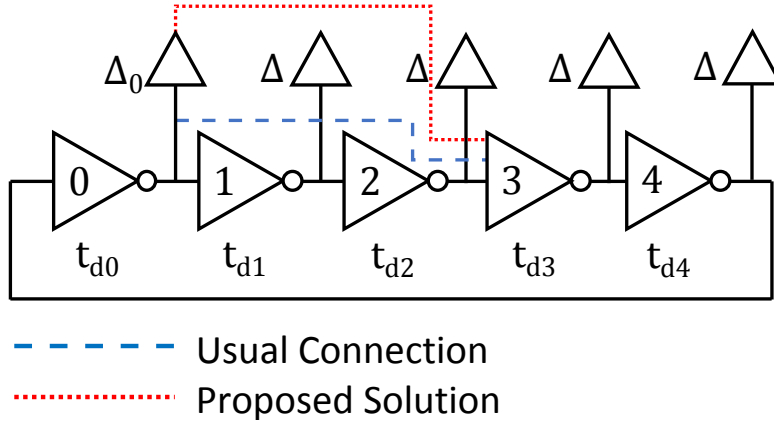


Figure 4.9: An example of a 5 stage multi-path RO with two types of feedforward connections (red dotted line: proposed solution). A mismatch on the delay of the first buffer Δ_0 is assumed for this analysis.

$t_d = t_d(\delta)$ in Taylor series and neglecting all the components after the linear one², we obtain

$$t_d(\delta) \approx t_d(0) + \frac{dt_d}{d\delta}(0)\delta. \quad (4.7)$$

From Figure 4.9, it is possible to see that in the dashed line case $\delta = -2t_d$. Replacing this relation in Equation 4.7 leads to

$$t_d = t_{dmax} - 2\eta t_d \longrightarrow t_d = \frac{t_{dmax}}{1 + 2\eta} \quad (4.8)$$

where $t_{dmax} = t_d(0)$ is the maximum value of t_d (in the case of no multi-path architecture implemented) and $\eta = dt_d(0)/d\delta$ is the feedforward parameter anticipated before. Simulations of the cell in Figure 4.5a justify the approximations of Equation 4.7 and 4.8 with values of $\eta \approx 0.25$. The star-marked curves of Figure 4.10 show the behavior of the maximum and the Root Mean Square (RMS) value of the DNL as function of η with $t_{dmax} = \Delta = 50$ ps, $\Delta_0 = 70$ ps. For what concerns the proposed solution (dotted line in Figure 4.9) a proper evaluation of the non-linearities in the case $\Delta_0 \neq \Delta$ can be performed analysing the distribution of the edge times in each node of the oscillator t_i . As done for Equation 4.7 and 4.8 and considering the presence of the delay buffers in the feedforward paths, these times can be expressed as

$$t_{i+1} = t_i + t_{dmax} - \eta[t_i - (t_{(i-2) \bmod 5} + \Delta_{(i-2) \bmod 5})]. \quad (4.9)$$

A numerical approach was used to calculate the values of t_i for enough oscillator cycles such that all delay cells t'_{di} reach their convergence values. At this point, the DNL can be calculated exploiting Equation 4.5, replacing t_d with the average value of the cell delays t'_{d-av} and taking

²The approximation of Equation 4.7, as it will be explained later in the Section, is justified by simulations. However, the analysis reported in this paper is general and can be easily extended to situations in which the non-linear terms are not negligible.

into account that $\Delta_0 \neq \Delta$ as done for Equation 4.6a. The plots in Figure 4.10 show that, for the proposed solution (dashed line curves), the RMS and the maximum of the absolute value of the DNL is smaller than the one related to the usual feedforward architecture (star-marked curves). The same parameters can also be compared as function of the cell delays (LSB). In Figure 4.11, it is possible to see that the non-linearity of the proposed solution has smaller values also when t_d and t'_{d-av} are comparable. The use of t'_{d-av} has been justified in Section 4.1.1. Indeed, the event-by-event calibration system in which the TDC is integrated is able to compensate potential variations in the oscillation period measuring the frequency of the RO through a comparison with an external reference signal.

A simplified approach can be used to analyze the behavior of the proposed solution. This approach is based on neglecting the variation of t'_{di} as function of the variation of other cell delays and considering for it only the impact of Δ . This simplification, as it will be later shown, will give similar results to the ones obtained with the more detailed approach explained before because, in this analysis, only the effect of the mismatches of the buffers have been evaluated. Following the same considerations that lead to Equation 4.8, it is possible to obtain the value

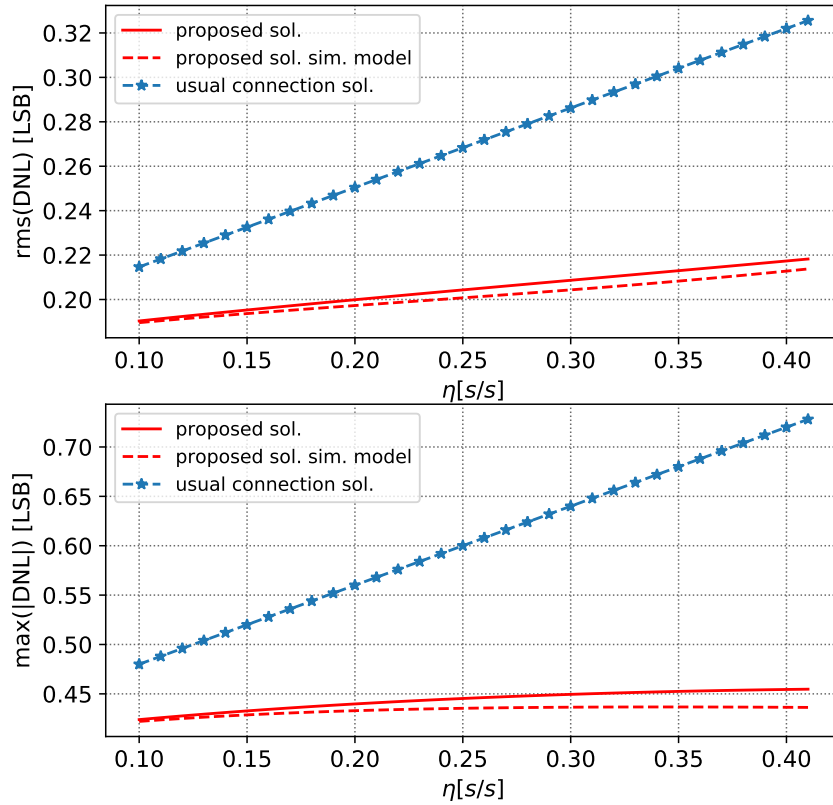


Figure 4.10: RMS (top) and maximum of the absolute value (bottom) of DNL as function of η of both of the solutions depicted in Figure 4.9 (calculated with Equation 4.6 for the usual connection case, with Equation 4.14 for the proposed solution scenario and exploiting the edge time distribution of Equation 4.9 for the more detailed model).

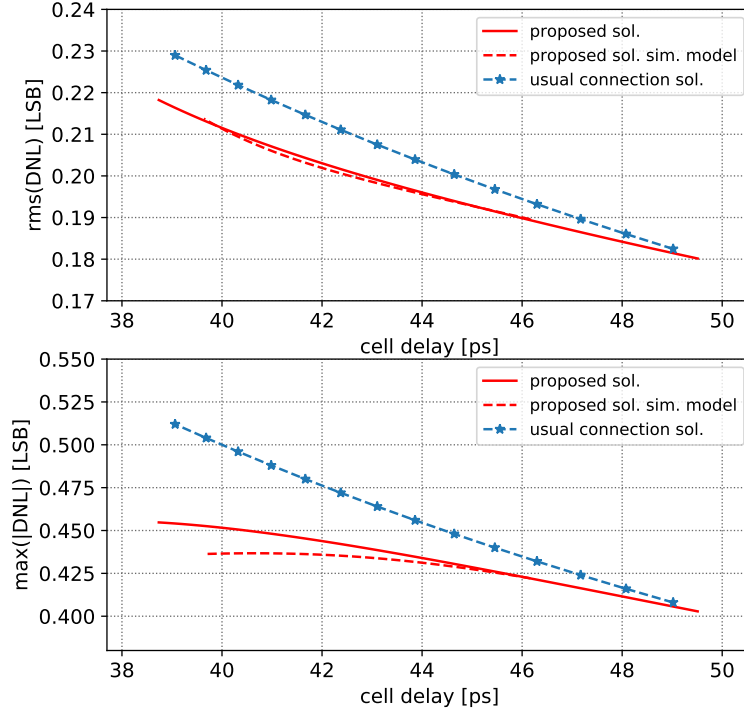


Figure 4.11: RMS (top) and maximum of the absolute value (bottom) of DNL as function of the cell delay (calculated with Equation 4.6 for the usual connection case, with Equation 4.14 for the proposed solution scenario and exploiting the edge time distribution of Equation 4.9 for the more detailed model).

of the cell delays t'_d as

$$t'_d = t_{dmax} - \eta(2t'_d - \Delta) \longrightarrow t'_d = \frac{t_{dmax} + \eta\Delta}{1 + 2\eta}. \quad (4.10)$$

However, the mismatch on the first buffer will also have an impact on the delay $t'_{d3} \neq t'_d$ that can be expressed as

$$t'_{d3} = t_{dmax} - \eta(2t'_d - \Delta_0) = t'_d + \eta(\Delta_0 - \Delta) \quad (4.11)$$

The new value of t'_{d3} will also cause a variation in the oscillation period of the RO

$$T_{RO} = 2[5t'_d + \eta(\Delta_0 - \Delta)]. \quad (4.12)$$

From Equation 4.12, it is possible to obtain the value of the equivalent LSB of the system (i.e. the average elementary delay of the cells) as

$$t'_{d-av} = \frac{T_{RO}}{10} = t'_d + \frac{\eta}{5}(\Delta_0 - \Delta). \quad (4.13)$$

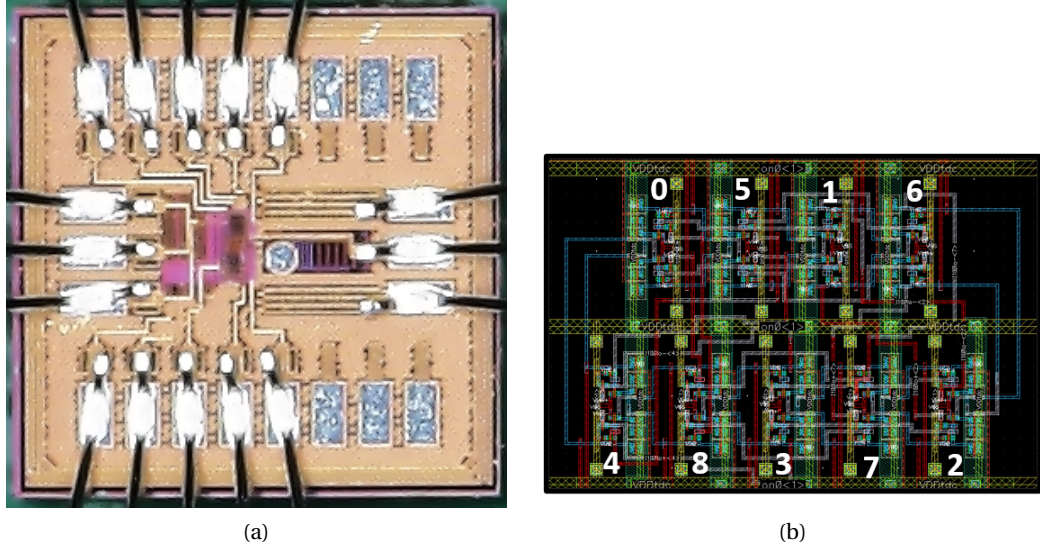


Figure 4.12: Picture of the test chip of the proposed TDC (total area: $0.9 \times 0.9 \text{ mm}^2$) (a) and layout of the RO with cells disposition (b).

Thus, the DNL of the architecture will be given by

$$DNL(i) = \begin{cases} \frac{(\Delta_0 - \Delta)(1 - \frac{\eta}{5})}{t'_{d-av}} & i = 0 \\ \frac{-\frac{\eta}{5}(\Delta_0 - \Delta)}{t'_{d-av}} & i = 1, 2, 4 \\ \frac{\frac{4}{5}\eta(\Delta_0 - \Delta)}{t'_{d-av}} & i = 3. \end{cases} \quad \begin{matrix} (4.14a) \\ (4.14b) \\ (4.14c) \end{matrix}$$

It must be clarified that in a N stages RO-based TDC, the total number of different codes the system is able to provide as output is $2N$. Hence, the $DNL(i)$ should be defined for $i = 0, 1, \dots, 2N - 1$. However, in this simplified analysis, assuming that the rise and fall times of the cells are perfectly equal, the mismatches affect the value of $DNL(i)$ for $i = j$ and $i = j + N$ with $j = 0, 1, \dots, N - 1$ in the same way. For this reason, it is possible to consider only half of the values of the DNL as done for Equation 4.6 and 4.14. In Figure 4.10 and 4.11, the solid lines represent the behavior of the non-linearities of the architecture with this more simplified approach. The approximation of the previous analysis are negligible for low values of η because of the reduced impact of the feedforward. However, even for larger η , the proposed solution shows better performance in terms of non-linearities.

4.2.3 Layout

A picture of a test chip for the proposed TDC is shown in Figure 4.12a while Figure 4.12b shows the layout of the RO. The position of the delay cells and buffer has been chosen to maximize

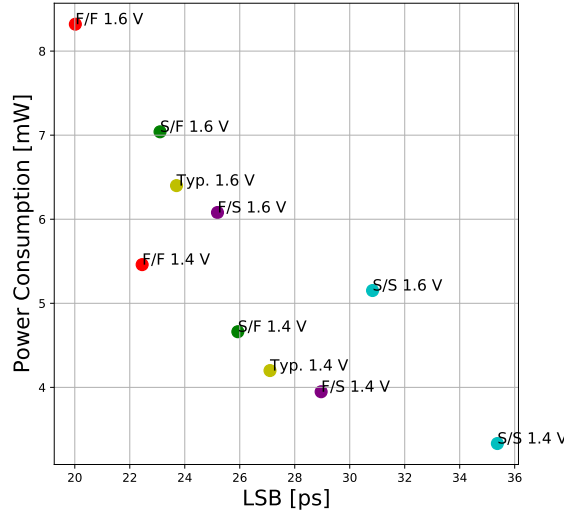


Figure 4.13: LSB and power consumption of the TDC for typical, Fast/Fast (F/F), Fast/Slow (F/S), Slow/Fast (S/F) and Slow/Slow (S/S) corners and for V_{DD} equal to 1.4 V and 1.6 V.

the symmetry of the connections. As it is possible to see in the figure, with this disposition the lengths of the feedforward paths are always one cell long while direct paths are two. The area of the RO core is $30.1 \mu\text{m} \times 20.9 \mu\text{m}$ and $30.1 \mu\text{m} \times 87.5 \mu\text{m}$ including the rest of the the system. Moreover, the outputs of the latches connected to the RO are routed on different metal layers (the pattern is 5-1-3-1-3-5 for the three inner stages) in order to reduce capacitive couplings and their effect on the oscillation frequency.

4.2.4 Post-layout simulations

The free-running frequency of the oscillator f_{RO} is highly dependent on the parasitics of the system. Simulations highlighted a 61 % drop (on average) of the f_{RO} going from schematics to post-layout netlist. Because of the sensitivity of the design to parasitics, the region occupied by the TDC is excluded from the metal filling³. The same solution was also adopted for the FASER front-end system (Chapter 3). The circuit has been analyzed for various supply voltages V_{DD} with a focus on 1.4 V and 1.6 V. Post-layout simulations show that the RO oscillates at a frequency f_{RO} equal to 2.05 GHz and 2.34 GHz for $V_{DD} = 1.4$ V and $V_{DD} = 1.6$ V respectively. Considering Equation 4.4, the system will be characterized by a nominal resolution (LSB) of 27.1 ps and 23.7 ps for the above-mentioned cases. As shown in Figure 4.13, corner simulations highlighted a variation of less than 30 % of the LSB compared to the typical case. Minimum values of the LSB are obtained in Fast/Fast corner (22.45 ps and 20.02 ps for $V_{DD} = 1.4$ V and $V_{DD} = 1.6$ V respectively) and the maximum in the Slow/Slow (30.38 ps and 35.37 ps for $V_{DD} = 1.4$ V and $V_{DD} = 1.6$ V respectively).

A preliminary analysis has been performed during the design process to evaluate the linearity of the system. The sampling of the RO was simulated sweeping the sampling time t_s in a time

³The converter layout respects the density rules of the adopted technology.

interval larger than T_{RO} , in order to be sure that the the system goes through all of its $2N$ states. The time step for t_s was set to 1 ps. For each step, several Monte Carlo simulations have been performed (using the same set of seeds for every value of t_s , in order to make the outputs coherent) to evaluate the effect of mismatch on the linearity of the system. At this point, it is possible to calculate the DNL and the Integral Non-Linearity (INL) in order to obtain the distribution of their maximum values and RMS. The INL can be defined as the integral of the DNL

$$INL(i) = \sum_{n=0}^i DNL(n). \quad (4.15)$$

The distribution of the DNL and INL obtained with this analysis for the case $V_{DD} = 1.6$ V is reported in Figure 4.14. Post-layout nominal simulations were also performed to analyze and extract the impact of the layout on the performance of TDC without considering device mismatch. The resulting maximum DNL was 0.28 LSB and 0.13 LSB for a 1.4 V and 1.6 V supply voltages respectively. Therefore, these simulations highlighted that the non-linearities of the system are mainly dominated by the mismatch of the MOSFETs and not by the asymmetries of the layout. Indeed, as described in Section 4.2.3, the cell disposition was chosen to improve the symmetry of the connection paths in the design.

Table 4.1 shows the values of nominal resolution, power consumption, DNL and INL (maximum value and RMS). The table also reports the simulated conversion time T_{conv} and a comparison with the performance of state-of-the-art solutions (that will be later presented). This parameter (equal to approximately 0.69 ns and 0.51 ns for $V_{DD} = 1.4$ V and 1.6 V respectively) only takes into account the time needed by the system to sample the state of the RO and the delay of the registers of the counters included in the converter. Thus, it represents

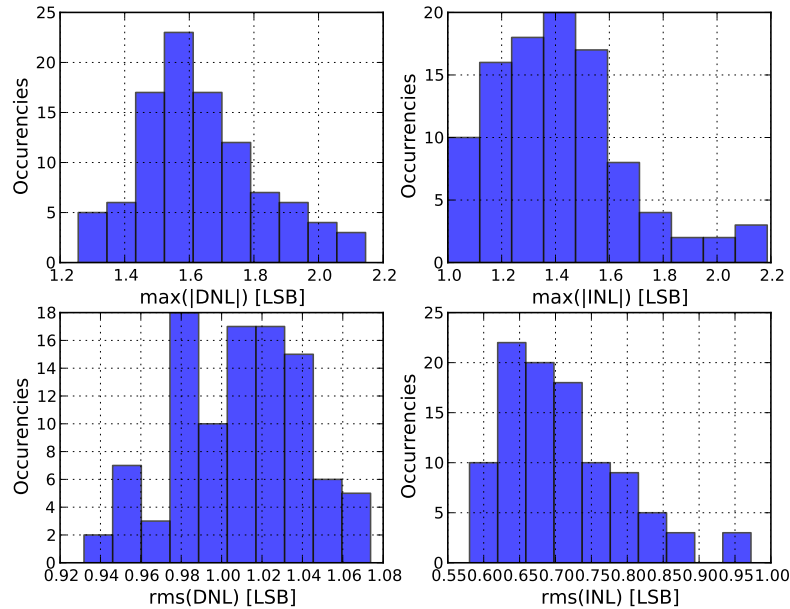


Figure 4.14: Maximum values and RMS distributions of DNL and INL calculated over various Monte Carlo simulations. In this case, the supply $V_{DD} = 1.6$ V.

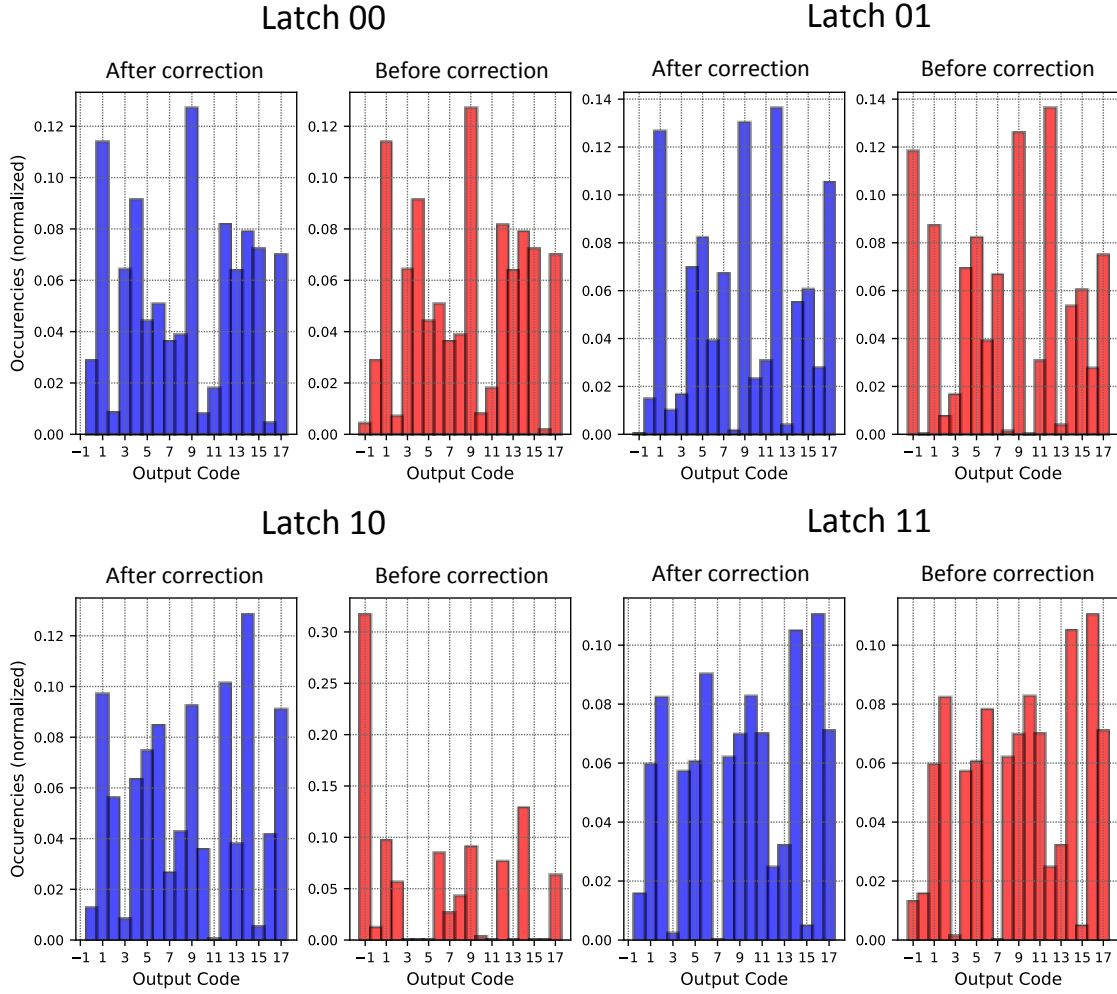


Figure 4.15: Measured output distribution (before and after correction) of the TDC for $V_{DD} = 1.6$ V and for all the latch stages connected to the RO.

the minimum ideal conversion time of the system. The measurement setup of the TDC, that will be described in the next subsection, did not allow a correct estimation of the conversion time since the system was limited by the readout logic. Hence, the aforementioned values of Table 4.1 just give an indication of the potential speed of the proposed TDC. Moreover, the T_{conv} of the converters presented in the cited work were simply extracted from the output data rate of the TDCs reported on the papers. Therefore, they simply represent upper limits of the real conversion times. The parameter E_{conv} in the table represents the conversion energy, calculated by multiplying the conversion time with the power consumption.

A test chip of the TDC featuring one channel (i.e. 4 latch stages) was submitted and its measurements will be presented in Section 4.2.5. A simulation analysis highlighted that the RO can be connected to more than one channel. Its oscillation frequency is reduced by 5.5% if 2 channels are connected and 23% in the case of a 4 channels configuration. In the applications in which such a drop is not acceptable, it is possible to add more ROs. The integration of multiple ROs is

usually problematic for area and power consumption. However, as it will be shown in Section 4.2.5, the area and the dissipated power of the proposed architecture is smaller or comparable to the ones of many state-of-the-art TDCs. Simulations and parasitic extraction analysis also showed that routing is the main limiting factor of the oscillation frequency. In the proposed architecture, the buffers of the RO drive sub-fF capacitances associated to MOS gates while their output lines are characterized by parasitics in the order of few fF. Therefore, designing the proposed solution in a more down-scaled technology can lead to an improvement of the performance only if more compact structures and routing lines are implemented.

4.2.5 Test-chip measurements and state-of-the-art comparison

The measurements of the test chip were performed using the UNIGE USB3 GPIO board (see Section 3.2.4).

Linearity Measurements

A firmware was loaded on the FPGA that the board features in order to handle the communication with the chip and send sampling signals for the analysis of the linearity of the TDC. The distribution of the output read from all the latch stages connected to the RO before and after bubble correction is shown in Figure 4.15 for $V_{DD} = 1.6$ V. The output codes have been reported along the x-axis using numbers from 0 to 17 while -1 indicates the amount of forbidden state outputs after the correction. As anticipated, applying the algorithm of bubble correction

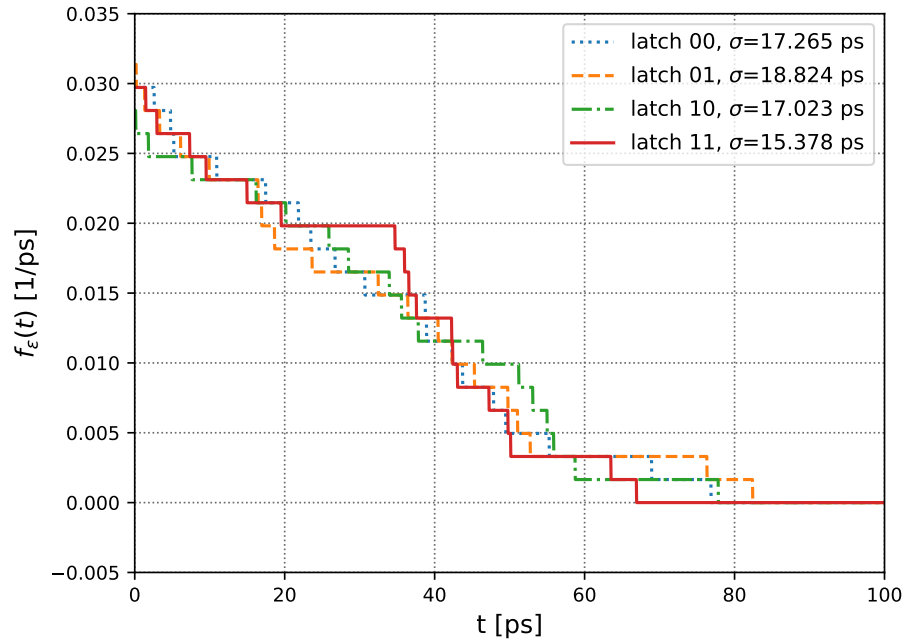


Figure 4.16: Probability density function of the quantization error for each latch stage ($V_{DD} = 1.6$ V).

explained in Section 4.1.2 to the outputs obtained during the measurements it is possible to see that only the 0.03 % of them is not corrected (latch 01 of Figure 4.15).

Table 4.1 reports the results of the measurements, compared to the ones obtained with post-layout simulations. The test chip shows a smaller oscillation frequency that turns in to a lower time resolution due to non-extracted substrate capacitances that reduced the speed of the system. The measured LSB is 38.7 ps for $V_{DD} = 1.4$ V and 33.6 ps for $V_{DD} = 1.6$ V. However, the behavior of the circuit in terms of linearity is in line with the simulation results.

The presented converter is characterized by a non-ideal output distribution in which some of the codes are missing (see Figure 4.15) leading to a $DNL > 1$ LSB. However, the TDC satisfies the specifications of the project for both equivalent time bin and linearity. The output distribution, as the one of Figure 4.15, allows calculating the standard deviation of the quantization error σ_q which represents the uncertainty on the measurements performed by the TDC. This parameter can not be calculated using Equation 2.10 because of the irregular and not ideal distribution of the bins of the system. The probability density function $f_\epsilon(t)$ of the error can be obtained using the law of total probability as

$$f_\epsilon(t) = \sum_{i=0}^{2N-1} f_\epsilon(t|C=i)P(C=i) \quad (4.16)$$

where $P(C=i) = t_{di}/T_{RO}$ is the probability that the output code C is equal to i . The behavior of the pdf for all the latch stages is reported in Figure 4.16 for $V_{DD} = 1.6$ V. The average value of the quantization error standard deviation is 21.1 ps (0.54 LSB) for $V_{DD} = 1.4$ V and 17.1 ps (0.51 LSB) for $V_{DD} = 1.6$ V. A ~ 0.5 LSB σ_q is ~ 1.7 times larger than the ideal quantization error given by Equation 2.10, for which $\sigma_q = 1/\sqrt{12}$ LSB ≈ 0.288 LSB. As anticipated, this contribution is in line with the specifications of the TT-PET project. Moreover, as reported in Table 4.1, the performance in terms of linearity of the proposed architecture is comparable with state-of-the-art solutions.

SSP and PN

The so-called Single Shot Precision (SSP), i.e. the jitter of repeated measurements of the same time interval, was measured exploiting the block diagram in Figure 4.17. A Ready signal, connected to the gating of the latches, activates a 8 bit divider. The rising edge of the output of

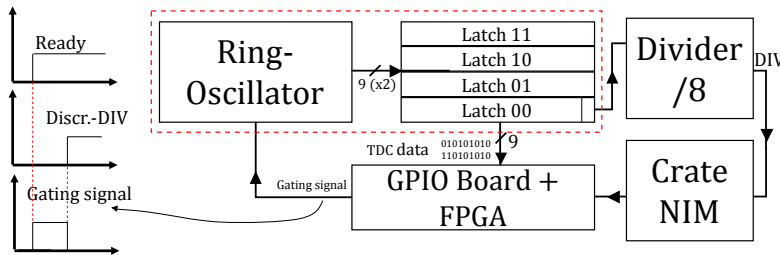


Figure 4.17: Block diagram of the measurement system to evaluate the SSP of the converter.

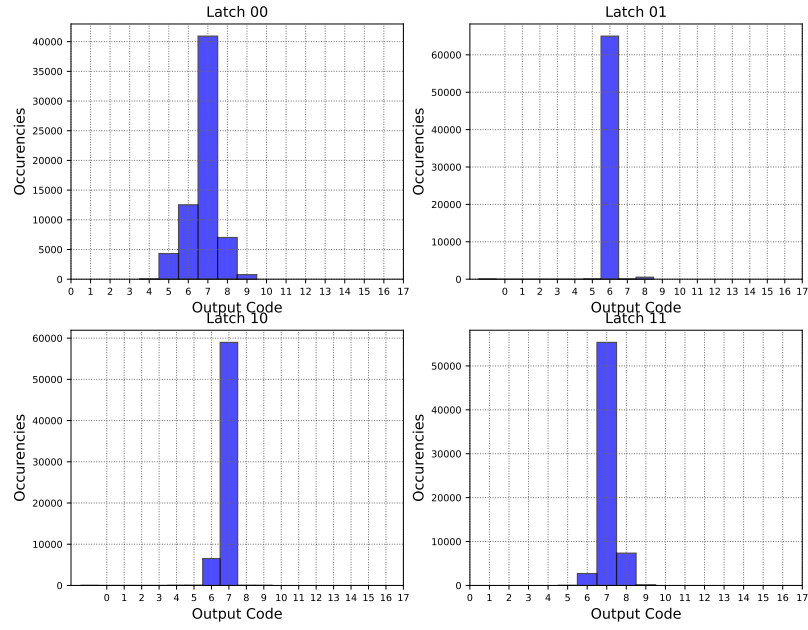


Figure 4.18: Output distribution of the data obtained with the measurement system depicted in Figure 4.17 for $V_{DD} = 1.4$ V.

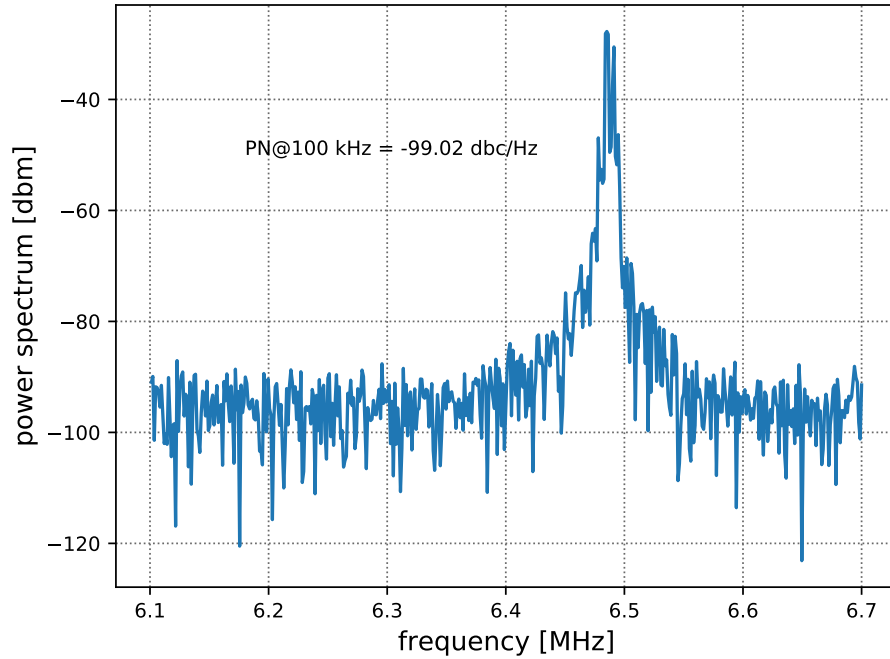


Figure 4.19: Zoom of the power spectrum of the divider output for $V_{DD} = 1.6$ V around the fundamental component of the signal.

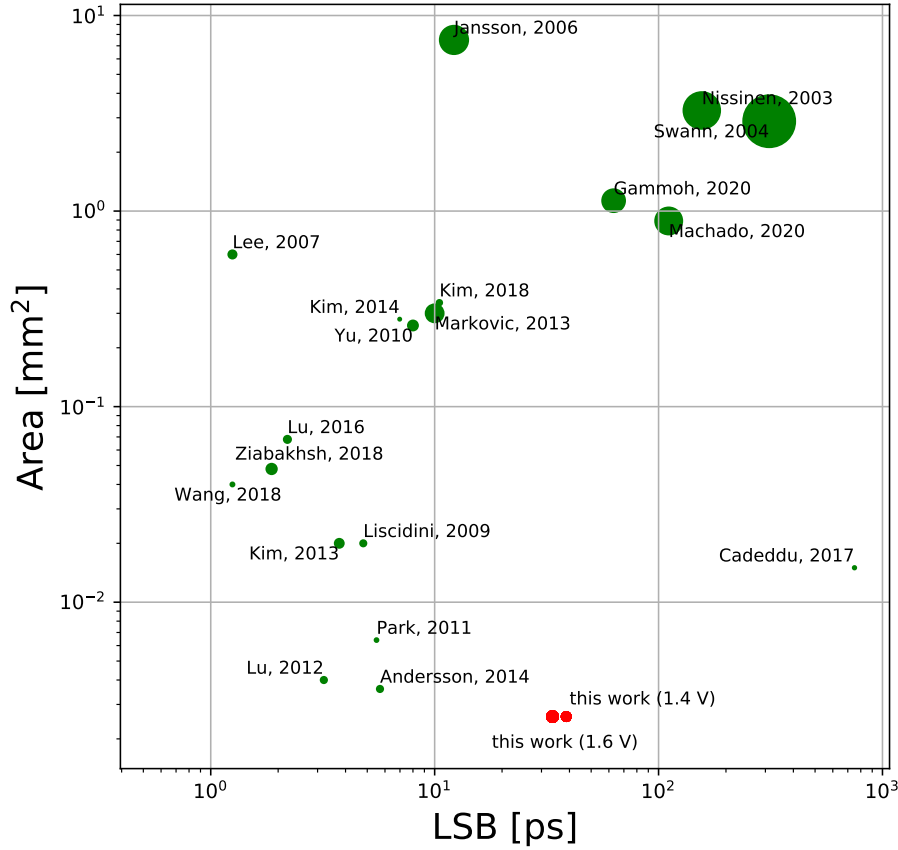


Figure 4.20: Area and LSB of the presented TDC compared to the works of Table 4.1 and the ones reported in [103–115]. The size of the dots on the plot is proportional to the power consumption of the analyzed TDCs (logarithmic scale).

this block (Discriminated-DIV in the figure) is sent, through a Crate NIM, to the GPIO Board, that will then turn off the gating signals sampling the oscillator. The value provided by the TDC should ideally be always the same. However, the standard deviation of the distribution of this outputs will represent the above mentioned SSP. The output distribution for a supply voltage $V_{DD} = 1.4$ V is reported in Figure 4.18. The average standard deviations are 15.8 ps (0.41 LSB) and 19.5 ps (0.58 LSB) for $V_{DD} = 1.4$ V and $V_{DD} = 1.6$ V respectively.

The output of the divider was also exploited to analyze the power spectrum of the RO in order to evaluate the Phase Noise (PN). Figure 4.19 shows a zoom of the power spectrum of this signal around its fundamental component for $V_{DD} = 1.6$ V. The measured value of PN at 100 kHz from this component is -99.02 dBc/Hz for a 1.6 V supply and -97.7 dBc/Hz for 1.4 V. The value of SSP and PN are reported in Table 4.1.

Comparison with state-of-the-art and final comments

Table 4.1 also offers a comparison between the TDC of this section and other works. As highlighted before, the main characteristics of the presented TDC is the compactness and

Table 4.1: Multi-path simulations and measurements results. A comparison with other works is also reported.

Architecture	Sim.			Meas.	[73]	[74]	[76]	[80]	[82]	[83]	[84]
	PLL-less Multi-path RO				TAC ¹	RO	RO-TA ²	Multi-path RO	Vernier line	Cyclic Vernier	2-D Vernier
V_{DD} [V]	1.4	1.6	1.4	1.6	5	3	1.8	3.3	1.2	1	1.2
Technology [nm]	130				500	350	180	180	65	65	65
Area [mm ²]	0.0006 (0.0026) ³				2.88	3.27	0.34	-	0.0036	0.0064	0.02
LSB [ps]	27.1	23.7	38.7	33.6	312	156	10.5	128	5.7	5.5	4.8
DNL _{max} [LSB]	1.41	1.63	1.34	1.26	0.2	-	0.7	5	<1.5	1	<1
INL _{max} [LSB]	1.28	1.41	1.77	2.02	0.3	0.23	0.5	2.4	<9	1	3.3
DNL _{RMS} [LSB]	0.87	1.01	0.68	0.66	-	-	-	-	-	-	-
INL _{RMS} [LSB]	0.57	0.66	1.15	1.31	-	-	-	-	-	-	-
Power [mW]	4.2	6.4	3.6	5.4	175	72	1.34	9 (1) ⁴	1.75	0.63	1.7
σ_q [ps] ⁵	-	-	21.1	17.1	100	-	-	-	-	-	-
SSP [ps]	-	-	15.8	19.5	-	78.5	-	57.6 - 98.6 ⁶	<17.1	2.31	-
Accuracy [ps]	-	-	40.9	31.0	-	-	-	-	-	-	-
T_{conv} [ns] ⁷	≈ 0.69	≈ 0.51	-	-	100	-	20 ⁸	-	10 ⁸	-	20 ⁸
E_{conv} [mW·ns]	≈ 2.9	≈ 3.3	-	-	17500	-	26.8	-	17.5	-	34
PN @ 100 kHz [dBc/Hz]	-	-	-97.7	-99.6	-	-	-	-	-	-	-

¹ Time-to-Amplitude Converter.² RO Time Amplifier.³ RO core (whole structure).⁴ Peak (standby).⁵ In [73] indicated as resolution.⁶ Min. and max. value reported on the paper.⁷ For the proposed solution, it does not take into account the counters.⁸ Extracted from the reported conversion rate.

the simplicity of the PLL-less architecture which makes it the solution with the smallest area among all the cited works (for [80] the area is not reported). Solutions [76, 82–84] are characterized by smaller power consumption and LSB but they have been developed in a more advanced technological node and, as explained in Chapter 2, the complexity and/or the limited maximum measurable time interval make them more difficult to be integrated in large pixel detector chips. The non-linearities of the presented architecture are comparable with the other works (only solutions [73, 74] have significantly better values of DNL and INL but their power consumption is one or two order of magnitude higher than the one of the PLL-less TDC). The performance of the proposed converter is compared to some of the works reported in Table 4.1, [73, 74, 76, 82–84], and the ones described in [103–115] in Figure 4.20. Even this plot highlights the compactness of our architecture compared to others with similar performance in terms of resolution and power consumption.

4.3 Low-power TDC for FASER experiment

This section describes the TDC designed and integrated in the first prototype chip for the upgrade of the pre-shower detector of the FASER experiment at CERN. Despite the specifications and the architecture of this prototype are slightly different from the specifications of the pre-production chip presented in Chapter 3, the TDCs designs are similar and, for this reason, the analysis and the results shown in this section can be considered valid also for the new version of the converter. In any case, the main difference will be described in the following sections.

As stated before, TDCs play a crucial role for the definition of the performance of state-of-the-art timing detectors [116]. Equation 2.10 highlights that even an ideal TDC can have a non-negligible impact on the overall resolution of the detector depending on the value of the LSB T_{LSB} . The measurements of the multi-path TDC of Section 4.2 underscored that, in a real implementation case, non-linearities, noise and other non-idealities sources can reduce the time resolution of the system of a significantly larger factor than the one given by Equation 2.10. For this reason and considering the timing specifications of the FASER experiment, the design process of the proposed TDC has been carefully dedicated to the implementation of a structure characterized by a nominal $T_{LSB} < 200$ ps. However, the accuracy of the converter was not the only parameter that needed to be optimized during the design process. Indeed, due to the large amount of TDCs in the final version of the chip, the proposed architecture must be characterized also by a power consumption P_{TDC} of few tens of μW per channel and a compact area.

4.3.1 Architecture and integration in pre-shower detector chip

The block diagram of the architecture of the proposed TDC is depicted in Figure 4.21. This converter follows the working principle of the RO-based TDCs explained in Section 2.2: a 5 stages oscillator is connected to 16 groups of three D-latch sets. The latter are used for the

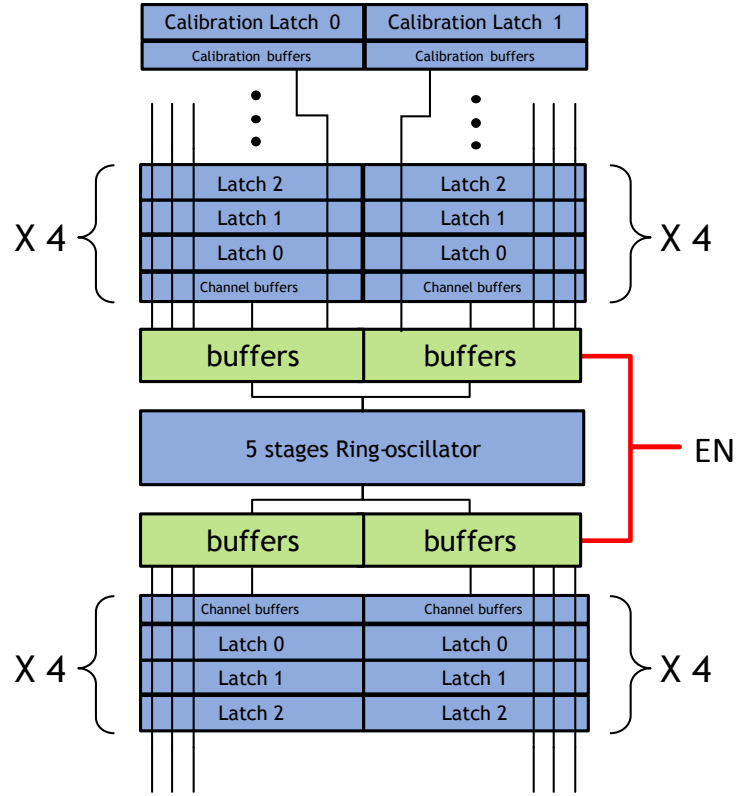


Figure 4.21: Block diagram of the TDC designed for the demonstrator chip.

sampling of the RO at the rising and trailing edge of the signal produced by the front-end electronics of the pixel. The third set of latches instead is used to sample the oscillator at the start of the readout process used as reference time instant. The TDC features also two additional latch stages that are used for the calibration of the converter exploiting the same PLL-less system described in Section 4.1.1: a test pulse can be sent to the system in order to measure the frequency of the RO that, as expressed through Equations 2.14, 2.15 and 2.16, is a fundamental parameter for the measurement process. Also in this case, this simple calibration system allows avoiding any kind of PLL-based synchronization structure and potential or unavoidable drifts of the free-running oscillation frequency can easily be compensated through a test measurement of a well-known external reference signal. Therefore, this solution makes the proposed design suitable for the presented applications since, as stated before, the power consumption, the simplicity and the compactness of the of the system represent fundamental requirements.

The power consumption is further reduced with the deactivation of the buffers that connect the RO with the latches of the 16 channels of the TDC. These buffers are implemented as pseudo-NMOS inverters in which the PMOS gate is driven by the EN signal of Figure 4.21. When no event is detected, these buffers will not continuously charge and discharge their capacitive load (associated to the gate of the D-latches) and consequently the power consumption will be significantly reduced.

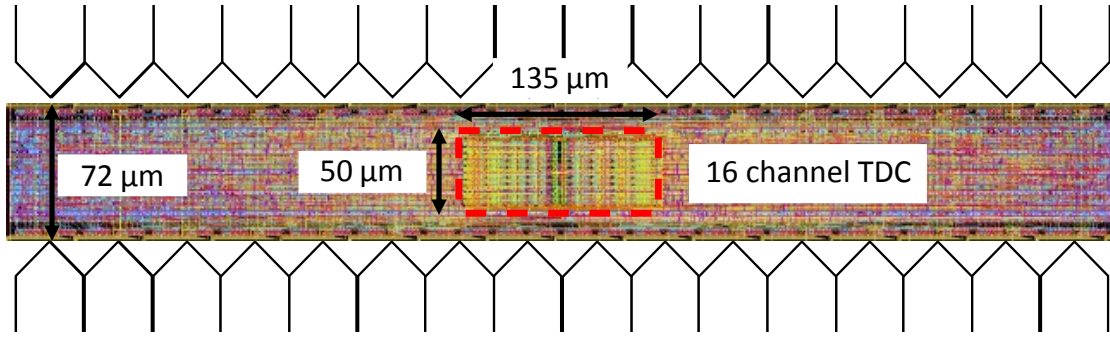


Figure 4.22: Integration of the TDC inside the readout logic a ~96 % cell density.

Integration in the first prototype

In the first prototype of the pre-shower detector ASIC, the TDC was designed and integrated within the super-column logic. The latter is 72 μm thick (it was reduced to 37.7 μm only in the new pre-production chip) in order to guarantee (considering a different shape and distribution of the pixels) a detection efficiency of approximately 92%. The specifications for the presented TDC highlighted the need of designing a compact structure to be integrated inside the 72 μm thick readout logic region. During the digital design flow, the converter has been included in the region as a routing and placement blockage, constraining the driving strength and the load of the TDC. The results of the digital flow showed that, with a 50 μm x 135 μm area converter placed in the middle of the region, the readout logic will be characterized by a ~96 % density (see Figure 4.22).

The architecture of the counters used for the evaluation of the coarse components of the measurements is based on Linear Feedback Shift Registers (LFSR). The digital readout logic integrates 16 LFSRs of 9 bit for the rising edge dynamics, 16 LFSRs of 7 bit for trailing edge range extension and one 6 bit LFSR for the calibration channels. All the counters have been synthesized, placed and routed during the digital implementation flow. The total area of the TDC, not including the LFSR counters is 50 x 135 μm^2 .

TDC design in the new FASER chip

In Figure 4.23, the block diagram of the new version of the TDC for the pre-production chip is depicted. The architecture is similar to the one of the first version of the converter: CMOS free-running RO-based structure with decoupling buffers that can be disabled to reduce the power consumption of the converter. The main differences with the previous version of the TDC are the following:

- the RO features 7 stages now. the TDC is not integrated in the super-column logic anymore (as also shown in the block diagram in Figure 3.26) but next to the end-of-column logic. For this reason, the TDC area is not a critical specification as in the previous version of the chip. Moreover, increasing the number of stages gave the possibility to

make the constraints of the LFSR counters design less stringent.

- the TDC channels are 24 and each of them features only one latch stage: the converter in the new prototype is only used for the evaluation of the ToA, while the TOT is measured with the analog memory system or the counters in the super-column logic for the charge evaluation. For this reason, additional D-latch stages in the TDC channels are not required for the pre-production chip.

The logic elements used for the new TDC, i.e. inverters of the RO, D-latches and buffers, have also been integrated in the 5-stage version of the converter. For this reason, the nominal binning is the same and the structure is slightly larger ($70\ \mu\text{m}$ instead of $50\ \mu\text{m}$). The converter also features a digital controller (left part of the layout of Figure 4.23) to generate the latching signals to the correspondent TDC channel.

4.3.2 Measurements

The first prototype of the FASER pre-shower chip was tested to perform a characterization of the front-end systems and the TDCs.

As explained before, the coarse measurement component of the converters presented in this chapter is provided by a set of LFSRs. In the case of the first prototype chip of the FASER experiment, the strict area specification of the super-column forced the counters to be integrated inside the rest of the logic by the digital-implementation tool (only the RO and the D-latches were implemented as full-custom blocks). However, because of a timing problem in the logic of this prototype, the counters were not able to provide a coherent output making the characterization process of the TDC more complicated. Without counters, it was not possible to calculate the frequency of the RO by sending a reference clock signal to the

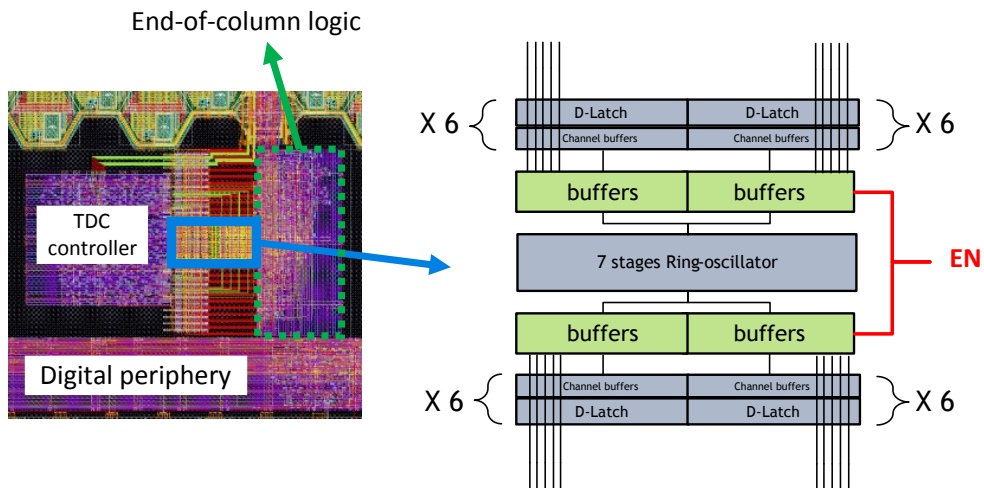


Figure 4.23: Integration of the TDC in the new pre-production ASIC for FASER pre-shower (left) and block-diagram (right).

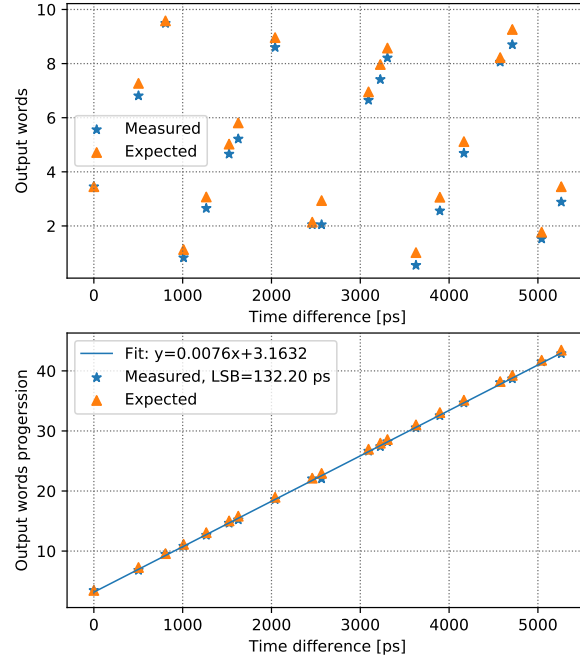


Figure 4.24: Average output words obtained with different clock frequencies as function of the time difference (top) and words progression with linear fit (bottom) for supply voltage $V_{DD} = 1.4$ V.

calibration channel because the system was only able to provide the fine component of the measurement. For this reason, the evaluation of the oscillation frequency and, therefore, of the LSB of the converter was performed with an alternative process. The reference clock signal used for the calibration is sent by an FPGA that handles the data acquisition from the chip. Modifying the multiplier of the PLLs inside the FPGA, it was possible to feed the prototype chip with clock signals with different oscillation frequencies. The evaluation of LSB was based on analysing the behavior of the fine measurement component (provided by the latches of the calibration channel) as function of the period of the clock sent to the TDC. In this way, since it

V_{DD} [V]	LSB [ps]	DNL_{max} [LSB]	DNL_{rms} [LSB]	$P_{EN} - P_{no-EN}$ [mW]
1.4	132.2	0.82	0.56	14.6
1.3	147.0	0.79	0.53	11.0
1.2	169.5	0.84	0.55	8.1
1.1	200.0	0.90	0.59	5.9
1.0	238.1	1.11	0.71	4.0

Table 4.2: Performance of the TDC integrated in the first prototype of the FASER pre-shower detector.

is possible to modify the clock period of the FPGA with steps of few hundreds of picoseconds, the behavior of the difference of the states of the RO provided by the calibration latches can be used to extrapolate the LSB. The blue star marks of Figure 4.24 on top show the average output word of the TDC as function of the clock oscillation period. The latter is reported as difference with the period of a 50 MHz clock used as reference. Making a prediction on the number of cycles of the RO within the clock period i.e. trying to make a raw guess of its frequency, it is possible to linearize the characteristics extending it for words over $2N = 10$ in the presented case (5-stages oscillator) and obtaining the plot of Figure 4.24 bottom. A linear fit of the characteristics is then performed to extract offset and slope m . The latter can be used to calculate the LSB of the TDC as $LSB = 1/m$. At this point, the expected values of average output words considering the computed LSB (orange triangles) are calculated and compared to the measured ones in order to crosscheck the characterization method.

The above-explained measurement method allowed producing the results reported in Table 4.2 in which the characteristics of the TDC are reported for different values of supply voltages V_{DD} . It is possible to see that the $LSB < 200$ ps specification is satisfied for $V_{DD} \geq 1.2$ V. A reduction of the power supply also worsens the DNL of the system. However, the non-linearities of this TDC are smaller than the ones showed by the architecture described in Section 4.2 due to the simplicity of the design proposed for the FASER pre-shower enhancement. Table 4.2 also highlights that when the enable signal is not activated, the current absorbed by the TDC is significantly smaller. Therefore, this feature will be implemented also in the final version of the FASER chip.

4.4 Possible improvements and new features

Section 2.2.2 proposed a brief overview on the most common TDC designs. The work of this thesis focused on the implementation of converters characterized by suitable architectures for timing detectors: in order to meet various specifications in terms of area, simplicity of the design and power consumption, RO-based architectures without complex calibration systems were chosen. These solution provides tens of picosecond level resolutions which satisfy the specifications of the projects for which they were designed. However, if smaller LSBs aim to be achieved still preserving compactness and power consumption, other architectures need to be adopted. For instance, Section 2.2.2 highlighted that Vernier lines or TACs can be used to obtain picosecond level resolutions as also shown in other works like [117]. Therefore, other architectures and new design methodologies must be developed for the implementation of future high-performance timing detectors. Section 4.2 presents a mathematical analysis that describes the non-linearity of the multi-path converter designed for the TT-PET project. This represents an attempt to model the characteristics of the TDC showing that it is possible to reduce the DNL with a simple re-distribution of some of the elements of the design. Therefore, a more extended analysis framework can be useful to improve the performance of state-of-the-art converters in order to highlight the aspects of the system on which the optimization process must be focused.

As done for the design of front-end systems (Chapter 3), the exploitation of other technologies can be considered as an interesting path to investigate for the development of new TDCs. In particular, the characteristics of the HBTs (like SiGe BJTs used for the front-end systems presented in this thesis) could be adopted for the implementation of fast ROs characterized by LSBs (i.e., cell elementary delays) of few picoseconds. However, the complexity of HBT-based architectures and the potentially significant increase in power consumption makes this kind of solutions difficult to use in the context of timing detectors without an comprehensive investigation and implementation of alternative designs.

5 Readout IC for Hybrid Pixel Detectors

This chapter describes the architecture of a hybrid pixel detector chip aimed to be connected to external particle sensors. Despite most of the thesis work being focused on the design and optimization of ASICs and systems for monolithic pixel detectors, the development of a hybrid architecture was considered to be useful for the testing of different sensing devices. The motivation behind this project and the goals of the presented readout chip are discussed more in detail in Section 5.1 while the design of the ASIC is reported in Section 5.2. Section 5.3 discusses potential new implementations and future developments of the chip. Also the ASIC presented in this chapter was designed in 130 nm SiGe BiCMOS technology.

5.1 Motivation

As explained in Chapter 2 and 3, the interest in monolithic pixel detectors for HEP experiments has been rising in the last years [88] but the associated design challenges often require creative and complex solutions. For this reason, hybrid architectures are still widely used for HEP and medical imaging applications: the flexibility of these detectors gives the possibility of a better optimization of the main components of the system without many critical issues associated to the integration of pixels and electronics in the same chip, e.g. area, reduced detection efficiency, front-end instabilities, increased noise.

The readout chip described in this chapter was designed to

- exploit the potential and the characteristics of hybrid architectures to explore the limits of the electronic systems designed for other projects. Without the constraints related to monolithic architectures, it is possible to study the behavior of various configurations of front-end amplifiers, TDCs and many other of critical circuits more accurately;
- test different kinds of sensors. The chip also aims to be used as a general purpose tester that can be useful for the development of new particle sensors. The sensors can be connected through wire bond or flip-chip assembling.

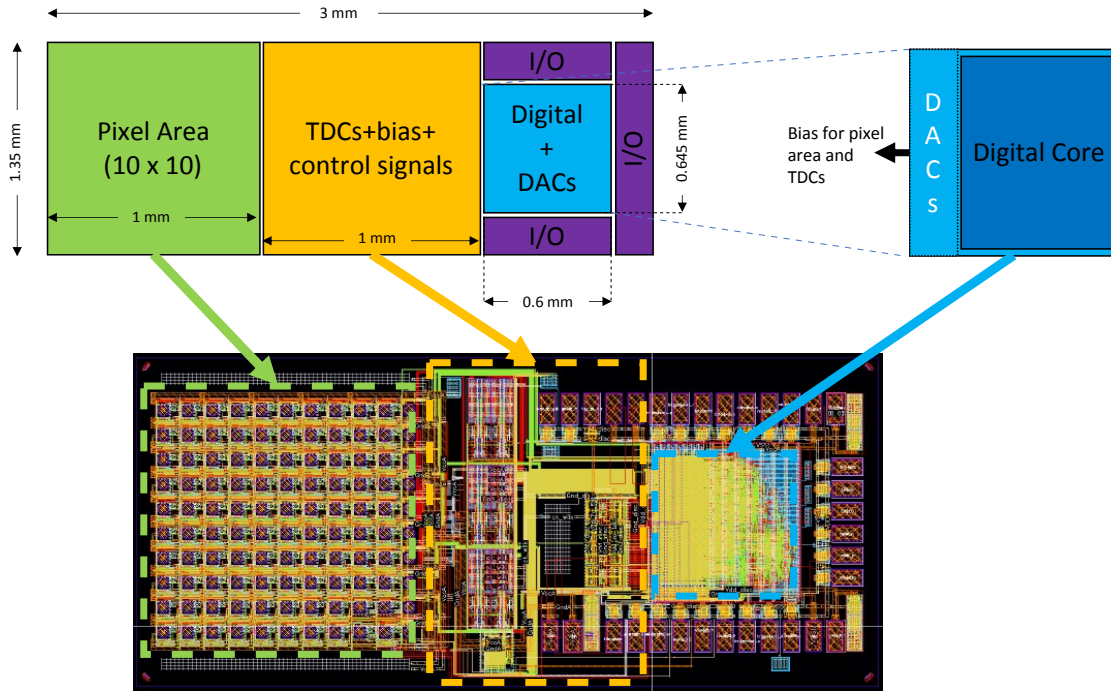


Figure 5.1: Block diagram of the readout chip (top) and layout (bottom).

For these reasons, the design of the ASIC was mainly focused on facilitating the testing of the circuits and the external sensors and making it as flexible as possible. Some of the sensors that will be tested with the presented chip are the AC-LGADs also indicated as Resistive Silicon Detector (RSD) [118–120].

5.2 System description and architecture

The design of this architecture was particularly focused on its flexibility in order to facilitate the analysis of the external sensors and the electronics integrated in the chip. For example, the ASIC features various analog channels for a direct evaluation of the outputs of the front-end

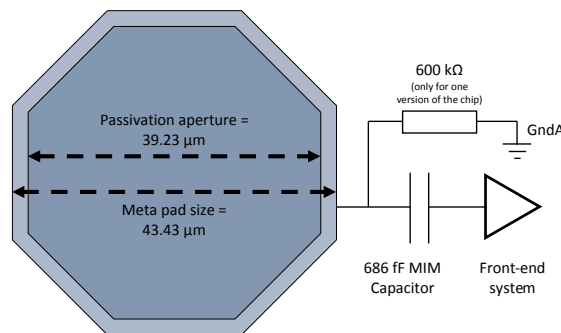


Figure 5.2: Shape and size (in scale) of the pads of the matrix and their passivation aperture.

and external test-pulse pads to stimulate the TDCs.

A block diagram of the chip is displayed in Figure 5.1. The system is composed of a 10 x 10 pixel matrix with a 100 μm pitch in both of the horizontal and the vertical directions. Each pixel is connected to an octagonal metal pad with a passivation opening (Figure 5.2) about 40 μm wide. These pads are AC-coupled with the front-ends through a 686 fF MIM capacitor. Two versions of the chip have been submitted for production, with the only difference being the 600 k Ω resistors connected to the pads and depicted in Figure 5.2. The version featuring these resistors is able to directly provide a bias to the external sensors. On the other hand, the readout ASIC configuration without the 600 k Ω resistors can only be connected to sensors that can be biased with a dedicated signal on the die in which they are integrated. The pads have been implemented as stacks of all available metal layers excluding the lowest two in order to reduce the capacitance with the substrate.

The front-end amplifiers, including discriminators, have been integrated in the pixel area, next to the respective pads. The output of the discriminators are then routed outside the matrix and connected to the rest of the chip.

The central portion of the ASIC (yellow in Figure 5.1) is featuring two different types of TDCs (Section 5.2.3) and a power gating system (Section 5.2.4).

Finally, the right part of the chip includes the digital logic, the DACs for the global bias and the I/O pad ring (Section 5.2.2).

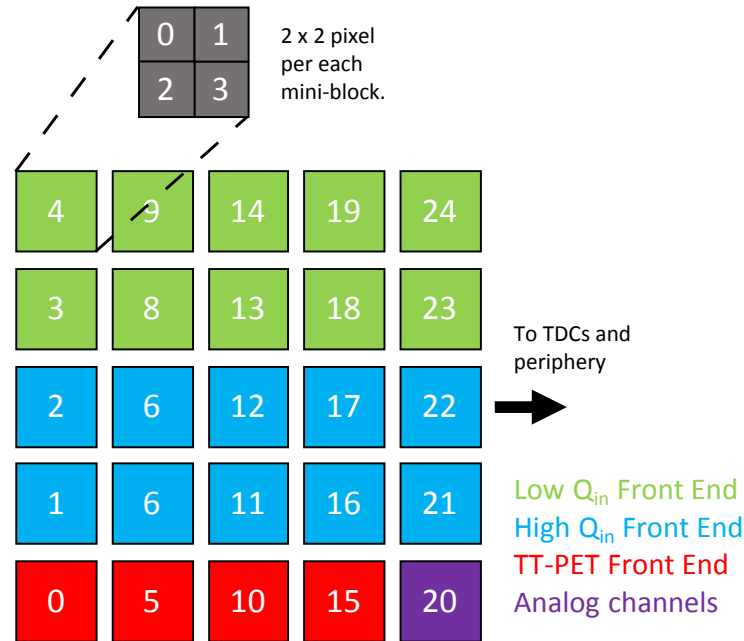


Figure 5.3: Front-end configurations for each mini-block in the pad matrix.

5.2.1 Front-end

The proposed ASIC features four different front-end architectures integrated in different sub-matrices. The main difference among these configurations is related to the range of input charges Q_{in} that the amplifiers can detect and their sensitivity. For the analysis of the front-end electronics, a Spectre simulation framework was developed in which the external sensor was modeled as an 80 fF capacitance, an estimation of the characteristics of the average detector that can be tested with the readout ASIC. All the simulations results reported below have been obtained considering this value of capacitance.

Figure 5.3 displays the arrangement of the configurations in the chip. The entire architecture can be divided in twenty-five sub-matrices indicated as mini-blocks, each of them is composed of 2 by 2 pixels. The reason behind this division is related to the number of independent TDC channels in the structure and will be better explained in Section 5.2.2 and 5.2.3.

The mini-blocks depicted in green feature a front-end designed for low input charges Q_{in} in the 0.1 - 1 fC range. This means that the system is sensitive to charges down to 0.1 fC and that over 1 fC the circuit saturates and is not able to distinguish different values of Q_{in} . The mini-blocks in blue feature a front-end configuration designed for higher charges in a 1 - 20 fC range. The ones in red, instead, are featuring an architecture with a similar sensitivity to the one for higher charges, since it is able to discriminate Q_{in} between ~ 1 fC to 20 fC but, as it will be shown later, it is characterized by a reduced Time-Over-Threshold (TOT) dynamic range. However, the power consumption of this solution is significantly smaller than the previous configurations. This architecture was already integrated in various TT-PET project ASICs. Finally the pixels of the mini-block in purple are used as analog channels since the outputs of their pre-amplifiers and discriminators are directly connected to output pads.

Pre-amplifiers architectures

The pre-amplifier schematics of the low and high Q_{in} architectures are reported in Figure 5.4. These circuits feature a common-emitter HBT-based first stage in similar fashion of the FASER ASIC pre-amplifiers. The first stage is AC-coupled with the pad making the feedback structure (on the left of each schematic in Figure 5.4) necessary for the biasing of the base current of the BJT. The main difference between these two architectures is the way the feedback MOSFET is connected to the HBT of the first stage. In order to guarantee a certain sensitivity to small charges (especially in the case of larger input rise times) in the low Q_{in} configuration, the connection of the feedback MOS is inverted with respect with the high Q_{in} variant as done for the FASER front-end (more details are reported in Section 3.2.1). The second stage of the low Q_{in} configuration was also sized to show a larger gain compared to the one of the other variant.

The TT-PET front-end configuration features a pre-amplifier whose schematic is displayed in Figure 5.5. The system is a single stage amplifier similar to the one designed for the other configurations but the feedback resistance is implemented with a different architecture. As reported in [121], this structure shows a high-value floating resistance that can be tuned in a

range of approximately one order of magnitude.

Finally, the analog channels, i.e. the mini-block number 20 in Figure 5.3, are characterized by the same architecture of the high Q_{in} configuration. In three of the four analog channels, the output of the pre-amplifier is sent to an I/O pad of the chip through a HBT-based cascode driver. The remaining channel features a discriminator (the same of the high Q_{in} front-end) connected to a MOS-based CML driver that will send two differential signal to a couple of output I/O pads. In this way, it is possible to deeply study the performance of different portions of the integrated architectures.

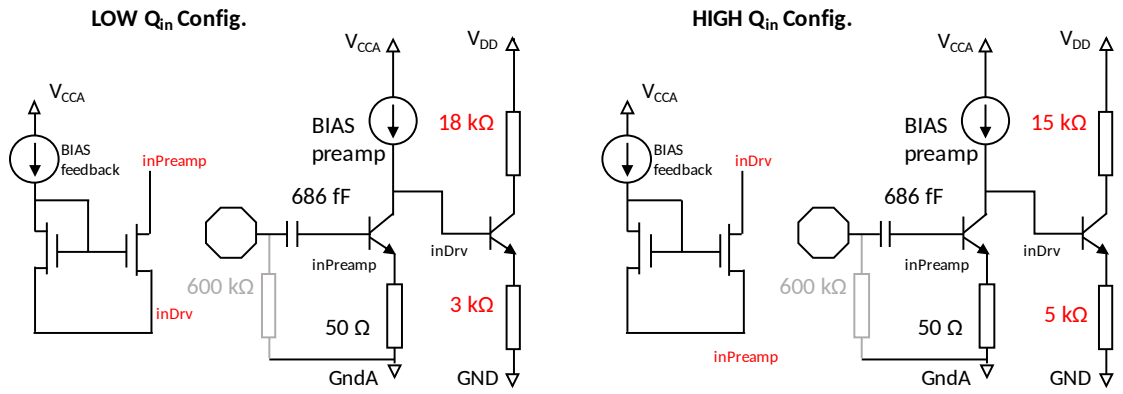


Figure 5.4: Schematic of the pre-amplifiers of low and high Q_{in} front-end configurations. Differences are highlighted in red.

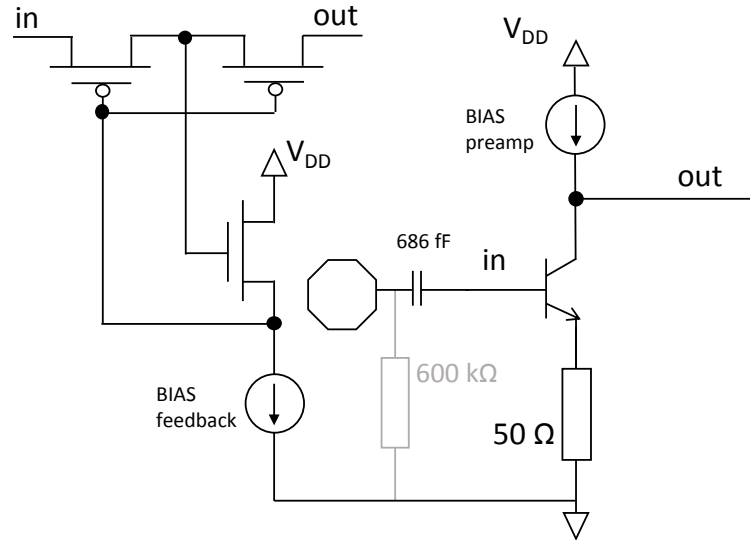


Figure 5.5: Schematic of the TT-PET pre-amplifier configuration of the readout chip.

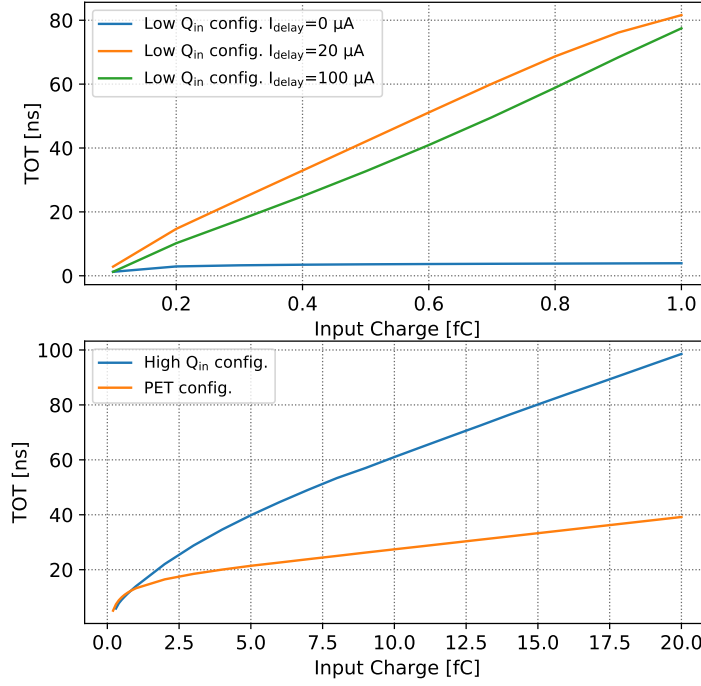


Figure 5.6: TOT dynamic range for the low Q_{in} front-end configuration for various values of delay bias current I_{delay} (top) and for the high Q_{in} and TT-PET variants (bottom).

Peak-measuring system and discriminators

The low and high Q_{in} pre-amplifier configurations integrated in the readout chip are connected to a peak-measuring system. The architecture is similar to the one described in Section 3.2.5 and studied for the future iterations of FASER ASICs front-end systems.

Simulations show a significant increase of the dynamic range of the TOT when the peak measuring system is activated, as highlighted by the top plot of Figure 5.6. This solution is also helpful to have a better recognition of the input charge Q_{in} .

It is important to highlight that the above-mentioned peak-measuring architecture is efficient when the output of the pre-amplifier does not saturate. In the high Q_{in} configuration, the input of the discriminator saturates for charges larger than 3 fC and it is characterized by the TOT dynamic range displayed in Figure 5.6 bottom. However, the peak measuring system has also been integrated in this configuration for testing purposes (since, as said before, it can be deactivated setting $I_{delay} = 0$ mA).

The discriminator used for the low and high Q_{in} configurations has the same BJT-based architecture used for the FASER experiment ASIC and reported in Chapter 3. On the other hand, the TT-PET variant is featuring a CMOS discriminator (see Section 2.2) with a Schmitt trigger stage to guarantee an hysteresis of ~ 30 mV.

Configuration	Low Q_{in}	High Q_{in}	TT-PET
Pre-amp Gain [mV/fC]	570	353	240 - 99
Power [mW/channel]	0.88	1.16	0.18
ENC [e^-]	72	79	88 - 214

Table 5.1: Expected performance of the three front-end configurations integrated in the readout chip. The gain and the ENC of the pre-amplifier have been calculated as the average gain in the input charge range of interest. However, since in the high Q_{in} configuration the pre-amplifier saturates for $Q_{in} > 3$ fC, the gain reported in the table was extracted from simulations performed with lower values of input charge. The gain and the ENC reported for the TT-PET configurations are the values for 1 fC and 5 fC Q_{in} respectively. For $Q_{in} > 5$ fC the front-end saturates.

Expected performance

The main performance parameters of the three front-end configurations integrated in the readout chip are summarized in Table 5.1. As anticipated, the low Q_{in} configuration has the largest average gain in the input charge range of interest and the best performance in terms of noise (ENC). Slightly larger ENC is obtained with the high Q_{in} configuration while the TT-PET front-end, due to a smaller charge gain, shows the worst equivalent noise charge.

The timing performance of the circuits is reported in Figure 5.7: the plots display the behavior of the standard deviation of the ToA as function of the input charge in the range of interest (this parameter is equivalent to noise contribution given by the electronic front-end in a detector of Equation 2.7). Since the system uses the TOT information to perform the extraction of Q_{in} for the time-walk correction (see Section 2.2.1), the uncertainty on the evaluation of this period of time has an influence on the ToA. For this reason, the total standard deviation of the time of arrival $\sigma_{ToA-total}$ can be expressed as

$$\sigma_{ToA-total} = \sqrt{\sigma_{ToA}^2 + (\sigma_{TOT} \frac{dToA}{dTOT})^2}, \quad (5.1)$$

where the second contribution in the square-root represents the propagation on the ToA of the uncertainty on the measure of the TOT σ_{TOT} ¹. Equation 5.1 is justified by the fact that the jitter on the ToA is significantly smaller than the one on the TOT. Thus, the two contribution in the square-root can be considered as independent. The analysis led to the results reported in the plots of Figure 5.7. It is possible to show that the configuration for low charges shows a $\sigma_{ToA-total}$ of several tens of picoseconds for small Q_{in} and ~ 27 ps for Q_{in} close to 1 fC. These values are also similar to the ones showed by the high charge configuration in which, because of the large TOT dynamics (i.e., small values of $\frac{dToA}{dTOT}$), the second contribution in Equation 5.1 is almost negligible. The higher values of $\sigma_{ToA-total}$ for small charges in the low Q_{in} configurations are justified by Equation 2.8 which highlights the impact of the SNR on the noise contribution of the electronics.

¹ It is important to underscore that the time-walk correction requires a calibration process to extract the behavior of the time-walk as function of the TOT.

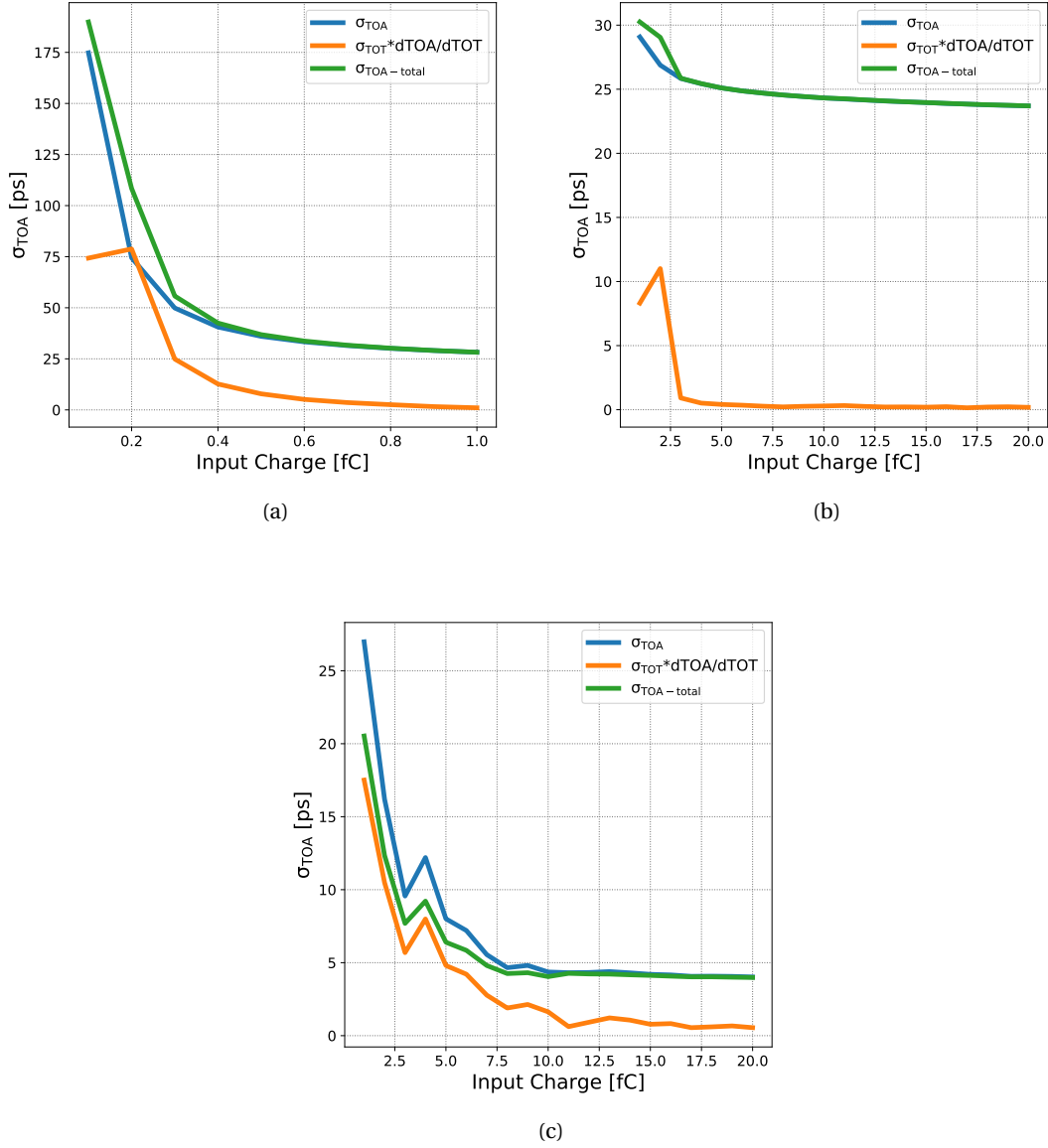


Figure 5.7: Standard deviation of the TOA as function of the input charge for the low Q_{in} (a), high Q_{in} (b) and TT-PET (c) front-end configurations.

The TT-PET configuration is the best one in terms of $\sigma_{TOA-total}$ which spans in the ~ 20 ps to ~ 5 ps range. These values are in line with a smaller average rise-time of the configuration (see Equation 2.8). However, the TT-PET variant is not able to sense small charges as the low Q_{in} configuration and its TOT dynamics is lower than the one of the high Q_{in} architecture.

The TT-PET front-end also has the lowest power consumption among the integrated architectures mainly due to the simplicity of the design (the architecture does not feature the peak-measuring system and the pre-amplifier is composed by a single stage). The power of

all the configurations can be adjusted by a set of global DACs programmed by the readout logic. Moreover, a power gating system is integrated in the chip to turn off the unused columns of the matrix during the tests. In this way, it is possible to increase the power delivered to the front-end systems above what it is allowed by specifications for testing purposes without compromising the stability of the entire chip. More details in Section 5.2.4

5.2.2 Digital logic

The digital logic of the ASIC aims to make the chip communicate with an external UNIGE GPIO board (briefly described in Chapter 3). It can be divided in two main sub-blocks indicated as pixel logic and periphery logic. The former handles the data extracted by the analog part of the ASIC and provides them in output during readout. The latter is used to define the settings of the chip (e.g. bias, working mode). These systems work on two different clock domains, as it will be explained in the rest of the section. However, pixel and periphery logic have been synthesized together using a similar implementation flow to the one adopted for the periphery of the FASER pre-production ASIC (Chapter 3).

Pixel logic

The main role of the pixel logic is to handle the readout. This logic was designed to work with a 100 MHz clock. The latter is provided to the chip through LVDS lines as well as the readout reset signal. Two receivers are integrated inside the ASIC to generate a single-ended version of the clock and the reset. The readout data are provided through CML drivers connected to output pads.

Figure 5.3 shows that the pad matrix is divided in twenty-five mini-blocks. Each of the pixels in these structures belongs to a different set of matrix elements indicated as Pixel Groups (PGs). Therefore, the matrix of the readout chip is composed of four PGs of twenty-five pixels each. As it will be clearer in Section 5.2.3, in one of the working mode of the ASIC, the front-ends of the PGs are connected to four different TDCs (see Section 5.2.3), i.e. each PG is associated to a different TDC channel. Therefore, the readout chip is able to acquire the timing data related to multiple hits on different PGs. This means that simultaneous events on pixels belonging to different PGs can be readout by the logic and the timing information provided by the TDCs related to those events can be correctly stored and sent to the GPIO board. However, if multiple hits occurs in pixels of the same PG, the logic is able to acquire the timing data only of the first matrix element that sensed a hit but it will also store the position of the other pixels that detected an event.

The sub-block of the pixel logic that handles the communication during readout and provides to the ASIC the above-mentioned multi-hit capability has been named Readout Control (RC). The output of the discriminators are latched by this block in four 25-bit words, one for each PG. A '1' in the PG word indicates that a hit on a pixel of the correspondent PG occurred. In the first phase of the readout, the data of the TDCs are collected in the pixel logic and then sent out on the output pad. A data pruning system allows skipping the readout of TDCs associated

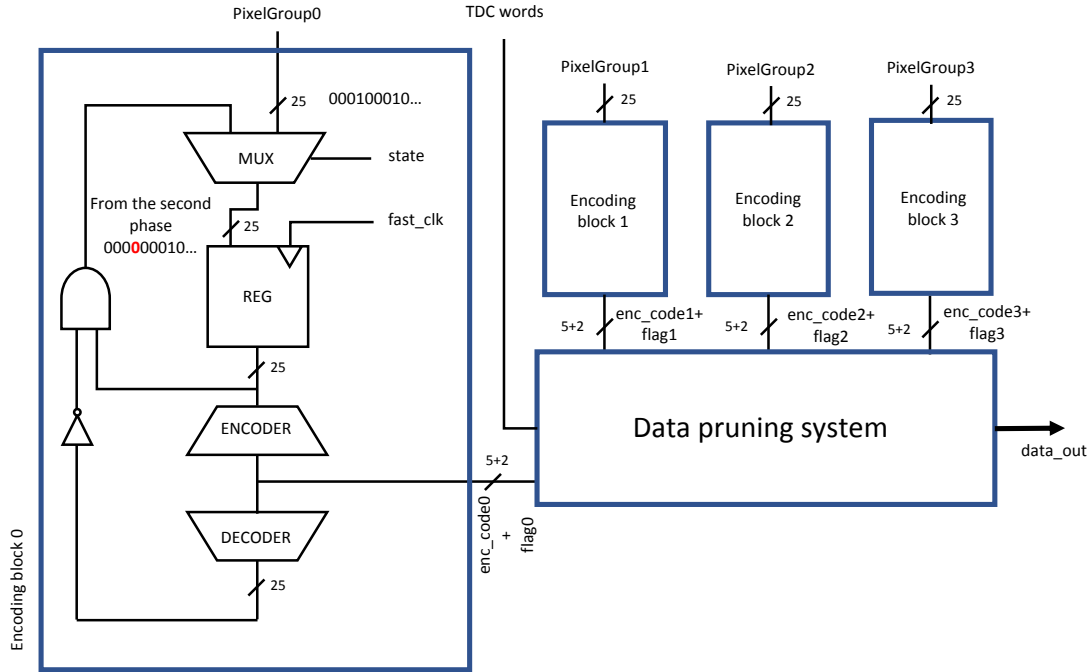


Figure 5.8: Simplified block diagram of the Readout Control system. The PG word sent to the encoding blocks in the first part of the readout are stored in a separate register since they are used to identify the first pixel that sensed a hit in the readout time frame.

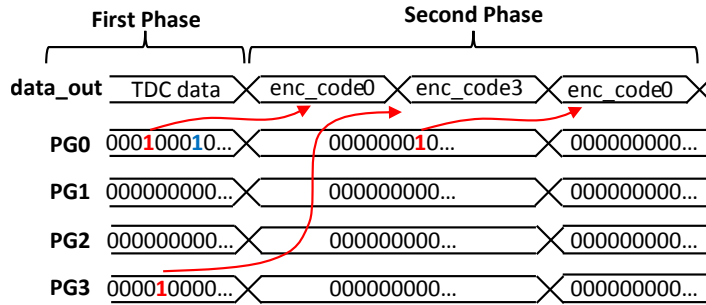


Figure 5.9: Time diagram example of the output data and PG words during readout.

to PGs which did not sense any hit. For this reason, at the beginning of every TDC word, a 2-bit flag is sent to identify the corresponding PG. As it will be explained in Section 5.2.3, this solution is useful to significantly reduce the readout time due to the large amount of data coming from the converters.

In the second phase of the readout, the identifiers of the pixels that detected an event are read. The same data pruning system for the output of the TDCs is adopted to compress the data and skip the identifiers of the inactivated PGs. The block in Figure 5.8 describes the behavior of the encoding system used to extract the position of the detected events from the four PG words. At the beginning of the readout, when an event is sensed, a trigger signal is activated and the first hit detected on each PG is stored in a separate register. These data are encoded by the

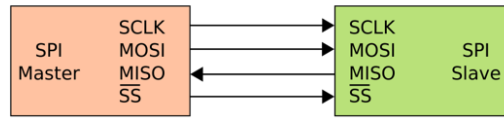


Figure 5.10: Chip and GPIO board communication for the periphery logic based on SPI standard.

ENCODER block, a priority encoder, and the correspondent identifiers are sent to the GPIO right after the outputs of the TDCs. At this point, if only one hit per PG is sensed, i.e. if there is only one '1' the PG words, the readout process is completed; otherwise, the feedback of the encoding blocks in Figure 5.8, are activated. The working principle is the following: a FSM checks if there is at least a '1' in the PG words and, in this case, saves the latter in a register REG. The priority encoder ENC generates a 5-bit pixel identifier from one of the '1' bits of the PG word and sends it to the data pruning system. At this point, the AND between the content of REG and the negative output of the decoder DEC will turn to '0' the bit correspondent to the last pixel whose identifier has been sent to the MISO pad. In this way, in the next phase of the readout, ENCODER can produce an identifier associated to another pixel (another '1' in the PG words). The process continues until all the hit positions have been acquired. An external signal can also be sent to stop the readout. A simplified example of the behavior of the output data and pixel group words during readout is reported in Figure 5.9.

Periphery logic

The periphery logic allows programming the global DACs for the bias of the analog circuits of the ASIC and generates the signals to drive the power gating system. Moreover, it allows configuring the working operation of the chip and choose which TDC architecture to use to acquire the timing information (more details in Section 5.2.3). The communication with the UNIGE GPIO board is based on SPI standard in which the ASIC is the slave as reported in Figure 5.10. A 10 MHz clock is used for the synchronization and all the I/Os are single-ended CMOS.

5.2.3 TDC

The readout ASIC features two different systems for time-to-digital conversion. As explained before, the chip can use a set of four TDCs and each of them is connected to a different PG. Their architecture is based on a free-running RO in a similar fashion to the one implemented for the TT-PET project (Section 4.2) and they also feature the PLL-less calibration system explained in Section 4.1.1. Therefore, the performance of these converters are expected to be similar to the one obtained for the TT-PET ASIC TDC. Each of the converters generates 43-bits output words plus 2-bit flags per event.

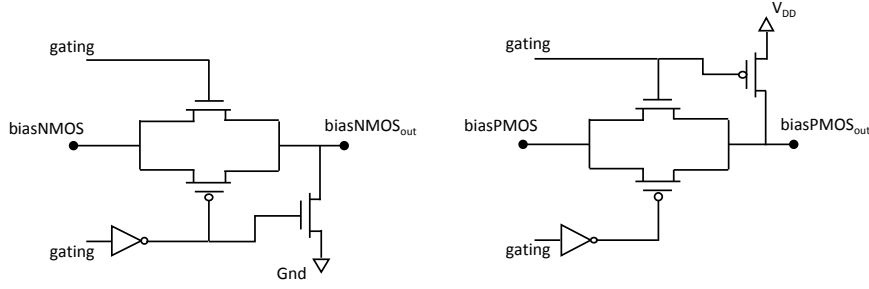


Figure 5.11: Switches used for the power gating of the readout chip matrix.

The readout chip also integrates another single channel TDC which can be used as alternative to the set of four converters mentioned before. The architecture of this system will not be reported in the thesis. The system is expected to be characterized by a picosecond level resolution with a ~ 1 ps single-shot precision. However, due to the size of the system (area optimization is still to be done), only one channel was integrated inside the readout chip. This means that, when the user of the ASIC decides to exploit the characteristics of this second TDC, the outputs of all the PGs is connected to the same channel. During the programming phase of the chip it is possible to choose which TDC configuration must be enabled. If the alternative converter is chosen, the system will be able to store the timing information only of one pixel in the case of a multiple-hit scenario but the position of all the pixels that detected a particle will still be readout. Also the data pruning system of the outputs of the encoding block will work as explained before. The second converter produces 177-bit words per event.

5.2.4 Power gating system and biasing

As explained before, the readout chip was designed for testing purposes focusing on the flexibility of its architecture and the possibility to evaluate the various configurations of front-end electronics integrated in the matrix. The chip features a set of fourteen 8-bit DACs used to bias the analog circuits in the ASIC such as pre-amplifiers, converters for the local threshold of the pixels, discriminators, CML drivers and some of the circuits included in the picosecond resolution TDC. The architecture of the DACs is based on a set of current mirrors with different division ratios and they are directly controlled by registers programmable with the SPI slow control bus. The output currents of these converters are collected in additional current mirror stages that will scale the output to match the bias needed by the correspondent analog circuit. The readout ASIC was designed to test the properties of the integrated front-end architectures in different operation modes with the possibility of increasing the power consumption of the systems as needed and to evaluate the limits of the amplifiers. In order to guarantee the normal operation of the chip and to prevent potential damage due to excessive heat, a simple power gating system was integrated in the ASIC. As said before, the bias of the analog circuits is provided through a connection between the gates of the master and slave MOSFETs of the output current mirrors of the DACs. In the readout chip, a set of switches was added on these

connection lines to give the possibility to disable the bias and to turn-off the analog circuits in the ASIC. The schematics of the switches is reported in Figure 5.11 which also highlights the presence of pull-down/pull-up transistors: when the switches are disconnected, these devices set the voltage at the gates of the slave MOSFETs of the current mirrors in the analog circuits in order to disable them. The power gating system is driven by the periphery logic and it is able to independently enable or disable the five mini-blocks rows displayed in Figure 5.3. In this way, it is possible to safely change the working point of some parts of the circuit while assuring that the parts currently not under test do not absorb a significant amount of power.

5.3 Future developments

A new revision of the readout chip will be designed and submitted for production. Various design solutions will be explored in the future including:

- larger matrix for the measurement of dies with larger sensors (to be defined on the basis of future specifications).
- different front-end configurations. Other architectures for pre-amplifiers and discriminators will be designed and alternative solutions for the evaluation of the ToA and TOT will be investigated (e.g., double threshold discriminators).
- new TDC architectures and optimization of the picosecond level converter.

Moreover, the possibility of exploiting more downscaled technology nodes will be taken into consideration.

6 Genetic Algorithms for Circuit Optimization

The design of high-performance circuits often requires the implementation of new architectures to optimize the usual trade-offs among various parameters, e.g. speed and power consumption. However, during this optimization process, tuning small elementary circuits can be crucial to further improve the performance of the system. This chapter describes how Genetic Algorithms (GAs) can be exploited for the automatic optimization of various performance parameters of analog and digital circuits. Despite state-of-the-art CAD tools for circuit optimization are already available on the market, this work shows the implementation of custom GAs in order to focus on specific performance parameters of the circuits that may be difficult to optimize with commercial tools. In this chapter, various GAs used for the tuning of the D-latches of the TT-PET project TDC (Chapter 4), and the minimization of their setup-time, power consumption and area are described. The development of dedicated and efficient GA-based codes represents a simple solution to efficiently address the optimization of particular circuit parameters that cannot be easily described in available CAD tools (e.g., setup time). This approach allows having greater insight on the inner working of the algorithms that can be useful to achieve suitable circuit solutions for the project specifications. Moreover, different types of GAs (e.g., linear aggregation or Pareto-based algorithms) were explored and compared to highlight their strengths and how to exploit them on the basis of the applications. In Section 6.1, the basics of GAs will be explained. Section 6.2 shows the way these algorithms can be used for the design of analog circuits while Section 6.3 describes more in detail some of the GAs used for the optimization of the aforementioned D-latches. A comparison and final discussion is reported in Section 6.4.

6.1 Introduction to Genetic Algorithms

GAs belong to a class of optimization procedures to minimize a certain number of functions. They mimic the same mechanisms that regulate evolution of species in nature. For this reason, they are also indicated as Evolutionary Algorithms (EA) [122] can be applied to a large variety of optimization problems.

A generic GA can be defined as a procedure in which the main goal is the minimization

of a function $F(\underline{x})$ with a certain set of constraints, where $\underline{x} = (x_1, \dots, x_N)^T \in I \subseteq \mathbb{R}^N$ is a N -dimensional variable and $F : I \subseteq \mathbb{R}^N \rightarrow \mathbb{R}$ [122]. The function F can often be defined as a linear combination of a set of M parameters (or costs) to optimize $F_i = F_i(\underline{x})$ with $i = 1, \dots, M$ as

$$F(\underline{x}) = \sum_{i=1}^M \alpha_i F_i(\underline{x}). \quad (6.1)$$

The values of α_i can be chosen to set the weight and the importance of each parameter F_i during the optimization process on the basis of the specifications of the problem. The cost functions F_i can be defined as the performance of a certain system to optimize as, for instance, power consumption, area and noise in the case of circuit design.

As it will also be explained in Section 6.2, a general GA starts with a first set of randomly generated solutions (population) that satisfy some pre-determined constraints. After that, a portion of this population is selected to form the so-called mating pool. The latter will be used for the generation of new solutions that will update the original population. In principle, with the mating pool the algorithm aims to select the best solutions in terms of fitness in order to create new individuals that could potentially replace the worst ones in the previous generation. This procedure is based on two operations indicated as crossover and mutation: the former allows calculating a new (son) solution \underline{x}_{son} combining the ones composing the mating pool (parents). In this way, it is possible to 'transmit' the good characteristics of a certain set of solutions to \underline{x}_{son} , which can potentially show a better fitness than the previous ones. For what concerns the mutation operator, it is used to modify the value of \underline{x}_{son} applying a certain random variability to it. For example, in some of the algorithms described in Section 6.3 the mutation operation is based on adding to \underline{x}_{son} the vector $\Delta \underline{x} = (\Delta x_1, \dots, \Delta x_N)^T$ in which each component is a Gaussian random variable. Other approaches consist in applying the mutation to each component only with a certain probability in order to reduce the research space of each 'offspring' solution [123]. The algorithm finishes when a certain termination condition is satisfied. For instance, a GA can be stopped when a fixed number of generations has been calculated or when the fitness value of the best individual is stable for a certain amount of cycles.

From this description of a generic GA, it is possible to highlight that the generation of new individuals requires the comparison of several solutions in order to establish the ones that are suitable for the creation of the mating pool. Moreover, once the offspring is generated, a new comparison with the previous generation is required to replace old and worse solutions. While in many algorithms this operation is easily implemented by a numerical comparison of the fitness of the various solutions $F(\underline{x})$ (checking also the feasibility of the latter, i.e. if they respect the set of predetermined constraints of the algorithm), in other GAs new comparison criteria are introduced, such as the concept of domination [122, 123]. For instance, as done for the algorithm discussed in Section 6.3.4, indicating with \underline{x}_1 and \underline{x}_2 two feasible solutions and $F_i = F_i(\underline{x})$ with $i = 1, \dots, M$ a set of M parameters to optimize (where \underline{x} represents the generic solution), \underline{x}_1 dominates \underline{x}_2 , $\underline{x}_1 \prec \underline{x}_2$, if both of the following conditions are respected

- $F_i(\underline{x}_1) \leq F_i(\underline{x}_2) \forall i = 1, \dots, M$

- $\exists j = 1, \dots, M \mid F_j(x_1) < F_j(x_2)$.

This means that a solution dominates another one if it is not worse than the other for every parameter to optimize and if it is strictly better for at least one of them. This definition can be extended to the case in which at least one of the solutions is unfeasible: if x_1 is feasible and x_2 is not then $x_1 \prec x_2$, while, if both of them are unfeasible, the constraint violation function needs to be calculated to define the domination. More details on the constrain violation function are reported in [123].

The definition of domination is useful to underscore the main goal of the so-called Pareto-based algorithms (as the one presented in Section 6.3.4) which is finding a group of feasible solutions of the Pareto-optimal front. The latter is defined as the set of all the individuals that are not dominated by any other solution [122]. In Figure 6.1 a simplified representation of a Pareto-based GA and the associated Pareto front is depicted. The algorithm aims to calculate a set of solutions that are as close as possible to the Pareto-optimal front. These solutions are not dominated by any others and each of them is handling the trade-offs among the various components to be minimized (or maximized) in a different way. The final solution of the problem can be chosen on the basis of the parameters that the user wants to optimize.

An alternative approach for the implementation of a GA for the research of the Pareto-front is based on defining a fitness function $F = F(x)$ as in equation 6.1 and solving a problem for different values of α_i in order to focus singularly on each component $F_i(x)$. However, this approach is not efficient under a computational point of view. Indeed, many GAs, as the one explained in Section 6.3.4, aim to find the individuals of the Pareto front in one single algorithm process. This goal is achieved also introducing new metrics for the evaluation of the individuals in each generation, like for example density functions that allow analysing the distances among solutions in the objective space. In this way, the algorithm is able to

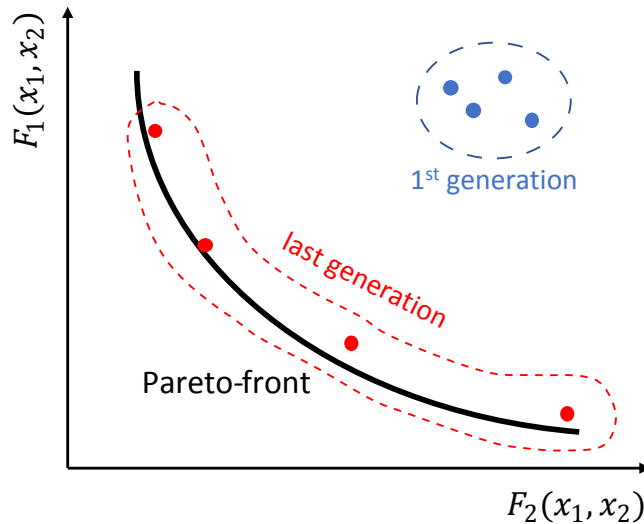


Figure 6.1: Representation of the Pareto-optimal front of an evolutionary algorithm with $M=2$ optimization parameters F_1 and F_2 .

calculate, generation after generation, sets of scattered and heterogeneous offspring in \mathbb{R}^M . The goal of this brief overview is to give an introduction to the most important concepts related to GAs implemented in this work and described in the rest of the chapter for the tuning of analog circuits. A more detailed dissertation about evolutionary algorithms can be found in various textbooks as [122, 124, 125].

6.2 Design framework for D-Latch optimization

Figure 6.2 depicts the block diagram of the GAs implemented in this work for the design of D-latches. The working principle is the same described in Section 6.1: a first random set of solutions is created and their fitness is evaluated. A group of them is selected to form the mating pool from which, applying the crossover and mutation operators, the offspring is generated. The latter is then evaluated (fitness calculation) and eventually, if a set of conditions is satisfied, the new solutions will replace a portion of the old ones to form the following generation. At the end of this cycle, if the predetermined termination condition is true then the algorithm stops and provides the best solution of the problem (or a set of individuals if a

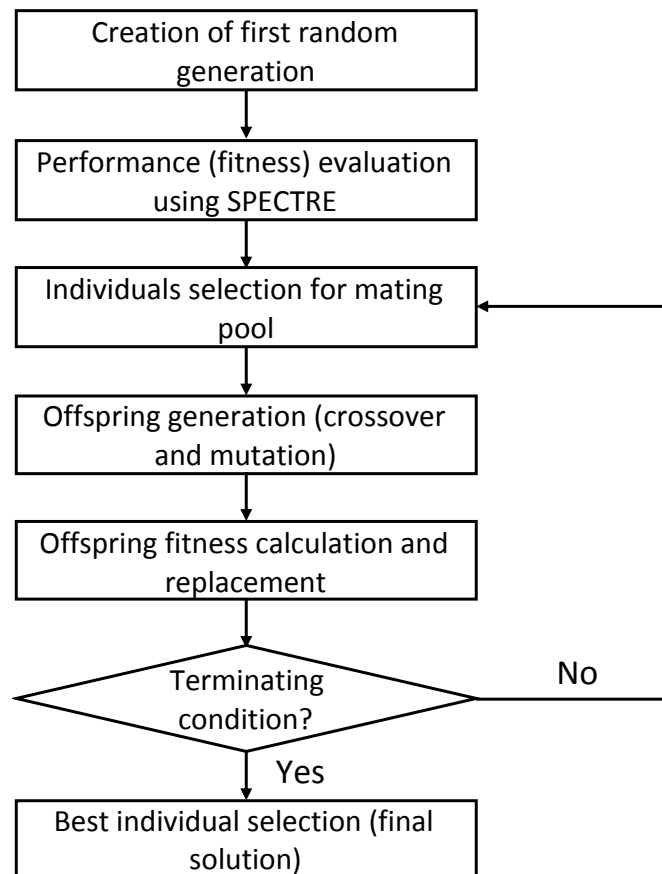


Figure 6.2: Block diagram of a generic GA for the optimization of D-latches.

Pareto-based GA was implemented).

The evaluation of the fitness is based on a set of circuit simulations. This means that when a new solutions is generated, the associated netlist is created and simulated for the characterization of its performance. Once the architecture of the circuit is chosen, as well as the driving strengths of the inputs and the capacitive load at the outputs, a solution of the problem can be defined as

$$\underline{x} = (W_1, L_1, W_2, L_2, \dots, W_N, L_N)^T \in I \quad (6.2)$$

where W_i and L_i are the width and length of the MOS transistors of the circuit and I is the space of the solutions in which these values are defined. It is worth to highlight that I is not a continuous subset of \mathbb{R}^{2N} because the size of a MOS in a given technology can not be set with infinite precision (in this case, it is 5 nm).

Once a new set of MOSFETs sizes \underline{x} is generated, a script will modify the netlist and call the Spectre simulator for the evaluation of the performance. For the design of D-latches, the parameters to be optimized include power consumption P_C , area A and setup time t_s . The latter can be split in t_{s-rise} and t_{s-fall} that indicate the setup time of the latch respectively for a positive or negative transition of the input. The setup time can be analyzed evaluating the behavior of the delay of the latch [126]. Assuming that the latch is in hold mode (the output does not change) when the gating signal G is 0 and transparent (the output follows the input) when G is 1, if the input signal D performs a 0-to-1 or 1-to-0 transition in a time window close to the falling edge of G then the propagation delay of the latch increases compared to its

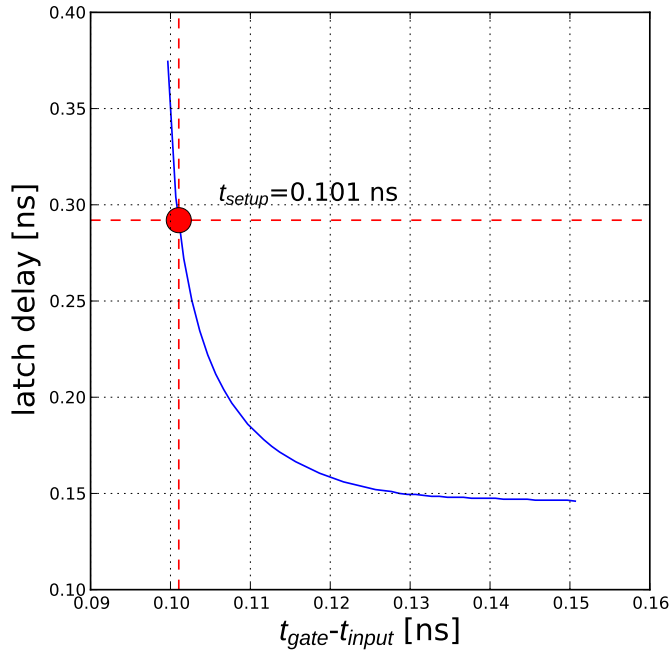


Figure 6.3: Setup time calculation for D-latches.

nominal value (eventually leading to metastability). For this reason, it is possible to evaluate t_s as the minimum time distance between the edge of D , t_{input} , and the falling edge of G , t_{gate} that makes the output delay larger than a certain threshold (see Figure 6.3). For the implementation of the GAs described in Section 6.3, this threshold was chosen equal to 2 times the delay of the latch when the circuit is not close to metastability $2t_d$ (i.e. when $t_{gate} - t_{input}$ is large). Finally, a set of constraints can be defined on all the above-mentioned parameters on the basis of the specifications of the design.

6.3 Chosen algorithms and implementation

In this Section four different implementations of evolutionary algorithms for the optimization of D-latches will be described and the performance of the final solutions will be analyzed. The design procedures are general and the characteristics of the GAs can be exploited for the enhancement of a large variety of analog or digital circuits.

6.3.1 GA with linear aggregation

The first evolutionary algorithm implemented for the design of D-latches is based on the minimization of a fitness $F(x)$ defined as in equation 6.1. In this example, as it will be shown later, F is mainly function of the setup times t_{s-rise} and t_{s-fall} defined in Section 6.2. The implemented algorithm is based on the so-called Generalized Generation Gap (G3) model, introduced in [127]. This model represents an optimized and a computationally faster variation of the Minimal Generation Gap (MGG) [127]. The algorithm starts with an initialization phase in which N individuals in I are randomly generated creating the set B . Each of the individuals will be associated to a fitness value. At this point, the model follows the steps:

- from the set B the best individual (smallest fitness), indicated with $BEST$ and other $m-1$ solutions are selected, obtaining the set P
- k offspring are generated from the set P , creating the set O
- c random individuals are then chosen from the set B , creating the set R
- the c individuals selected from set B in the previous step are replaced by the best c solutions of the group given by $(R \cup O)$.

In this implementation $N=5$, $m=3$, $k=2$ and $c=2$. The main difference between the G3 and the MGG model is on the last step: indeed, in a GA based on the MGG model, of the c individuals of the set B selected at the third step, one is replaced by the best of the set $(R \cup O)$ while the others with a group of solutions obtained with a roulette-wheel selection. The latter is a process in which individuals are selected with a probability that depends on the value of their fitness [122]. As stated before, this variation in the model makes the algorithm more efficient

and faster from a computational point of view.

This first algorithm was mainly focused on the optimization of the performance of the latch in terms of setup time. Indeed, the fitness function F was defined as

$$F = F(t_{s-rise}, t_{s-fall}) = \max(t_{s-rise}, t_{s-fall}) + \alpha |t_{s-rise} - t_{s-fall}|. \quad (6.3)$$

For this implementation of the algorithm α was chosen equal to 0.2 (arbitrary). Equation 6.3 defines an optimization process focused on the minimization of the maximum setup time and also on the reduction of the difference between t_{s-rise} and t_{s-fall} . However, since this parameter is multiplied to $\alpha < 1$, the most important factor that the algorithm will take into account for the minimization process will still be the maximum of the setup times.

As it is possible to see, power consumption P_C and area A are not included in Equation 6.3. However, the algorithm evaluates these parameters to check the feasibility of the solutions, i.e.

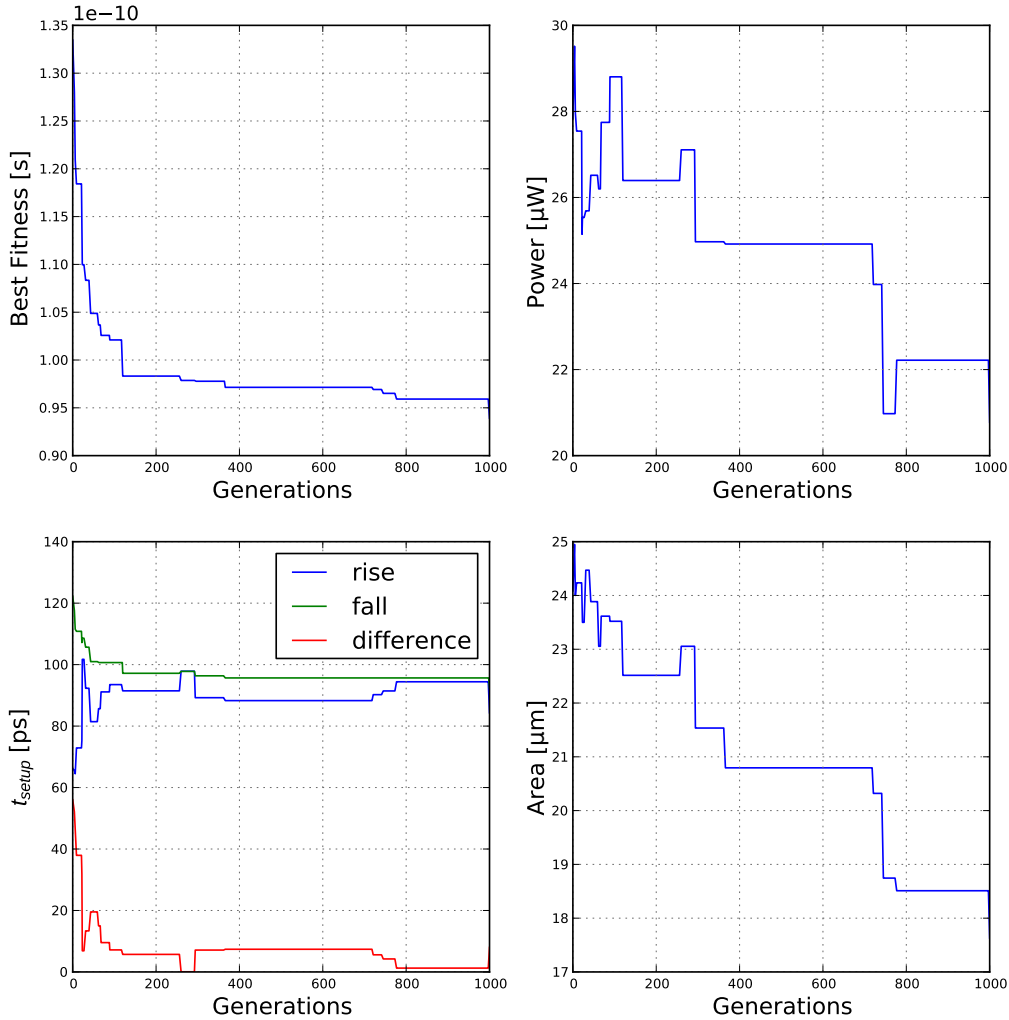


Figure 6.4: behavior of the best solution over generations in terms fitness, area, setup time and power consumption with the first algorithm based on G3 model.

if it respects a set of constraints on power and total area: when a new individual is generated if $P_C > P_{Cmax}$ or $A > A_{max}$ (or both) the associated fitness will be equal to

$$F = a\Omega, \quad (6.4)$$

where Ω is a large number with respect to usual values of F (in this implementation it is equal to 10^9 s) and a is the number of non-respected constraint. In a minimization problem, this technique allows easily avoiding the unfeasible solutions. The values of P_{Cmax} and A_{max} depends on the specifications of the system in which the circuits will be integrated.

The crossover operator, that will be the same also for the other algorithms described in this chapter, is based on a linear combination of the parents x_1 and x_2 to generate x_{son} in which each component x_{son-i} is given by

$$x_{son-i} = x_{1i} - r|x_{2i} - x_{1i}|, \quad (6.5)$$

with $r = 0.2$ and $F(x_1) < F(x_2)$. In this way the offspring will be more similar to the parent with better characteristics. The mutation operator is implemented adding to the offspring a vector w in which each component is a random Gaussian variable with a predetermined standard deviation σ equal to the range in which the individuals are defined divided by 4. For all of the algorithms presented in this chapter, this range stretches from 150 nm to 5 μ m. The termination condition is set to a fixed number of generations that in this case is equal to 1000. Figure 6.4 shows the performance parameters of the best solutions over generations. It is possible to see that the fitness F follows a monotonic behavior and reaches its best value around the 800-th generation. The setup time plot highlights the impact of the factor $|t_{s-rise} - t_{s-fall}|$ in the fitness function: at the beginning of the algorithm, t_{s-rise} was almost 40% smaller than its final value but it started to increase to follow the value of t_{s-fall} . The value of the latter has been reduced by 22.5% instead. There is also an improvement in terms of area (simply defined as the sum of the components of the solutions x) and power consumption. However, since the fitness is not function on these values (except in the case of constraint violation) this behavior is mainly dependent on the chosen architecture of the system. The area is also following a decreasing trend because smaller transistors are characterized by reduced parasitic capacitors.

6.3.2 Second implementation of a linear aggregation-based GA

A second version of the evolutionary algorithm based on the G3 model has been implemented. The main difference with the previous one is the way the fitness function is defined since, this time, F is also function of the power consumption and area. In this case, F was defined as

$$F = F(\max(t_{s-rise}, t_{s-fall}), P_C, A) = c_t \max(t_{s-rise}, t_{s-fall}) + c_P \frac{P_C}{c'_P} + c_A \frac{A}{c'_A}. \quad (6.6)$$

The weight parameters were set as $c_t=0.7$ and $c_P = c_A=0.15$ in order to mainly focus on the minimization of the setup time. The parameters c'_P and c'_A are only used to normalize power consumption and area at the same order of magnitude of the setup time. In Figure 6.5 the behavior of the performance of the best solutions is reported. Since the fitness function is not dependent only on the setup time but also on P_C and A , it is possible to notice that the values of t_{s-rise} and t_{s-fall} are worse with respect to the previous solution. In this case, the algorithm provides a final solution characterized by $t_{s-rise} = 114.6$ ps and $t_{s-fall} = 119.1$ ps while the previous implementation of the G3 model resulted in a final individual with $t_{s-rise} = 84.3$ ps and $t_{s-fall} = 92.3$ ps. However, it is possible to highlight an improvement in terms of power consumption and area: P_C and A are $11.56 \mu W$ and $4.38 \mu m$ respectively, a significant improvement with respect to the values obtained with the previous algorithm ($20.77 \mu W$ and $17.60 \mu m$).

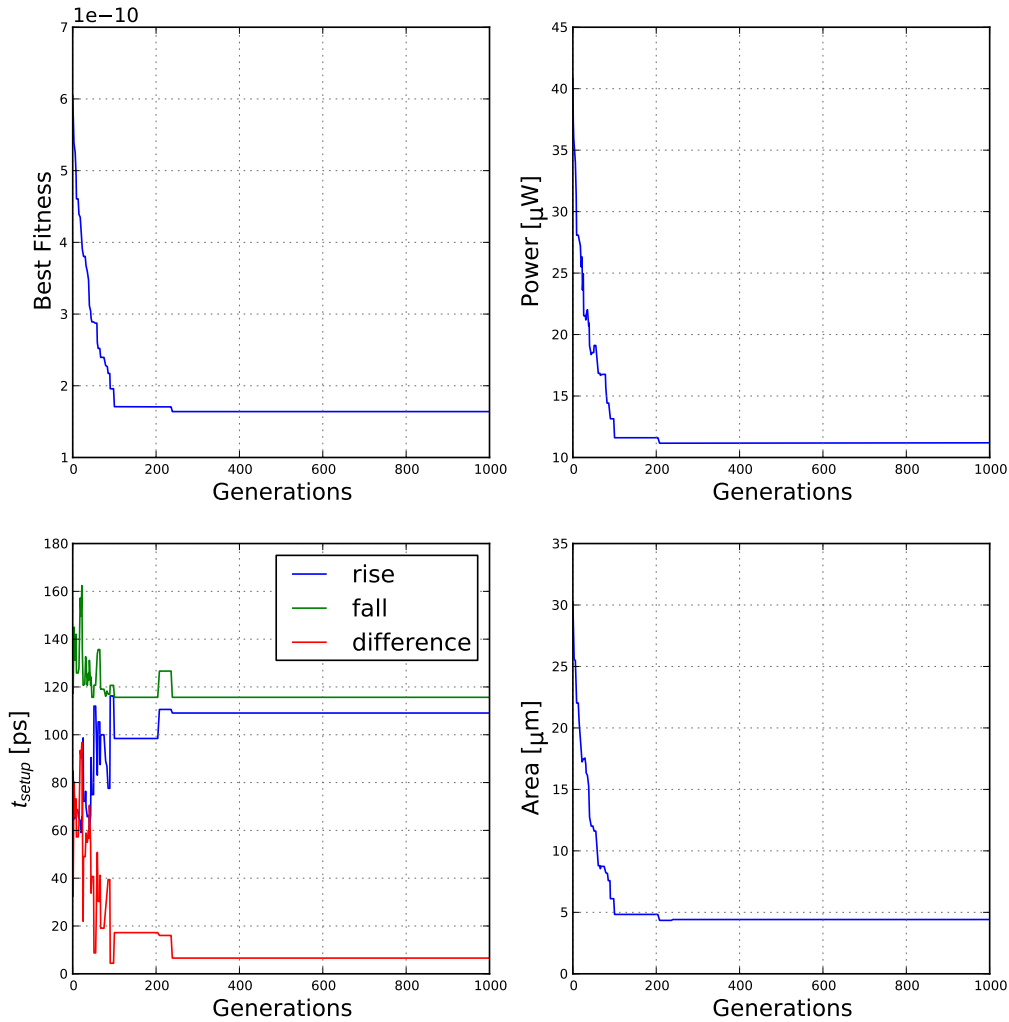


Figure 6.5: behavior of the best solution over generations in terms fitness, area, setup time and power consumption with the second algorithm based on G3 model. This time the fitness function directly includes P_C and A .

6.3.3 Third implementation of a linear aggregation-based GA

The mutation operator of the previous implementations of the GAs was simply based on adding to each component of the \underline{x}_{son} vector a random Gaussian variable with a predetermined standard deviation. However, this operator can be improved by analysing the distribution of the populations generation by generation. A third version of an evolutionary algorithm for SOOP has been implemented using the same G3 model explained before but the mutation operator is characterized by a different value of standard deviation σ_i for each component of \underline{x}_{son} and calculated as

$$\sigma_i = M\sigma(x_i^1, \dots, x_i^N) = M \sqrt{\frac{1}{N} \sum_{j=1}^N (x_i^j - \mu_i)^2}, \quad (6.7)$$

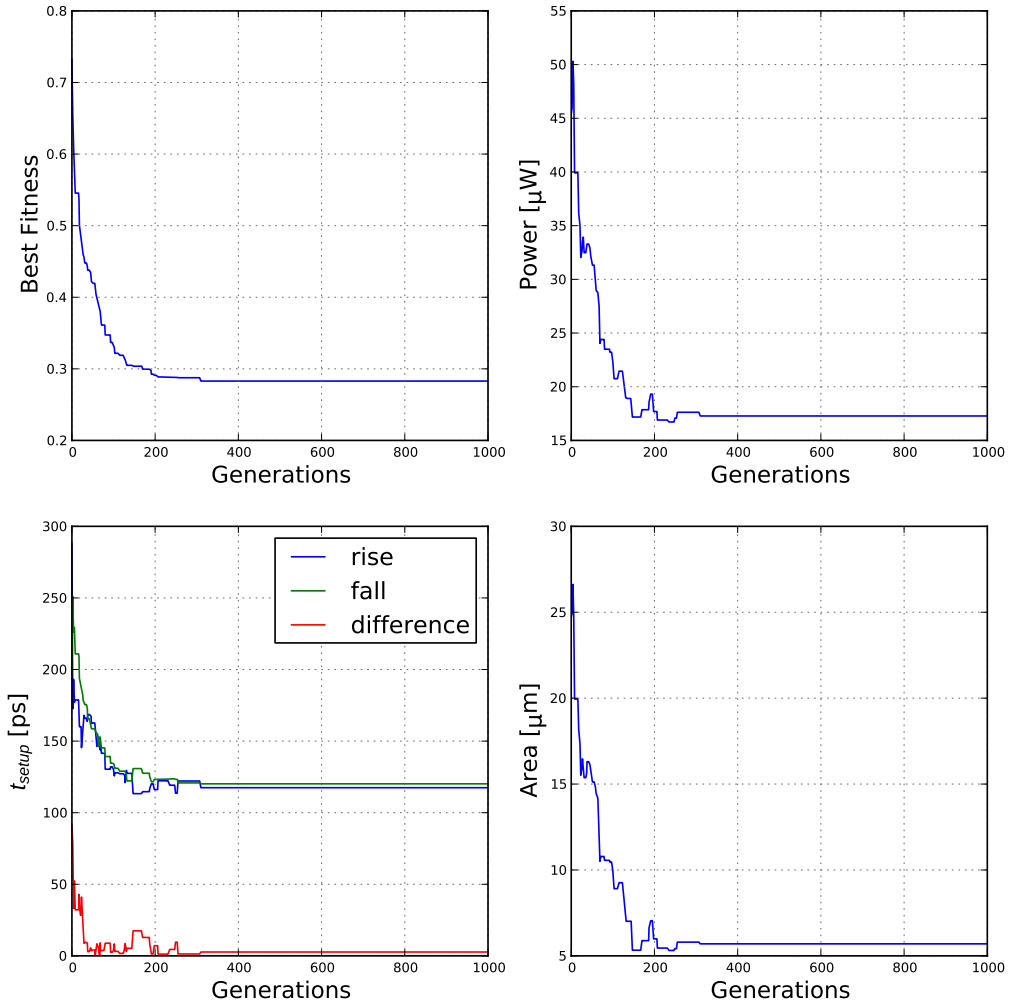


Figure 6.6: behavior of the best solution over generations in terms fitness, area, setup time and power consumption with the third algorithm based on G3 model with a different mutation operator.

where x_i^1, \dots, x_i^N indicate the i -th component of all the individuals in a certain generation and μ_i is their average. This means that the value of σ_i of the random variable to be added to the i -th element of the offspring vectors is proportional to the standard deviation of the respective components of the population of the set B . The parameter M has been set equal to 3 if the best solution does not change for the previous 100 generations and 1 otherwise. With this operator, a more efficient research of the best solutions in the objective space is performed because a larger variability is added to those components characterized by more heterogeneous values among the ones of the individuals of the set B . Moreover, the parameter M is added to increase the variability of all components if the best solution is not updated after more than 100 generations. In this way, if the algorithm found a local minimum, this procedure allows searching in a wider space to look for potential better solutions. The fitness function exploited for this version of the algorithm is similar to the one used before

$$F = F(\max(t_{s-rise}, t_{s-fall}), P_C, A) = c_t \frac{\max(t_{s-rise}, t_{s-fall})}{T_R} + c_P \frac{P_C}{P_R} + c_A \frac{A}{A_R}, \quad (6.8)$$

where the parameters T_R , P_R and A_R represent the maximum acceptable values of setup time, power and area respectively. Figure 6.6 shows the behavior of the performance of the best solutions over generation considering $c_t = 0.7$ and $c_P = c_A = 0.15$ (as done before). The best solution shows similar performance to the previous version of the algorithm since $t_{s-rise} = 115.6$ ps, $t_{s-fall} = 118.3$ ps, $P_C = 17.2$ μ W and $A = 5.7$ μ m. In Section 6.4 a more detailed comparison of the results of described algorithms will be presented.

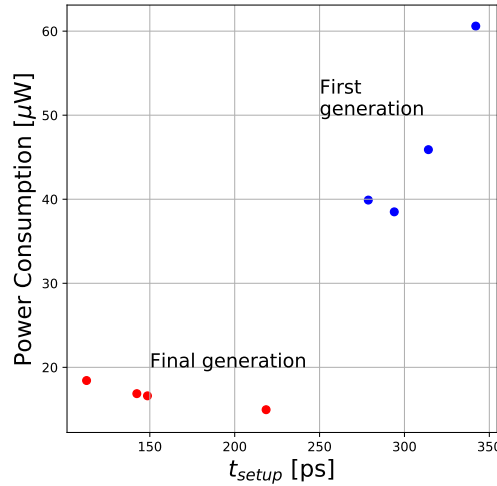


Figure 6.7: Comparison between the first and last generation of the Pareto-based evolutionary algorithm. For the sake of simplicity, only power consumption and setup time have been considered in the plot.

6.3.4 Pareto-based evolutionary algorithm

In this Subsection, an evolutionary Pareto-based algorithm for the optimization of D-latch will be described. In this example, the objectives to be minimized are $\max(t_{s-rise}, t_{s-fall}, P_C$ and A . The procedure used for this GA is based on the Strength Pareto Evolutionary Algorithm 2 (SPEA2) that was introduced in [128]. This approach represents an upgrade of the Strength Pareto Evolutionary Algorithm (SPEA), developed in 1999. As introduced in Section 6.1, GAs for the research of the Pareto-optimal front require the concept of dominance to evaluate and compare the generated solutions. In SPEA2, each individual \underline{x} is associated to a function $F(\underline{x})$ that can be indicated as fitness and it is given by

$$F(\underline{x}) = R(\underline{x}) + D(\underline{x}). \quad (6.9)$$

$R(\underline{x})$ is the raw fitness and gives an indication of the dominance of the solution in a certain population P . It is calculated as

$$R(\underline{x}) = \sum_{\underline{y} \in P | \underline{y} \prec \underline{x}} S(\underline{y}), \quad (6.10)$$

where $S(\underline{y})$ is the so-called strength of the solution \underline{y} given by

$$S(\underline{y}) = \{ \underline{x} | \underline{x} \in P \wedge \underline{y} \prec \underline{x} \}. \quad (6.11)$$

Equation 6.10 and 6.11 indicate that the raw fitness of \underline{x} returns the sum of the number of solutions dominated by the individuals that dominate \underline{x} . This means that if \underline{x} is a dominant solution its raw fitness is equal to 0.

The second component of the total fitness, $D(\underline{x})$, is the density function of the solution \underline{x} , calculated as

$$D(\underline{x}) = \frac{1}{d_n(\underline{x}) + 2}, \quad (6.12)$$

where $d_n(\underline{x})$ is the distance in the objective space from the parameters of the solution \underline{x} and the n -th nearest neighbour. In this implementation the value of n is 2. The combination of $R(\underline{x})$ and $D(\underline{x})$ defines a fitness function $F(\underline{x})$ that is able to take into account both the dominance of the solution \underline{x} and its distance from other ones inside the objective space in order to obtain a final population that gives a good description of the Pareto-optimal front. Equations 6.10 and 6.12 also highlights that dominant solutions are characterized by a fitness $F(\underline{x}) < 1$.

The steps of the SPEA2 procedure implemented in this work are also based on the algorithm proposed in [129]. These steps are summarized as follows:

- in the initialization phase, 2 sets of solutions are generated: P_0 , indicated as main population and with a length of $N = 8$, and \bar{P}_0 , indicated as archive set and with a length of $\bar{N} = 4$. Fitness values of all the individuals are calculated;
- dominant solutions ($F(\underline{x}) < 1$) are chosen to create a new archive set \bar{P}_1 . The length

of this new set must be \overline{N} . For this reason the environmental selection process is implemented: if the number of dominant solutions is different from \overline{N} the archive needs to be either populated with other individuals from $P_0 \cup \overline{P}_0$ or truncated. In the first case, the best solutions in terms of fitness are chosen. In the second case, some dominant solution has to be deleted from the set. This process is implemented analysing the minimum distances of each individual from the closest $d_1(\underline{x})$. If some of this distances are equal, $d_2(\underline{x})$ are analyzed and so on;

- 2 offspring solutions are generated by crossover and mutations of 2 couples of parents chosen with binary tournament selection: at each stage of this process, 2 random solutions are chosen and from those the best one is selected [122]. Implementing this algorithm 4 times will provide the parents to generate two son solutions \underline{x}_{son1} and \underline{x}_{son2} ;
- the worst 2 individuals in set P_0 , \underline{x}_1 and \underline{x}_2 are selected in order to generate the set; $G = \{\underline{x}_{son1}, \underline{x}_{son2}, \underline{x}_1, \underline{x}_2\}$. The strength $S(\underline{x})$ will be calculated for all the individuals of this set.
- the best 2 solutions of set G will be used to replace \underline{x}_1 and \underline{x}_2 in set P_0 , creating the set P_1 ;
- the index of the sets are increased by 1 and the algorithm restarts from the second step.

The archive set at the last generation \overline{P}_{FINAL} represents the final set of solutions of the problem. It is important to highlight that, as explained in [128], SPEA2 represents an efficient improvement of the original SPEA especially because of the introduction of the density function $D(\underline{x})$ of Equation 6.12. Indeed, it is possible that a certain set is composed by similar or equal solutions in terms of dominance. The absence of a function that is able to analyze the density of the individuals of a certain population in the objective space can potentially lead to situations in which the selection processes are purely random, decreasing the efficiency of the optimization process.

In Figure 6.7, a comparison among the solutions of the first archive set and the final is reported (for the sake of a better visualization, only power consumption and setup time are considered). The plot shows a significant improvement of the performance and a sort of migration of the population toward a zone of the objective space in which it is possible to notice the characteristic inverse proportionality between power consumption and setup time. As for the previous implementation, the total number of generation was set equal to 1000.

6.4 Discussion and comparisons

Table 6.1 reports the performance of the latches obtained with the presented evolutionary algorithms. It is possible to notice that the best values in terms of setup time are obtained with the first version of the algorithms based on the G3 model. As anticipated before, this behavior is explained considering that, in this case, the fitness to minimize is only function of

Table 6.1: Comparison of the presented GAs

Algorithm	G3 ¹	G3 ²	G3 ³	SPEA2			
	Best solution			Sol. 1	Sol. 2	Sol. 3	Sol. 4
t_{s-rise} [ps]	84.3	114.6	115.6	103.4	135.9	142.4	135.6
t_{s-fall} [ps]	92.3	119.1	118.3	112.7	142.3	148.6	218.5
$ t_{s-rise} - t_{s-fall} $ [ps]	8.0	4.5	2.7	8.8	6.4	6.2	82.9
P_C [μ W]	20.8	11.6	17.2	18.4	16.9	16.6	14.9
A [μ m]	17.6	4.4	5.7	4.1	3.1	3.2	1.9

¹ P_C and A are considered only in the constraints

² Linear aggregation fitness including P_C and A

³ Linear aggregation fitness including P_C and A , different mutation operator

the timing characteristics of the latches while P_C and A are only limited by the constraints of the problem. Better performance in terms of power and area are obtained with the second and third versions of the algorithm since these parameters are directly included in the fitness function. However, solving this problem a linear aggregation of t_s , P_C and A for the fitness function requires a careful definition of the weights of the single objectives and (c_t , c_P and c_A in this case) and also the normalization parameters. For this reason, as highlighted in Table 6.1, Pareto-based evolutionary algorithms as SPEA2, despite the higher computational complexity for the calculation of a dominance and density-based fitness, may be considered as an efficient methodology for problems like the one proposed in this work. Indeed, the solutions obtained in the last generation are characterized by a heterogeneous set of performance parameters, handling the trade-offs among power, area and setup time in different ways. This gives the possibility to obtain a group of final solutions that can be used for various implementations of the circuit on the basis of requests and specifications.

It is worth highlighting that the algorithms proposed in this chapter are general and they can be applied to a large variety of analog and digital circuit optimization processes, exploiting the same methodology described in Section 6.2. However, in optimization problems in which there are many parameters to take into consideration an exhaustive analysis of the objective space can be complicated or impossible. In this case, in order to avoid the algorithm to search only in neighbouring space around a potential local minima of the defined fitness, the implementation of several algorithms based on different search criteria can be useful. An alternative solution is based on adopting genetic algorithms featuring mechanisms to avoid local minima of the fitness, e.g. simulated annealing [122]. This kind of solutions can be added to GAs like the ones presented in this work.

Conclusions

This thesis focused on time-resolved particle detectors for a large spectrum of applications, mainly HEP and medical imaging. In particular, the particle physics community is setting a new trend based on the development of high-luminosity particle colliders (as HL-LHC and the planned FCC at CERN). High-luminosity accelerators require new tracking systems able to cope with intense collision rates and the consequent increase of pile-up. An efficient reconstruction of the events inside the colliders can be performed with high-resolution timing detectors that give the possibility to improve tracking accuracy, reduce background noise and increase pile-up suppression. Timing detectors can also be exploited for the enhancement of the performance of PET scanners. The SNR of the reconstructed PET image is directly affected by time resolution of detectors installed in the system. This allows lowering the dose of radioisotopes injected in the body under exam, thus decreasing the radiation exposure of patients and personnel.

This thesis attempted to describe the numerous challenges related to the design of high-performance electronics used for different types of timing detectors. Various ASICs were developed and fabricated in a 130 nm SiGe BiCMOS technology. The advantages of SiGe HBTs were exploited to obtain fast and low-noise front-end systems designed for HEP and medical imaging detectors such as the ones related to FASER and the TT-PET project.

The development of ASICs for the FASER experiment detecting system required a preliminary study on the front-end and the effects of different levels of integration inside the pixel. Measurements showed that the integration of the entire front-end inside the active area represents a feasible solution for the stability of the architecture and the specifications of the FASER experiment. In particular, unwanted self-induced oscillation could be prevented with a careful layout design that aimed to increase the coupling of those output nodes that showed a negative feedback with the pixel in which the system was integrated. Measurements with a laser source showed that this front-end, when the switching system is activated, is able to significantly increase the TOT dynamic range and reduce its relative jitter. This solution is also characterized by a <90 ps jitter on the ToA. Because of its performance, this architecture will be further investigated for future ASICs.

A description of the pre-production chip of the FASER experiment was provided. The main challenges were related to the design of a large ASIC (the final chip will occupy an area of 23.2×15.3 mm² area) which had to satisfy strict specifications in terms of dead-area ($<3\%$).

Because of this, the physical implementation of the super-column logic required a deep analysis of the characteristics of the architecture. The digital design tool was able to complete the routing of the system only by forcing the floorplan according to the periodicity of the sub-blocks of the super-column logic. In the FASER ASICs, the charge associated to each pixel is needed to be recorded and acquired for a good reconstruction of the cluster of hits. Because of the above-mentioned dead-area specifications, the implementation of an analog memory system for the measurement of the input charge was designed. A logic integrated in the central part of the super-column was designed to avoid unwanted charge sharing problems. Finally, the digital periphery of the system features a readout block characterized by a frame-based architecture with a super-pixel level compression. This solution allows a significant reduction of the average amount of data to readout compared with the case of a packet-based architecture. The latter, would have also required the implementation of a thicker super-column logic.

Two RO-based TDCs were presented. The first was designed in the context of the TT-PET project. This converter is characterized by a ~ 33 ps LSB, a ~ 3 mW power consumption and is featuring a 9 pseudo-differential stages RO in pseudo-NMOS architecture. A feedforward system was designed to increase the frequency of the oscillator (simulations showed a $\sim 30\%$ improvements). An event-by-event calibration system is able to measure, through an external reference signal, the free-running frequency of the RO each time a measurement needs to be performed. This system is useful to avoid any kind of PLL-based synchronization system. Moreover, a non-linearity model was developed to analyse the proposed TDC. The model demonstrated that the introduction of the output buffers inside the feedforward path is able to reduce the impact of their mismatch on the linearity of the converter. This solution did not add any complexity to the architecture but only required a different distribution of the internal connections. The above-mentioned characteristics and the compact area of the TDC make the proposed solution suitable for the many applications in which the integration of a large number of converters is often required.

The TDC for the FASER experiment ASIC was designed to satisfy a $\text{LSB} < 200$ ps specification. A 7 CMOS stages RO was chosen for the implementation of the converter. The design of the TDC was focused on the reduction of power consumption when the system is in idle (i.e., no event is detected) through a gating system. The measurements of the converter showed that the specifications in terms of resolution, linearity and power consumption were respected. For this reason and the compactness of the architecture, the presented TDC will be integrated in the final chip for the FASER experiment detector.

An IC for hybrid pixel detectors was designed and produced. Most of the thesis was focused on the development of dies for monolithic detectors, integrating pixels and electronic front-end in the same chip. However, the ASIC for hybrid detectors aims to work as a test chip for the evaluation of the limits of the electronic systems designed for other experiments without all the constraints related to monolithic architectures. Therefore, the architecture was designed focusing on the flexibility and the testability of its components. Four different front-end

Conclusions

flavours were integrated and a power-gating system allows an independent activation of the macro-columns of the matrix. In this way, it is possible to push the performance of certain sub-systems without compromising the rest of the ASIC. Considering sensors characterized by ~ 80 fF capacitances and input charges $Q_{in} > 0.6$, all of the front-end configurations are expected to have a $\sigma_{elec} < 30$ ps, where σ_{elec} is the contribution of the electronic front-end on the time-resolution of the system. The readout logic is based on a packet-based architecture that is able to provide only the data associated to the pixels that sensed an event.

This thesis also illustrated how to use GAs for the optimization and the tuning of analog circuits. Various algorithms were presented and compared for the implementation of low-setup times D-latches. However, this approach can be used for the design of a large variety of circuits. Linear aggregation GAs produce the best results in terms of setup times depending on the weight of this parameter in the fitness function to optimize. Pareto-based algorithms are able to provide a heterogeneous set of final solutions that are handling the trade-offs among the parameters to optimize in different ways. This allows obtaining a group of various implementations of the circuit that can be used for different purposes on the basis of the specifications.

A Ionization and Bethe-Bloch formula

The detection of particles is performed with an interaction of the latter with the sensing system. When a particle crosses a certain material, it can interact with it through an energy exchange that in the case of charged particles can lead to the ionization of the atoms of the detecting medium.

It is possible to analyse the behaviour of a ionizing particle crossing a certain material through the Bethe-Bloch formula [1]. The latter expresses the behaviour of the average energy loss (also indicated as stopping power) of a particle that passes through a material. The energy loss

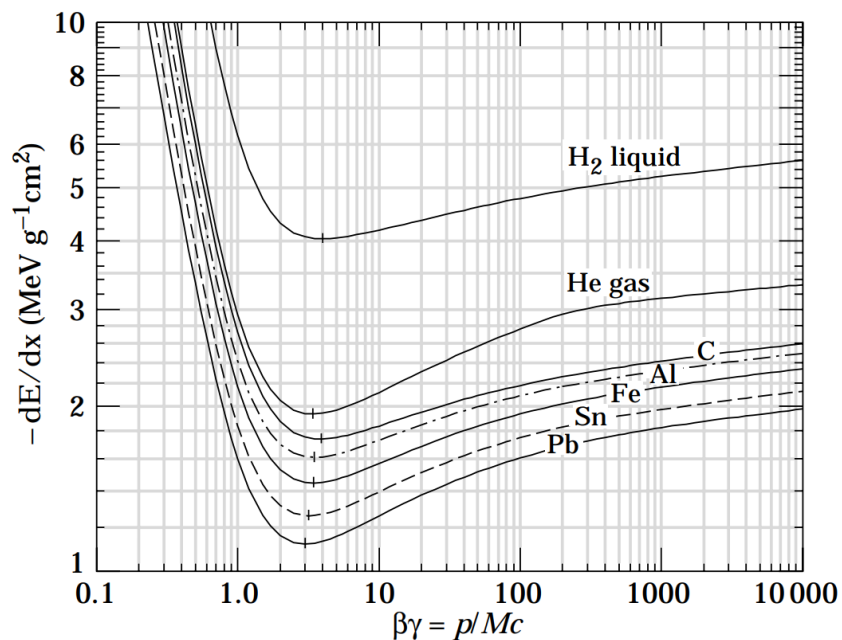


Figure A.1: Average stopping power (energy loss) in various materials (from top: liquid hydrogen, helium, carbon, aluminium, iron, tin and lead). Source: [130].

Appendix A. Ionization and Bethe-Bloch formula

can be obtained resolving the following [1]

$$-\left\langle \frac{dE}{dx} \right\rangle = n \int_{T_{min}}^{T_{max}} T \frac{d\sigma_{mat}}{dT} dT \quad (A.1)$$

where $\langle dE/dx \rangle$ indicates the average energy variation (loss) of the particle inside the material, n is the number density of the target (ration between number of particles in the medium and its volume), T is the transferred kinetic energy in the collision and σ_{mat} is the cross section of the crossing material. Its derivative in T is function of the characteristics of the medium, the mass of the incoming particles M and their speed β , $d\sigma_{mat}/dT = d\sigma_{mat}/dT(M, \beta, T)$. As described in [1], it is possible to calculate $\langle dE/dx \rangle$ resolving the integral for different energy transfer ranges (comparable or not to the binding energy of the electrons in the material) obtaining the Bethe-Bloch formula

$$-\left\langle \frac{dE}{dx} \right\rangle = \rho K \left(\frac{z}{\beta} \right)^2 \frac{Z}{A} \left[\frac{1}{2} \ln \left(\frac{2m_e c^2 (\beta\gamma)^2 T_{max}}{I^2} \right) - \beta^2 - \frac{\delta(\beta\gamma)}{2} - \frac{C(\beta\gamma, I)}{Z} \right]. \quad (A.2)$$

where z is the charge of the incoming particle, γ is the Lorentz factor, Z is the atomic number of the material, A its atomic mass, $\delta(\beta\gamma)$ and C/Z represent two correction factors relevant for high energies and low values of particle speed β respectively. As explained in [1] [26] [130], a minimum of the the stopping power is obtained for particles with $\beta\gamma \approx 3 - 4$ usually indicated as Minimum Ionizing Particles (MIPs). For higher values of $\beta\gamma$ the energy loss approximately follows a logarithmic behaviour and the ionizing process is the dominant one in the particle-medium interaction, as also shown in Figure A.1. From this picture, it is also possible to highlight that the value of $\beta\gamma \approx 3 - 4$ for the MIPs is valid for a large range of materials. However, the values of stopping powers are depending on the characteristics of the materials like density. A more detailed analysis of this formula can be found in textbooks [1] [56] and papers [130]. In particular, Equation A.2 is mainly valid for heavy particles (i.e. protons) while, in the case of electrons or positrons, the energy loss needs to be corrected due to the influence of their spin, kinematics and other characteristics (like $e^- e^+$ annihilation) [1]. Most of the detectors developed to sense charged particles are exploiting the ionization process to generate electrical signals readable by the readout electronics. The Bethe-Bloch formula represents a fundamental starting point for the optimization of the sensor characteristics.

B The Shockley-Ramo theorem

The generation of the electrical signal of a silicon detector for ionizing radiation sent to the readout electronics can be modeled with the outcomes of the Shockley-Ramo theorem [131] [132]. The latter is useful to analyse a situation like the one depicted in Figure B.1. Considering a charge q moving according to the velocity vector \underline{v} in a space surrounded by N metallic electrodes (in the example of Figure B.1 $N=5$). The movement of the charge will induce a current in the m -th electrode i_m being described by the following relation

$$i_m = -q \underline{v} \cdot \underline{E}_{wm} \quad (\text{B.1})$$

where \underline{E}_{wm} is the so-called weighting field and it represents the electric field obtained setting to 1 the potential of the m -th electrode and to 0 the potential of the others (a detailed treatment of the theorem and the analysis of the results is reported in [133]). This means that, in a sensor, when the ionization occurs, the movement of the charges induces a current proportional to their speed. Equation B.1 also gives important hints for the optimization of sensors. As underlined in [26], the timing performance of a detector can be improved designing a system in which the response to an event is as independent as possible on the position in which the

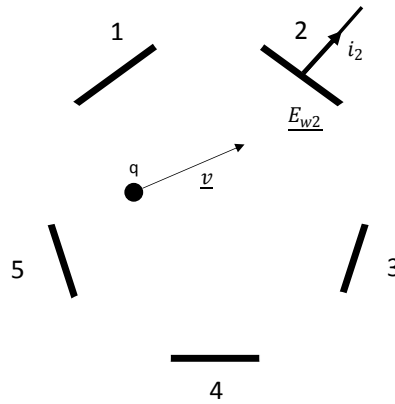


Figure B.1: Representation of the problem solved by the Shockley-Ramo theorem.

Appendix B. The Shockley-Ramo theorem

ionization process started. The goal can be achieved designing a sensor in which the weighting fields are approximately constant and parallel to the direction in which the charges move, like a planar and thin detector with large electrodes. In this case, it is possible to approximate the induced current of Equation B.1 with [26]

$$i = -q \frac{|\nu|}{D} \quad (\text{B.2})$$

where D is the width of the detector system (i.e., the distance between the electrodes). From Equation B.1 and B.2, it is possible to highlight the role of charge velocity in the induced current: higher values of ν can improve improve the timing performance of the sensor. However, the drift velocity of electrons and holes in silicon tends to saturate at electric fields close to $\approx 2 \text{ V}/\mu\text{m}$. Saturated sensors represents a good solution for the optimization of the time resolution of a detecting system [26].

C Chip Gallery



Figure C.1: FASER first prototype ASIC.

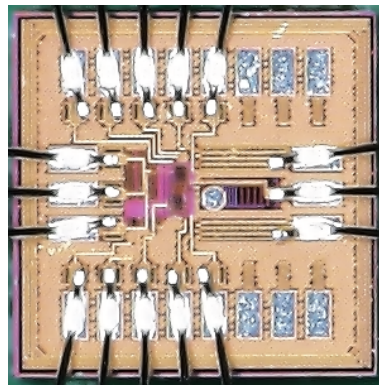


Figure C.2: Feedforward TDC test-chip.



Figure C.3: Layout of the three FASER pre-production ASIC variants.

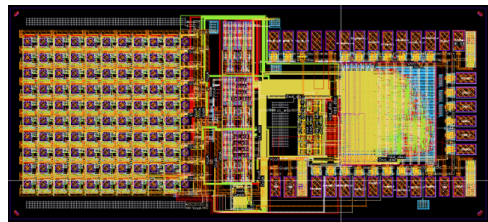


Figure C.4: Readout ASIC layout.

Bibliography

- [1] H. Kolanoski and N. Wermes. *Particle Detectors: Fundamentals and Applications*. Oxford University Press, USA, 2020.
- [2] G. Aad et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Physics Letters B* 716.1 (2012), pp. 1–29.
- [3] J. T. Bushberg and J. M. Boone. *The essential physics of medical imaging*. Lippincott Williams & Wilkins, 2011.
- [4] J. Beutel, H. L. Kundel, and R. L. Van Metter. *Handbook of medical imaging*. Vol. 1. Spie Press, 2000.
- [5] R. W. Brown et al. *Magnetic resonance imaging: physical principles and sequence design*. John Wiley & Sons, 2014.
- [6] J. Rossignol et al. “Time-of-Flight Computed Tomography a Proof of Principle Study”. In: *Journal of Nuclear Medicine* 60.supplement 1 (2019), pp. 47–47.
- [7] M. E. Phelps. *PET: molecular imaging and its biological applications*. Springer Science & Business Media, 2004.
- [8] D. Hayakawa. “Development of the Thin TOF-PET scanner based on fast timing monolithic silicon pixel sensors”. PhD thesis. University of Geneva, 2020.
- [9] R. Duara et al. “Positron emission tomography in Alzheimer’s disease”. In: *Neurology* 36.7 (1986), pp. 879–879.
- [10] S. Strother, M. Casey, and E. Hoffman. “Measuring PET scanner sensitivity: relating countrates to image signal-to-noise ratios using noise equivalents counts”. In: *Ieee transactions on nuclear science* 37.2 (1990), pp. 783–788.
- [11] M. Conti. “Focus on time-of-flight PET: the benefits of improved time resolution”. In: *European journal of nuclear medicine and molecular imaging* 38.6 (2011), pp. 1147–1157.
- [12] E. Ripicini et al. “Expected performance of the TT-PET scanner”. In: *arXiv preprint arXiv:1811.12381* (2018).
- [13] M. Conti et al. “First experimental results of time-of-flight reconstruction on an LSO PET scanner”. In: *Physics in Medicine & Biology* 50.19 (2005), p. 4507.

- [14] B. Jakoby et al. "Physical and clinical performance of the mCT time-of-flight PET/CT scanner". In: *Physics in Medicine & Biology* 56.8 (2011), p. 2375.
- [15] J. W. Cates and C. S. Levin. "Evaluation of a clinical TOF-PET detector design that achieves ≤ 100 ps coincidence time resolution". In: *Physics in Medicine & Biology* 63.11 (2018), p. 115011.
- [16] S. Gundacker et al. "High-frequency SiPM readout advances measured coincidence time resolution limits in TOF-PET". In: *Physics in medicine and biology* (2019).
- [17] P. Lecoq. "Pushing the limits in time-of-flight PET imaging". In: *IEEE Transactions on Radiation and Plasma Medical Sciences* 1.6 (2017), pp. 473–485.
- [18] P. Lecoq et al. "Roadmap toward the 10 ps time-of-flight PET challenge". In: *Physics in Medicine & Biology* 65.21 (2020), 21RM01.
- [19] R. Beeson and K. T. Network. "Time of Flight Positron Emission Tomography The potential benefits of early engagement between CERN and the TOF-PET industry." In: (2017).
- [20] L. Paolozzi et al. "Test beam measurement of the first prototype of the fast silicon pixel monolithic detector for the TT-PET project". In: *Journal of Instrumentation* 13.04 (2018), P04015.
- [21] P. Valerio et al. "A monolithic ASIC demonstrator for the Thin Time-of-Flight PET scanner". In: *arXiv preprint arXiv:1811.10246* (2018).
- [22] Y. Bandi et al. "The TT-PET project: a thin TOF-PET scanner based on fast novel silicon pixel detectors". In: *Journal of Instrumentation* 13.01 (2018), p. C01007.
- [23] G. Iacobucci. *TT-PET: Thin Time-of-Flight PET with depth of interaction measurement capability based on very-low noise Silicon-Germanium BJT electronics and semiconductor detector*. URL: <http://p3.snf.ch/project-160808>.
- [24] L. Paolozzi et al. "Characterization of the demonstrator of the fast silicon monolithic ASIC for the TT-PET project". In: *Journal of Instrumentation* 14.02 (2019), P02009.
- [25] L. Rossi et al. *Pixel detectors: From fundamentals to applications*. Springer Science & Business Media, 2006.
- [26] L. Paolozzi. "Development of particle detectors and related Front End electronics for sub-nanosecond time measurement in high radiation environment." PhD thesis. Rome U., 2015.
- [27] L. Evans and P. Bryant. "LHC machine". In: *Journal of instrumentation* 3.08 (2008), S08001.
- [28] J. Va'vra. "Picosecond timing detectors and applications". In: *Journal of Physics: Conference Series*. Vol. 1498. 1. IOP Publishing. 2020, p. 012013.
- [29] J. Va'vra. "Picosecond timing detectors and applications". In: *arXiv preprint arXiv:1906.11322* (2019).

Bibliography

- [30] G. Apollinari et al. *High-luminosity large hadron collider (HL-LHC): Preliminary design report*. Tech. rep. Fermi National Accelerator Lab.(FNAL), Batavia, IL (United States), 2015.
- [31] *Quantum Diaries*. URL: <https://www.quantumdiaries.org/tag/pile-up/>.
- [32] A. Collaboration. *Technical Proposal: A High-Granularity Timing Detector for the ATLAS Phase-II Upgrade*. Tech. rep. 2018.
- [33] H. F. Sadrozinski, A. Seiden, and N. Cartiglia. “4D tracking with ultra-fast silicon detectors”. In: *Reports on Progress in Physics* 81.2 (2017), p. 026101.
- [34] G. A. Rinella et al. “The NA62 GigaTracker: a low mass high intensity beam 4D tracker with 65 ps time resolution on tracks”. In: *Journal of Instrumentation* 14.07 (2019), P07010.
- [35] E. C. Gil et al. “The Beam and detector of the NA62 experiment at CERN”. In: *Journal of instrumentation* 12.05 (2017), P05025.
- [36] F. Zimmermann et al. *Future circular collider*. Tech. rep. CERN-ACC-2018-0059, 2018.
- [37] M. Benedikt and F. Zimmermann. *Status of the future circular collider study*. Tech. rep. FCC-DRAFT-ACC-2016-030, 2016.
- [38] G. Borghello. “FCC radiation environment: an unprecedented challenge for MOS transistors”. In: *FCC week 2018* (2018).
- [39] F. Martinelli et al. “Measurements and analysis of different front-end configurations for monolithic SiGe BiCMOS pixel detectors for HEP applications”. In: *Journal of Instrumentation* 16.12 (Dec. 2021), P12038. DOI: 10.1088/1748-0221/16/12/p12038.
- [40] J. L. Feng et al. “Axionlike particles at FASER: The LHC as a photon beam dump”. In: *Physical Review D* 98.5 (2018), p. 055021.
- [41] A. Ariga et al. “Technical proposal for FASER: forward search experiment at the LHC”. In: *arXiv preprint arXiv:1812.09139* (2018).
- [42] J. L. Feng et al. “Dark Higgs bosons at the forward search experiment”. In: *Physical Review D* 97.5 (2018), p. 055034.
- [43] J. L. Feng et al. “ForwArd search ExpeRiment at the LHC”. In: *Physical Review D* 97.3 (2018), p. 035001.
- [44] K. Ehret et al. “New ALPS results on hidden-sector lightweights”. In: *Physics Letters B* 689.4-5 (2010), pp. 149–155.
- [45] *Medipix Collaboration*. URL: <https://medipix.web.cern.ch/home>.
- [46] R. Ballabriga, M. Campbell, and X. Llopart. “Asic developments for radiation imaging applications: The medipix and timepix family”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 878 (2018), pp. 10–23.

-
- [47] J. Carreira et al. "Direct integration of micro-LEDs and a SPAD detector on a silicon CMOS chip for data communications and time-of-flight ranging". In: *Optics express* 28.5 (2020), pp. 6909–6917.
- [48] A. Cavalleri et al. "Femtosecond melting and ablation of semiconductors studied with time of flight mass spectroscopy". In: *Journal of Applied Physics* 85.6 (1999), pp. 3301–3309.
- [49] N. Moffat et al. "Low Gain Avalanche Detectors (LGAD) for particle physics and synchrotron applications". In: *Journal of Instrumentation* 13.03 (2018), p. C03014.
- [50] F. Zappa et al. "Principles and features of single-photon avalanche diode arrays". In: *Sensors and Actuators A: Physical* 140.1 (2007), pp. 103–112.
- [51] S. Gundacker and A. Heering. "The silicon photomultiplier: fundamentals and applications of a modern solid-state photon detector". In: *Physics in Medicine & Biology* 65.17 (2020), 17TR01.
- [52] S. M. Sze, Y. Li, and K. K. Ng. *Physics of semiconductor devices*. John Wiley & sons, 2021.
- [53] S. Mandai and E. Charbon. "A $4 \times 4 \times 416$ digital SiPM array with 192 TDCs for multiple high-resolution timestamp acquisition". In: *Journal of Instrumentation* 8.05 (2013), P05024.
- [54] W. U. Boeglin. *Scintillation Detector — Modern Lab Experiments documentation*.
- [55] N. I. of Standards and T. (NIST). *XCOM: Photon Cross Sections Database*. URL: <https://www.nist.gov/pml/xcom-photon-cross-sections-database>.
- [56] G. F. Knoll. *Radiation detection and measurement*. John Wiley & Sons, 2010.
- [57] C. Dujardin et al. "Needs, trends, and advances in inorganic scintillators". In: *IEEE Transactions on Nuclear Science* 65.8 (2018), pp. 1977–1997.
- [58] B. Wang et al. "Novel light-guide-PMT geometries to reduce dead edges of a scintillation camera". In: *Physica Medica* 48 (2018), pp. 84–90.
- [59] H. Hirayama. "Lecture note on photon interactions and cross sections". In: *KEK, High Energy Accelerator Research Organization, Oho, Tsukuba, Ibaraki, Japan* (2000).
- [60] A. Owens et al. "High resolution x-ray spectroscopy using GaAs arrays". In: *Journal of Applied Physics* 90.10 (2001), pp. 5376–5381.
- [61] T. Takahashi and S. Watanabe. "Recent progress in CdTe and CdZnTe detectors". In: *IEEE Transactions on nuclear science* 48.4 (2001), pp. 950–959.
- [62] E. Bossini and N. Minafra. "Frontiers: Diamond Detectors for Timing Measurements in High Energy Physics". In: *Front. Phys.* 8 (2020), p. 248.
- [63] R. Szczygiel. "Krummenacher feedback analysis for high-count-rate semiconductor pixel detector readout". In: *Proceedings of the 17th International Conference Mixed Design of Integrated Circuits and Systems-MIXDES 2010*. IEEE. 2010, pp. 412–415.

Bibliography

- [64] M. De Gaspari et al. "Design of the analog front-end for the Timepix3 and Smallpix hybrid pixel detectors in 130 nm CMOS technology". In: *Journal of Instrumentation* 9.01 (2014), p. C01037.
- [65] I. Filanovsky and H. Baltes. "CMOS Schmitt trigger design". In: *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications* 41.1 (1994), pp. 46–49.
- [66] E. Gatti and P. F. Manfredi. "Processing the signals from solid-state detectors in elementary-particle physics". In: *La Rivista del Nuovo Cimento (1978-1999)* 9.1 (1986), pp. 1–146.
- [67] A. S. Sedra et al. *Microelectronic circuits*. New York: Oxford University Press, 1998.
- [68] H. Murakami. "On the Equivalent Noise Charge (ENC) of a CR-(RC) n prefilter into a gated integrator system including the 1f noise source". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 234.1 (1985), pp. 132–141.
- [69] J. D. Cressler and G. Niu. *Silicon-germanium heterojunction bipolar transistors*. Artech house, 2003.
- [70] A. K. Singh et al. "Analysis of Si/SiGe heterostructure solar cell". In: *Journal of Energy* 2014 (2014).
- [71] F. Martinelli et al. "A massively scalable Time-to-Digital Converter with a PLL-free calibration system in a commercial 130 nm process". In: *Journal of Instrumentation* 16.11 (Nov. 2021), P11023. DOI: 10.1088/1748-0221/16/11/p11023.
- [72] S. Henzler. *Time-to-digital converters*. Vol. 29. Springer Science & Business Media, 2010.
- [73] B. K. Swann et al. "A 100-ps time-resolution CMOS time-to-digital converter for positron emission tomography imaging applications". In: *IEEE Journal of Solid-State Circuits* 39.11 (2004), pp. 1839–1852.
- [74] I. Nissinen, A. Mantyniemi, and J. Kostamovaara. "A CMOS time-to-digital converter based on a ring oscillator for a laser radar". In: *ESSCIRC 2004-29th European Solid-State Circuits Conference (IEEE Cat. No. 03EX705)*. IEEE. 2003, pp. 469–472.
- [75] R. Krishna, A. K. Mal, and R. Mahapatra. "Time-domain smart temperature sensor using current starved inverters and switched ring oscillator-based time-to-digital converter". In: *Circuits, Systems, and Signal Processing* 39.4 (2020), pp. 1751–1769.
- [76] M. Kim et al. "A two-step time-to-digital converter using ring oscillator time amplifier". In: *2018 International SoC Design Conference (ISOC)*. IEEE. 2018, pp. 143–144.
- [77] J. A. Michaelsen. *Lecture on Ring Oscillators - University of Oslo*. 2012. URL: https://www.uio.no/studier/emner/matnat/ifi/INF4420/v12/undervisningsmateriale/INF4420_12_Ringoscillators.pdf.
- [78] A. Ramazani, S. Biabani, and G. Hadidi. "CMOS ring oscillator with combined delay stages". In: *AEU-International Journal of Electronics and Communications* 68.6 (2014), pp. 515–519.

- [79] F. H. Gebara et al. "4.0 GHz 0.18/spl mu/m CMOS PLL based on an interpolate oscillator". In: *Digest of Technical Papers. 2005 Symposium on VLSI Circuits, 2005*. IEEE. 2005, pp. 100–103.
- [80] A. Muntean et al. "Blumino: the first fully integrated analog SiPM with on-chip time conversion". In: *IEEE Transactions on Radiation and Plasma Medical Sciences* (2020).
- [81] P. Keränen and J. Kostamovaara. "A wide range, 4.2 ps (rms) precision CMOS TDC with cyclic interpolators based on switched-frequency ring oscillators". In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 62.12 (2015), pp. 2795–2805.
- [82] N. U. Andersson and M. Vesterbacka. "A Vernier time-to-digital converter with delay latch chain architecture". In: *IEEE Transactions on Circuits and Systems II: Express Briefs* 61.10 (2014), pp. 773–777.
- [83] Y. Park and D. D. Wentzloff. "A cyclic Vernier TDC for ADPLLs synthesized from a standard cell library". In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 58.7 (2011), pp. 1511–1517.
- [84] A. Liscidini, L. Vercesi, and R. Castello. "Time to digital converter based on a 2-dimensions Vernier architecture". In: *2009 IEEE Custom Integrated Circuits Conference*. IEEE. 2009, pp. 45–48.
- [85] Y. Cao, P. Leroux, and M. Steyaert. *Radiation-tolerant Delta-sigma Time-to-digital Converters*. Springer, 2015.
- [86] V. Gromov et al. "Development and applications of the Timepix3 readout chip". In: *PoS (Vertex 2011)* 46 (2011), p. 1.
- [87] P. Valerio. "Electronic systems for radiation detection in space and high energy physics applications". PhD thesis. Rome U., 2013.
- [88] W. Snoeys. "Monolithic pixel detectors for high energy physics". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 731 (2013), pp. 125–130.
- [89] S. Sze and G. Gibbons. "Effect of junction curvature on breakdown voltage in semiconductors". In: *Solid-State Electronics* 9.9 (1966), pp. 831–845.
- [90] J. Ueda and N. Totsuka. "Breakdown voltage analysis of planar pn junctions taking into account the radius of curvature of the corners in the patterning mask". In: *Solid-state electronics* 28.12 (1985), pp. 1245–1249.
- [91] F. M. et al. "A monolithic silicon pixel sensor in SiGe BiCMOS for the FASER high granularity pre-shower detector." Presentation at PSD12: The 12th International Conference on Position Sensitive Detectors. 2021.
- [92] J. Beringer et al. "Particle data group". In: *Phys. Rev. D* 86.010001 (2012).
- [93] M. J. Pelgrom, A. C. Duinmaijer, and A. P. Welbers. "Matching properties of MOS transistors". In: *IEEE Journal of solid-state circuits* 24.5 (1989), pp. 1433–1439.
- [94] E. Noah et al. "Readout scheme for the Baby-MIND detector". In: *PoS* (2016), p. 031.

Bibliography

- [95] M. P. Kennedy. “On the robustness of R-2R ladder DACs”. In: *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications* 47.2 (2000), pp. 109–116.
- [96] L. N. H. Becquerel. *Table de Radionucléides - Cd-109*. http://www.nucleide.org/DDEP_WG/Nuclides/Cd-109_tables.pdf.
- [97] L. N. H. Becquerel. *Table de Radionucléides - Fe-55*. http://www.nucleide.org/DDEP_WG/Nuclides/Fe-55_tables.pdf.
- [98] J. K. Ousterhout et al. *Tcl: An embeddable command language*. Citeseer, 1989.
- [99] W. W. Moses. “Time of flight in PET revisited”. In: *IEEE Transactions on Nuclear Science* 50.5 (2003), pp. 1325–1330.
- [100] J. Jakubek et al. “Pixel detectors for imaging with heavy charged particles”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 591.1 (2008), pp. 155–158.
- [101] R. Cardarelli et al. “European Patent Application”. In: *Filing-UGKP-P-001-EP, Europe Patent EP 18181123.2* (2018).
- [102] D. Z. Turker, S. P. Khatri, and E. Sánchez-Sinencio. “A DCVSL delay cell for fast low power frequency synthesis applications”. In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 58.6 (2011), pp. 1225–1238.
- [103] J.-P. Jansson, A. Mantyniemi, and J. Kostamovaara. “A CMOS time-to-digital converter with better than 10 ps single-shot precision”. In: *IEEE Journal of Solid-State Circuits* 41.6 (2006), pp. 1286–1296.
- [104] H. Wang, F. F. Dai, and H. Wang. “A Reconfigurable Vernier Time-to-Digital Converter With 2-D Spiral Comparator Array and Second-Order $\Delta\Sigma$ Linearization”. In: *IEEE Journal of Solid-State Circuits* 53.3 (2018), pp. 738–749.
- [105] M. Lee and A. A. Abidi. “A 9b, 1.25 ps resolution coarse-fine time-to-digital converter in 90nm CMOS that amplifies a time residue”. In: *2007 IEEE Symposium on VLSI Circuits*. IEEE, 2007, pp. 168–169.
- [106] S. Cadeddu et al. “A time-to-digital converter based on a digitally controlled oscillator”. In: *IEEE Transactions on Nuclear Science* 64.8 (2017), pp. 2441–2448.
- [107] K. Gammoh et al. “Linearity Theory of Stochastic Phase-Interpolation Time-to-Digital Converter”. In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 67.12 (2020), pp. 4348–4359.
- [108] R. Machado et al. “Technology Independent ASIC Based Time to Digital Converter”. In: *IEEE Access* 8 (2020), pp. 195820–195831.
- [109] S. Ziabakhsh, G. Gagnon, and G. W. Roberts. “A Second-Order Bandpass $\Delta\Sigma$ Time-to-Digital Converter With Negative Time-Mode Feedback”. In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 66.4 (2018), pp. 1355–1368.

-
- [110] Y. Kim and T. W. Kim. "An 11 b 7 ps resolution two-step time-to-digital converter with 3-D Vernier space". In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 61.8 (2014), pp. 2326–2336.
- [111] P. Lu, A. Liscidini, and P. Andreani. "A 3.6 mW, 90 nm CMOS gated-Vernier time-to-digital converter with an equivalent resolution of 3.2 ps". In: *IEEE Journal of Solid-State Circuits* 47.7 (2012), pp. 1626–1635.
- [112] J. Yu, F. F. Dai, and R. C. Jaeger. "A 12-Bit Vernier Ring Time-to-Digital Converter in 0.13 μ m CMOS Technology". In: *IEEE journal of solid-state circuits* 45.4 (2010), pp. 830–842.
- [113] B. Markovic et al. "A high-linearity, 17 ps precision time-to-digital converter based on a single-stage vernier delay loop fine interpolation". In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 60.3 (2013), pp. 557–569.
- [114] K. Kim et al. "A 7 bit, 3.75 ps resolution two-step time-to-digital converter in 65 nm CMOS using pulse-train time amplifier". In: *IEEE Journal of Solid-State Circuits* 48.4 (2013), pp. 1009–1017.
- [115] P. Lu, Y. Wu, and P. Andreani. "A 2.2-ps two-dimensional gated-Vernier time-to-digital converter with digital calibration". In: *IEEE Transactions on Circuits and Systems II: Express Briefs* 63.11 (2016), pp. 1019–1023.
- [116] Z. Cheng et al. "Recent developments and design challenges of high-performance ring oscillator CMOS time-to-digital converters". In: *IEEE Transactions on Electron Devices* 63.1 (2015), pp. 235–251.
- [117] P. Keranen, K. Maatta, and J. Kostamovaara. "Wide-range time-to-digital converter with 1-ps single-shot precision". In: *IEEE transactions on instrumentation and measurement* 60.9 (2011), pp. 3162–3172.
- [118] N. Cartiglia et al. *Signal formation and designed optimization of Resistive AC-LGAD (RSD)*. Tech. rep. 2020.
- [119] N. Cartiglia et al. "LGAD designs for future particle trackers". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 979 (2020), p. 164383.
- [120] M. Tornago et al. "Resistive AC-Coupled Silicon Detectors: Principles of operation and first results from a combined analysis of beam test and laser data". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1003 (2021), p. 165319.
- [121] A. Tajalli, Y. Leblebici, and E. J. Brauer. "Implementing ultra-high-value floating tunable CMOS resistors". In: *Electronics letters* 44.5 (2008), pp. 349–350.
- [122] C. A. C. Coello, G. B. Lamont, D. A. Van Veldhuizen, et al. *Evolutionary algorithms for solving multi-objective problems*. Vol. 5. Springer, 2007.

Bibliography

- [123] K. Deb. "Multi-objective optimisation using evolutionary algorithms: an introduction". In: *Multi-objective evolutionary optimisation for product design and manufacturing*. Springer, 2011, pp. 3–34.
- [124] M. Mitchell. *An introduction to genetic algorithms*. MIT press, 1998.
- [125] S. Sivanandam and S. Deepa. "Genetic algorithms". In: *Introduction to genetic algorithms*. Springer, 2008, pp. 15–37.
- [126] V. Stojanovic and V. G. Oklobdzija. "Comparative analysis of master-slave latches and flip-flops for high-performance and low-power systems". In: *IEEE Journal of solid-state circuits* 34.4 (1999), pp. 536–548.
- [127] K. Deb, A. Anand, and D. Joshi. "A computationally efficient evolutionary algorithm for real-parameter optimization". In: *Evolutionary computation* 10.4 (2002), pp. 371–395.
- [128] E. Zitzler, M. Laumanns, and L. Thiele. "SPEA2: Improving the strength Pareto evolutionary algorithm". In: *TIK-report* 103 (2001).
- [129] K. Amouzgar. *Multi-objective optimization using genetic algorithms*. 2012.
- [130] D. E. Groom and S. Klein. "Passage of particles through matter". In: *The European Physical Journal C-Particles and Fields* 15.1 (2000), pp. 163–173.
- [131] W. Shockley. "Currents to conductors induced by a moving point charge". In: *Journal of applied physics* 9.10 (1938), pp. 635–636.
- [132] S. Ramo. "Currents induced by electron motion". In: *Proceedings of the IRE* 27.9 (1939), pp. 584–585.
- [133] M. Dris and T. Alexopoulos. "Signal formation in various detectors". In: *arXiv preprint arXiv:1406.3217* (2014).

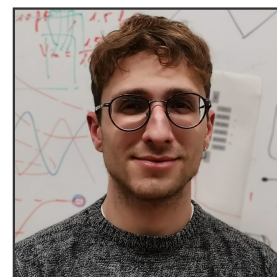
Fulvio Martinelli

PHD STUDENT · ANALOG AND MIXED-SIGNAL ASIC DESIGNER

Geneva, Canton of Geneva, Switzerland

☎ (+39) 334 12 85 642 | ✉ fulvio.martinelli@cern.ch / fulvio.martinelli@epfl.ch | 📧 fulvio-martinelli-b8b370109

| 💬 join.skype.com/invite/hfX38qBWWitM



Summary

PhD student at the Conseil Européen pour la Recherche Nucléaire (CERN) and at Ecole Polytechnique Fédérale de Lausanne (EPFL).

Master's degree summa cum laude in Electronic Engineering at University of Naples "Federico II".

Winner of the 2017 edition of the "Happy Birthday Federico II" Award as one of the best 33 students of the University of Naples "Federico II".

Main interests: Analog and Mixed Signal Design, Silicon Sensors, Timing Sensors, Integrated Circuits Design, Microelectronics, Semiconductor Devices, Power Electronics, Digital Systems Design, Solid State Physics, Time-to-Digital Conversion.

Work Experience

Conseil Européen pour la Recherche Nucléaire (CERN)

Geneva, Switzerland

PHD STUDENT

Sep. 2018 - Mar. 2022

- I work, as PhD student, on the development of fast timing detectors for medical imaging and high energy physics applications. My tasks mainly involve mixed-signal and digital electronics design, validation and characterization.
- Main projects: TT-PET project (CERN/University of Geneva); FASER experiment (CERN).

SK Hynix

Agrate Brianza (MB), Italy

ANALOG DESIGN ENGINEER

Apr. 2018 - Jul. 2018

- I used to work as Analog Design Engineer in the Core Team of 3D NAND FLASH division of the company. My duties were mainly related to the design and testing of the systems and circuits that needed to interface with the memory stack.

Conseil Européen pour la Recherche Nucléaire (CERN)

Geneva, Switzerland

GRADUATE ENGINEER TRAINEE

Jun. 2017 - Apr. 2018

- Traineeship for master thesis in which I dealt with metrological and functional design of analog and control circuits for a monolithic pixel detector in SiGe Bi-CMOS technology for TT-PET project and in collaboration with University of Geneva.

Education

Ecole Polytechnique Fédérale de Lausanne (EPFL)

Lausanne, Switzerland

DOCTOR OF PHILOSOPHY, MICROSYSTEMS AND MICROELECTRONICS

Sep. 2018 - Mar. 2022

- Thesis title: "SiGe Time Resolving Pixel Detectors for High Energy Physics and Medical Imaging".
- Supervisor (EPFL): Prof. Edoardo Charbon. Co-Supervisor (CERN): Prof. Marzio Nessi.

University of Naples "Federico II"

Naples, Italy

MASTER'S DEGREE IN ELECTRONIC ENGINEERING

Sep. 2015 - Dec. 2017

- Final Grade: 110/110 summa cum laude
- Thesis title: "Development of analog and control circuits for high time resolution sensors integrated in monolithic pixel detectors in SiGe Bi-CMOS technology".
- Supervisor: Prof. Pasquale Arpaia. Co-Supervisors: Prof. Giuseppe Iacobucci, PhD. Pierpaolo Valerio.

University of Naples "Federico II"

Naples, Italy

BACHELOR'S DEGREE IN ELECTRONIC ENGINEERING

Sep. 2012 - Sep. 2015

- Final Grade: 110/110 summa cum laude
- Thesis title: "Design and simulation of SRAM Content-addressable (CAM) in 65 nm CMOS".
- Supervisor: Prof. Davide De Caro

Liceo Scientifico R. Livatino

Naples, Italy

SCIENTIFIC HIGH SCHOOL DIPLOMA

Sep. 2007 - Jun. 2012

- Final Grade: 100/100

Honors & Awards

2017 **Winner**, "Happy Birthday Federico II" Award 2017

Naples, Italy

Presentations

12th International Conference on Position Sensitive Detectors

Birmingham, United Kingdom

SPEAKER

Sep. 2021

- Conference on Position Sensitive Devices and their applications.
- Presentation Title: A monolithic silicon pixel sensor in SiGe BiCMOS for the FASER high granularity pre-shower detector.

SNF md-NUV PET PROJECT MEETING @ EPFL

Neuchatel, Switzerland

SPEAKER

May. 2020

- Topical Workshop on image sensing and PET systems.
- Presentation Title: Development of a monolithic pixel sensor with 100 ps time-resolution for the pre-shower detector of the FASER experiment at CERN.

SNF md-NUV PET PROJECT MEETING @ EPFL

Neuchatel, Switzerland

SPEAKER

May. 2019

- Topical Workshop on image sensing and PET systems.
- Presentation Title: High Resolution TDC Design and Synchronization for TT-PET project.

Publications

- F. Martinelli et al. "Measurements and analysis of different front-end configurations for monolithic SiGe BiCMOS pixel detectors for HEP applications". In: Journal of Instrumentation 16.12 (Dec. 2021), P12038. DOI: 10.1088/1748-0221/16/12/p12038.
- F. Martinelli et al. "A massively scalable Time-to-Digital Converter with a PLL-free calibration system in a commercial 130 nm process". In: Journal of Instrumentation 16.11 (Nov. 2021), P11023. DOI: 10.1088/1748-0221/16/11/p11023.
- L. Paolozzi, et al. "Time resolution and power consumption of a monolithic silicon pixel prototype in SiGe BiCMOS technology." Journal of Instrumentation 15.11 (2020): P11025.
- Iacobucci, G., et al. "Efficiency and time resolution of monolithic silicon pixel detectors in SiGe BiCMOS technology." arXiv preprint arXiv:2112.08999 (2021).
- G. Iacobucci et al. "MonPicoAD-a Monolithic Picosecond Avalanche Detector."

Skills & Tools

MAIN TOOLS

- Cadence Virtuoso
- Innovus
- CST
- LabView
- HSpice
- Spectre
- Simetrix
- Microsoft Excel
- LaTeX
- Quartus II
- UVM
- Arduino IDE

MAIN PROGRAMMING LANGUAGES & HDLS

- Python
- Matlab
- Verilog
- SystemVerilog
- VHDL
- C++

OPERATIVE SYSTEMS

- Linux (Manjaro, Redhat)
- Windows

LANGUAGES

- Italian (mother tongue)
- English (professional working proficiency)
- French (elementary proficiency)