Time Resolution Simulations of Monolithic CMOS Sensors with Internal Gain

Time Resolution Simulations of Monolithic CMOS Sensors with Internal Gain

Umberto Follo
dep. of Electronics and Telecommunications
Politecnico di Torino
Torino, Italy
umberto.follo@polito.it

***Abstract*—Low Gain Avalanche Detectors represent one of the most mature technologies of solid-state detection for time measurements in particle physics experiments. However, due to their hybrid approach and their non-commercial grade process, the production and assembly costs are considerable. A possible solution is to develop monolithic CMOS sensors designed for timing applications which has been one of the goals of the 3rd run of the ARCADIA project. Indeed, a first prototype of monolithic CMOS sensor in 110nm with internal gain has been developed with the support of simulations to predict the performance of the designed system. A Monte Carlo software, Garfield++, has been employed to generate realistic signals coming out from the sensor while a Computer Aided Design software has been used to simulate the electronics response. The total time resolution extracted from the simulations is around 100ps rms.**

***Index Terms*—Monte Carlo simulation, Radiation Detectors, Monolithic CMOS Sensors**

I. Introduction

Particle detection is an essential component of any modern particle physics experiment. Specifically, the reconstruction of the particle trajectory and type is of crucial importance to study the physics behind fundamental interactions. Thus, an extensive R&D phase on radiation detectors has been carried out but, more importantly, is still ongoing.

In this section the monolithic sensors and the time resolution of a read-out system for radiation detector are presented.

A. Monolithic sensors

Usually, in radiation detection devices, the electronics and the sensor are two different elements designed and manufactured separately. Only in a later stage they are joined together through welding techniques such as bump or wire bonding. This second step is usually very expensive and leads to a yield reduction. Instead, monolithic sensors are devices in which the same substrate hosts both the electronics and the sensor. This approach represents an evolution of the standard "hybrid" style described above. Furthermore, an additional advantage lies in the material budget, i.e. the amount of medium a particle must cross to be detected. Indeed, this parameter is very important in modern detectors where the radiation needs to be scattered as low as possible. Thus, having just one layer for electronics and sensor is a major improvement.

979-8-3503-8630-1/24/$31.00 ©2024 IEEE

Several R&D projects led to the development of different monolithic technologies implemented in commercial CMOS processes where the major problem lies in avoiding competitive charge collection between the collection electrode and the transistors with the same type of doping as the electrode. Different solutions were proposed and developed, all described in [1].

B. Time resolution in radiation detectors

In high-energy physics experiments, measuring the crossing time of a particle with a high degree of precision is a critical challenge. Indeed, by measuring the Time Of Flight (TOF), i.e. the time taken by a particle to travel a given distance, the particle momentum might be reconstructed and used to discriminate the particle type e.g. electron, proton, and pion. Thus, it is mandatory to parameterize and study the different terms contributing to the total time resolution σ_t (1):

$$\sigma_t^2 = \sigma_L^2 + \sigma_D^2 + \sigma_J^2 + \sigma_{TW}^2 + \sigma_{TDC}^2 \tag{1}$$

The first two addends are intrinsic of the medium and geometry of the radiation sensor, while the last three arise from the read-out electronics. A brief description of each term is presented:

- The Landau term (σ_L) arises from the non-constant charge deposition in a medium. Indeed, the number of electron-hole pairs created in a fixed thickness of material by an impinging radiation can be described by the Landau distribution [2]. This contribution is intrinsic of the sensor and can be reduced lowering the active thickness of the sensor.
- The Distortion term (σ_D) parameterizes the non-uniformity of the Weighting Potential (WP) which is used in the Schottky-Ramo theorem to compute the signal induced in the collection electrode. This term can be optimized by a thorough study of the sensor geometry, in particular the thickness and width of the electrode are the key parameters to be considered [3].
- The Jitter (σ_J) is due to the front-end electronics, and it is defined in (2) where the σ_N is the electronic noise while the SR is the Slew Rate. This term can be minimized

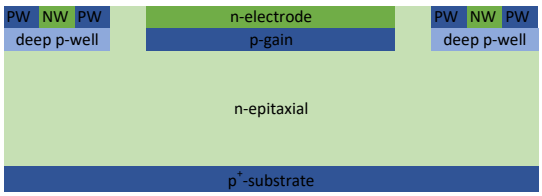

Fig. 1. The ARCADIA sensor with the gain layer placed underneath the collection electrode. At the border the deep p-wells host the electronics.

by carefully sizing the input transistor according to the desired power consumption [4].

$$\sigma_J = \frac{\sigma_N}{SR} \tag{2}$$

- The σ_{TW} is the Time Walk contribution. This term occurs during signal discrimination because different amplitudes cross a fixed threshold at different times. This term can be corrected a posteriori using software algorithms based on total signal amplitude or on time over threshold measurements. It can also be minimised implementing electronics architectures such as the Constant Fraction Discriminator (CFD) [4].
- The last contribution (σ_{TDC}) is due to the quantization error. Indeed, it is necessary to assign a time stamp to a discriminated signal using a Time to Digital Converter (TDC). This term can be neglected in respect to the others [5].

II. Sensor and Electronics Design

Among the different R&D projects, the ARCADIA project developed, in collaboration with LFoundry, a technology for fully depleted monolithic CMOS sensors in the 110 nm technology node with different active thicknesses [6]. The sensor design led to large collection electrode allowing the integrated electronics to be placed between-pixel and the fast charge collection in the order of ns [7]. However, for precise time measurements, the power consumption required to achieve resolutions below 1 ns in standard silicon sensor is too high due to the very small Signal to Noise Ratio (SNR). Indeed, the most well-established technology in silicon sensor for TOF measurements is represented by Low Gain Avalanche Detectors (LGADs). They are PIN diodes with an additional high doped p-type layer creating an electric field high enough to trigger a multiplication process. Simulations and measurements showed outstanding time resolution with these devices [8]. From this evidence, an in-pixel gain layer has been introduced into the technology developed in the ARCADIA project, like in the LGAD one.

The cross section of the ARCADIA sensor is shown in Fig. 1. Starting from the bottom, the first region is the p-substrate, followed by the epitaxial layer of 48 μm grown on the underlying silicon. Then, the highly doped gain layer is implanted underneath the collection electrode (in green). Surrounding the pixel, the deep p-well are used to shield the in-pixel CMOS electronics. In the design, the deep p-well

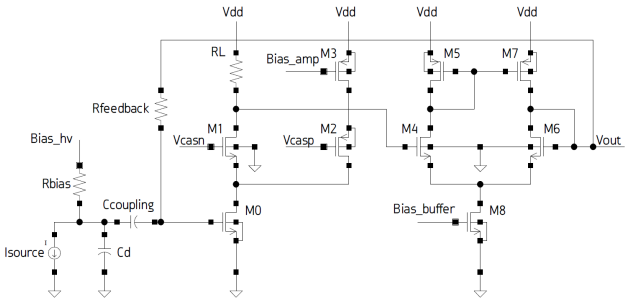

Fig. 2. The in-pixel front-end architecture.

is detached from the gain layer in such a way that signals generated by particles arriving at the edges, i.e below the electronics, are not multiplied. This enables the discrimination of signals below the collection electrode where the WP is constant from those at the edges where the collection of the charges is slower due to a longer drift path. Indeed, choosing the largest possible pixel is necessary not only to make the WP as uniform as possible, thus making the distortion term as low as possible, but also to maximize the ratio between collection electrode area and pixel area. For a constant total sensor area, the higher this ratio, the smaller the area where the timing performance is poor. However, by increasing the pixel area, the capacity of the structure grows, yet this effect can be mitigated by raising the intrinsic gain of the sensor.

Fig. 1 depicts the technology used for active thickness of 48 μm. However in the ARCADIA technology the production of thicker sensor is also possible using a high resistivity n-substrate, nevertheless, for time measurement, the Landau contribution (1) should be as low as possible, hence the use of thinner substrate, i.e. 48 μm.

The sensor must be reversely biased in order to work, with a negative bias (around -30 V) on the p-side which contributes to the depletion, and a positive bias (around 40 V) on the n-type electrode which, instead, drives the multiplication field.

To read the sensor, an in-pixel front-end electronics working at 1.2 V has been designed. The architecture is based on a common source amplifier with a resistive load. In addition, a split bias is used to boost the current fed to the input transistor. The front-end is connected to a differential amplifier used as a buffer stage to decouple the front-end from the capacitance load. A feedback loop closes the output of the buffer and the gate of the input transistor which is AC coupled to the collection electrode. This capacitor ($C_{coupling}$) is required to separate the medium voltage of the sensor from the 1.2V power domain of the electronics. The schematic of the in-pixel electronics is shown in Fig. 2, where a biasing resistor R_{bias} is used to properly polarize the collection electrode. The total power consumption of the front-end is 200 μW.

The designed pixel is 250 μm $\times$ 100 μm, which was the largest structure satisfying the density rules of the foundry in a non-dedicated engineering run. 4 μm $\times$ 100 μm of this space, from each border, have been exploited to allocate the front-end

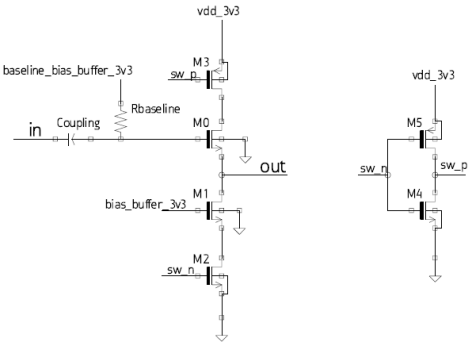

Fig. 3. The output buffer architecture on the left and the inverter used to drive the switches on the right.

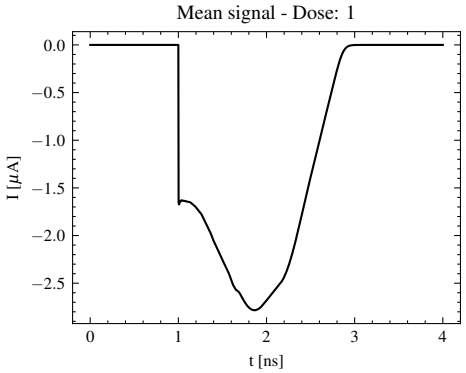

Fig. 4. Mean of a set containing 1k signals obtained from the MC simulation.

electronics. This base cell has been then copied in a matrix of 8×8 pixels and a total of 8 matrices have been placed inside the chip.

Since the main goal was the characterization of the raw performance of these first monolithic sensors with gain, the discrimination stage has been not implemented, so a source follower buffer has been added to each channel, located outside the matrix, and operating at 3.3 V. This buffer has been designed to drive a 10 pF load capacitor with a power consumption of 1.6 mW and it has been enclosed in two transistors employed as switches (Fig. 3). In this way, the 512 channels have been then reduced shorting the output of 4 different buffers, and thus, only 128 analog lanes exit the chip. Two shift registers and 8 enable signals allow to read simultaneously every pixel of one matrix or just four pixels.

III. Time resolution simulations

To predict the time resolution of the silicon sensor, a Monte Carlo (MC) software has been used: Garfield++ [9]. This tool allows to simulate realistic signals in semi-conductor particle detectors using probabilistic models.

The signal formation follows the Shockley-Ramo theorem [10]. This theorem binds the current induced in an electrode with the Weighting Field (WF) which can be extrapolated placing the collection electrode at two slightly different biasing points, e.g. 1% of the voltage applied on it [9].

Thus, using the Technology Computer Aided Design (TCAD) tool, the 2D section of the electric field at 45 V and 45.01 V has been simulated for the device under study and the WF extrapolated. The other biases are the backside voltage, which is placed at -40 V, and the guard-ring structures, that are at 9 V. These values were extracted from TCAD simulations where the depletion voltage and the gain dependence on the n-type electrode could be evaluated. Next, the WF is inserted in Garfield++.

Therefore, the charge deposited in silicon by a Minimum Ionizing Particle, i.e. pion of 180 GeV, is then simulated to obtain MC signals. The primary electron-hole pairs generated by the crossing particle are drifted toward (electrons) or away from (holes) the electrode using the WP map obtained previously. The parameters used for the charge transport inside the medium follow the Masetti-Canali model [11] while the avalanche ionization of the medium observes the van Overstraeten and de Man model [12]. The path of each electron-hole pair generated inside the silicon is then evaluated through a MC simulation where the drift velocity and the random diffusion are calculated every 0.5 μm. Additionally, the avalanche phenomenon can be activated, i.e. the drift and multiplication of charges generated in avalanche processes are simulated or not. This feature allows the estimation of the gain.

The average signal computed on a set of 1k MC events is plotted in Fig. 4 for a gain layer of unitary dose (nominal profile) with the biasing voltages previously outlined. Here, the particles have been simulated to impinge perpendicular to the sensor in its center. From the shape, the major contribution to the total signal can be extrapolated, indeed the multiplied holes in the gain layer contribute to the most because of their drift along all the depth of the sensor (peak at 2 ns). Instead, the electrons are responsible for the steep fall time at 1 ns.

Each MC current signal is then integrated to assess the gain which is about 5.2. This value has been obtained comparing the two Most Probable Values (MPVs) of the Landau distributions extracted with and without the activation of the avalanche. Furthermore, from the integrated signals, a first evaluation of the time resolution of the sensor can be obtained.

In Fig. 5 the rms of the time evaluated at a given Constant Fraction (CF) of the total integrated charge is presented. The results extracted from the data described above are plotted with a solid line, while the values derived from a set in which the particles have been randomly impinged between the center and the edge of the sensor, perpendicularly to it, are plotted with a dotted line. The worsening of the time resolution shown in Fig. 5 arise from the distortion term in (1). Indeed, the random set takes into account the non-uniformity of the WF.

Next, the signals generated by a random incident position are inserted into the electronics simulation and the average output is displayed in Fig. 6. Specific Computer Aided Design (CAD) tools provided by the foundry have been used to extract parasitic components from the channel layout to obtain a more realistic simulation.

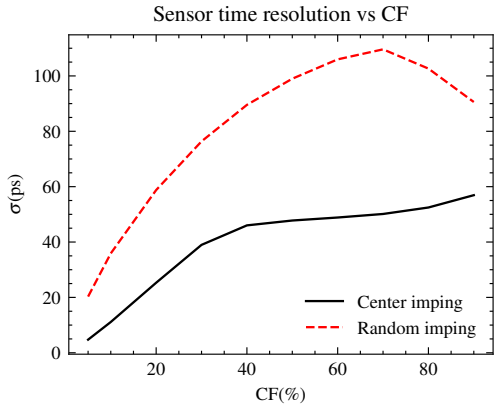

Fig. 5. Time resolution of the sensor depending on the fraction of the charge at which the arrival time is extrapolated.

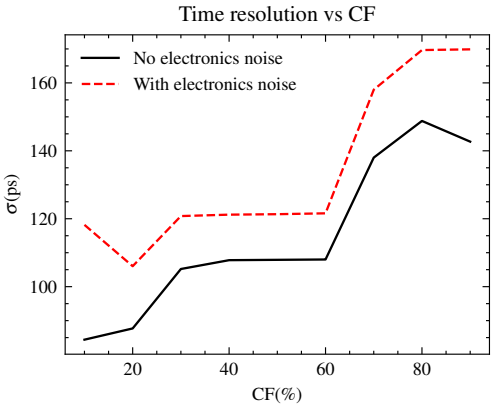

Fig. 7. Time resolution with and without the jitter contribution obtained for particles impinging randomly on the sensor.

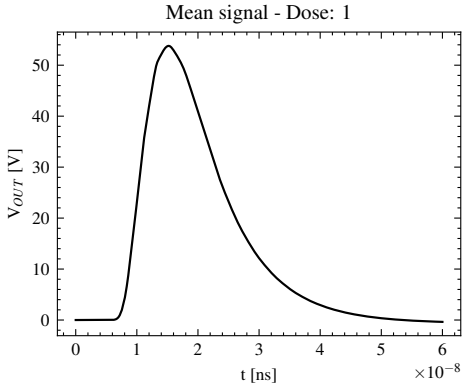

Fig. 6. Mean of a set containing 1K signals after the electronics simulation.

The rms of the time evaluated at a certain CF of the maximum amplitude allows to estimate the resolution of the overall system if the electronic noise is absent. Fig. 7 shows in black the results: a significant increase in the time resolution at low CF has been observed due to the convolution with the electronics. Finally, also the jitter term in (2) has been evaluated. Indeed, post layout simulations of noise have been carried out and at each CF, the mean slope has been evaluated and the square of the jitter summed to the square of the values previously obtained. A realistic estimation of the time resolution depending on the CF is shown in red (Fig 7) where a 106 ps can be achieved for a CF of 20% of the output amplitude.

IV. Conclusion

A first prototype of a monolithic CMOS sensor with gain layer for particle detection has been designed in 110 nm. The total resolution of this system has been simulated starting from a Monte Carlo program, i.e. Garfield++, used for generating realistic signals in the sensor. Then a Computer Aided Design program has been employed to convolve the sensor signals with the response of the electronics channel. Thanks to this procedure, the simulations for a reference dose implanted in the substrate showed that a total time resolution of 106ps rms, evaluated at a constant fraction of the signal amplitude, could be achieved.

The prototype is now in production and it should be ready for testing in mid 2024.

Acknowledgment

This research has received funding from the INFN CSN5 Call ARCADIA and INFN CSN3 ALICE Timing Layer projects.

References

[1] N. Wermes, "Depleted cmos pixels for lhc proton–proton experiments," *Nuclear Instruments and Methods in Physics Research Section A*, vol. 824, pp. 483–486, 2016.

[2] L. D. Landau, "On the energy loss of fast particles by ionization," *J. Phys.*, vol. 8, no. 4, pp. 201–205, 1944.

[3] W. Riegler and G. A. Rinella, "Time resolution of silicon pixel sensors," *Journal of Instrumentation*, vol. 12, p. P11017, nov 2017.

[4] A. Rivetti, *CMOS: Front-End Electronics for Radiation Sensors*. CRC Press, 06 2015.

[5] J. Szyduczyński, D. Kościelnik, and M. Miśkowicz, "Time-to-digital conversion techniques: a survey of recent developments," *Measurement*, vol. 214, p. 112762, 2023.

[6] L. Pancheri, R. A. Giampaolo, A. D. Salvo, S. Mattiazzo, T. Corradino, P. Giubilato, R. Santoro, M. Caccia, G. Margutti, J. E. Olave, M. Rolo, and A. Rivetti, "Fully depleted maps in 110-nm cmos process with 100–300-μm active substrate," *IEEE Transactions on Electron Devices*, vol. 67, no. 6, pp. 2393–2399, 2020.

[7] C. Neubüser, T. Corradino, G.-F. Dalla Betta, and L. Pancheri, "Arcadia fd-maps: Simulation, characterization and perspectives for high resolution timing applications," *Nuclear Instruments and Methods in Physics Research Section A*, vol. 1048, p. 167946, 2023.

[8] V. Sola, R. Arcidiacono, M. Boscardin, N. Cartiglia, G.-F. Dalla Betta, F. Ficorella, M. Ferrero, M. Mandurrino, L. Pancheri, G. Paternoster, and A. Staiano, "First fbk production of 50 μm ultra-fast silicon detectors," *Nuclear Instruments and Methods in Physics Research Section A*, vol. 924, pp. 360–368, 2019.

[9] H. Schindler, *Garfield++ User Guide*. CERN.

[10] W. Shockley, "Currents to Conductors Induced by a Moving Point Charge," *Journal of Applied Physics*, vol. 9, pp. 635–636, 10 1938.

[11] G. Masetti, M. Severi, and S. Solmi, "Modeling of carrier mobility against carrier concentration in arsenic-, phosphorus-, and boron-doped silicon," *IEEE Transactions on Electron Devices*, vol. 30, no. 7, pp. 764–769, 1983.

[12] R. Van Overstraeten and H. De Man, "Measurement of the ionization rates in diffused silicon p-n junctions," *Solid-State Electronics*, vol. 13, no. 5, pp. 583–608, 1970.