Università degli Studi di Torino
Scuola di Dottorato

The Endcap Timing Layer of the CMS experiment: detector development and impact on physics analyses

Federico Siviero

Università degli Studi di Torino
Scuola di Dottorato

Dottorato in Fisica

The Endcap Timing Layer of the CMS experiment: detector development and impact on physics analyses

Federico Siviero

Tutor: Marco Costa

A Claudia

"ricorda le piccole cose"

Abstract

The CMS experiment will be upgraded with a MIP Timing Detector (MTD) for the high-luminosity phase of the Large Hadron Collider (HL-LHC). The MTD will allow measuring the time of passage of charged particles with a resolution of 30-40 ps, disentangling in this way particles coming from interesting events from those originated in uncorrelated, but overlapping, scattering processes. Thanks to the MTD, the CMS detector will maintain its excellent performance while operating in an environment with an integrated luminosity a factor 6-7 higher than the present one.

This work focuses in particular on the sub-detector that will instrument the endcap region of MTD, the Endcap Timing Layer (ETL). ETL will be equipped with thin planar silicon sensors based on the Low-Gain Avalanche Diode (LGAD) technology, the Ultra-Fast Silicon Detectors (UFSDs), which combine a moderate internal gain (10-30) with an active thickness of only 50 μm to provide excellent timing performance.

The first chapter introduces the HL-LHC and the physics motivation for it, along with an overview of the MTD. The second chapter describes in detail the principles of operation of the UFSDs, while a summary of the productions can be found in chapter 3. The experimental techniques used to characterize the sensors are reported in chapter 4.

Chapters 5 and 6 present the results on two measurement campaigns that led to the definition of an almost final design for the UFSDs to be installed at the ETL. The first campaign focused on the development of the inter-pad region of multi-pads UFSDs: it demonstrated the importance of having low-doped and small-area p-stops to avoid the premature breakdown of the arrays and the micro-discharge effect. The grid guard-ring design of the p-stops proved to be the most reliable and the least sensitive to floating pads.

The second measurement campaign, instead, proved that the UFSD gain layer design most suited for the ETL needs is a deep carbonated boron implant, with both the boron and the carbon diffused at low thermal load: such design is able to achieve the target time resolution up to a radiation fluence of $2.5 \cdot 10^{15}$ n_{eq}/cm^2, delivering a charge ≥ 5 fC, with low sensitivity to both non-uniform biasing conditions and non-uniform

irradiation.

Finally, the last chapter presents the analysis, using simulated data, of a Higgs bosons pair decaying in two $b\bar{b}$ pairs, in the vector-boson-fusion (VBF) production mode. This is one of the most interesting physics channels for CMS at the HL-LHC, as it allows measuring the Higgs self-coupling. The analysis focuses, in particular, on the impact of the MTD timing, which improves the primary vertex tagging by $\sim 2\%$, and the jet reconstruction efficiency and purity by 0.5 and 6 %, respectively.

Contents

Chapter 1

The High-Luminosity LHC and the CMS MIP Timing Detector

The subject of this work is the MIP Timing Detector (MTD) that will be installed at the Compact Muon Solenoid (CMS) experiment for the High-Luminosity phase of the Large Hadron Collider (HL-LHC, also known as the Phase-2 of LHC) [1]. MTD will bring a completely new capability to CMS: the precise measurement of the production time of minimum ionizing particles (MIPs). Timing will allow disentangling the approximately 200 nearly-simultaneous interactions that will occur in each bunch crossing of the HL-LHC, playing a fundamental role in events reconstruction and background rejection for several physics channels to be studied in the new LHC era [2].

The main focus of this work will be on the MTD sub-detector that will be installed in the CMS endcap regions, the Endcap Timing Layer (ETL), which will be instrumented with Ultra-Fast Silicon Detectors (UFSDs), a new technology of thin silicon sensors with excellent timing precision and radiation resistance.

The developments of the UFSD design towards its final version for ETL and the impact of MTD on a selected HL-LHC physics analysis will be the main topics of the following chapters.

In this chapter, the HL-LHC and its physics motivations will be presented, along with a description of the upgrades of the CMS detector for Phase-2, with a highlight on MTD and, in particular, ETL.

1.1 The High-Luminosity LHC

The Large Hadron Collider (LHC), a particle accelerator operating at the European Organization for Nuclear Research (CERN) in Geneva, was commissioned in 2010 for

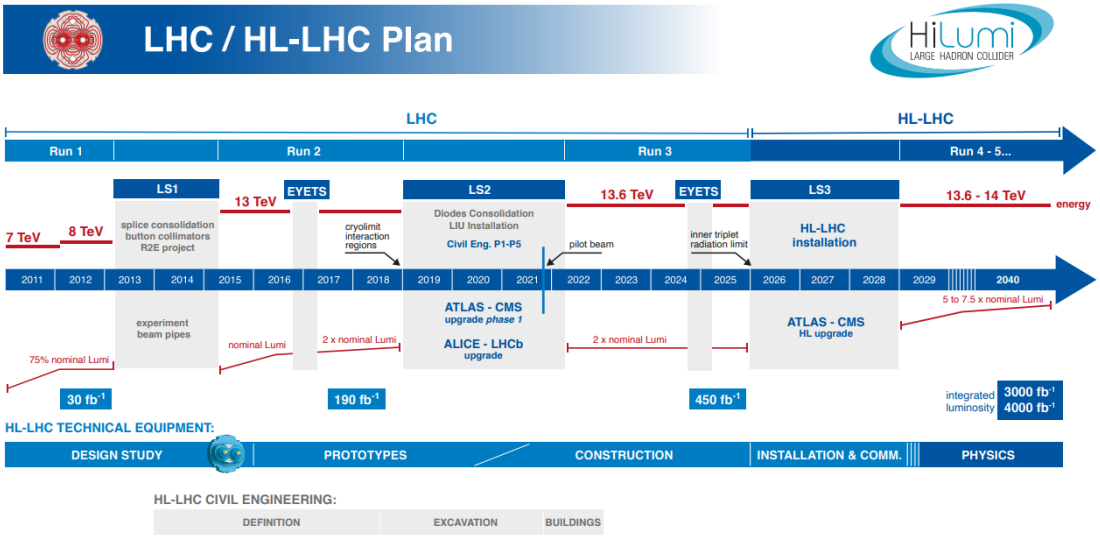


Figure 1.1: LHC / HL-LHC plan. Taken from [3].

proton-proton collisions with a center-of-mass energy of 7 TeV, which then increased to 13 TeV in 2015. In 2016-2018, the LHC delivered a peak instantaneous luminosity of 1.8×10^{34} cm^{-2}s^{-1} and an integrated luminosity of 40 fb^{-1} per year.

Since 2020, the statistic gain in running the accelerator without an increment of luminosity has become marginal: therefore, to maintain scientific progress and extend the sensitivity of the search for physics beyond the standard model (BSM), LHC will be upgraded. A long shutdown of approximately 3 years, Long Shutdown 3 (LS3), will start in 2026, aiming at the upgrade of the optics in the interaction region, in order to produce more tightly focused and overlapping beams at collision. The LHC will resume operations in 2029, beginning its High-Luminosity phase [1, 4, 5], with the following targets in terms of luminosity:

- Peak instantaneous luminosity of 5×10^{34} cm^{-2}s^{-1}
- Integrated luminosity of 250 fb^{-1} per year, for a total of 3000 fb^{-1} in a dozen year

Those are the HL-LHC nominal running conditions, whereas the ultimate operational scenario foresees an increase of the instantaneous luminosity to 7.5×10^{34} cm^{-2}s^{-1} for a total of 4000 fb^{-1} in a dozen years (300-350 fb^{-1} per year).

Phase-2 LHC will pose two main challenges to its experiments: (i) the radiation damage to the detectors due to the high integrated luminosity; (ii) the large number of concurrent interactions (pileup interactions) occurring at each bunch crossing, due to

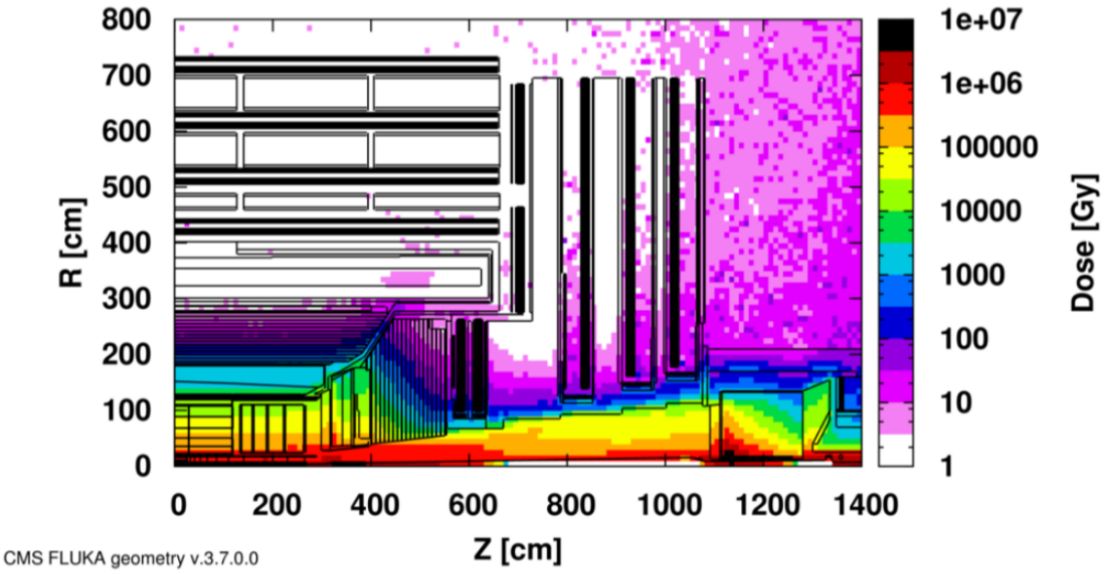


Figure 1.2: FLUKA [6] estimates of the absorbed dose in the CMS detectors after an integrated luminosity of 3000 fb^{-1}. R is the transverse distance from the beamline, z is the distance along the beamline from the interaction point.

the high instantaneous luminosity [4]. Addressing these challenges will be of key importance to maintaining the current performance of the detectors in terms of efficiency, resolution, and background rejection.

The particles generated in the collisions at the LHC cause damages to the detectors and the readout electronics, degrading their performance; at the HL-LHC, in particular, the radiation environment will be extremely harsh, with an annual dose absorbed by the detectors comparable to the total dose absorbed from the beginning of LHC operation until LS3.

The proton-proton collisions are the primary radiation source, generating a myriad of particles of various kinds:

- Charged particles (mainly pions): they produce ionization into the detectors and undergo nuclear interactions producing showers of particles, which are added to the radiation load.
- Photons: they either interact with the material of the beam pipe and with the tracker, producing electron-positron pairs, or are absorbed by the calorimeter, where they produce electromagnetic showers.

- Neutrons: they can cross long distances undergoing several scattering into the detectors and producing photons and electrons.

The high number of pileup interactions is the other major challenge for the experiments: at LHC, each bunch of protons is spatially distributed with an RMS of ~ 5 cm, and each beam contains about 2700 bunches crossing at a rate of 40 MHz; since the number of bunches cannot be increased by much, the increase of luminosity in HL-LHC will require an increase in the number of interactions per bunch crossing, i.e. in pileup. Indeed, the number of pileup interactions will raise to 140 (200) in the nominal (ultimate) HL-LHC scenario, to be compared with the ~ 40 pileup interactions of the present LHC.

The hard interactions of interest to discover new physics, those that probe energy scales ranging from a few GeV to several TeV, will occur in far fewer than 1% of the total beam crossings, but will always overlap with 140 (200) pileup interactions. The consequent spatial overlap of tracks [1] and energy deposits from the additional collisions could degrade the identification and the reconstruction of the hard interaction and could increase the rate of false triggers.

In order to avoid such degradation, and address the large radiation damage expected, the detectors of the LHC experiments will be upgraded.

1.2 Physics motivations for HL-LHC: the Higgs pair production

The HL-LHC program aims at carrying out studies that involve the recent discovery of the Higgs boson [7, 8] and the main target is to perform precise measurements of its properties to validate the Standard Model. The ATLAS and CMS collaborations estimate a comparable precision, with an uncertainty of 2-5%, for many of the Higgs boson couplings to elementary fermions and bosons, in a scenario with an integrated luminosity of 3000 fb^{-1} [9].

Moreover, the huge luminosity at the HL-LHC will make it possible to investigate rare processes that need high statistics: in this work, a highlight is put on the production of a Higgs bosons pair decaying in two bottom anti-bottom quark pairs ($\text{b}\overline{\text{b}}$), in the vector-boson fusion (VBF) production mode, which is one of the most interesting processes to be studied during Phase-2.

[1] In the context of this work, a track refers to the trajectory described by a charged particle, as determined by the tracker; hence, track and charged particle will be used as synonyms.

In this section, the process and its importance to discover new physics are described, while the discussion on the results obtained from simulated data and the impact of timing on the analysis are left to chapter 7.

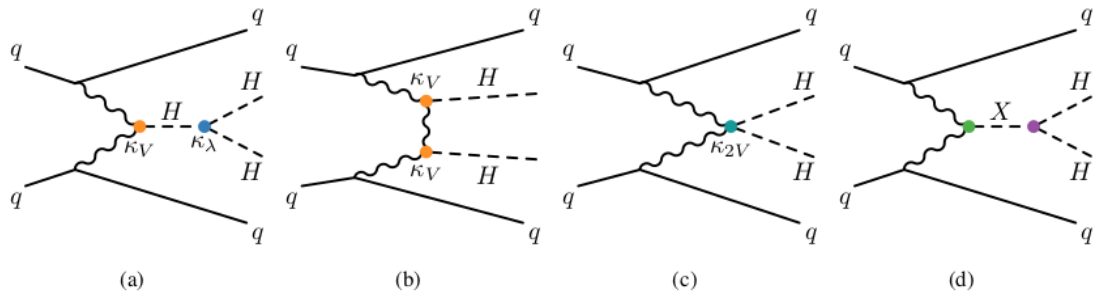


Figure 1.3: Tree-level Feynman diagrams contributing to Higgs boson pair production via VBF. Diagrams (a), (b) and (c) illustrate the non-resonant production modes scaling with $c_V c_\lambda$, c_V^2, c_{2V}, respectively. Diagram (d) illustrates the resonant production mode.

The Higgs scalar potential, in the SM, is fixed by just two low-energy parameters: the Higgs mass, $m_H \sim 125$ GeV, and the vacuum expectation value, $\nu \sim 246$ GeV. At the weak scale, the potential can be written in terms of the Higgs self-coupling, λ [10]:

$$V(H) = \frac{1}{2} m_H^2 H^2 + \lambda \nu H^3 + \lambda H^4 \tag{1.1}$$

in the SM, the self-coupling is predicted to be $\lambda = \frac{m_H^2}{2\nu^2} \sim 0.13$. A direct measurement is therefore crucial to test whether λ matches the SM expectation or deviates from it.

A deviation of the self-coupling from the SM expectation would hint at the presence of physics beyond the standard model, since a large departure of λ from its SM value is predicted by several BSM scenarios [11–13].

The possible deviations of the Higgs couplings from their SM values are quantified by the coupling strength modifiers, $c_i = k_i^{obs}/k_i^{SM}$, given by the ratio between the observed coupling (k_i^{obs}) and the SM expectation (k_i^{SM}). The couplings to fermions and bosons are usually denoted by a k, while only the self-coupling is identified with a λ, therefore the self-coupling modifier is: $c_\lambda = \lambda^{obs}/\lambda^{SM}$.

The self-coupling can be measured in particular through the Higgs pair production, which is a rare process in the SM, with a cross section a factor 1000 lower than single Higgs production. Because of its unique sensitivity to λ, this process is sometimes described as the *ultimate* test of the standard model [10]. This work, in particular, focuses on the VBF production of the Higgs pair.

In this production mode ($pp \to HHjj$) two vector bosons ($V = W, Z$) are emitted softly by two incoming protons, and then interact hardly ($VV \to HH$). The quarks

from which the vector bosons are emitted result in two jets (j) with a large rapidity gap.

In the SM, three different types of couplings are involved in the HH non-resonant production via VBF, visible in the three leftmost Feynman diagrams in figure 1.3: the Higgs boson self-coupling (HHH), the Higgs-boson–vector-boson coupling (VVH) and the di-vector-boson–di-Higgs-boson ($VVHH$) coupling. The resonant mode, in which an unknown state X decays into the Higgs pair, is also possible, diagram (d) in the figure.

The coupling strength modifiers concerning VBF HH are: c_V, the modifier of the Higgs coupling to the vector bosons; c_λ, the Higgs self-coupling modifier; the di-Higgs-di-vector-bosons coupling modifier is c_{2V}. Indeed, the three leftmost diagrams in figure 1.3 scales with $c_\lambda c_V$, c_V^2, and c_{2V}, respectively. That is a key aspect, since it underlines that the VBF mode, differently from other production modes, is sensitive not only to deviations in c_λ and c_V, but also to c_{2V}, because it is the only process involving the $VVHH$ interaction [12]. That is one of the reasons to choose VBF for the study of the Higgs pair production.

The VBF mode features a cross section of 1.726 ± 0.036 fb, more than an order of magnitude lower than the main production mode, the gluon-gluon fusion (ggF). Despite that, the VBF mode is particularly interesting in the BSM physics search since even small deviations of the couplings from their SM values can induce a striking increase of its cross section with respect to SM predictions. Indeed, in the SM, diagrams (b) and (c) in figure 1.3 interfere and cancel, since $c_V^2 = c_{2V}$; whereas, in BSM scenarios where $c_V^2 \neq c_{2V}$, the cancellation does not occur and the VBF production cross section is expected to grow with the partonic center-of-mass energy, see [13] and references therein for a detailed discussion. This is a second reason to choose VBF instead of other production modes.

This work focuses on a particular decay channel for VBF HH, the $\mathrm{b\bar{b}b\bar{b}}$ final state, which has $\sim 34\%$ branching ratio. While this final state benefits from increased signal yields due to the large branching fraction of a Higgs boson decaying to a $\mathrm{b\bar{b}}$ pair (~ 0.582 in the SM), it also suffers from an overwhelming QCD multijet backgrounds. In this respect, the remarkable VBF topology, characterized by two forward jets well separated in rapidity and with a large invariant mass, together with reduced hadronic activity in the central region, provides an essential handle to disentangle signal events from the QCD background.

Searches for Higgs boson pair production in the $\mathrm{b\bar{b}b\bar{b}}$ channel have already been carried out both in the ggF and VBF production modes by CMS and ATLAS [12, 14–19], and limits were set for both the resonant and the non-resonant production.

Other rare processes to be investigated at the HL-LHC are: the dark matter observation, fundamental forces unification, and QCD behavior under extreme conditions, just to mention some of the most important ones [4, 9].

1.3 The upgraded CMS detector for HL-LHC

The central feature of the present CMS detector is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter, each composed of a barrel and two endcap sections. Forward calorimeters extend the pseudorapidity coverage provided by the barrel and endcap detectors. Muons are detected in gas-ionization chambers embedded in the steel flux-return yoke outside the solenoid. The coordinate system adopted by CMS has the origin centered at the nominal collision point, the y-axis pointing vertically upward, and the x-axis pointing radially inward toward the center of the LHC. Thus, the z-axis points along the beam direction. The azimuthal angle ϕ is measured from the x-axis in the x-y plane and the radial coordinate in this plane is denoted by r. The polar angle θ is measured from the z-axis. The pseudorapidity is defined as $\eta = -\ln(\tan(\frac{\theta}{2}))$. A more detailed description of the CMS detector can be found in [14, 20].

The HL-LHC running conditions will far exceed the capabilities of the existing CMS detector, which will consequently require significant upgrades, the so-called Phase-2 of CMS, to continue to function efficiently. The primary goal of the CMS Phase-2 upgrade is to maintain the current excellent performance of the CMS detector in efficiency, resolution, and background rejection for all final state particles and physics observables used in data analyses [21], in an environment with much harsher irradiation and very high pileup.

The main upgrades to the CMS sub-detectors are reported below:

- *Tracker*: the tracker will suffer radiation damage effects, and it will be completely replaced for Phase-2. To maintain excellent tracking efficiency, at much higher pile-up levels, the pixel system and outer tracker granularity will be increased by a factor of 4. In the outer tracker, this will be achieved by shortening the lengths of silicon sensor strips; in the pixel system, it will be implemented by having smaller pixels and thinner sensors. Moreover, in forwards regions, the coverage of the pseudorapidity region will be extended to close to $|\eta| \sim 4$.

- *Calorimeter endcaps*: the calorimeter upgrade, the High Granularity Calorimeter (HGC), will be installed with electromagnetic and hadronic sections in the endcap regions. The electromagnetic section consists of 30 tungsten and copper plates interleaved with silicon sensors as the active material. The hadronic section has a front section of 12 brass and copper plates interleaved with silicon sensors.

- *Muon endcaps*: currently, the muon system covers the pseudorapidity region 1.5 $\leq \eta \leq$ 2.4 and consists of four stations composed of Drift Tube (DT), Cathode Strip Chambers (CSC), and Resistive Plate Chambers (RPC). To maintain the good trigger acceptance in the endcap regions, it has been proposed to extend the coverage with four additional chambers: the first two stations, where the magnetic field is still high, will be Gas Electron Multiplier (GEM) chambers with excellent spatial resolution; the last ones will be low-resistivity Resistive Plate Chambers (RCP), with lower granularity but with good time resolution to mitigate background effects. By implementing the GEM in the free region behind the endcaps of the calorimeter, the coverage of the muon detector will be close to $|\eta| = 3$.

Along with these upgrades, Phase-2 CMS will be characterized by the new timing detector, MTD, which is described in detail in the following sections, with a particular focus on the impact of timing on the CMS event reconstruction.

1.4 The CMS MIP Timing Detector

The CMS detector will be instrumented with a timing detector, with hermetic angular coverage up to $|\eta| = 3$, that will provide accurate timing of charged tracks with 30-40 ps resolution at the beginning of operations, degrading slowly as a result of radiation damage to 50–60 ps at the end of the HL-LHC lifetime. The purpose of MTD is to maintain the performance of the Phase-1 CMS detector in the challenging conditions of HL-LHC. A comprehensive description of the detector's details can be found in the MTD Technical Design Report (TDR) [2].

The issue posed by Phase-2 LHC to CMS is the unprecedented number of pileup collisions per bunch crossing, which will increase to 140-200 from the current $\sim$ 40. Pileup mitigation in CMS builds upon particle-flow event reconstruction [22], a technique that combines information from different detectors to establish a list of particle candidates: charged leptons, charged hadrons, photons, and neutral hadrons. It improves the quality of the objects included in the primary vertex (i.e. the hard-scatter

interaction, usually abbreviated to PV, probing the high-energy scale, the one interesting to discover new physics) by removing charged tracks that are inconsistent with originating from that vertex, as well as neutral deposits in the calorimeters that might belong to a different interaction, based on statistical inference techniques, as described in [23].

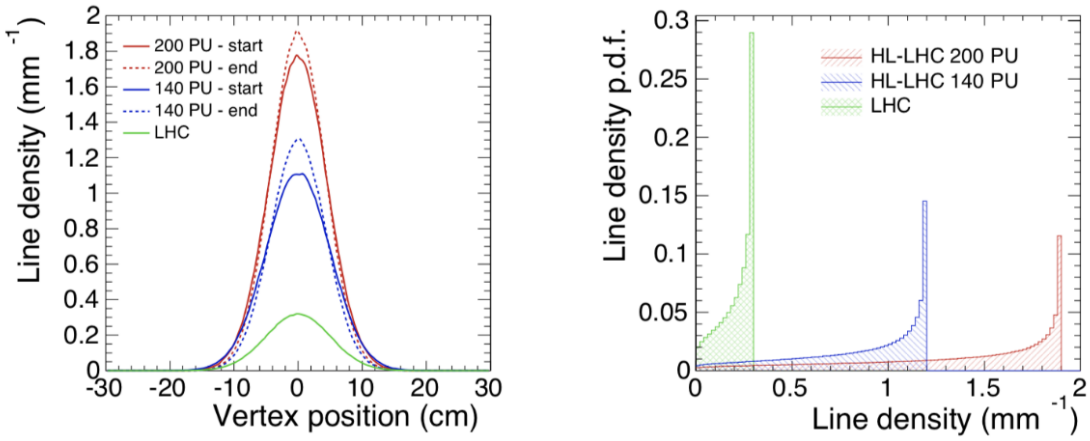


Figure 1.4: *Left*: density of vertices along the beam axis at the LHC with about 30 pileup interactions (Run-1 and early Run-2) and at the HL-LHC with 140 and 200 pileup interactions. The solid (dashed) line refers to the start (end) of the LHC fill. *Right*: probability density function of the line density for pileup values of about 30, 140 and 200. The modes and the means of the three distributions are respectively 0.3, 1.2, 1.9 mm^{-1} and 0.2, 0.9, 1.4 mm^{-1}. Taken from [2].

The high spatial granularity of the tracking sub-detectors will enable the upgraded CMS detector to separate vertices, identify the hard collision, and measure signal particles with good efficiency in the offline analyses, considering a 140 pileup scenario. In the transition from 140 to 200 pileup interactions, however, the peak *line density* along the beam axis, dN_V/dz (N_V is the number of vertices), grows from 1.2 to 1.9 mm^{-1}, assuming a root-mean-square (RMS) spread of the vertices along the beam axis of 4.5 cm, as seen in figure 1.4. According to simulation, the optimal selection window to associate a track with a primary vertex is of the order of 1 mm, therefore, for line densities of the PV of about 1 mm^{-1} or larger, the contamination of pileup tracks into the PV is not negligible; as a consequence, at 200 pileup, the probability of spatial overlaps in the CMS sub-detectors is so large that the particle-flow reconstruction begins to fail at a substantial rate [2].

In the time domain, the RMS spread of the vertices will be 180-200 ps within the 25 ns bunch crossing structure of the colliding beams, approximately constant during the fill and largely uncorrelated with the spatial distribution. The addition of timing to the CMS event reconstruction allows slicing the bunch crossing in different time

exposures of 30-40 ps, the minimum time interval that can be resolved by the MTD. The number of pileup vertices per time exposure would drop, in this way, to about the present LHC conditions, with 40-60 concurrent interactions.

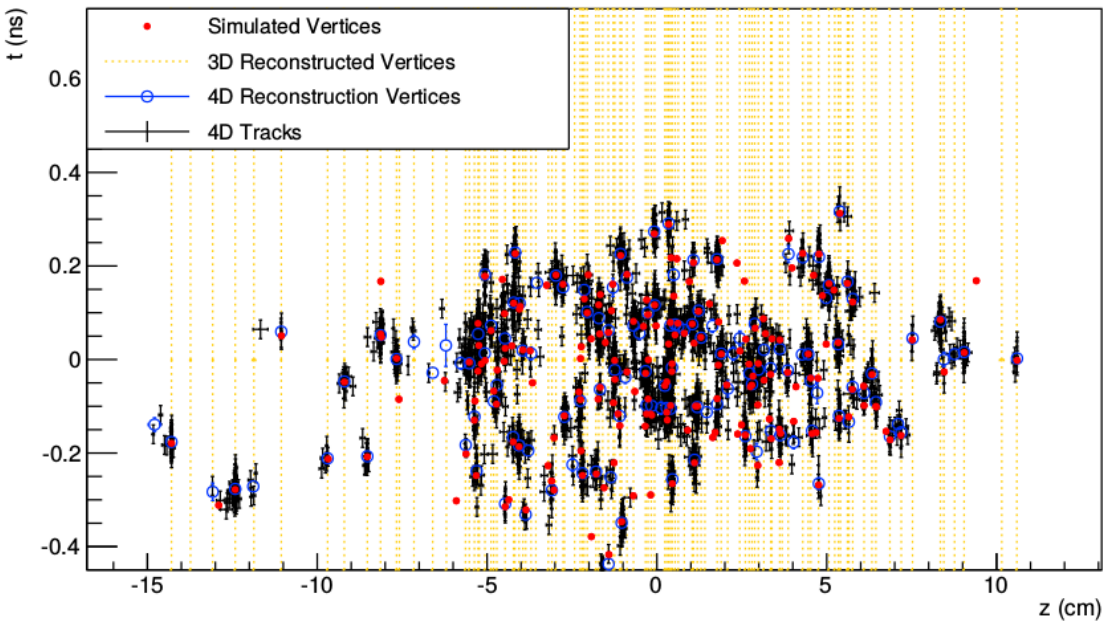


Figure 1.5: Simulated and reconstructed vertices in a bunch crossing with 200 pileup interactions assuming a MIP timing detector with $\sim$ 30 ps time resolution. The horizontal axis is the z-position along the beamline, where the "0" is the center of the interaction region. The vertical axis is the time with "0" being the point in time when the beams completely overlap in z. The simulated vertices are the red dots. The vertical yellow lines indicate 3D-reconstructed (i.e. no use of timing information) vertices, with instances of vertex merging visible throughout the display. The black crosses and the blue open circles represent tracks and vertices reconstructed using a method that includes the time information and is therefore referred to as $4D$. Many of the vertices that appear to be merged in the spatial dimension are clearly separated when time information is available. Taken from [2].

The impact of adding the time domain can be noticed in the event display in figure 1.5, where several vertices overlap in space (x-axis), but not in time (y-axis), so they can be correctly separated using timing.

According to simulation, the fraction of merged vertices at 200 pileup drops from 15% to 1% when time information is added, while the number of pileup tracks wrongly associated to the PV is halved with respect to the no-timing scenario [2].

Figure 1.6 shows, in the plot on the left, that, assuming an MTD with 35-40 ps resolution, the number of pileup tracks wrongly associated to the PV at 1.9 mm^{-1} line density is equal to the scenario where no MTD is present and the line density is ~ 0.8 mm^{-1}, which is a condition where the particle-flow algorithm can work efficiently,

as previously reported. Moreover, the plot on the right illustrates how, with timing, the number of pileup tracks associated to the leading vertex, assuming a 3σ (where $\sigma = 35$ ps) selection (green histogram), is similar to the condition where no timing information is present and the pileup is 60 (black histogram), which is similar to the condition experienced by CMS during the LHC Run-2 when the detector has been fully efficient. Hence, figure 1.6 demonstrates the capability of MTD to restore the Phase-1 conditions, when CMS has been delivering excellent performance.

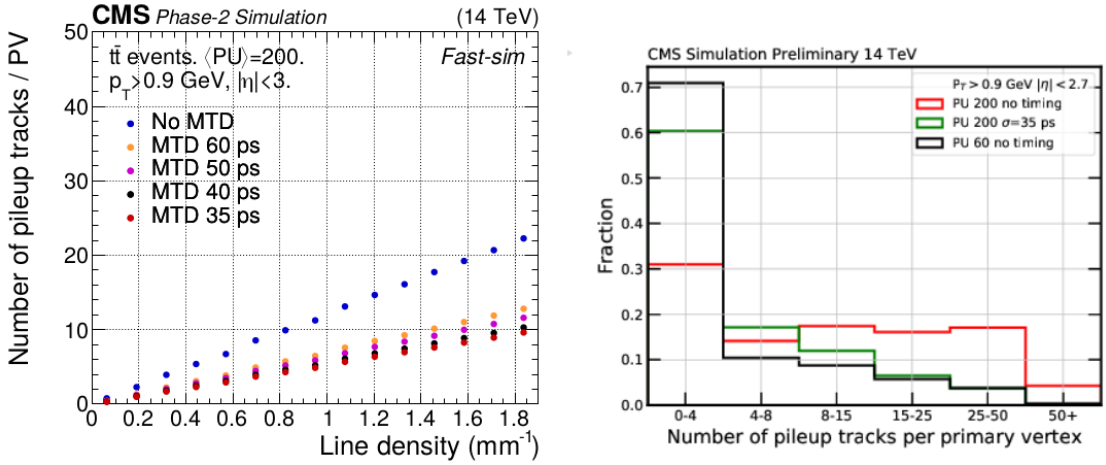


Figure 1.6: *Left*: number of pileup tracks incorrectly associated with the hard interaction vertex as a function of the collision line density for different time resolutions. *Right*: Distribution of the number of incorrectly associated tracks with the use of a 3σ (where $\sigma = 35$ ps) selection on timing information and without use of timing information. The vertical axis is the fraction of primary vertices which have the number of pileup tracks shown on the horizontal axis associated to them. Taken from [2].

The removal of pileup tracks inconsistent with the hard-interaction improves the reconstruction of many final state observables. For example, removing pileup tracks from the isolation cones improves the identification efficiency for isolated leptons, especially τ leptons, and photons, which are key signatures of many processes of interest for the HL-LHC program. The performance of b-jet identification, which relies on vertex reconstruction, is enhanced. The reconstruction of spatially extended objects and global event quantities that are vulnerable to the high pileup, such as jets and missing transverse momentum, are also improved significantly. A detailed discussion of the MTD-related improvements on the object reconstruction and on selected physics channels can be found in [2].

The addition of the MTD is also a matter of costs: the integrated luminosity×efficiency increase due to timing is equivalent to collecting data at the HL-LHC for three additional years beyond the ten-year period presently foreseen.

Approaching the end of this section, it is worth stressing out that timing in the CMS Phase-2 will not only be provided by MTD: the time of electromagnetic showers will also be determined to a precision of 30–50 ps above a transverse momentum of a few GeV in the upgraded calorimeters [24, 25]. These timestamps can then be used to associate photons to the correct charged particle vertex, based on compatibility with the time obtained from the MTD.

1.5 The MTD design

Mechanical constraints, performance, radiation tolerance, cost, and the upgrade schedule led CMS to develop a detector design for MTD consisting of a thin layer between the Tracker and the calorimeters, divided into a barrel ($|\eta| < 1.45$) and two endcap sections covering $1.6 < |\eta| < 3.0$. The requirements on the MTD are rather different in the barrel and endcap regions since (i) the surface area of BTL is about 2.5 times that of ETL, (ii) the radiation environments are significantly different, as illustrated in table 1.1, with the most irradiated part of ETL (the inner region) that will absorb a far larger radiation dose than the most irradiated one in BTL (the high $|\eta|$ part); because of these differences, BTL and ETL will be instrumented with different technologies.

Region	$\|\eta\|$	r [cm]	z [cm]	Φ n_{eq}/cm^2	Dose [kGy]
BTL	0.0	116	0	$1.65 \cdot 10^{14}$	18
	1.15	116	170	$1.8 \cdot 10^{14}$	25
	1.45	116	240	$1.9 \cdot 10^{14}$	32
ETL	1.6	127	303	$1.5 \cdot 10^{14}$	19
	2.0	84	303	$3 \cdot 10^{14}$	50
	2.5	50	303	$7.5 \cdot 10^{14}$	170
	3.0	31.5	303	$1.6 \cdot 10^{15}$	450

Table 1.1: Radiation doses and fluences (Φ) expected in the MTD (both barrel and endcap regions) after an integrated luminosity of 3000 fb^{-1}. Values taken from [2].

The best available technology for BTL is a crystal scintillator read out by Silicon Photo-Multipliers (SiPMs) [26, 27], whereas for ETL is the Low-Gain Avalanche Diode (LGAD) technology [28], which is the founding technology of the UFSDs. BTL and ETL cannot be instrumented with the same devices because the SiPM technology is not sufficiently radiation tolerant for the ETL needs and, conversely, the cost of equipping BTL with UFSDs is prohibitive. Moreover, the schedule constraints are different: BTL will have less time for development and construction (end detector installation in summer 2025), requiring the choice of a technology that needs relatively little R&D and

for which production in the industry is well-established; while ETL can benefit from a later installation (end of installation summer 2027), leaving more room for R&D and development of large-scale production.

The MTD detector is designed to function efficiently throughout the HL-LHC lifetime, up to an integrated luminosity of 3000 fb^{-1}. The radiation levels predictions for either the BTL and ETL reported in table 1.1 are based on the FLUKA Monte Carlo multi-particle transport code [6], which uses a simplified version of the CMS Phase-2 geometry. The table reports both the absorbed dose and the *1 MeV neutron equivalent* fluence [n_{eq}/cm^2]. The *1 MeV neutron equivalent* fluence (Φ) will be often used hereafter: it represents the fluence of 1 MeV neutrons producing the same damage in a detector material as induced by an arbitrary particle fluence with specific energy distribution [29–31] (see chapter 2 for more details).

For radiation tolerance qualification, all the MTD components are required to stand radiation levels at least a factor 1.5 (safety factor) larger than the nominal prediction, covering, in this way, the uncertainties in the predictions from the geometry model, the *pp* inelastic cross section, and possible sensor-to-sensor variations.

The ETL, exploiting its two-disks design (see section 1.6) and the characteristics of UFSDs (chapter 2), will provide a constant time resolution for charged tracks of 30-40 ps up to the nominal maximum radiation level; whereas, the BTL resolution will slowly degrade because of the radiation damage, reaching a track time resolution of 50-60 ps at the nominal maximum radiation level.

For the BTL, no maintenance access for repairs is possible for the lifetime of the HL-LHC while the ETL is designed to be accessible for repairs in situ and capable of being removed from the collision hall, repaired, and reinstalled during an extended technical stop.

1.5.1 The Barrel Timing Layer (BTL)

The Barrel Timing Layer is a thin, cylindrical detector that will be housed inside the Tracker Support Tube (TST), shown in figure 1.7, at its outer radius. The inner boundary of its radial envelope is 1148 mm from the beam and the outer boundary is at 1188 mm, for a maximum radial extent of 40 mm. Its overall active length in z is about 5000 mm.

Both LYSO:Ce scintillating crystals and SiPM devices instrumenting BTL are technologies for which there are well-established production and assembly procedures and facilities in industry. Both the crystals and the SiPMs are proven to be radiation tolerant up to a fluence of at least $2 \cdot 10^{14}$ n_{eq}/cm^2 and a total integrated dose of

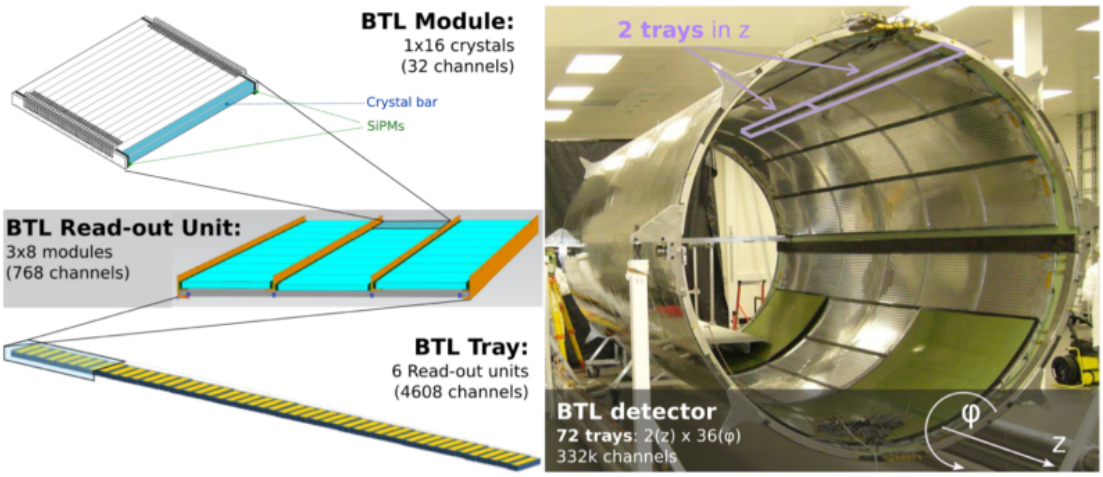


Figure 1.7: Overview of the BTL detector and the hierarchical arrangement of its components: bars, modules, read-out units and trays

25 kGy [26].

The BTL will cover the pseudorapidity region up to $|\eta| = 1.48$. The fundamental detecting cell will consist of a thin LYSO:Ce crystal bar coupled to a SiPM. The readout of both ends of the bar provides two measurements of the time of arrival of a MIP that are combined to eliminate the effect of the time delay of the light traveling along the crystal.

The BTL is read out by a dedicated ASIC, named the TOFHIR (Time-of-flight, High Rate) chip [32].

This is just a short introduction to BTL: many additional details on the detector design and its performance, whose description goes beyond the purpose of this work, can be found in [2, 33].

1.6 The Endcap Timing Layer

The Endcap Timing Layer is the MTD sub-detector on which this work is focused, therefore it will be described in details in this section.

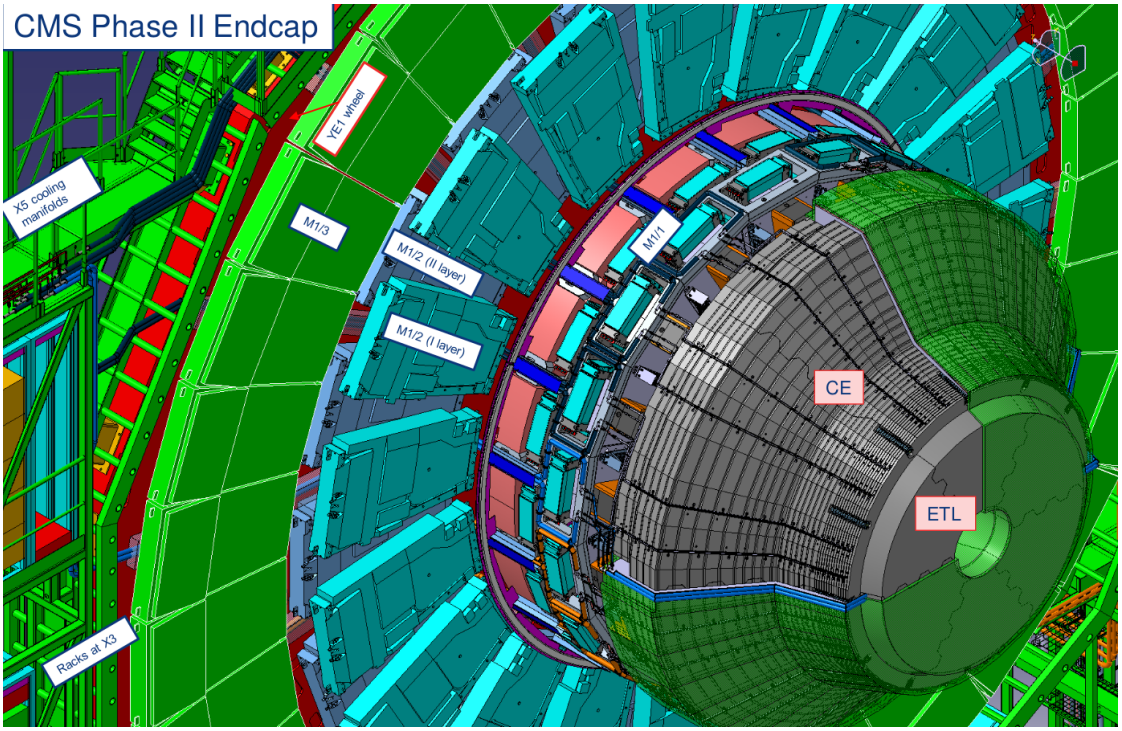


Figure 1.8: The CMS Phase-2 Endcap region, with the Endcap Calorimeter (CE) and the ETL.

ETL will be installed in its own, independent, thermally isolated enclosure on each side ($\pm$ z) of the interaction region (IR), with a hermetic, two-disk system of MIP-sensitive silicon devices, the UFSDs. Each pair of disks is located between the CMS Endcap Calorimeter (CE) and the end of the Tracker, about 3 m from the IR, and covers the radial region $315 < r < 1200$ mm, corresponding to a pseudorapidity acceptance of $1.6 < |\eta| < 3.0$. The longitudinal space allowed for ETL is about 45 mm. More details on the ETL geometry are given in appendix A.

Silicon sensors are mounted on all four faces of the two disks in each endcap in an x-y layout. The sensors are placed in a staggered way such that areas for read-out, power, and cable infrastructure, arranged in channels along each line of sensors on one face, are covered by the sensors on the opposite face. The fractional area of each disk that is sensitive to MIPs (usually called fill factor or occupancy) is requested to be greater than 85%. The use of two such disks per endcap, adjacent to each other with 20 mm z-separation, provides hermetic coverage and an average of about 1.7 hits per

track; hence, a significant fraction of the tracks passing through the ETL will have its timing measured twice, improving the final resolution. The use of two disks, instead of a single one, is the solution followed by ETL to achieve 30-40 ps resolution up to the end of the HL-LHC lifetime.

The ETL placement within CMS will allow accessibility throughout the HL-LHC operation because it is subject to high radiation dose and may need repairs or upgrades. It follows that the ETL must occupy its own independent, thermally isolated volume mounted on the nose of the CE detector (figure 1.8). To facilitate maintenance, the disks are split down the center vertically so that they form a *clam shell* around the beam pipe and can be removed and reinstalled during a technical stop without the need to remove the beam pipe. For each ETL endcap, the active sensor area (sum of both disks) is about 7.2 m^2 [2].

The radiation dose that will be received by ETL is much greater than for the BTL and highly non-uniform in η (figure 1.9), not allowing the same sensor technologies to be used. The maximum radiation fluence expected at ETL is $1.6 \cdot 10^{15}$ n_{eq}/cm^2, which further grows to $2.5 \cdot 10^{15}$ n_{eq}/cm^2 when including the safety factor that was previously mentioned. In the rest of this work, $\Phi = 1.5 \cdot 10^{15}$ n_{eq}/cm^2 will be the maximum expected fluence at ETL, while $2.5 \cdot 10^{15}$ n_{eq}/cm^2 will refer to the maximum fluence including the safety factor.

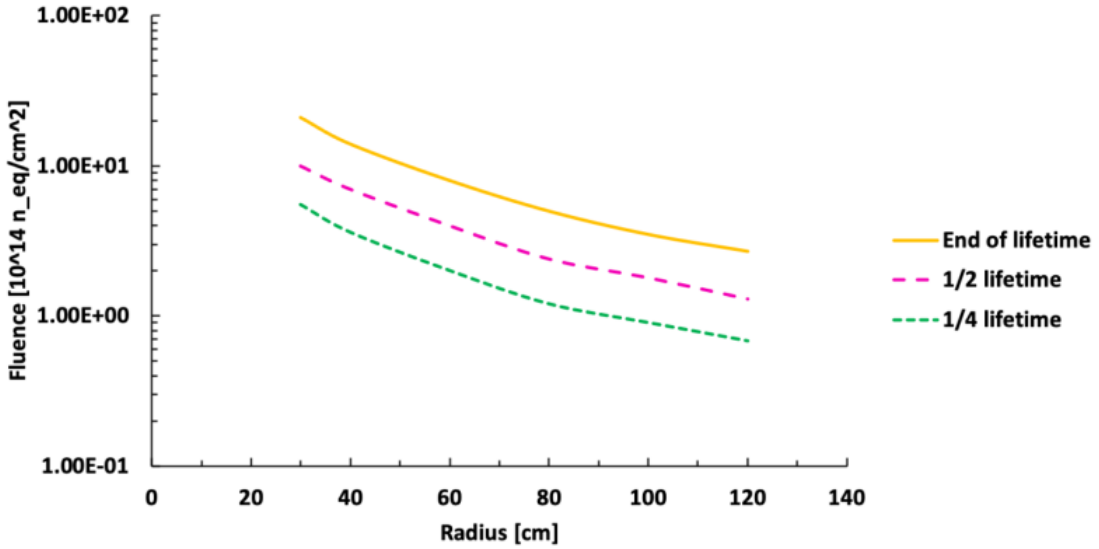


Figure 1.9: Expected radiation fluence, in n_{eq}/cm^2, in ETL regions, as a function of radius, at three time moments during the operation period of HL-LHC.

ETL will be instrumented by UFSDs, thin silicon planar sensors, based on the LGAD technology, with excellent timing performance and radiation resistance. UFSDs

feature a moderate (10-30) internal gain and low noise, achieving in this way low-jitter and fast-rising pulses, which are key to provide a time resolution of about 30 ps. UFSDs are also radiation tolerant, able to maintain almost unchanged performance up to the end of the ETL lifetime. Because of these characteristics, they have been chosen also by the ATLAS collaboration for the fast-timing layer of the ATLAS detector [34].

Small UFSD pads are required for ETL sensors so that the sensor capacitance does not impact the achievable time resolution. A very large number of pads will be thus needed to cover the $\sim$ 14 m^2 of ETL. The final sensor module (a prototype is shown in figure 1.10) will be a 16×16 array with 1.3×1.3 mm^2 pads, for a total surface of 21×21 mm^2. The TDR reported a different size for the final module, which was designed to be a 32×16 array with 3×1 mm^2 pads, but subsequent optimization studies on the module design led to the decision of decreasing either the pad size and the total module dimensions: the former choice lowers the pad capacitance, improving the time resolution; the latter increases the production yield, since the probability of having a module with one pad not working increases with the total number of pads in the module.

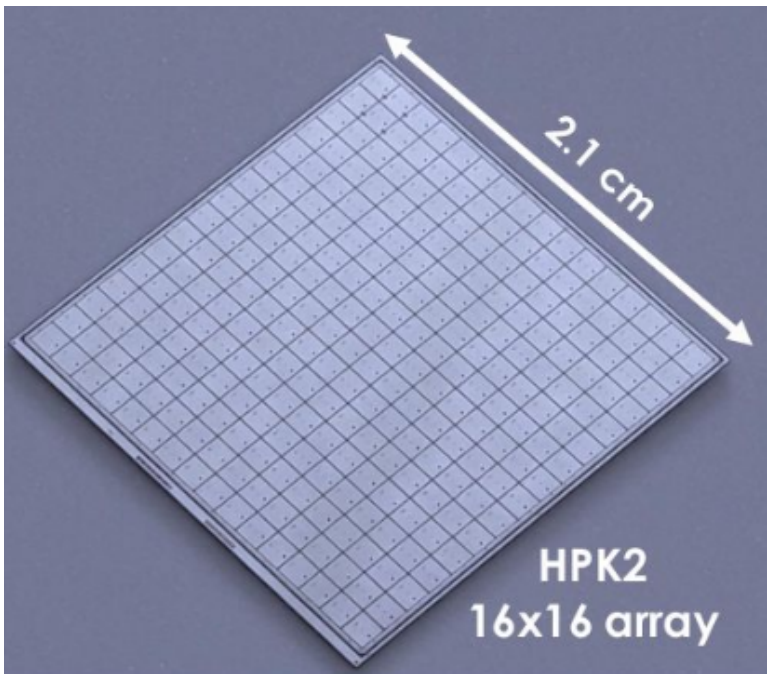


Figure 1.10: A 16×16 prototype UFSD manufactured by Hamamatsu Photonics (HPK, Japan).

Each ETL module will be read out by the Endcap Timing Readout Chips (ETROC) ASIC [35, 36], having a dimension of approximately 20×20 mm^2.

ETROC contains amplifiers and discriminators, followed by circuits to measure the

Time-of-Arrival (TOA) and Time-over-Threshold (TOT) of each particle: in this way, it can use the timing of the leading edge of the UFSDs pulses to provide timestamps corrected for time walk.

A group of readout chips communicates with an on-detector board, called the service hybrid, that reads out the ETROCs on receipt of a Level-1 trigger and sends the data out for further processing. The service hybrid also supplies DC power, bias voltage, and monitoring to the ETROCs. The service hybrid contains two boards, a readout board, and a power board.

It was previously stated that ETL aims at achieving a track time resolution of 30-40 ps up to the end of the detector lifetime, which means that the bump-bonded UFSD+ETROC pair must reach $\sim$ 50 ps resolution per single hit, considering that most tracks will have two hits in ETL. Consequently, the target resolution per single hit for either the bare UFSD and the bare ETROC lays in the 30-40 ps range up to a radiation fluence of $1.6 \cdot 10^{15}$ $\mathrm{n_{eq}/cm^2}$ [2,37] [2]. The bare UFSD resolution is measured by bonding the sensor to a fast analog read-out board (see chapter 6), whereas, the "bare" ETROC resolution is the resolution measured by injecting a fast pulse directly into the electronics, when the sensor is not bonded to the ASIC [38].

1.6.1 Impact of timing on the CMS physics program

The physics program of CMS, which includes the precision measurement of standard model processes, especially the characterization of the Higgs boson as well as searches for BSM particles and processes, will significantly benefit from the increased luminosity at the HL-LHC. The MTD will be key to maintain a good resolution and reconstruction efficiency; some of the benefits coming from timing are reported below [2]:

- One of the highest priorities of the HL-LHC physics program is the measurement of Higgs boson self-coupling, previously discussed. Increase in signal yields of 22% for constant background has been predicted, when adding timing, for several HH decay channels.

- A 10-15% improvement in the missing transverse momentum (p_T^{miss}) resolution, in the case of final states such as H$\rightarrow$ $\tau\tau$ in the vector-boson fusion (VBF) production mode.

[2]The track resolution is given by: $\sigma_{track} = \sigma_{hit}/\sqrt{n}$, with n being the number of hits in ETL. The resolution of the UFSD+ETROC pair, instead, is given by the squared sum of the resolutions of the UFSD and ETROC.

- The track-timing reconstruction opens a new future in searches of Long-Lived-Particles (LLPs), postulated in many extensions of the Standard Model. The space-time information associated with the displaced decay vertex, reconstructed from the decay daughters detected, provides the kinematic constraints necessary to get a direct measurement of the LLP mass.

- The use of the Time Of Flight (TOF) technique for charged particles identification (PID) of charged hadrons, which is essential in heavy ions physics and in low mass QCD and flavour physics.

1.7 The vector-boson fusion production of an Higgs pair in the $b\bar{b}b\bar{b}$ final state

The VBF $HH \to b\bar{b}b\bar{b}$ process has been selected as a benchmark in this work not only because it is particularly sensitive to BSM physics and carries a rather clear signature, but also because the timing provided by the MTD could particularly impact on the signal yield. Indeed, due to its peculiar topology, this channel can greatly benefit from the MTD timing in terms of (i) enhanced b-tagging efficiency, and (ii) improved VBF tagging.

For what concerns (i), in very high pileup conditions secondary vertex b-tagging is degraded by the formation of spurious secondary vertices caused by pileup tracks, reducing the ability to distinguish signal from background, as demonstrated in figure 1.11. The degradation depends upon the average pileup, and pileup density. In order to mitigate this problem, the secondary vertexing algorithms for Phase-2 have been updated to be aware of timing information from the MTD: by requiring tracks to be within 3.5σ [3] (σ = 30–40 ps) of the selected PV, the number of spurious reconstructed secondary vertices is reduced by 30% [2]. This causes a significant improvement of the separation between b-jets and jets from light-flavor quarks, as quantified by the receiver operating characteristics (ROC) curve in figure 1.11 [2].

For what concerns (ii), instead, the VBF production mode, as previously explained, carries a clear signature, with two forward jets characterized by a large rapidity gap and invariant mass. However, the challenging HL-LHC pileup conditions might make the VBF tagging difficult because of the large QCD background, which can cause pileup jets to be wrongly considered as signal (i.e. coming from the vector-boson fusion). The MTD timing can help improve the correct tagging of the signal jets by matching them

[3] A similar selection is presented in chapter 7 for jet reconstruction, requiring tracks to be within 3σ from the PV.

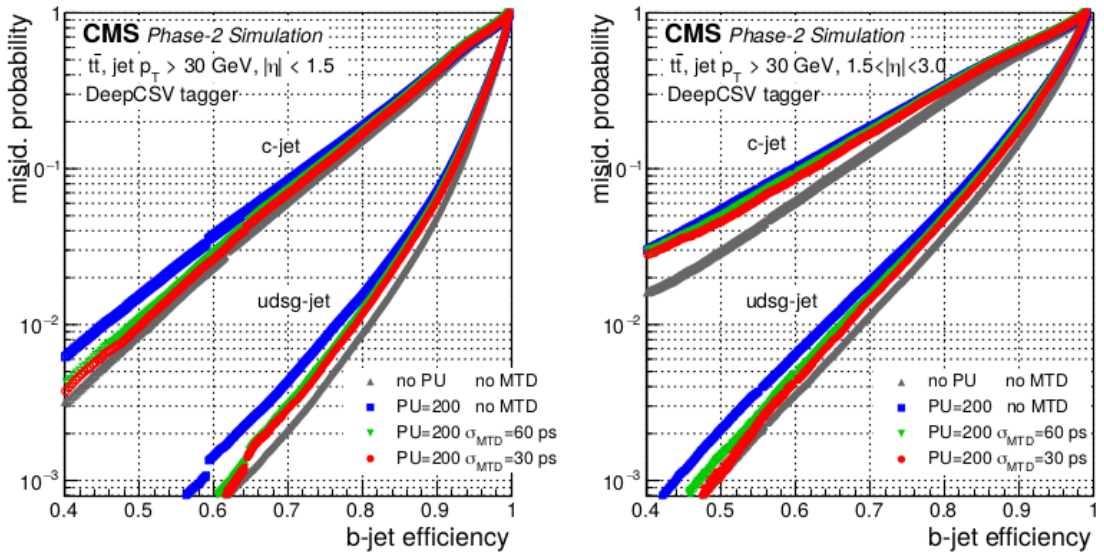


Figure 1.11: Secondary vertex tagging ROC curves for light and charm jets for $|\eta|<$ 1.5 (left) and for $1.5 < |\eta|< 3.0$ (right). Results without (blue) and with timing for 30 (red) and 60 ps (green) resolution hypotheses are compared to the zero pileup case (grey). Taken from [2].

in time: indeed, since they are simultaneously produced, the signal jets should have timings compatible within the measurement uncertainty. Therefore, if more than two jets pass the spatial and invariant mass selections, their timing is crucial to select the correct pair. This analysis strategy is further discussed in chapter 7, along with the analysis results relative to simulated data.

Chapter 2

Ultra-Fast Silicon Detectors

The CMS ETL will be equipped with thin planar silicon sensors based on the Low-Gain Avalanche Diode (LGAD) technology, the Ultra-Fast Silicon Detectors (UFSDs) [28]. The UFSDs combine an optimized sensor geometry with a moderate internal gain to achieve a time resolution of 30-40 ps, and they are able to maintain their performance up to a radiation fluence (Φ) larger than $1 \cdot 10^{15}$ n_{eq}/cm^2.

This chapter gives an overview of the UFSDs, while a description of the productions and of the experimental techniques used to characterize the sensors is given in chapters 3 and 4. The experimental results relative to the measurement campaigns, which led to an almost-final UFSD design for the ETL, are in chapters 5 and 6.

In the first section of this chapter, the principles of operation of a silicon detector are briefly summarized. The LGAD technology and how it has been implemented in the UFSD design are then presented, followed by a description of the UFSD segmentation technology. Finally, the effects of irradiation on the operation of the sensors and the possible technological solutions to improve their radiation hardness are reported.

2.1 Principles of operation of a silicon detector

As shown in figure 2.1, the silicon detector considered (also "PIN diode" [1] in the following) has a p-doped bulk, a p^{++}, and a n^{++} electrodes. A negative external bias voltage is applied to the p^{++} electrode to inversely polarize the pn junction, leaving the n^{++} electrode at ground: in this way, a large depleted volume is created. This is the typical biasing scheme of ETL sensors, with the signals being read out from the n^{++} side.

[1] A PIN diode is actually a device made of p-doped / intrinsic / n-doped silicon: despite that, silicon sensors, although having a lightly doped bulk, are usually referred to as PIN as well.

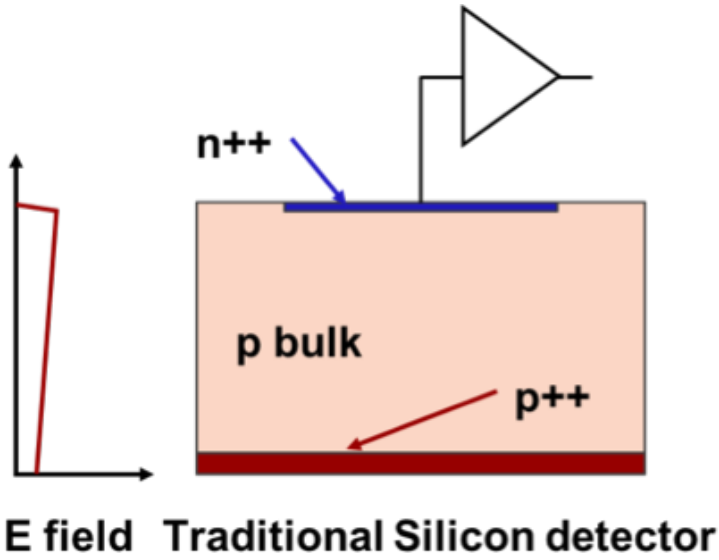


Figure 2.1: Cross-sectional view of a p-doped silicon sensor. The electric field profile is also reported on the left.

The bias voltage required to fully deplete the sensor is:

$$V_{FD} = \frac{qN_{A,eff}d^2}{2\epsilon_{Si}} \tag{2.1}$$

where q is the elementary charge, $N_{A,eff}$ the effective acceptor density, d the sensor active thickness, and ϵ_{Si} the silicon permittivity.

When an ionizing particle crosses the sensor, it creates electron-hole $(e-h)$ pairs, which then move towards the electrodes under the influence of the electric field, according to their polarity. Because of their movement, the amount of induced charge on the electrodes changes with time, generating a current signal which begins when the first charge starts drifting and ends when the last one is collected at the electrode.

The energy lost by an ionizing particle crossing the sensor follows the Landau distribution, which is an asymmetric distribution with a not negligible tail at high energies due to δ-rays, which appear when a particle loses a large amount of its energy during a single interaction, and the electrons produced have enough energy to ionize other atoms.

The induced current signal generated by the charges drift is quantified by the Shockley-Ramo's theorem [39, 40]: the current $i_k(t)$ induced on the electrode k by a charge q is given by the scalar product of its drift velocity $\vec{v}$ with the weighting field $\vec{E_w}$:

$$i_k(t) = -q\vec{v} \cdot \vec{E_w} \tag{2.2}$$

The weighting field E_w ($[L^{-1}]$) describes the coupling between the charge q and the k-th electrode.

E_w corresponds numerically, although the dimensions are different, to an electric field calculated setting at 1 the read-out electrode and at 0 all other electrodes. For this reason, E_w has the same dependence upon the geometry of the electrodes of an electric field: in the sensors studied in this work, which can be seen, in a simplified view, as parallel plate capacitors with electrodes much larger than the sensor thickness, E_w is constant and equal to $\frac{1}{d}$.

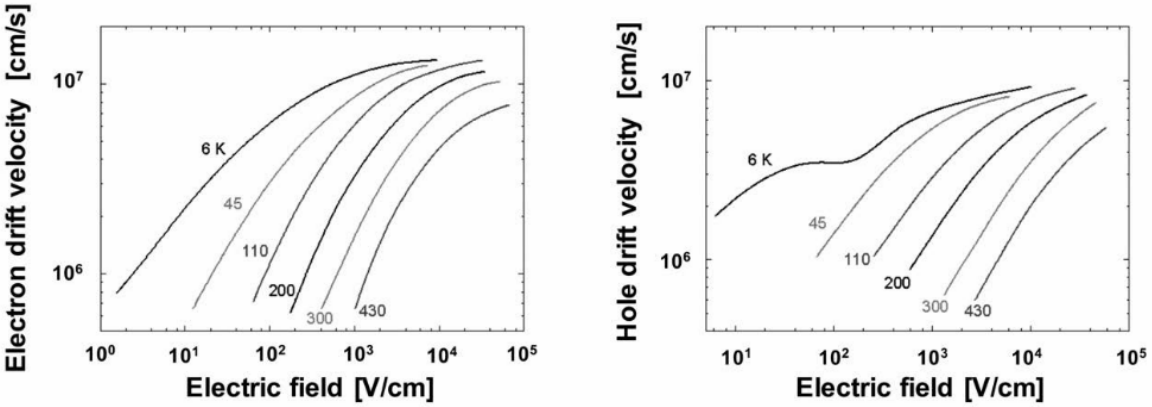


Figure 2.2: Electrons (left) and holes (right) drift velocities as a function of the electric field at several temperatures.

If E_w is constant, then it follows from equation 2.2 that the induced current at a given time depends uniquely upon the charge drift velocity, $\vec{v}$, and not upon its position in the sensor.

The drift velocity, which depends upon the electric field and the temperature, determines the shape of the current signal: the larger it is, the shorter and sharper would be the signal. The electrons drift velocity saturates at room temperature at an electric field of 30 kV/cm, whereas the holes drift velocity never really saturates, but can be seen as almost constant above an electric field of 50-100 kV/cm. Both the electrons and holes reach drift velocities of the order of 10^7 cm/s, as illustrated in figure 2.2.

2.2 Low-Gain Avalanche Diode Technology

The moderate internal gain is one of the most defining features of the UFSDs, a key ingredient to achieving an excellent time resolution. Internal gain is possible because of the LGAD technology.

Charge multiplication (responsible for the internal gain) occurs in silicon when the

carriers travel in a region with an electric field higher than 300 kV/cm (the high-field region): in such a high field, the electrons acquire enough energy to produce secondary $e-h$ pairs by impact ionization [41]. A general expression for gain G is:

$$G(x) = \frac{e^{\int_0^x (\alpha_n - \alpha_p) dx'}}{1 - \int_x^d \alpha_p e^{\int_0^{x'} (\alpha_n - \alpha_p) dx''} dx'} \tag{2.3}$$

which describes the gain of the charge carrier produced at position x in a sensor with thickness d. $\alpha_{n,p}$ [L^{-1}] are the ionization coefficients for electrons or holes, respectively. For a given electric field, $\alpha_n > \alpha_p$, which explains why impact ionization works very well in silicon detectors: other semiconductors have a smaller difference between α_n and α_p.

The inverse of $\alpha_{n,p}$, $\lambda_{n,p}$ [L], is the mean free path necessary to an electron/hole to acquire enough energy to achieve charge multiplication. The mean free path is shorter for electrons than for holes, therefore the electric field can be tuned so that electrons multiplication can occur, but holes multiplication cannot: in this way, a moderate and controlled gain can be achieved, whereas holes multiplication would lead to a full development of the charge avalanche, producing a high and uncontrolled multiplication (which is, for instance, typical in silicon photomultipliers).

Both α and λ are functions of electric field, temperature, and irradiation fluence, Φ. Leaving all the other parameters fixed, λ decreases with increasing electric field, since the carriers need a shorter path to reach the energy necessary to start the avalanche.

If (i) hole multiplication does not occur, (ii) α_n is constant across the high-field region, and (iii) the generation of secondary/tertiary particles is neglected, then the number of generated pairs can be written as:

$$N_{e-h}(d, E, T, \Phi) = N_{0;e-h} \cdot e^{\alpha_n(E,T,\Phi) \cdot d_{High}} \tag{2.4}$$

$N_{0;e-h}$ is the number of pairs generated in absence of gain, d_{High} is the length of the high-field region.

If (i), (ii), (iii) hold, then equation 2.4 is valid and the denominator in equation 2.3 is equal to one, given that $\alpha_p = 0$; in addition, if the position x in 2.3 is not in the high-field region, the numerator becomes $e^{\alpha_n \cdot d_{High}}$. Equation 2.3 thus becomes, under these conditions, equal to the ratio between the number of $e-h$ pairs produced over the number of pairs created in absence of gain:

$$Gain = \frac{N_{e-h}}{N_{0;e-h}} = e^{\alpha_n \cdot d_{High}} \tag{2.5}$$

Given that the above conditions are usually valid and the equation is simple, 2.5 is often used to define the LGAD internal gain.

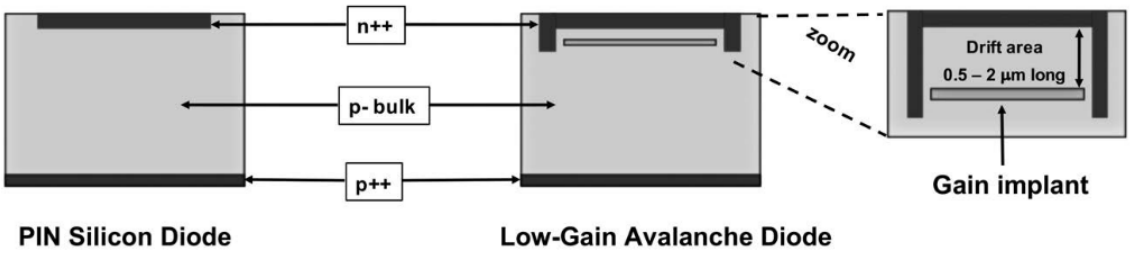


Figure 2.3: A schematic view of: (left) an n-in-p PIN diode; (right) a Low-Gain Avalanche Diode. The LGAD design is characterized by the presence of an additional p^+ implant underneath the pn junction. Taken from [42].

The first solution that one can think of to achieve an internal gain could be to bias the PIN diode at voltages high enough to make the electric field reach the 300 kV/cm threshold. However, this choice is not feasible, as it would cause an electrical breakdown in the sensor periphery, due to the extremely high fields: a different path has to be followed to reach the internal charge multiplication.

The LGAD technology is based on the implantation of a narrow (0.5-1 μm) p^+-doped ($\sim 10^{16}$ atoms/cm^3) layer close to the n^{++} electrode, at a depth of 0.5-2 μm; such layer is named gain layer (GL) or gain implant, and it is the most recognizable feature of this technology.

Once the GL is depleted, an electric field high enough to start the avalanche multiplication is locally created close to the pn junction. Dedicated structures, which are described in detail in the following, prevent the high-field region to extend to the periphery of the sensor, avoiding the electrical breakdown that would instead occur in a standard PIN. A comparison between a standard PIN and an LGAD is provided in figure 2.3.

From an experimental point of view, a PIN and an LGAD can be distinguished by the measurement of their current as a function of the bias voltage: the PIN diode has an almost constant current, whereas the LGAD features the characteristic knee (corresponding to the GL depletion) followed by an exponential trend, due to the internal charge multiplication. This is shown in figure 2.4 and further discussed in chapter 4.

The depth of the gain layer is a highlight of this work. Figure 2.5 illustrates the difference between an LGAD with a shallow (left) or a deep (right) gain layer, depicting, on the left, also the electric field profiles.

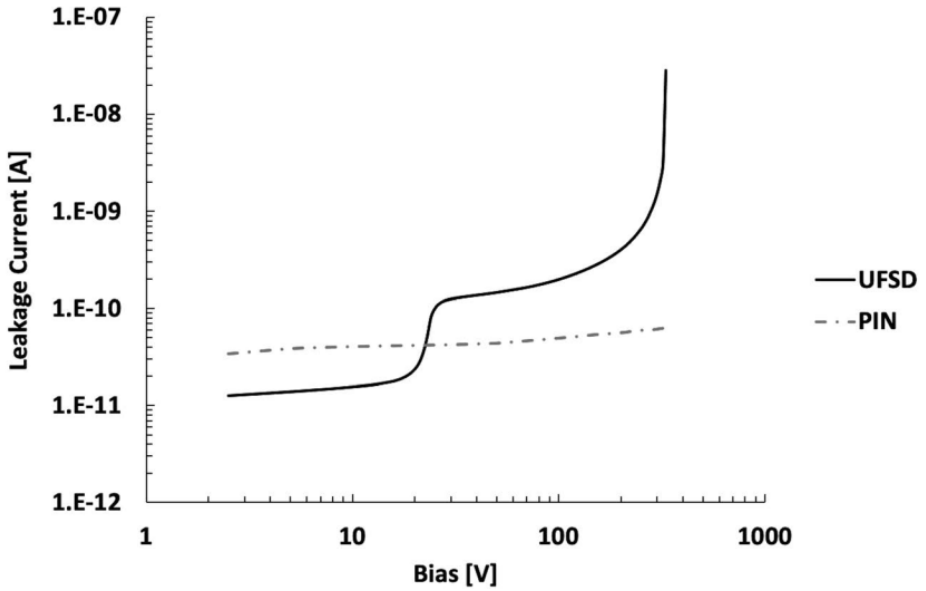


Figure 2.4: $I(V)$ characteristics of a 50 μm-thick UFSD (solid) and a PIN diode (dashed). The PIN diode presents an almost flat curve, while the UFSD features the characteristic knee at a few tens of volts, corresponding to the GL depletion, after which an exponential trend begins, due to the internal charge multiplication.

For a given bias voltage, the device with a deep implant gains more than the one with a shallow GL, since the length of the high-field region, d_{High} in equation 2.4, is larger. Hence, in order to achieve the same gain at a given voltage, sensors with a deep implant must have a lower doping of the GL, i.e. they are operated at a lower electric field.

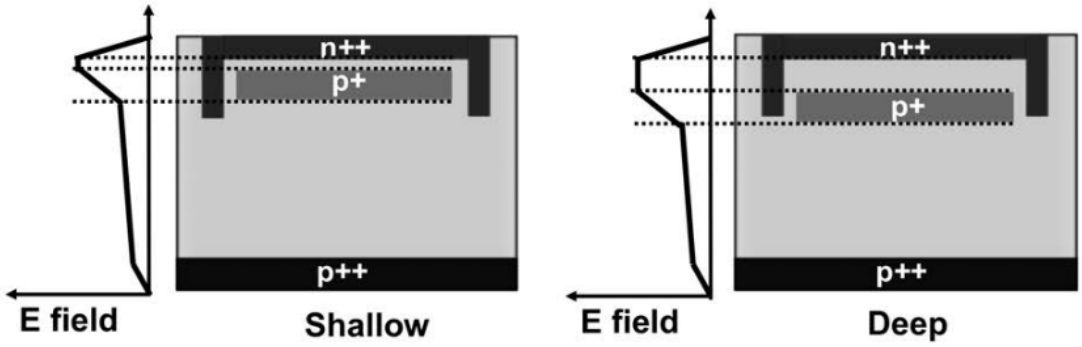


Figure 2.5: Schematic cross section of a UFSD with shallow (left), and deep (right) gain implants and their respective electric field profiles. Taken from [42].

The LGAD technology can be applied to several different devices to achieve charge multiplication, but it does not ensure excellent timing performance, other additional characteristics are needed:

- A fast signal, meaning that the sensor needs to be thin
- A uniform E_w, typical of a parallel plate geometry
- The ability of withstanding very high voltages
- An electric field large enough to saturate the electrons drift velocity
- A fine segmentation along with a high fill factor [2].

The reason why all these characteristics need to be implemented is the subject of the next sections.

The UFSDs are based on the LGAD technology, but, in addition, they implement all the above features: in other words, they are LGADs optimized for timing.

2.3 UFSD signal formation and noise sources

The current signal in a PIN diode can be calculated, assuming that the ionizing particle hits the sensor at normal incidence, using equation 2.6, in the case where the dimensions of the electrodes are much larger than the sensor thickness and the electric field is high enough for the drift velocity to be saturated [3]:

$$I_{Max} \propto Nq\frac{1}{d}v_{sat} = (n_{e,h} \cdot d)q\frac{1}{d}v_{sat} = n_{e,h}qv_{sat} \tag{2.6}$$

N is the number of charges created by the ionizing particle, which can be written (assuming uniform ionization) as $n_{e,h} \cdot d$, where d is the sensor thickness, and $n_{e,h}$ is the number of $e-h$ pairs generated per unit length; q is the elementary charge; v_{sat} is the saturated drift velocity of charge carriers. Given that the sensor considered has parallel electrodes much larger than the sensor thickness, the weighting field can be written as $E_w \propto \frac{1}{d}$.

The key aspect of equation 2.6 is that the current generated by a PIN diode does not depend upon the sensor thickness, but only on the drift velocity: since $\vec{v_{sat}}$ saturates, the current saturates as well, reaching the maximum value reported in the equation, corresponding to ~ 1.5 μA. A higher current cannot be produced in a PIN, regardless of its thickness: indeed, in a thicker sensor, more charges are generated, but their contribution to the signal is lower, because of the lower weighting field. The impossibility of having larger signals is the ultimate limit to the time resolution of PIN diodes.

[2] The fill factor is the ratio between the sensor active and total areas

[3] This is a reasonable choice since ETL sensors will have these characteristics.

Signal formation in a UFSD sensor follows a different dynamic because of the gain mechanism: as in a PIN diode, the primary electrons and holes drift towards the n^{++} and p^{++} electrodes, respectively. The primary electrons enter the gain layer and start the avalanche multiplication mechanism, producing secondary $e-h$ pairs. Since the multiplication happens very near to the cathode, the electrons are immediately collected, whereas the holes must travel across the (almost) whole bulk before being collected by the anode, generating most of the signal.

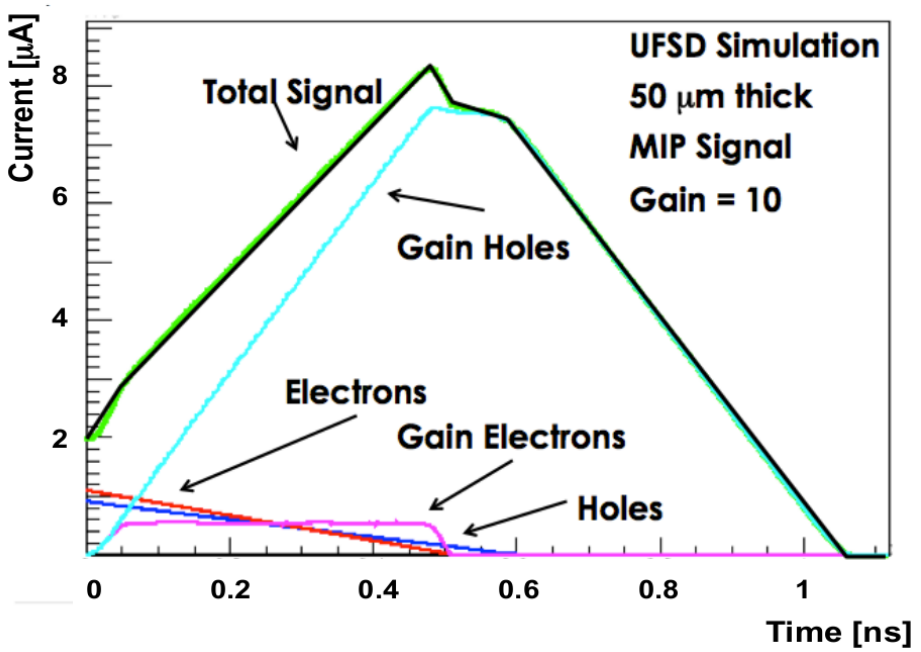


Figure 2.6: Simulation of the total current, with the various contributions due to primary and secondary electrons/holes, generated by a MIP traversing a 50 μm-thick UFSD. Simulation performed with Weightfield2 [43]. Picture taken from [42].

Since the electrons drift velocity is higher than the holes drift velocity, when the last primary electron reaches the cathode, the first secondary hole is still drifting: for this reason, the signal in UFSD increases up to the collection of the last primary electron, then it stays almost constant [4] until the first secondary hole reaches the anode, and then decreases to zero in a period controlled by the holes drift velocity. A simulated UFSD current signal is presented in figure 2.6.

The holes-generated current is dominant in a UFSD since holes have to drift for the whole sensor bulk, differently from electrons which quickly recombine into the cathode.

The duration of a UFSD signal is longer than that of a PIN. In PIN diodes, the

[4] It is not constant, because the electrons are recombining in the meantime: however, electrons generate a tiny fraction of the signal, therefore the signal is not affected much, remaining almost constant.

signal length is determined by the holes drift time while in UFSD by the sum of the electrons and holes drift times. The risetime is also very different: in a PIN diode it is almost instantaneous since it is the time it takes for the ionized $e-h$ pairs to reach their drift velocity, while in UFSD the risetime is equal to the electrons drift time.

It is possible to calculate the number of primary electrons entering the GL in a time dt, and the consequent number of secondary electrons (dN_{gain}) produced, assuming a gain G_I:

$$dN_{gain} \propto n_{e,h} v_{sat} dt \cdot G_I \tag{2.7}$$

and the corresponding current, dI, can be found using equation 2.2:

$$dI = dN_{gain} q \cdot v_{sat} \cdot \frac{1}{d} \propto (n_{e,h} v_{sat} dt \cdot G_I) q \cdot v_{sat} \cdot \frac{1}{d} \tag{2.8}$$

which leads to:

$$\frac{dI}{dt} \sim \frac{dV}{dt} \propto \frac{G_I}{d} \tag{2.9}$$

this result highlights a key aspect: in UFSD, a fast signal slew rate ($\frac{dI}{dt}$) is achieved with a high gain and a thin sensor [5]. The slew rate as a function of the UFSD thickness for different gain levels is presented in figure 2.7.

The gain introduced in the above equations, G_I (the current gain), differs from that defined in equation 2.3 (G): indeed, G (the signal gain) does not consider the trapped charges, which do not undergo multiplication; whereas, G_I is calculated from the measurement of the leakage current, which takes place on a much longer time scale, therefore the charges are already de-trapped and contribute to the multiplication. Hence, $G_I \geq G$. In the following chapters, the signal gain will be mostly used.

In every silicon sensor, the flow of electrons over the pn junction generates a fluctuation of the current commonly known as shot noise. The shot noise is proportional the square-root of the surface and bulk generation current:

$$\sigma_{shot} \propto \sqrt{I_{surface} + I_{generation}} \tag{2.10}$$

The shot noise is generally small in PIN diodes, but it can increase in devices with a large internal gain. The UFSDs operate with a relatively small gain, therefore the shot noise is usually subleading with respect to the electronic noise; however, in heavily irradiated sensors, the leakage current increases significantly, as it will be discussed in

[5] A high electric field is also needed to obtain a slew rate as fast as possible, as it will be illustrated in the following chapters.

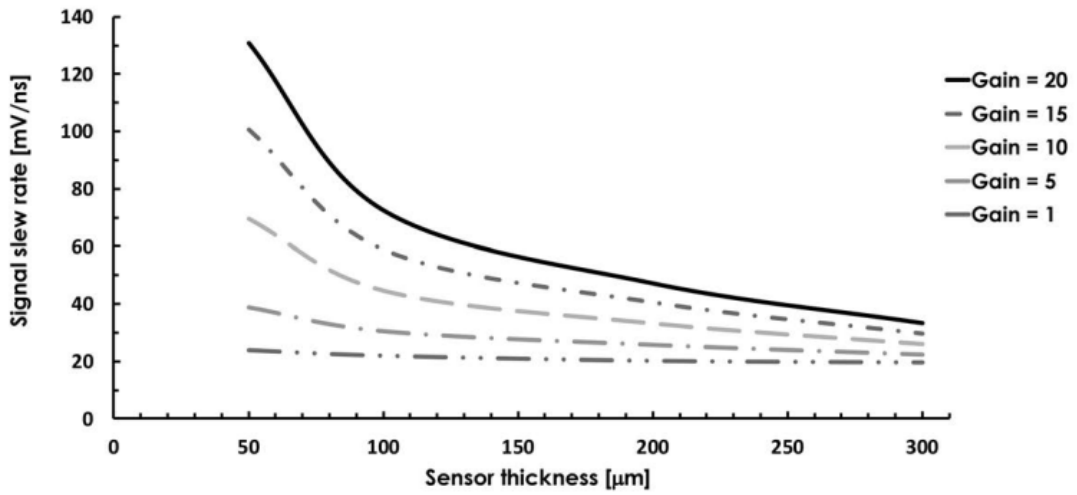


Figure 2.7: Signal slew rate as a function of the UFSD thickness, considering different gain levels.

the following, and the shot noise can become the main source of noise.

A second effect, the so-called excess noise, contributes making the shot noise term larger in the UFSD [44]. The excess noise is an additional noise induced by the multiplication mechanism: each primary electron entering the gain layer generates a number of secondary charges that, on average, is equal to G; however, since the multiplication is a random process, the number of generated secondary electrons can be different from G. This variability increases the shot noise by a factor F, the so-called excess noise factor, which is a function of the gain:

$$F \sim G_I^x \tag{2.11}$$

with x being the excess noise index. When considering also the excess noise factor, the UFSD shot noise becomes:

$$\sigma_{shot} \propto \sqrt{I_{surface} + I_{leak} G_I^2 F} \tag{2.12}$$

In this way, it is possible to derive the equation of the UFSD signal-to-noise ratio (SNR), considering that the UFSD leakage current can be written as:

$$I_{UFSD} \propto I_{PIN} \cdot G_I \tag{2.13}$$

hence:

$$SNR_{UFSD} = \frac{I_{UFSD}}{\sigma_{shot}} \propto \frac{I_{PIN} \cdot G_I}{\sqrt{I_{surface} + I_{leak} G_I^2 F}} \propto \frac{1}{\sqrt{F}} \tag{2.14}$$

This equation demonstrates that, if the sensor is the only source of noise, adding gain would worsen the SNR: in other words, the internal gain increases the noise more than the signal. However, in real life, the sensor noise is subleading with respect to the electronic noise, which is larger and constant, not dependent on the gain: hence, the gain increases the signal while the dominant source of noise remains constant, improving the SNR. That is true up to the gain level where the shot noise becomes the dominant source of noise: at this point, the SNR begins worsening and keeps worsening at higher gain levels, as shown in figure 2.8. This is the reason why UFSDs are operated with a gain of 10-30, a "moderate" gain if compared to other devices with internal multiplication.

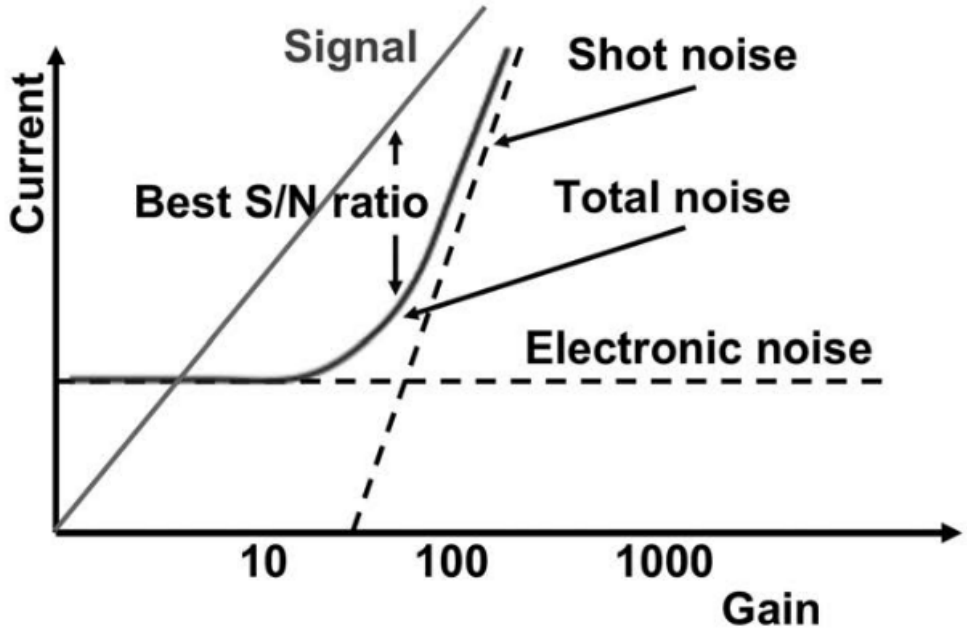


Figure 2.8: Signal and shot noise growth as a function of the sensor internal gain.

A similar discussion can be made also on the effect on the SNR of an external amplifier: it would increase the signal, but the noise would be increased by the same amount, with no improvements on the SNR. This is why reading out a PIN with an external amplifier does not improve its timing performance.

2.4 The ingredients for a timing detector

Figure 2.9 shows a simplified model of a sensor and the associated electronics needed to measure the time of arrival of a particle.

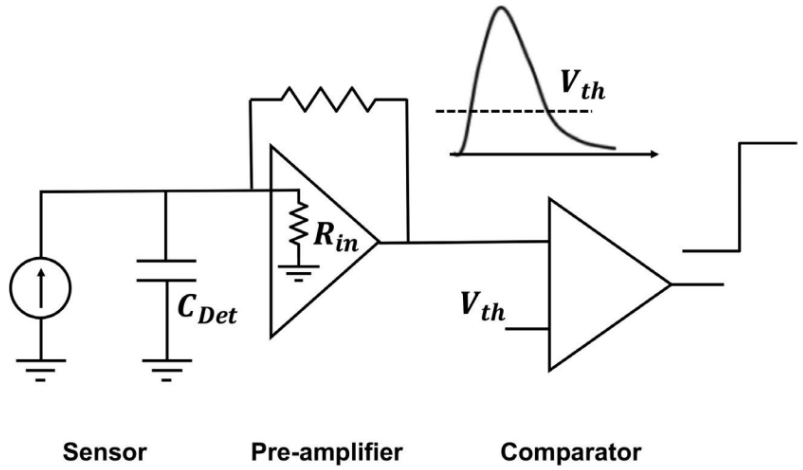


Figure 2.9: Schematic block diagram of a time-tagging detector.

The sensor, modeled as a capacitance (C_{det}) with a current generator (I_{in}) in parallel, is read out by a preamplifier that shapes the signal. A comparator fires when the preamplifier output exceeds a given voltage value (V_{th}). The output of the comparator is digitized by a Time-to-Digital Converter (TDC). Any effect that changes the shape of the signal near the V_{th} value can anticipate or delay the firing of the comparator, smearing the determination of the time of the hit. The time resolution σ_t is given by several terms [45, 46]:

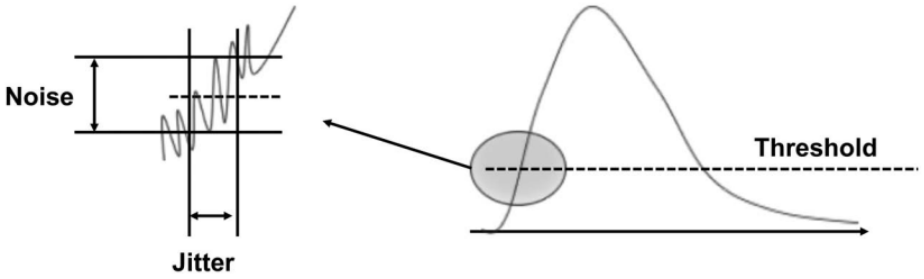


Figure 2.10: The noise causes the early or late firing of the comparator. The uncertainty in time tagging introduced by this effect is called jitter.

- σ_{TDC}: this term describes the uncertainty introduced by the TDC. It can be neglected given that TDCs commonly used in high-energy physics have a resolution lower than 10 ps.

- Jitter: The presence of noise on the signal, either coming from the sensor, added by the preamplifier electronics or on the discriminator threshold V_{th} , shifts the firing time of the comparator to an earlier or later time. The effect is illustrated in figure 2.10. The jitter is given by the ratio between the noise, N and the signal slew rate, $\frac{dV}{dt}$:

$$\sigma_{jitter} = \frac{N}{\frac{dV}{dt}} \tag{2.15}$$

  Assuming a constant slope of the signal, the slew rate can be approximated as the ratio between the signal and the risetime:

$$\frac{dV}{dt} \sim \frac{S}{t_{rise}} \tag{2.16}$$

  therefore, equation 2.15 can be rewritten as:

$$\sigma_{jitter} = \frac{N}{\frac{dV}{dt}} \sim \frac{t_{rise}}{SNR} \tag{2.17}$$

  A low jitter term is achieved by keeping the noise low (in particular keeping the sensor noise below the electronic noise, as previously explained) and increasing the signal slew rate as much as possible. As it will be discussed in detail in the following, in order to increase the slew rate, a high gain, along with a high electric field, are needed.

- Ionization: a charged particle crossing a silicon sensor generates along its path $e-h$ pairs, whose density varies on an event-by-event basis, producing two effects: (i) non-uniform current signals (Landau noise), and (ii) changes in signal amplitude (time walk).

  - The energy deposition of MIPs crossing a thin layer of silicon follows a Landau distribution. Events in the lower tail of the distribution tend to have low and uniform energy-per-unit-length deposits, while events in the upper tail have very large localized deposits. Therefore, the position of an event in the Landau distribution is an indicator of the uniformity of the energy deposition: the higher the energy deposited, the higher the non-uniformity. The Landau noise term (σ_{Landau} in the following) accounts for this variability in the energy deposition.

Thicker devices show a larger σ_{Landau}: this is shown in chapter 6, along with a study on how σ_{Landau} can be reduced, in order to improve the overall time resolution.

- The time walk term arises since larger signals cross a fixed threshold earlier than smaller ones. It can be almost completely corrected by using an appropriate Constant-Fraction Discriminator (CFD) instead of a fixed threshold, i.e. the comparator fires when the signal crosses a certain fraction of the total signal amplitude.

- Distortion: in a silicon sensor, the shape of the current signal is given by Shockley-Ramo's theorem. The two elements, appearing in equation 2.2, determining the time resolution are the drift velocity, v, and the weighting field, E_w, which need to be as constant as possible in order to generate uniform signals (which is key for a good time resolution).

  For what concerns v, it must be saturated everywhere in the sensor, since a non-uniform drift velocity leads to variations in the current signal that depend upon the particle hit position. The electrons drift velocity saturates at an electric field of $\sim$ 30 kV/cm, while the holes velocity never really saturates, but, for fields larger than 50-100 kV/cm, it can be considered constant.

  The uniformity of the weighting field, instead, depends upon the dimension of the electrodes and the sensor active thickness: if the electrodes are much larger than the thickness, E_w is constant almost everywhere in the sensor, equal to $1/d$, with d being the thickness. Moreover, in multi-pads devices, the dimension of the electrodes must be almost equal to the device pitch, see figure 2.11, to ensure good uniformity. Those reasons prevent the development of UFSDs with very small pixels, which would severely affect the timing performance of the sensor; solutions to cope with this issue are being studied: a comprehensive review can be found in [42].

This analysis points out that the optimum sensor geometry resembles as much as possible that of a parallel plate capacitor, with the electrodes much larger than the sensor thickness: under these conditions, the electric and weighting fields are as uniform as they can be, assuring constant drift velocity and coupling between a charge and the read-out electrode. The distortion term can be neglected in sensors with this geometry (i.e. all sensors presented in this work), therefore σ_{jitter} and σ_{Landau} are the most important contributors to the final resolution. Given that the two terms are

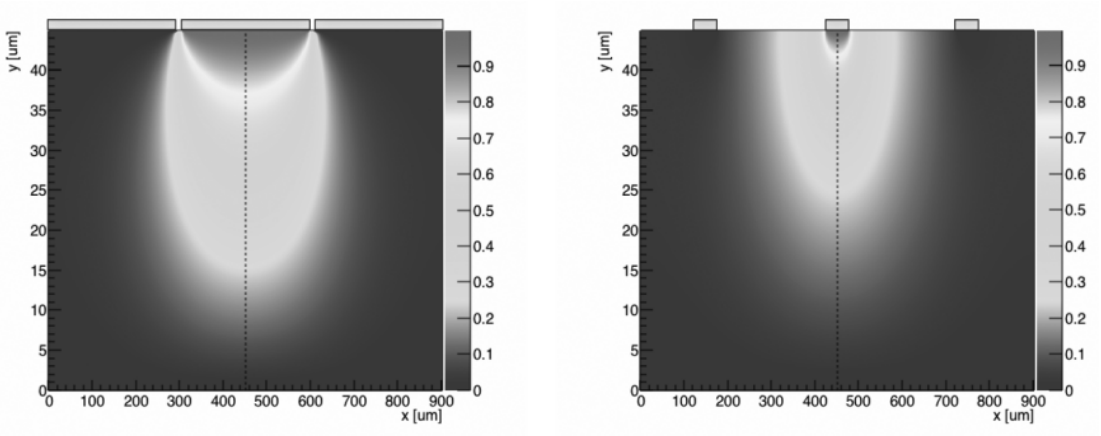


Figure 2.11: The weighting field E_w maps for a 300 μm pitch sensor with an electrode/-gain width of 290 μm (left), and 50 μm (right). In the narrow implant case (50 μm), the weighting field is not uniform along the x-axis. In this condition, the current (t) depends on the impact point and its variability increases the timing uncertainty.

uncorrelated, the UFSD time resolution is well approximated as:

$$\sigma_t = \sqrt{\sigma^2_{jitter} + \sigma^2_{Landau}} \tag{2.18}$$

In a well-designed system, the minimum time resolution is reached when the jitter term is minimized, leaving the Landau noise, which is constant for a given thickness of the sensor, as the dominant contribution to the total time resolution.

2.4.1 The impact of the n^{++} metalized electrode

The UFSD signal forms on the n^{++} metalized electrode and then propagates to the read-out electronics. Due to the n^{++} resistivity, the signal is delayed during the propagation through the electrode; the delay is ~ 0.5 ps/μm: not much, but enough to spoil the time resolution, since signals formed in different parts of the electrode would have different delays.

This is demonstrated in figure 2.12, which presents a 2d-map (right), obtained with the TCT laser setup described in chapter 4, with the signal Time-of-Arrival (ToA) of a 2×2 mm^2 UFSD represented in the colored axis. The tested region is framed in green in the picture of the sensor (left in the figure): it is limited to the non metalized region of the device, since the laser used for the test is absorbed elsewhere. The map clearly shows that the signal is more delayed (higher ToA) in the central part of the tested region, which is further away from the metalization, while signals originated closer to the metalization have a lower delay. The same effect is reported in [47].

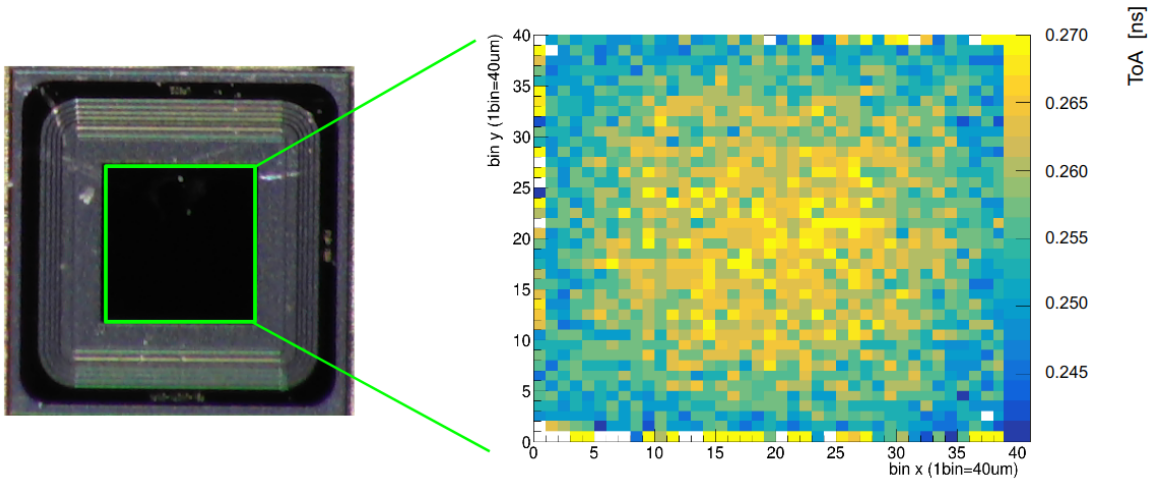


Figure 2.12: A 2d-map (right) obtained with the TCT laser setup described in chapter 4, illustrating the signal Time-of-Arrival (ToA) in a a 2×2 mm^2 UFSD (left). The tested region is framed in green. The map shows that signals originating far from the metalization are more delayed.

In order to fix this issue, a metalization is added on top of the device, separated from the n^{++} electrode by an oxide, since metal does not yield significant delays during the signal propagation.

In the UFSD design, the metalization directly contacts the n^{++} electrode only at the edges of the pad, because in such regions the electrode is thicker due to the presence of the JTE: this solution prevents from the risk of puncturing the electrode when opening the electrical contacts during the fabrication process, which would cause the shorting of the metal with the p-bulk.

The UFSD readout, therefore, is a mixed AC- and DC-coupled system: the signal is AC-coupled to the metal to avoid propagation delay, and the n^{++} electrode and the metal are in contact at the edge of the pad to avoid a bipolar signal, typical of an AC-coupled readout scheme.

All UFSDs for ETL will be fully metalized.

2.5 Multi-pads UFSD

The UFSDs covering the $\sim$ 14 m^2 of the two ETL endcaps will be multi-pads arrays, for a total of $\sim$ 8.5 million channels. In particular, the final sensor module will be a 16×16 array with 1.3×1.3 mm^2 pads, for a total surface of 21×21 mm^2. The ultimate target of the UFSD project is, therefore, the development of large-area sensors with excellent timing performance, even after heavy irradiation.

The main aspects to be considered are the arrays stability, their uniformity in terms of delivered performance, and the sensitivity to premature breakdown of the region between the pads (inter-pad region). The measurement campaign that led to the definition of the design of the ETL multi-pads UFSDs is reported in detail in chapter 5.

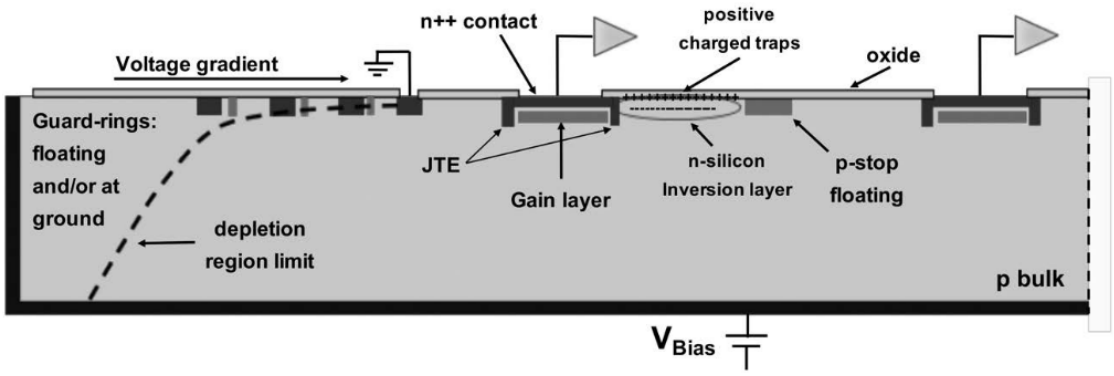


Figure 2.13: Cross cut of a multi-pads UFSD (not to scale) with a schematic view of the building blocks of the device. From the device physical edge: guard-rings, pad with JTEs, inter-pad region with p-stop. Taken from [42].

Figure 2.13 shows a simplified cross-section of a portion of a multi-pad UFSD (not-to-scale). At the physical edge of the sensor, there are the guard-rings.

The guard-rings have the task of grading the voltage from the sensor edge, held at the bias voltage, to the first read-out pad, held at virtual ground by the read-out electronics. Each guard-ring consists of an n^{++} doped implant, equipped with metal field plates. A p^{+} implant (p-stop) is (in most designs) interposed between each pair of guard-rings, with the outer ones left floating, and the inner one generally grounded in order to collect the leakage current generated outside the core region of the device. It is also possible to leave floating even the inner guard-ring: in this configuration, the pads next to the guard-ring will have a higher current.

The gain region, as shown on the left side of figure 2.13, is surrounded by a deep n^{++} implant called Junction Termination Extension (JTE), equipped with a metal field plate. The JTE is located around each pad and it ensures that the $e-h$ pairs generated by particles impinging in the region between the pads do not reach the gain layer. When the impact point is where the gain layer is implanted, the electrons initiate the multiplication mechanism without delay; on the contrary, if the $e-h$ pairs are generated in the inter-gap region, the electrons will have to drift to the gain implant, and the multiplication process will start with a considerable delay, see figure 2.14 (left). Considering a drift velocity of 10 ps/μm, the multiplication can easily be delayed by

hundreds of ps, causing a completely wrong assignment of the particles time of arrival. This problem can be avoided by inserting the JTE at the periphery of each pad, as shown on the right part of figure 2.14. Therefore, the JTE structure delimits the active area of the sensor to the regions where the gain implant is present.

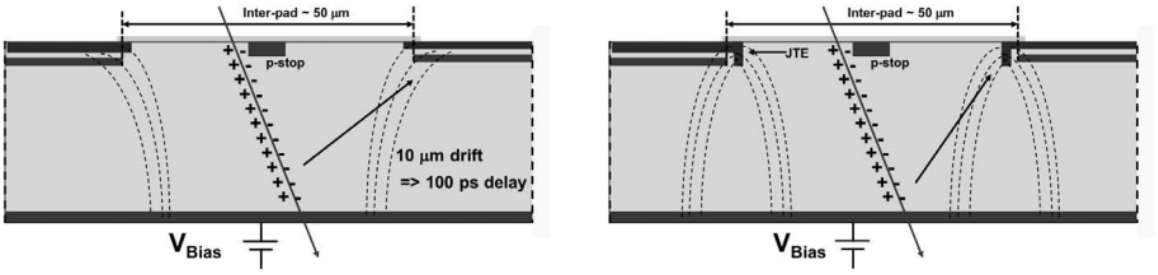


Figure 2.14: Sketch of the inter-pad region. The dashed lines show the $e-h$ pairs drift lines. In a design without JTE (left), the $e-h$ pairs generated in the inter-pad will reach the gain layer after a long drift. On the contrary, when the JTE is present (right), these charges are collected without reaching the gain layer. Taken from [42].

Common to every silicon sensor using the n-in-p design, the n^{++} implants need to be isolated from each other by an extra p^{++} implant, the so-called p-stop. The p-stop has a doping concentration in the range 10^{15}–10^{17} atoms/cm^3 . Without p-stops, the *inversion layer* created by the positive charges at the Si-SiO_2 interface would short together all n^{++} implants. The JTE and p-stop terminations introduce a no-gain region, tens of μm wide, between adjacent pads. This no-gain region, which is the main subject of chapter 5, decreases the fill factor of the device, therefore it needs to be kept as narrow as possible ensuring, at the same time, stability and the proper operation of the UFSD arrays.

2.6 Effect of irradiation on UFSDs

The UFSDs instrumenting the ETL will be required to deliver excellent timing performance up to the end of the HL-LHC lifetime, when they will have received a radiation fluence of $1.6 \cdot 10^{15}$ n_{eq}/cm^2.

A particle impinging on a silicon sensor loses energy via either ionizing or non-ionizing processes. The radiation damage induced can be classified into two distinct categories: surface damage, due to ionization, and bulk damage, due to the non-ionizing processes. In both cases, the amount of damage produced depends upon the type and energy of the impinging particles.

The ionizing processes are responsible for the surface damage, which causes a large number positive charges to be trapped at the Si-SiO_2 interface (responsible for the *inversion layer* previously mentioned). On the contrary, the non-ionizing processes occur when a hadron interacts with the silicon atoms of the crystal lattice, producing silicon interstitials (Si_i) and vacancies (V), called Frenkel pair. A fraction of Frenkel pairs recombine, causing no damages, while the remaining interstitials and vacancies migrate through the lattice and react with other impurities present in the silicon bulk producing point defects.

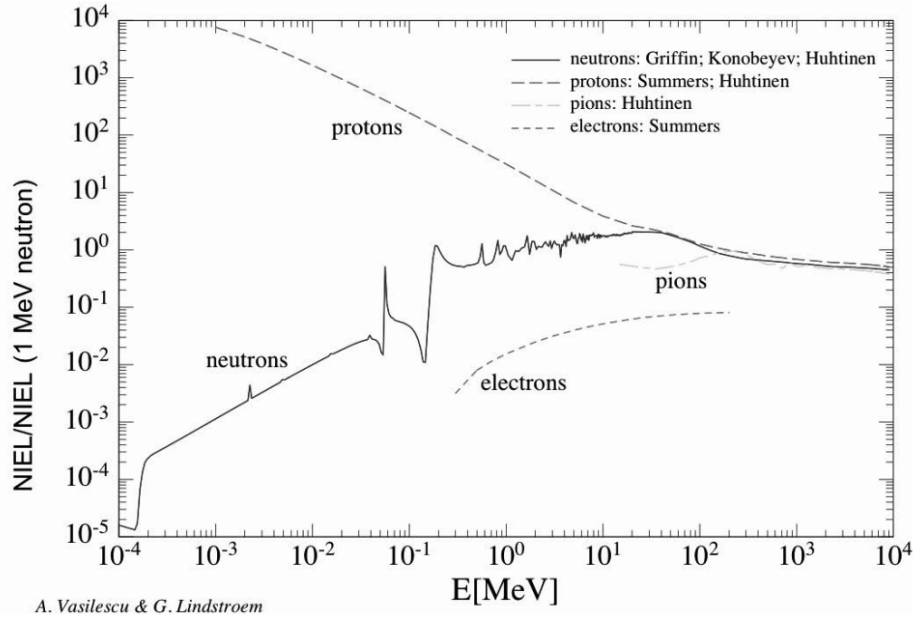


Figure 2.15: Relative $NIEL$ factor in silicon for neutrons, protons, pions, and electrons, as a function of the particle energy. The $NIEL$ factor of 1 MeV neutrons is used as normalization value.. Taken from [29].

The energy lost by the impinging particle which is not due to ionizing processes is called *Non Ionizing Energy Loss* ($NIEL$) [29]. A fraction of this energy produces lattice excitation, while the other fraction is responsible for bulk damages. In order to compare the effects induced by different particles (in type or energy), the hypothesis that the radiation damage scales with their $NIEL$ factors was introduced ($NIEL$ *scaling hypothesis*). The reference $NIEL$ value has been chosen to be that of 1 MeV neutrons. Using the $NIEL$ hypothesis, the studies carried out on irradiated devices are compared to each other by converting each fluence to their *1 MeV neutron equivalent* fluence.

Figure 2.15 presents the $NIEL$ factors in silicon for different particles: neutrons and protons have very different factors at low energies, where the displacement damage

caused by protons is much larger due to their very high energy loss per unit length (dE/dx), while the factors become similar above 50 MeV.

The first radiation-related effect is the leakage current increase. Indeed, as reported in [42], the leakage current of a PIN diode increases with irradiation since the lattice defects that are created by non-ionizing radiation can be generator centers of $e-h$ pairs; in the UFSD, such increase is particularly important, given that the leakage current generated in the UFSD bulk is multiplied by the gain, G, causing a significant increase with respect to an equivalent PIN diode, with a direct impact on the power consumption of the device.

This unavoidable effect can be kept under control by reducing the sensor thickness (the leakage current is proportional to the sensor's volume) and operating the device at low temperature.

Another important consequence of irradiation is the variation of the doping concentration, N_{eff}, as a function of the 1 MeV neutron equivalent fluence, Φ_{eq} [48–52]:

$$N_{eff} = N_{D_0} e^{-c_D \Phi_{eq}} - N_{A_0} e^{-c_A \Phi_{eq}} - g_{eff} \Phi_{eq} \tag{2.19}$$

with N_{D_0} (N_{A_0}) and c_A (c_D) being the donor (acceptor) concentration and the donor (acceptor) removal coefficients; $g_{eff} = 0.02$ cm^{-1} is the coefficient of proportionality between the fluence and the density of new acceptor-like defects.

The equation describes, with the first term, the donor removal, while the second one describes the acceptor removal, and the third stands for the acceptor creation.

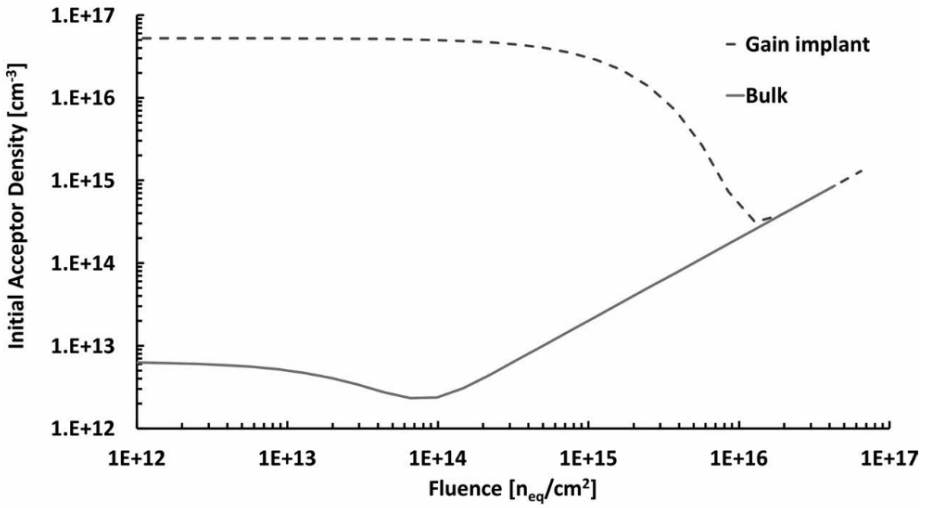


Figure 2.16: Evolution of the boron density for gain implant and p-doped bulk, computed using equation 2.19, as a function of the irradiation fluence, with initial boron densities of $3\cdot 10^{16}$ atoms/cm^3 and $5\cdot 10^{12}$ atoms/cm^3, respectively. Taken from [42].

Both the acceptor removal and the acceptor creation mechanisms apply to the gain implant and the bulk of a UFSD (both p-doped).

Figure 2.16 shows the evolution of the boron concentrations as a function of fluence in the gain implant and bulk, with initial boron densities of $3{\cdot}10^{16}$ atoms/cm^3 and $5{\cdot}10^{12}$ atoms/cm^3, respectively. At fluences higher than $1 \cdot 10^{16}$ n_{eq}/cm^2, the acceptor density of the gain implant matches the bulk acceptor density, indicating complete removal of the initial gain implant doping.

Using equations 2.19 and 2.1, it is possible to compute the full depletion voltage as a function of the irradiation fluence for either the gain layer and the bulk: this is shown in figure 2.17. The plot underlines as the acceptor removal is leading in the gain implant, whereas the acceptor creation dominates in the bulk.

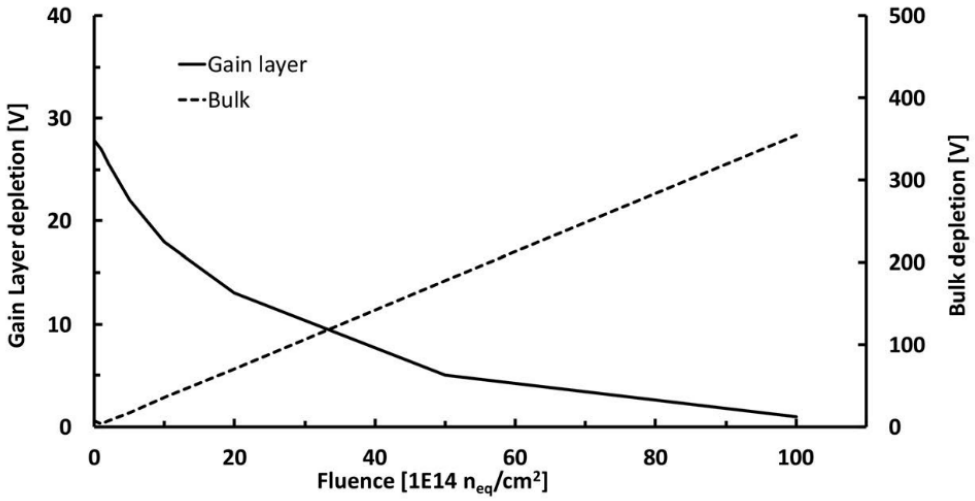


Figure 2.17: Evolution of the depletion voltage of the gain layer and bulk as a function of fluence, for a 50 μm-thick UFSD. Taken from [42].

The acceptor removal is the process affecting most UFSD operation, because it is the leading process in the GL. As discussed in detail in [53], a solution to slow it down is to implant carbon in the gain layer volume. Carbon replaces boron in becoming interstial, reducing its deactivation [54–56]. Moreover, carbon atoms in interstitial position tend to pair with boron interstitials and form centres with energy approximately 80% of the boron acceptor level energy, further recovering from the boron deactivation. The impact of carbon on ETL sensors is discussed in chapter 6.

Another radiation-related effect is charge trapping. In irradiated sensors, the charge carriers are subject to trapping, a mechanism that leads to charge collection decrease and affects the output signal shape. Trapping increases with the irradiation fluence and with the drift time of the charge carriers. Figure 2.18 shows the signal (compared

to the case where no trapping is present) for the irradiation levels $1 \cdot 10^{15}$ n_{eq}/cm^2 and $2.5 \cdot 10^{15}$ n_{eq}/cm^2: even at the highest fluence, the signal shape does not change dramatically, and, interestingly, the signal rising edge, which is key for a good time resolution, does not change.

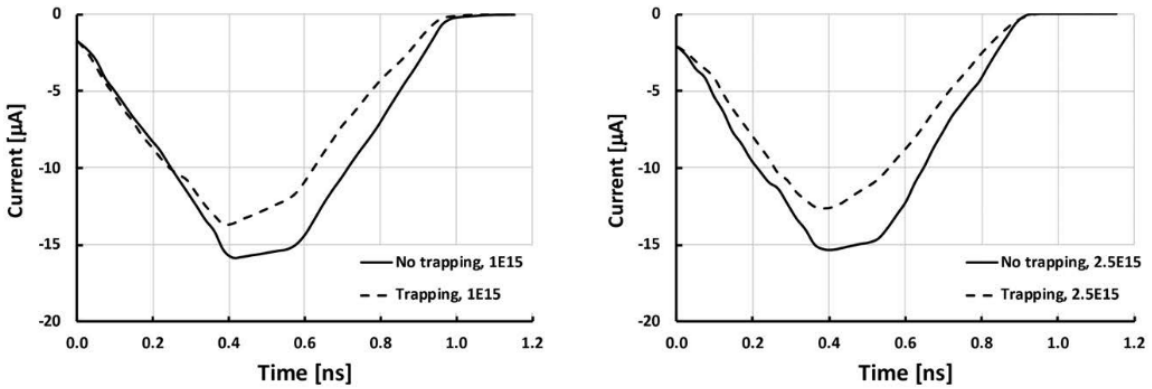


Figure 2.18: Weightfield2 simulation of the signal shape from a 50 μm-thick UFSD, with and without trapping. Left: current signals with and without trapping at $\Phi = 1 \cdot 10^{15}$ n_{eq}/cm^2 . Right: current signals with and without trapping at $2.5 \cdot 10^{15}$ n_{eq}/cm^2. Taken from [42].

Due to the acceptor removal effect, which causes a deactivation of the boron doping and, consequently, of the electric field in the gain layer, the gain in UFSD decreases rapidly with increasing fluence. This field loss can be compensated by raising the bias voltage, as shown in the Weightfield2 [43] simulation of figure 2.19. In this example, before irradiation, the gain layer generates about 90% of the electric field. After a fluence of $8{\cdot}10^{14}$ n_{eq}/cm^2 , this fraction is reduced to about 60%, and at $1.5{\cdot}10^{15}$ n_{eq}/cm^2 becomes 45%.

When the bias voltage reaches about 500–750 V over 50 μm ($E \sim$ 100–150 kV/cm), the multiplication mechanism also begins in the bulk, an effect called bulk gain. Bulk gain has a very sharp turn-on, and it leads rapidly to a breakdown condition.

2.6.1 Gain layer depth

Another parameter that affects the gain recovery mechanism is the depth of the gain layer. As previously explained, the same gain level in UFSDs with shallow or deep gain implants is achieved at different electric fields and, consequently, with different mean free paths λ. The value of the electric field in a shallow (deep) gain implant is $\sim$ 400 kV/cm ($\sim$ 300 kV/cm).

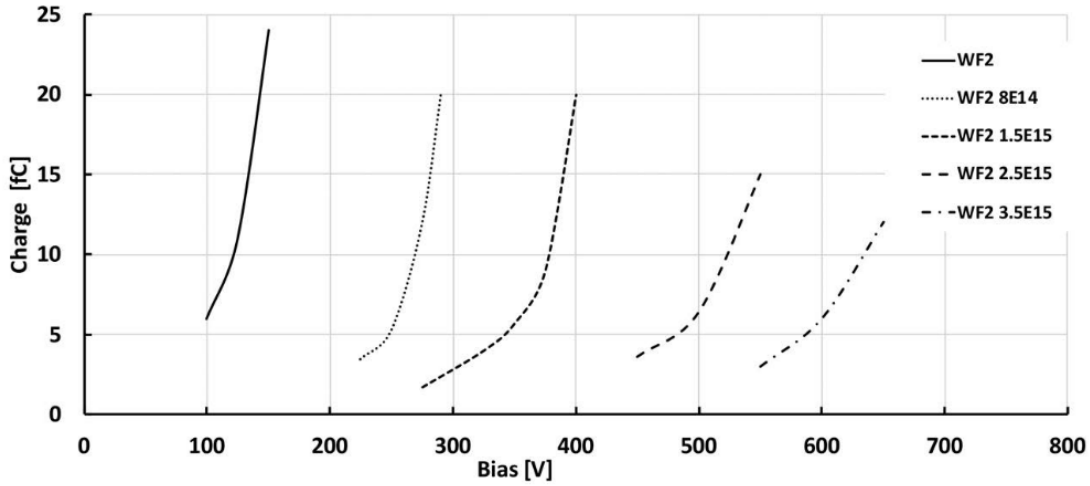


Figure 2.19: Example of the evolution of the collected charge in a given 50 μm-thick UFSD, as a function of bias voltage and for different values of the irradiation fluence (Weightfield2 simulation). Taken from [42].

As the field generated by the gain layer drops due to irradiation, λ increases, lowering the gain. Moreover, when the radiation fluence raises, the mean distance between lattice defects in the sensor is lowered [57]. Defects are scattering centers that slow the carriers down: the larger their number, the more slowly would the carriers move; if the mean distance between two defects is smaller than λ, multiplication in the sensor cannot happen.

The gain can be restored by raising the bias voltage, which in turn raises the electric field within the sensor, shortening λ. However, the effect on λ of a given bias increase is not equal for a shallow or deep gain implant since the electric fields are not the same.

Figure 2.20 presents λ and $d\lambda/dE$ as a function of the electric field: the deep gain layer is operated at a lower electric field and therefore features a higher $d\lambda/dE$. A higher $d\lambda/dE$ leads to a higher recovering power with bias, meaning that the bias increase needed to compensate for the effects of a given radiation fluence (i.e. to restore the value of λ before irradiation) is smaller for the device with a deep gain implant. Such an important aspect is further discussed in chapter 6.

2.7 ETL requirements for the UFSDs

Some of the most important requirements for the UFSDs to be installed in the ETL are listed below:

- Single-hit time resolution for the bare UFSD (no bump-bonded ASIC): 30-40 ps. To be maintained up to the end of the ETL lifetime.

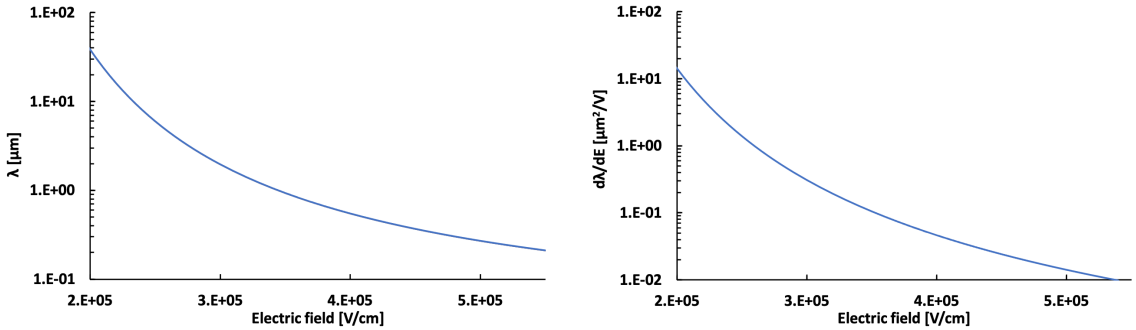


Figure 2.20: Electrons mean free path λ between two subsequent scattering events producing secondary charges (*left*) and $d\lambda/dE$ (*right*) at 300 K as a function of the electric field E, according to the Massey impact ionization model [58].

- Maximum expected fluence: $\Phi = 1.5 \cdot 10^{15}$ n_{eq}/cm^2.
- Delivered Charge > 5 fC up to the end of the ETL lifetime.
- Module arrays with 21×21 mm^2 area, corresponding to 16×16 matrix with 1.3×1.3 mm^2 pads (3-4 pF capacitance per pad).
- Inter-pad gap width $\leq$ 65-70 μm.
- Inter-pad resistance [6] ≥ 0.1 GΩ.

The measurement campaigns presented in the following chapters address such requirements and attempt to define the optimal UFSD design for ETL, focusing in particular on the improvement of the inter-pad region (chapter 5) and the gain layer (chapter 6).

[6] The measurement of this quantity is explained in chapter 5.

Chapter 3

UFSD productions

The first production and measurements of LGADs were presented by Centro Nacional de Micro-electronica (CNM) in Barcelona, Spain in 2014 [59]; CNM itself manufactured the first production of UFSDs (active thickness of 50 μm) in 2016 [28].

Hamamatsu Photonics (HPK) in Japan was the second vendor [60]. HPK presented the results of its first production, ECX20840, at the TREDI 2017 conference [61], which was then followed by the EXX298995 (2018, also called HPK1), EXX30327-EXX30328-EDX30329 (3 small batches of UFSDs), and HPK2 (2019) productions.

Finally, FBK [62] manufactured its first production of LGADs in 2016 [63], followed by four other UFSD productions: UFSD2 (2017), UFSD3 (2018), UFSD3.1 (2019), UFSD3.2 (2020).

Experimental measurements and results discussed in this work involve sensors from the HPK1 and HPK2 productions, for what concerns HPK, and from the FBK UFSD3, UFSD3.1, and UFSD3.2 productions. Their main features are presented below.

3.1 HPK ECX20840

HPK, in 2017, completed the ECX20840 run. This production comprises 16 high-resistivity p-type wafers, with active thicknesses of 50 and 80 μm. Four different p-gain doping concentrations have been implanted: split "A" corresponds to the lowest dose, "D" to the highest one. This production aimed at achieving very good time resolution. For this reason, very simple device geometries (single-pad and 2×2 arrays) have been implemented in the wafer layout.

3.2 HPK EXX28995 (HPK1)

EXX298995-HPK1 was the first production oriented to the R&D of the ATLAS and CMS timing detectors, aiming to demonstrate the feasibility of segmented large area UFSDs.

HPK1 consists of twenty, 6-inch, p-type wafers with a high-resistivity bulk and an active thickness of ~ 45 μm. A single p-gain dose is implanted. Two designs of the gain layer have been used: one shallow and wide (Type 3.1, wafers 1-10), and one deeper and narrower (Type 3.2, wafers 11-20). Only type 3.1 has been used in this work.

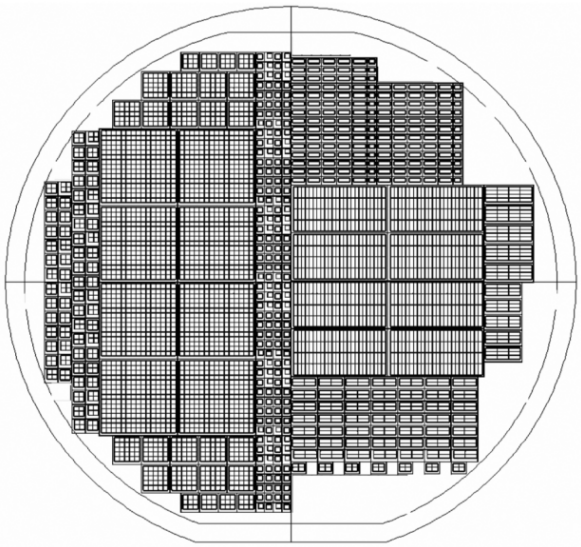


Figure 3.1: HPK1 wafer layout. The left block consists of sensors for ATLAS, while the CMS block is on the right.

The wafer layout of this production (figure 3.1) is equally divided into two main blocks, one for ATLAS (left block) and one for CMS (right block). The CMS devices (the interesting ones in this work) consist of single pads, 2 × 2, 3 × 3, 4 × 4, and 4 × 24 arrays (pad size of 1 × 3 mm^2). The single pads only have been considered in this work.

3.3 HPK2

HPK2 is the second HPK production focused on the R&D for the ATLAS and CMS timing detectors. It features sixteen wafers with two layouts (figure 3.2): one with small sensors to study the radiation resistance and the inter-pad design (8 wafers), and one with large sensors to study the uniformity (8 wafers).

The small sensors layout includes single pad devices and arrays of 2 × 2, 3 × 3, and 5 × 5 pads (pad-size of 1.3 × 1.3 mm^2). The large sensors layout includes arrays of 5 × 5, 8 × 8, 15 × 15, 16 × 16, 30 × 15, and 32 × 16 pads (pad size of 1.3 × 1.3 mm^2).

This production has a deep and narrow multiplication layer, implanted in a high-resistivity bulk. Four doses of the gain implant have been implemented: split 1 has the most doped gain layer, whereas split 4 is the least doped one. Split 1 is designed to have a breakdown voltage at room temperature of about 160 V, whereas for split 4 is about 240 V. Only split 4 single pads have been measured in this work.

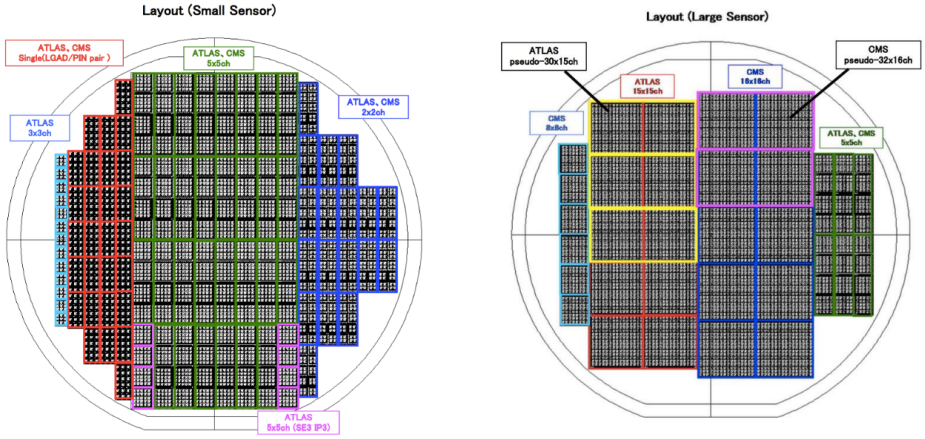


Figure 3.2: Left: HPK2 layout featuring small sensors. Right: HPK2 layout featuring large sensors.

Regarding the inter-pad strategy, the production implements four designs, named IP3, IP4, IP5, IP7. IP3 has the narrowest inter-pad, whereas IP7 is the widest. Two distances between the gain region and the edge of the sensor have been implemented, SE3 (300 μm) and SE5 (500 μm).

3.4 FBK UFSD3

The third UFSD production by FBK, named UFSD3, has been designed to investigate specific features such as the radiation hardness, the feasibility of a narrow inter-pad region, and the uniformity of segmented, large-area sensors.

This production consists of twenty 6-inch wafers (active thickness 55 μm), sixteen of which with a Float-Zone (FZ) substrate (similar to the previous FBK production, UFSD2), while the remaining four with an Epitaxial (Epi) one.

There are four gain layer configurations implemented in UFSD3: Boron High and

Low Diffusion (B HD/LD) and Boron High and Low Diffusion with co-implantation of carbon (B HD/LD + C).

The carbon enrichment has been implemented in UFSD3 because the previous UFSD2 production demonstrated the mitigation of radiation damage in UFSD with a carbonated gain layer [53].

The wafers of this production have five splits of gain dose in steps of 2% (1.00 is the dose reference, the same as UFSD2) and they have four splits of carbon dose labeled with A, B, C, and D. Carbon dose A corresponds to the lowest dose in UFSD2, while the next three correspond to x2, x3, and x5 dose A, respectively. The wafers of UFSD3 are listed in table 3.1.

Table 3.1: Wafers of the UFSD3 production

Wafer number	Substrate	Gain dose	Carbon dose	Diffusion
1	FZ	0.98		Low
2	Epi	0.96		Low
3	FZ	0.96	A	Low
4	Epi	0.96	A	Low
5	FZ	0.98	A	Low
6	FZ	0.96	B	Low
7	FZ	0.98	B	Low
8	FZ	0.98	B	Low
9	FZ	0.98	C	Low
10	FZ	1.00	C	Low
11	FZ	1.00	D	Low
12	FZ	1.02		High
13	Epi	1.00		High
14	FZ	1.02	A	High
15	Epi	1.00	A	High
16	FZ	1.02	B	High
17	FZ	1.02	B	High
18	FZ	1.04	B	High
19	FZ	1.02	C	High
20	FZ	1.04	C	High

UFSD3 has been produced using the innovative stepper technology for the lithographic process, instead of the mask-aligner technique previously used. The stepper technology uses an area 25 $\times$ 19 mm^2 called reticle, which is repeated several times on the surface of the wafer: it is substantially different from the mask aligner used in UFSD2, where a single pattern covers the entire surface of the wafer.

The strength of the stepper is the great spatial precision, which allows creating sharp implant borders and reducing the distance between implants with respect to previous

productions. Moreover, the stepper allows the photocomposition of the devices to be performed, that is, to compose a device combining images from different exposures. This technique is necessary to produce sensors with an area larger than the maximum area that can be covered in a single exposure by the stepper: that is key for the final full-size CMS sensors for ETL.

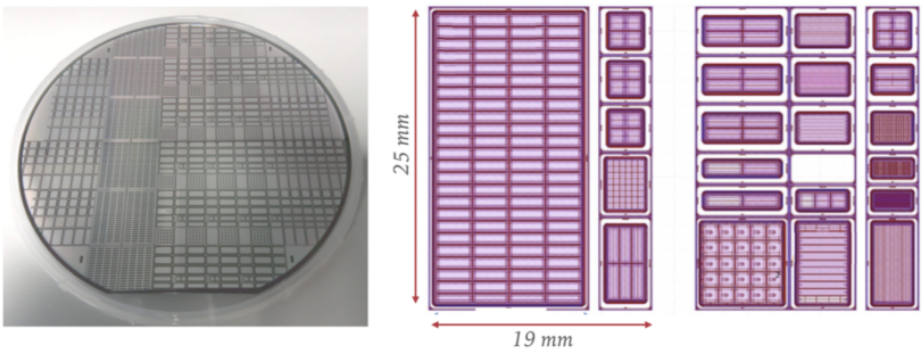


Figure 3.3: Left: picture of a UFSD3 wafer. Right: layout of the two UFSD3 reticles.

Finally, in UFSD3 different innovative strategies for the gain layer termination implants (JTE and p-stop) have been pursued, to minimize the no-gain region between pads in multi-pads sensors. Four different inter-pad designs have been implemented: Aggressive, Intermediate, Safe, and Super-Safe, with a nominal distance between gain layers of ~ 10 μm, ~ 20 μm, ~ 30 μm, and ~ 40 μm, respectively. All inter-pad designs feature an inter-pad nominal width lower than 50 μm, which is the ETL target, and significantly lower the inter-pad distance with respect to UFSD2 (70 μm), thanks to the innovative strategies employed by FBK.

3.5 FBK UFSD3.1

The UFSD3.1 production was developed to study the effect of different inter-pad designs and p-stop doses and the capability of sensors to hold high bias voltages. The aim was to improve the pad isolation in multi-pad arrays, an aspect that caused some issues in the previous production, UFSD3, as discussed in chapter 5.

UFSD3.1 consists of seven 6-inch epitaxial wafers with a high-resistivity bulk and an active thickness of 55 μm. All wafers have a shallow B HD gain layer with the same p-gain dose and without carbon enrichment. Only 2x2 pads matrices are present, with pad size of 1.3x1.3 mm^2. UFSD3.1, as UFSD3, has been produced using the stepper lithographic technology.

The production comprises six wafers with six different p-stop dopings (table 3.2),

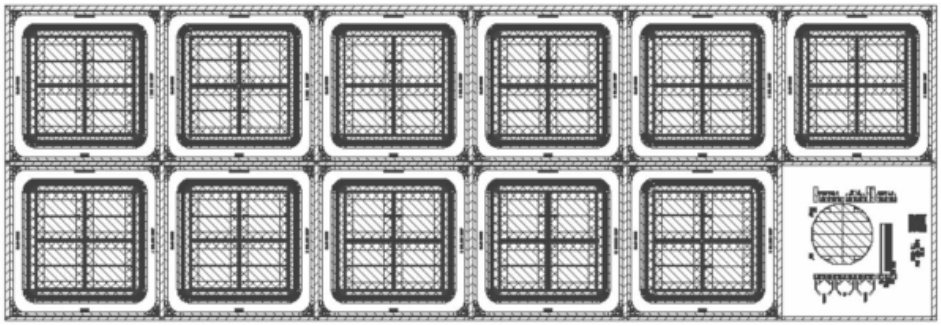


Figure 3.4: UFSD3.1 reticle.

reported in arbitrary units (a.u.), with doping 1 being the UFSD2 and UFSD3 reference p-stop dose.

Table 3.3 reports the 11 different inter-pad designs implemented in UFSD3.1 with the respective nominal inter-pad distances. The area of the regions called *Region A* and *Region B* are also reported in arbitrary units, along with their design.

Region A and *Region B* are the name of those regions, in the UFSD3.1 design, where the p-stops intercept. In UFSD3.1 2x2 arrays, the possible intersections are:

1. The matrix central region, where 4 p-stops join, named *Region A*.

2. The regions where 2 or 3 p-stops join. It will be called *Region B* in the following.

Those regions are important because they have an impact on the sensors' breakdown, as discussed in chapter 5. In the same chapter, a picture of *Region A* and *Region B* is also presented.

A variety of designs are implemented for both regions:

1. The *grid* design: the p-stops intersect creating a cross-shaped structure with a small area.

2. The *full circle* design: the p-stops join and create a circular structure with a large area and smooth corners.

3. The *empty circle* design: a circular structure equivalent to the previous one, but empty in its central part, in order to reduce the area.

4. Type 9 is the UFSD3 Super-safe design, with a smaller curvature radius of the p-structures in the corners. It is the only design with a double p-stop.

5. Type 10 implements the so-called grid guard-ring, namely a guard-ring structure that, differently from the standard UFSD design, not only surrounds the active area, but also separates the JTEs of neighbouring pads.

The grid structures have all the same area and sharp corners, whereas the circles are designed with different sizes: their area is mainly determined by the radius of curvature of the *p*-stops at the intersection, with large curvature radii resulting in large areas and smooth corners and small radii yielding small areas and sharper corners.

Because of their layout, the *p*-stops in types 9 and 10 never intercept, therefore *Region A* and *Region B* are not defined.

Table 3.2: Wafers of the UFSD3.1 production

Wafer number	*p*-stop doping [a.u.]
W12	0.02
W13	0.05
W14	0.1
W16	0.15
W17	0.2
W18	1

Table 3.3: Inter-pad design of the UFSD3.1 arrays.

Type	Nominal inter-pad width [μm]	*Region A/B* design	*Region A* area [a.u.]	*Region B* area [a.u.]
1	16	Grid	A	2B
2	20.5	Full	100A	10B
3	20.5	Full	10A	B
4	23.5	Grid	2A	2B
5	25	Full	20A	2B
6	27.5	Full	10A	B
7	27.5	Full	20A	2B
8	27.5	Grid	2A	2B
9	38	2 *p*-stops		
10	49	Grid guard-ring		
11	25	Empty	100A	10B

3.6 FBK UFSD3.2

The UFSD3.2 gained from the knowledge acquired from the previous productions, both in terms of radiation resistance and inter-pad strategy. The main focuses of the

production are the optimization of the gain layer, to maximize the radiation resistance and the finalization of the inter-pad design.

The production consists of nineteen, 6-inch, p-type epitaxial wafers, with a high-resistivity bulk and active thicknesses of $\sim$ 55 μm(3 wafers), $\sim$ 45 μm(14 wafers), $\sim$ 35 μm(1 wafer). A $\sim$ 25 μm-thick wafer is also present in the production but was not included in this work.

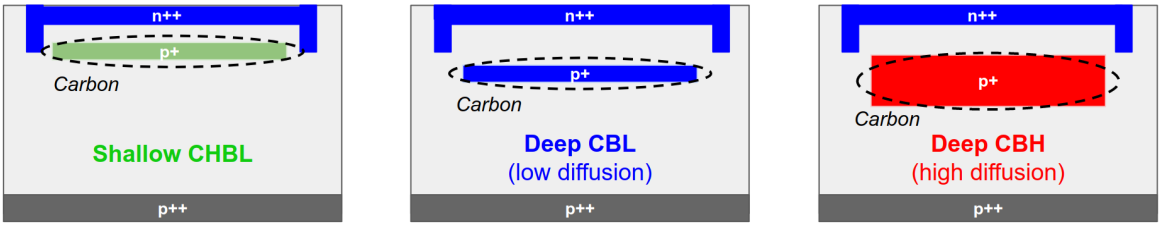


Figure 3.5: The different types of gain implants of the UFSD3.2 production.

Several variations of the gain implant design have been implemented in this production (see figure 3.5): (i) carbonated shallow gain implant diffused at high and low thermal load (Shallow B HD/LD + C); (ii) carbonated deep gain implant diffused at high and low thermal load (Deep B HD/LD + C). Prior to UFSD3.2, FBK only used shallow gain implants: in this geometry, the high electric field region is narrower than 1 μm; whereas, for the deep implants, the high-field region extends for about 1.5-2 μm. A detailed discussion on shallow and deep gain implants can be found in chapter 2.

The amount of carbon infusion was varied to map the performances as a function of the amount of carbon: for the shallow implants, the carbon doses (in arbitrary unit) 0.4, 0.8, 1 were used, while for the deep implants, the carbon doses 0.6 and 1 were used. Five different boron doses (again in arbitrary units, with 1.00 being the UFSD2 reference dose) were also used: 0.98, 0.94 in the shallow implants; 0.70, 0.74, and 0.78 in the deep ones. The characteristics of each measured wafer are reported in table 3.5.

FBK activated carbon and boron implants using different diffusion schemes: (i) for the shallow implants, carbon is implanted and activated at high thermal load, then boron is implanted and activated at low thermal load, a scheme called CHBL. (ii) Carbon and boron implants are activated together at low or high thermal load (CBL or CBH): deep implants have been activated with both schemes.

The production features either single pad and 2 $\times$ 2 arrays, as well as PIN-LGAD pairs, i.e. 2 $\times$ 1 array where one device is an LGAD and one is a standard PIN diode with no gain implant (figure 3.6). 5 $\times$ 5 arrays are also present. The lithographic stepper technology has been used.

The 2 $\times$ 2 arrays feature nine different types, which correspond to UFSD3.1 types 1,

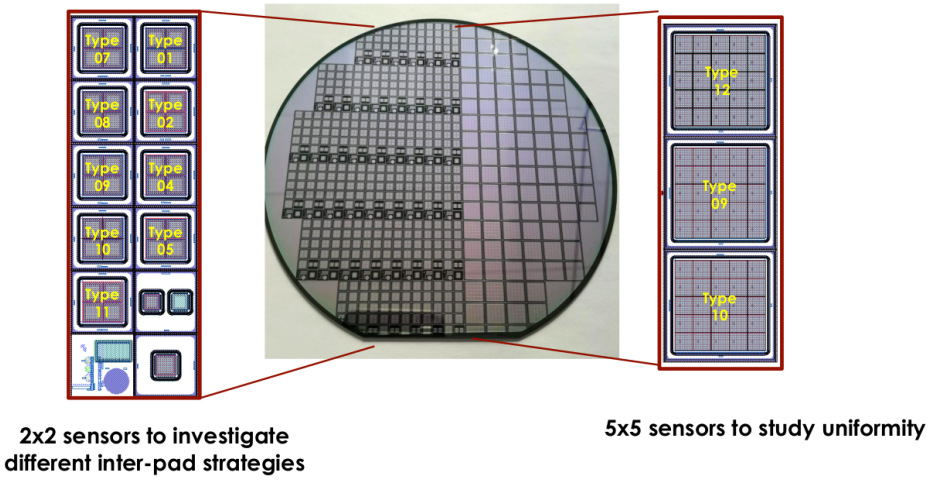


Figure 3.6: Picture of a UFSD3.2 wafer. Both small and large sensors layouts are shown.

2, 4, 5, 7, 8, 9, 10, 11; whereas the 5×5 arrays have been produced with the inter-pad layouts that have been considered particularly robust, according to the results on the previous productions:

1. Type 9, which is the UFSD3 Super-safe design.
2. Type 10, same as UFSD3.1. The grid guard-ring provides a significant improvement in the inter-pad design. The downside is a rather large width (measured to be ~ 62 μm).
3. Type 12 is an improved version of the UFSD3.1 type 4, with an interpad nominal width of 28 μm. It is the most aggressive design for what concerns the 5×5 matrices.

The p-stop implant dose in UFSD3.2 is 0.1 a.u., same as UFSD3.1 Wafer 14. Such choice, which avoids the sensors premature breakdown, is the result of the measurement campaign presented in chapter 5.

Table 3.4: Inter-pad design of the UFSD3.2 devices.

Device size	Type	Nominal inter-pad width [μm]
2×2	1	16
	2	20.5
	4	23.5
	5	25
	7	27.5
	8	27.5
	9	38
	10	49
	11	25
5×5	9	38
	10	49
	12	28

Wafer	Thickness	GI depth	p-dose [a.u.]	Carbon dose [a.u.]	Diffusion scheme
1	45	Shallow	0.98	1	CHBL
2	45		0.98	0.8	
3	45		0.98	0.8	
4	45		0.98	0.4	
6	35		0.94	1	
7	55		0.98	1	
8	45	Deep	0.70	1	CBL
9	55		0.70	1	
10	45		0.70	0.6	
11	45		0.70	-	
12	45		0.74	1	
13	45		0.74	0.6	
14	45	Deep	0.74	1	CBH
15	55		0.74	1	
16	45		0.74	0.6	
17	45		0.74	-	
18	45		0.78	1	
19	45		0.78	0.6	

Table 3.5: Wafers of the UFSD3.2 production.

Chapter 4

Experimental techniques

The measurements presented in this work were all performed in the Laboratory of Innovative Silicon Sensors in the Physics Department of the University of Torino and at INFN Torino. In this chapter, the laboratory setups employed are described, along with the main measurement techniques. Examples of typical measurements are also provided.

4.1 Experimental setup for the $I(V)$, $C(V)$, and $C(f)$ characterization

The static characterization of UFSDs is performed by measuring the current-voltage ($I(V)$), capacitance-voltage ($C(V)$), and capacitance-frequency ($C(f)$) characteristics of the devices in absence of external particles.

A probe station, connected with a curve tracer analyzer, is the setup used for the characterization. The Torino laboratory is equipped with two probe stations, each with a Keysight power device analyzer B1505 [64]. Figure 4.1 (left) shows one of the two setups with the probe station and the power device analyzer.

The probe station setup is equipped with a support called chuck, where the device-under-test (DUT) is placed; the DUT is kept still on the chuck by a vacuum system. An optical microscope equipped with different magnifications and a video camera positioned in a dedicated opening on top of the microscope is used to visualize the DUT.

The DUT biasing occurs via the chuck and coaxial tungsten-rhenium needles placed on specific contact points on the front side of the device. The needles are positioned using manipulators equipped with micrometric screws, as shown in figure 4.1 (right). Chuck and needles are connected with triaxial cables to the modules of the curve tracer. Generally, the chuck is supplied with a negative bias, while the needles are grounded.

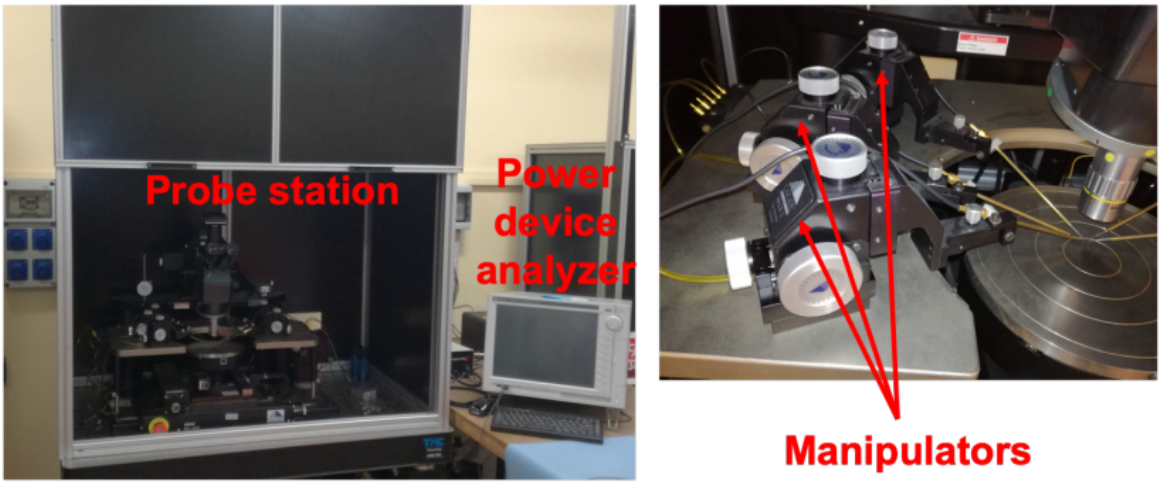


Figure 4.1: Probe station and Keysight B1505A power device analyzer in the Torino Laboratory of Innovative Silicon Sensors (left). A detail of the manipulators used to contact the DUT with the needles (right).

The Keysight power device analyzers employed for this work are equipped with three different types of modules:

- High Voltage Source Monitor Unit B1513C (HV-SMU), with a range up to 3000 V and 8 mA.
- Medium Power Source/Measure Unit B1511B (MP-SMU), with a range up to 100 V /0.1 A and a minimum measurement resolution of 10 fA/0.5 μV.
- Multi-Frequency Capacitance Measurement Unit Module B1520A (MF-CMU), with a frequency range from 1 kHz to 5 MHz and an AC signal level of amplitude up to 250 mV.

4.1.1 $I(V)$ characterization

When a reverse bias is applied to a *pn* junction, a steady current, called *leakage* or *dark* current, is flowing through the device, even in absence of an external stimulus.

Figure 4.2 depicts the $I(V)$ of an UFSD (solid) and a PIN diode (dashed). The PIN diode, subject to a reverse polarization, reaches a plateau, temperature-dependent, and remains roughly constant up to the breakdown; whereas the UFSD has an exponential trend above a certain voltage due to the presence of the gain layer. The exponential trend of the UFSD $I(V)$, which depends upon the gain of the DUT, begins after the depletion of the gain layer, which occurs at a few tens of Volts, and is represented in

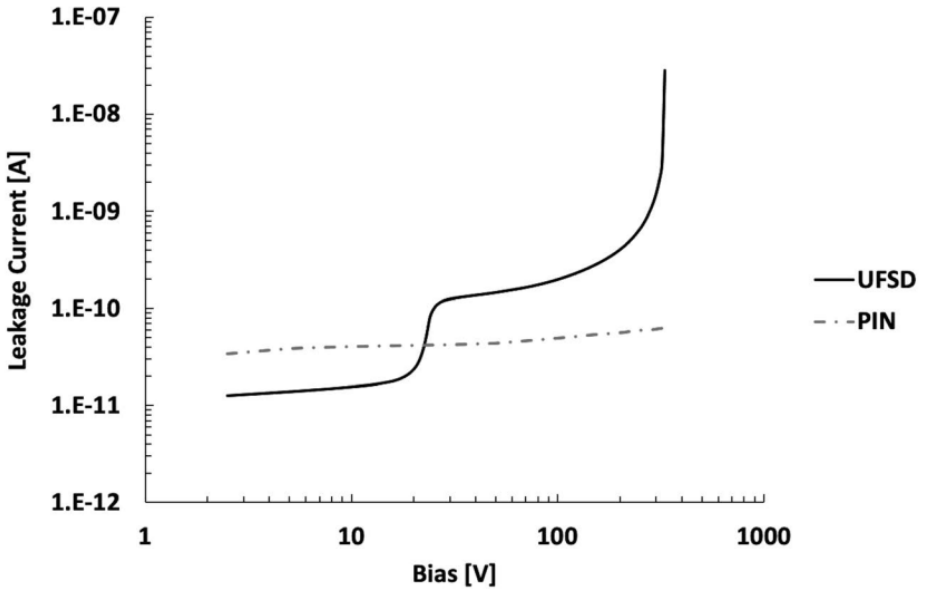


Figure 4.2: $I(V)$ characteristics of a 50 μm-thick UFSD (solid) and a PIN diode (dashed). The PIN diode presents an almost flat curve, while the UFSD features the characteristic knee at a few tens of volts, corresponding to the GL depletion, after which an exponential trend begins, due to the internal charge multiplication.

the curve by the characteristic knee. After the knee, the current keeps increasing with the bias voltage because of the increasing gain, until the sensor goes into a breakdown.

The description of the current-voltage characterization considers, for simplicity, a single pad device as DUT. The chuck and a needle positioned on the guard-ring are connected to the HV-SMU, while a second needle, positioned on the pad, is connected to the MP-SMU: this is shown in figure 4.3. The MP-SMU has a higher resolution and it is therefore used to measure the current flowing in the pad.

In multi-pad devices, all the pads are contacted with needles: the pads under test are wired to MP-SMUs, while the remaining pads are short-circuited to the guard-ring. This configuration allows measuring with high precision several pads during a single $I(V)$ scan.

4.1.2 $C(V)$ and $C(f)$ characterizations

A silicon sensor with parallel plate geometry can be modelled as a parallel capacitor-resistor or as a series capacitor-resistor equivalent circuit. Both circuits are good models for pre-irradiation sensors: the leakage current of these devices is very low, implying that the conductivity is low and that the measured admittance or impedance are given almost exclusively by the capacitance of the device. In irradiated sensors, the high

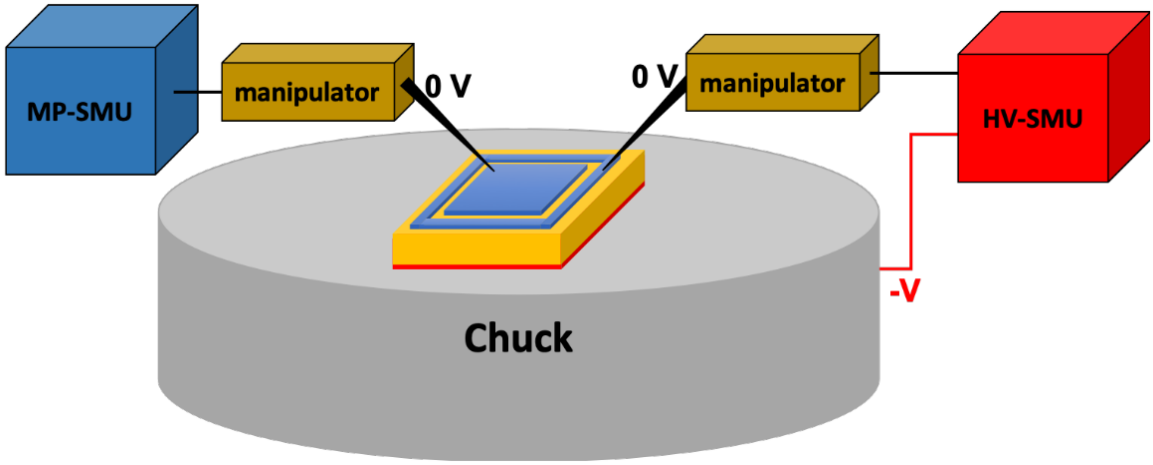


Figure 4.3: Sketch of electrical connections for current-voltage measurement on a single pad device.

leakage current implies that the series capacitor-resistor model cannot describe the correct behaviour of the device since it does not allow for the presence of leakage current [42].

$C(V)$ and $C(f)$ are performed using the HV-SMU and MF-CMU modules, interfaced with each other via a bias-T, which receives as input the DC bias voltage from the high voltage module and the AC signal from the capacitance bridge. The bias-T output is the sum of the two inputs, split in high ($\pm$V) and low (0 V) voltage levels. For a single pad device testing, the chuck and the needle contacting the pad are wired to the high and low voltage levels, respectively, while the guard-ring is grounded, as shown in figure 4.4.

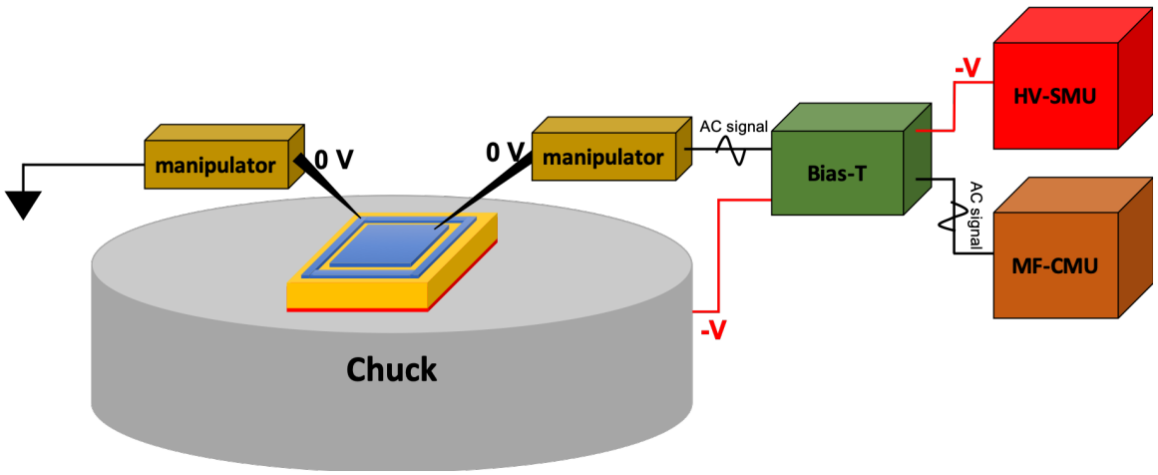


Figure 4.4: Sketch of electrical connections for capacitance-voltage and capacitance-frequency measurements on a single pad device.

Figure 4.5 shows the $C(V)$ curves of a PIN diode and an LGAD. The LGAD curve in black shows a capacitance drop (knee) at few dozen volts, corresponding to the depletion voltage V_{GL} of the gain layer, which is proportional to the active doping concentration and to the square of the width of the gain layer implant:

$$V_{GL} \propto \frac{qN_{A,eff}w^2}{2\epsilon_{Si}} \tag{4.1}$$

where q is the elementary charge, $N_{A,eff}$ the effective acceptor density, w the gain layer width, and ϵ_{Si} the silicon permittivity.

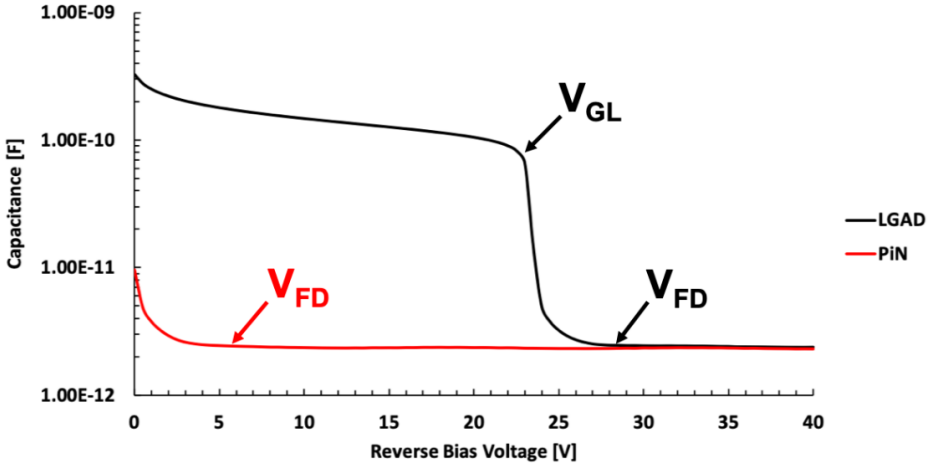


Figure 4.5: $C(V)$ characteristics of an LGAD (black) and a PIN diode (red).

The full depletion voltage V_{FD} of the device occurs when the capacitance becomes constant; it can be written as:

$$V_{FD} = \frac{qN_{A,eff}d^2}{2\epsilon_{Si}} \tag{4.2}$$

with d being the sensor active thickness.

The difference between V_{FD} and V_{GL} is the depletion voltage of the bulk, called V_{bulk}, which corresponds to the full depletion voltage of the PIN (red curve in figure 4.5). Given that, in pre-irradiation sensors, the bulk is much less doped than the GL, V_{bulk} is usually much smaller than V_{GL}.

Thanks to the $C(V)$ curves, it is possible to: (i) extract the GL profile (from which it is possible to obtain pieces of information on amplitude, width and depth of the implanted profile), (ii) measure the GL active thickness and its resistivity, (iii) determine the uniformity of a production. See [42] for a comprehensive report on the $C(V)$ measurements on UFSDs.

The $C(V)$ measurements are performed using a fixed frequency of the probing AC-signal provided by the MF-CMU: the selection of the optimal frequency of measurement is done by performing a $C(f)$ scan, using the same setup of the $C(V)$. The optimal frequency depends upon several parameters, such as the temperature, the sensor active area, or the resistivity.

The sensor can be approximated to an RC network with a frequency-dependent behaviour similar to that of a low-pass filter. The $C(f)$ measurement in figure 4.6 shows a capacitance, for the pre-irradiation device, more or less constant up to the frequency of $\sim$ 10 kHz, while it decreases above this value. The drop indicates the use of a too high frequency leading to a reduction of the area probed by the measurement. The optimal measurement frequency belongs to the low frequency range, where the capacitance values are constant. Things are different for the irradiated sensors, as discussed in detail in [42] and illustrated in figure 4.6, because the optimal frequency range drops, at room temperature, to about 1 kHz.

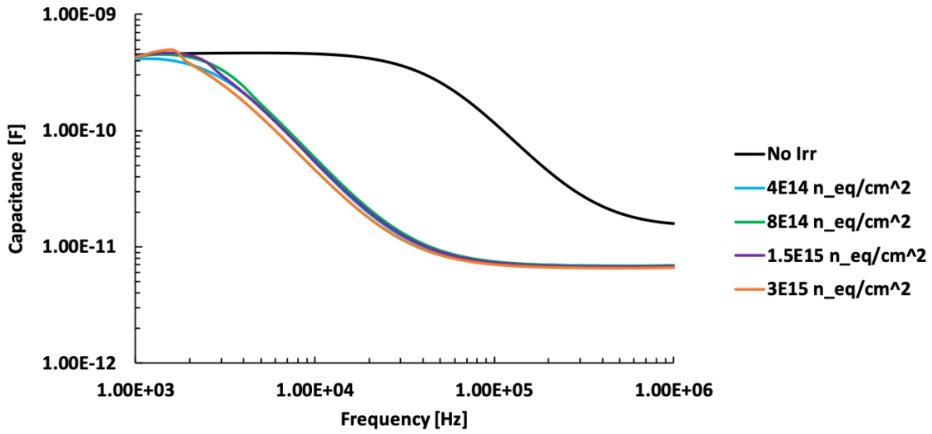


Figure 4.6: $C(f)$ characteristics of LGADs before and after irradiation.

4.2 The Particulars Transient-Current-Technique Setup

Particulars developed in Ljubljana an experimental setup, the Transient Current Technique (TCT) setup, which is widely used to characterize Silicon detectors, UFSDs included [65]. In this section, the principles of operation of the TCT are discussed, along with a description of the experimental setup and some typical measurements that can be performed.

The TCT technique exploits the motion of non-equilibrium $e-h$ pairs created by a

laser entering the DUT. The free charge carriers drift towards the electrodes according to their polarity, inducing a current signal which is a function of time. The analysis of the evolution in time of such current signal is the basis of the TCT technique, providing a wide number of detailed information about the DUT.

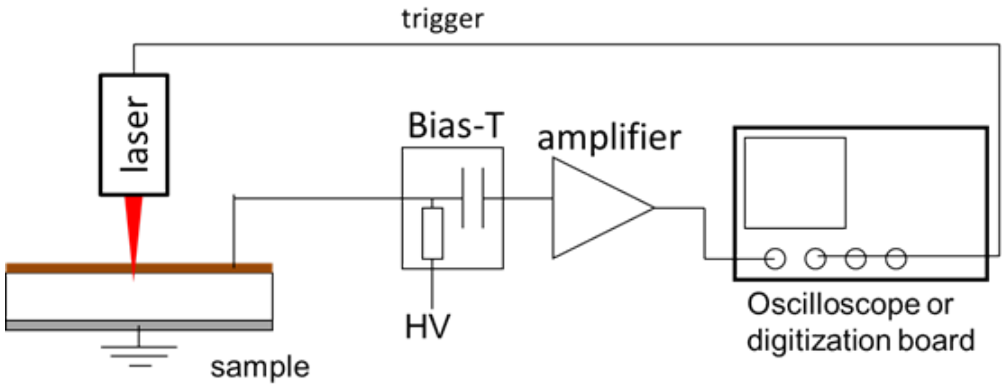


Figure 4.7: Princple of operation of the Transient Current Technique setup.

The basic scheme of a TCT system is shown in fig. 4.7: the laser lights the sensor, producing a signal which is amplified by an external trans-impedance amplifier (a 40 dB Cividec amplifier with 2 GHz bandwidth) and then fed to a fast oscilloscope (a Teledyne-Lecroy HDO9404 oscilloscope with a 20 GS/s sampling rate), where it is recorded for the offline analysis. The whole system sits within a metal box that can be closed to perform the measurements in dark. The Torino setup also includes a near-IR camera (Aven 26700-209-PLR Mighty Scope), which is mounted close to the DUT and controlled by a dedicated software: such addition allows spotting the IR laser.

The TCT measurements are performed with a negative bias voltage brought to the backside of the DUT, while the front side is grounded and reads out the signals. The backside of the DUT is glued with a conductive glue or tape to a custom PCB designed in the Electronics laboratory of INFN Torino (figure 4.8); the front side of the device is wire-bonded to the output connectors on the board.

Two main operation modes of the TCT are possible: (i) the Top-TCT (the one used in this work), when the laser shots on the surface of the sensor, and (ii) the Edge-TCT, when the laser shots on the edge of the sensor active thickness.

An important feature of the Particulars system is the possibility of mounting the DUT on a Standa 8MTF-102LS05/8MT175-100 translator stage (x-y stage), which can be moved with sub-micron precision over a range of tens of centimeters; in this way, the laser shot position can be accurately chosen, and the whole surface of the DUT can be mapped. In addition, the laser is mounted on a z-translator as accurate as the

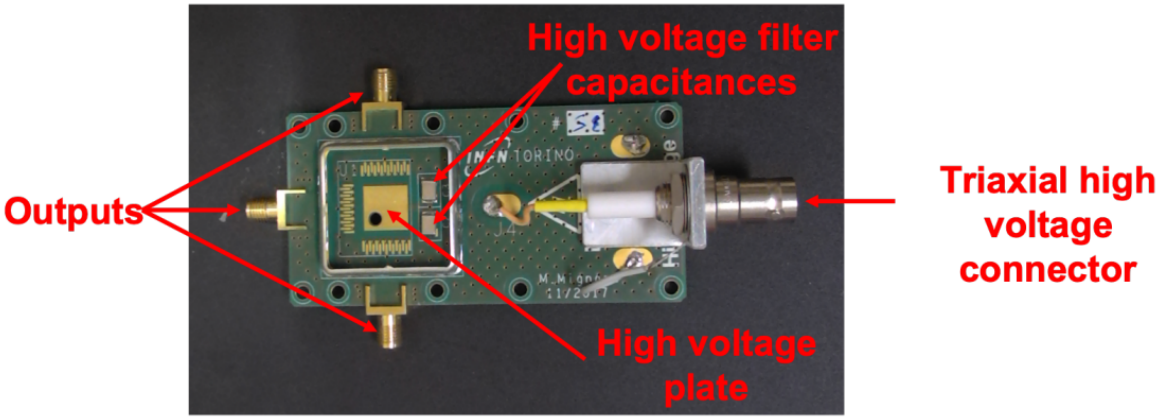


Figure 4.8: PCB designed at INFN Torino for the UFSD testing.

x-y stage, thus allowing a 3D translation of the device. The z-translator is necessary for the laser focusing procedure, described in the following.

Measurements at cold are mandatory when studying irradiated devices or to reproduce the expected operating conditions of the sensors at the experiments: in the TCT setup, the temperature of the x-y stage can be controlled by means of 4 Peltier elements placed underneath, which can vary their temperature in the -20 - +80 °C range. The temperature can be set on an external temperature controller (a Belektronig BTC benchtop) connected to the Peltier elements. The stability of the set temperature is monitored by a Pt100 element attached to the DUT, which operates as feedback to the temperature controller.

The TCT setup in the Torino Laboratory is also equipped with a chiller that fluxes a coolant in the coil of the Peltier elements, operating as a heat sink. Given the large volume of the metal box in which the system is enclosed, the Peltier elements alone are not enough to lower the temperature in the box below 0 °C, therefore all the inner walls of the box have been covered with nitrile rubber, allowing better isolation. In addition, an enclosure of polyurethane covers the x-y stage and the DUT providing further isolation. Such double isolation allows reaching -20 °C on the DUT. Figure 4.10 shows the nitrile rubber on the inner walls (black) and the polyurethane cover on the x-y stage (white).

The temperature on the DUT and the humidity in the metal box are measured with a thermo-hygrometer placed very close to the sensor: the relative humidity in the metal box can be lowered down to 10% by fluxing dry air at 8 bar by means of air inlets on a wall of the metal box, in order to lower the dew point.

The laser usually employed in the UFSD characterization is infrared with a wavelength $\lambda = 1060$ nm: its absorption depth in silicon is about 1 mm [66], so it manages

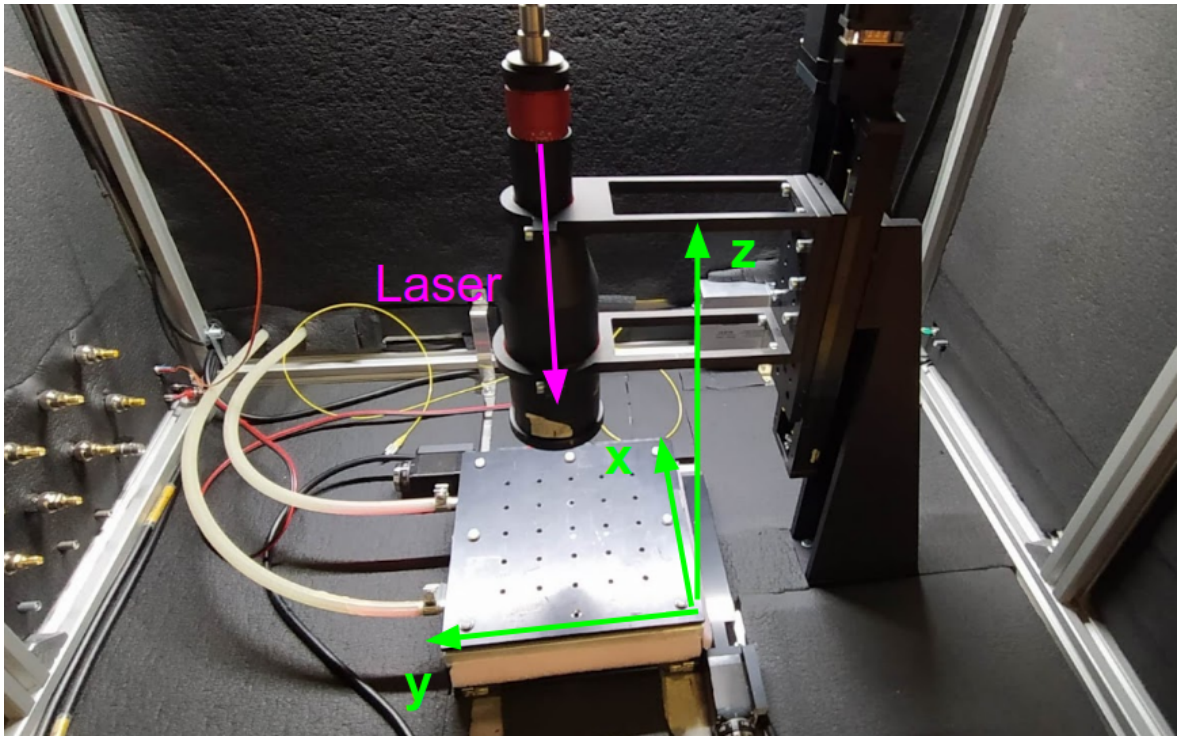


Figure 4.9: TCT setup in the Torino Laboratory for Innovative Silicon Detectors.

to fully cross the DUT, uniformly creating charges along its path and, thus, well simulating the passage of a MIP.

The lasers used in this work, produced by Particulars, are single-mode pulsed lasers with a core diameter of $\sim$ 6 μm. The laser intensity can be varied depending on the measurement: it can be set to a level corresponding to a signal lower than that generated by a MIP, up to many hundreds MIPs. The laser is usually auto-triggered, but it is possible to use an external trigger pulse.

The lasers used in this work have frequencies in the range 50 Hz - 1 MHz and their pulse durations range from $\sim$ 50 ps to 4 ns, with symmetrical pulses. A very short pulse is fundamental when measuring thin UFSDs so that it can reproduce the time of passage of a MIP in $\sim$ 50 μm of silicon.

The speed of the components of the TCT setup is one of its key features: drift velocities in silicon are in the range 10^6 - 10^7 cm/s, which sets the typical time scale of the observed induced currents from $\sim$ ns to a few tens ns. In most state-of-the-art systems the rise time of the measured induced current is determined by the capacitance of the electrode and input impedance of the amplifier: typical values result in rise times around 500 ps, which requires the bandwidth of the amplifier, and the oscilloscope/digitizer to be at least of the same order, >1 GHz, as those used in this work [67].

Despite well simulating the passage of an ionizing particle, it is worth pointing out that there are several differences in the creation of $e-h$ pairs in the sensor bulk by the laser beam or by a MIP: (i) the total number of $e-h$ pairs generated by a laser beam is almost constant, while that of a MIP follows the Landau distribution, (ii) the density of $e-h$ pairs is constant along the laser track while along the MIP track is non-uniform, and (iii) the laser beam creates $e-h$ pairs within a cylinder of about 10 μm diameter (depending on the laser optics), while for a MIP this diameter is much smaller, resulting in more substantial screening effects.

Another fundamental property of the TCT setup is the data acquisition software that allows controlling all the aspects of data taking from a computer. The software is involved in the movement of the 3D stages; it controls and sets the laser, the bias voltage, and the oscilloscope. All these operations can be done by means of a Labview interface specifically developed. Particulars also provides ROOT libraries for the offline analysis of taken data [68].

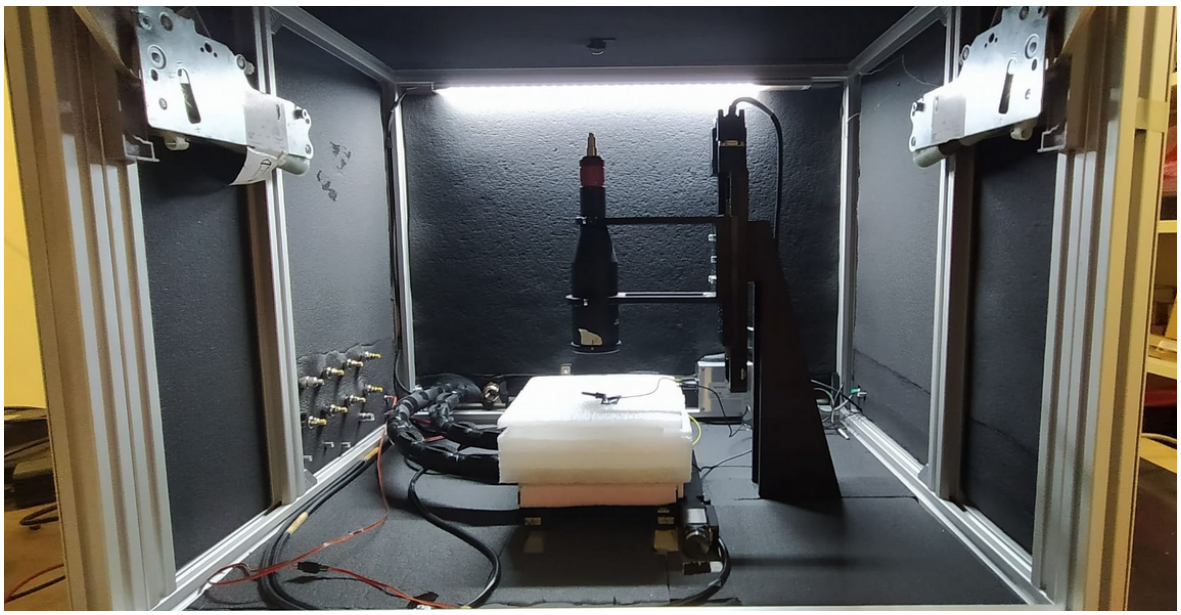


Figure 4.10: Interior of the metal box used for TCT measurements at cold. Nitrile rubber covers the inner walls, while a polyurethane enclosure is placed on top of the x-y stage.

A key issue when using the TCT setup is the calibration of the laser intensity to reproduce a MIP. The calibration setup comprises a 90-10 laser splitter, with the 90% branch reaching a PIN diode and the 10% one reaching an InGaAs reference diode: the laser intensity is varied until the charge collected in the PIN reaches the theoretical value predicted for a MIP [69]. The chosen laser intensity corresponds to a certain amplitude value of the signal generated by the reference diode, which is then used to

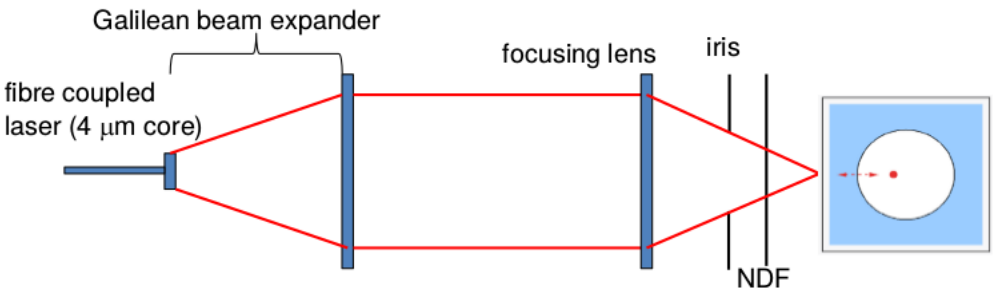


Figure 4.11: Schematic view of the optical system.

calibrate and monitor the laser intensity in the following measurements.

4.2.1 Focus Finding Procedure

Particulars lasers have a narrow core diameter that allows precise focusing, reaching a spot diameter of less than 8 μm. This is a highlight in the TCT setup: a small spot allows taking high-quality data, shooting with high precision on the detector area, and observing in detail its characteristics.

In order to reach a small spot, the laser has to be focused. The laser is connected by means of an optical fiber to a laser beam, which moves with the z-translator stage. As shown in fig. 4.11, the laser beam is an optical system provided with a Galilean beam expander, focusing lenses, and an iris which is used to calibrate the light intensity.

The focus finding procedure is the first operation that should be done before starting measurements and relies on the presence of a metalized region on the detector surface and another one non-metalized. To achieve the best result is convenient to have at least some tens of microns in both regions.

Using the DAQ software previously described, the focusing procedure is performed through an x-z scan. The scan in z should be made with steps of 20-50 μm, whereas the scan in x in 5-10 μm steps, starting on the active region and moving towards the inactive, metalized region. The results of the scan is a set of s-curves representing the charge collected by the sensor as a function of the x-position for different values of z (figure 4.12). The full-width at half maximum (FWHM) of the s-curves is a measurement of the laser spot size for a certain value of the z-position of the translator. The different FWHMs are thus plotted as a function of the z-position, resulting in a parabolic curve (figure 4.13): the position corresponding to the minimum of the parabola is the z-value that gives the minimum laser spot.

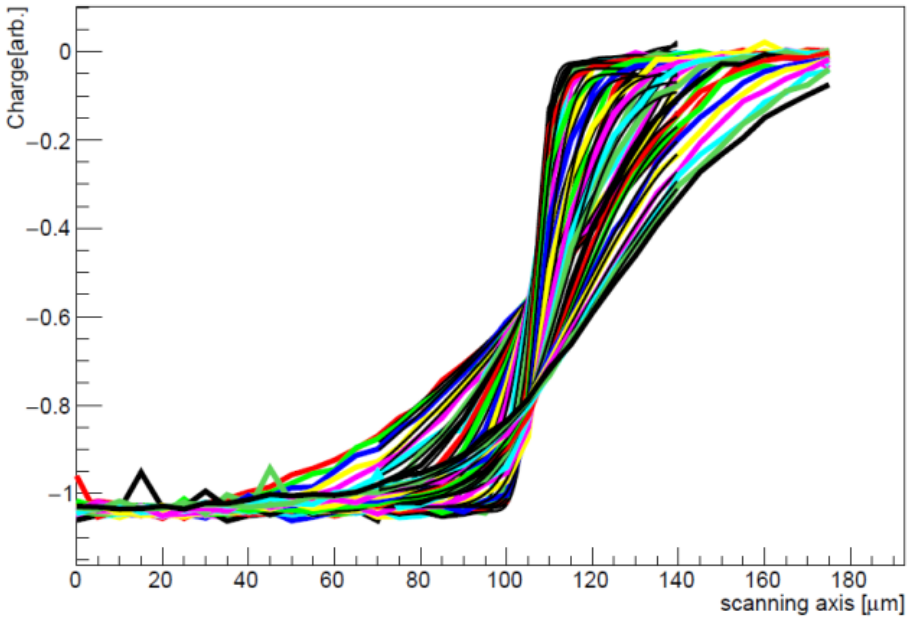


Figure 4.12: S-curves for different z-values.

4.2.2 Examples of TCT measurements: 2D-scan of the active area and the gain measurement

The 2D measurement of a sensor active area is a typical example of a TCT measurement since it exploits the main features of the setup: the small laser spot size, the accuracy of the x-y stage, the possibility of scanning very large areas. When performing the scan, the laser moves across the region to be scanned, both in x and y directions, in steps of fixed size (a step size ranges from less than one to hundreds μm). At each step, the laser shots and generates a signal, which is recorded by the oscilloscope. The acquired data are then analyzed offline to produce an x-y map representing different signal features: figure 4.14, for instance, displays the time of arrival of a single pad 2 $\times$ 2 mm^2 UFSD.

A second measurement, which is key when testing the UFSDs, is the measurement of the gain, given by:

$$Gain = \frac{Q_{UFSD}(V)}{Q_{PIN}(V)}, \tag{4.3}$$

where $Q_{UFSD}(V)$ and $Q_{PIN}(V)$ refer to the charge collected by a UFSD and a PIN with the same active thickness and geometry, and in the same generation condition (same laser intensity, same pulse duration, same temperature).

For this measurement, the IR laser intensity (monitored using the 90-10 splitter

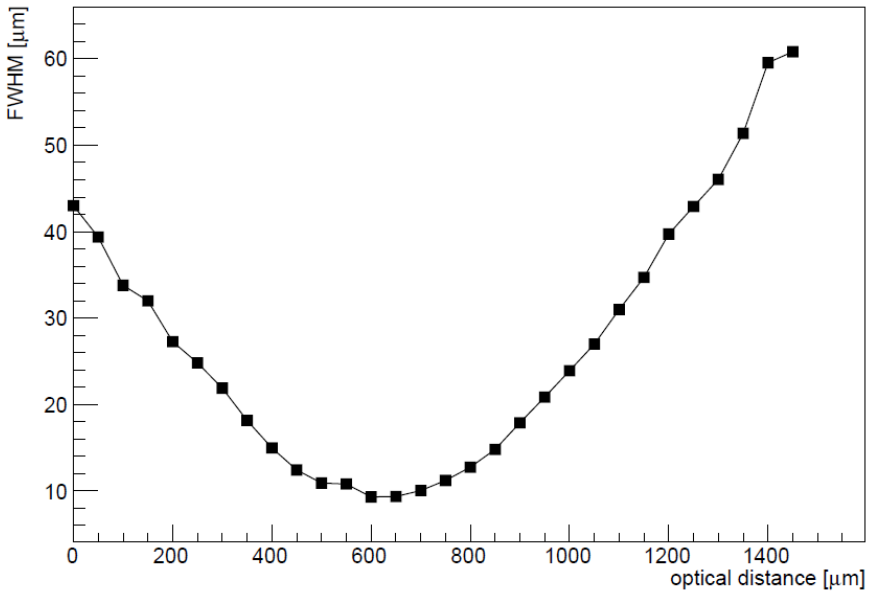


Figure 4.13: FWHM of the s-curves as a function of z-position.

and the reference diode) is usually set to a few MIPs, so that $Q_{PIN}(V)$ is sufficiently high and its relative uncertainty lower than with lower laser intensities. Usually, a 2×1 array with a UFSD on one side and an equivalent PIN on the other is employed: in this way, the laser can be shot on both devices by simply moving the x-y stage of a few millimeters, ensuring uniform measurement conditions. A bias scan is usually performed, to produce the gain-voltage characteristic of the UFSD under test.

Despite using a rather high laser intensity and a low-noise read-out chain, the uncertainty on $Q_{PIN}(V)$ is large, and the relative uncertainty on a measured gain of 20-30 is $\geq 20\%$.

4.3 The Hamamatsu ORCA2 CCD-camera

The Hamamatsu ORCA2 C11090-22B is an EM-CCD camera able to perform ultra-low light imaging [70]. It consists of a 1024×1024 pixel matrix providing 1M pixel resolution and high quantum efficiency from near-IR to UV, and it is particularly suitable for applications requiring long exposure times and low noise.

In this work, the camera is employed to take pictures of the UFSD inter-pad region (see chapters 2 and 5). For such purpose, the camera is mounted on top of the microscope of one of the probe stations in the Torino laboratory by means of a specific ocular, as shown in 4.15; this allows framing the inter-pad region in detail.

The camera is controlled by an external module driver and dedicated software.

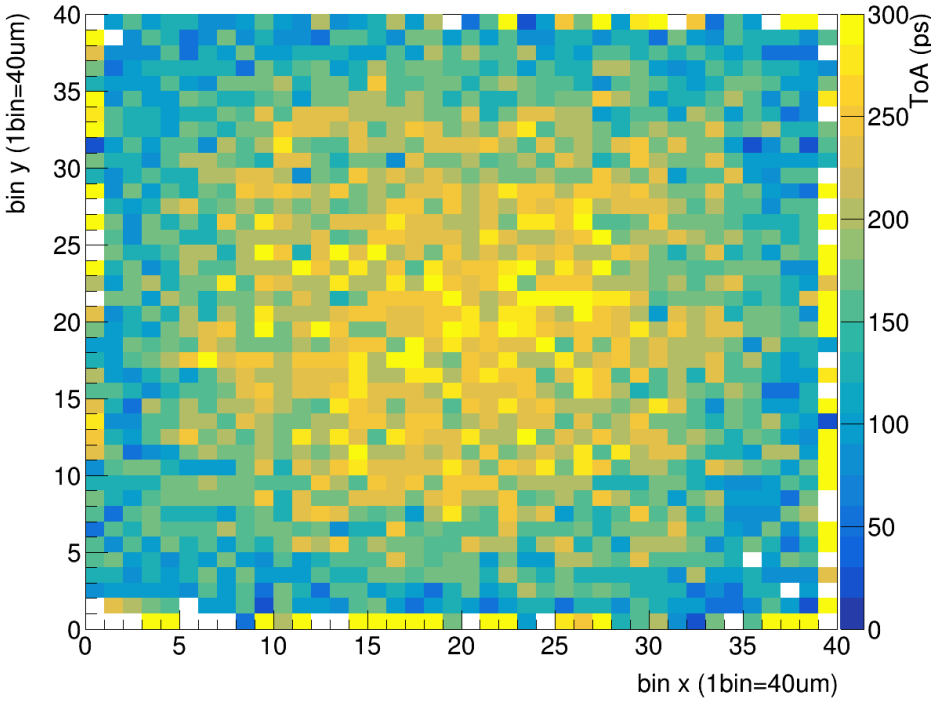


Figure 4.14: A 2D-map representing the collected charge of a 2 × 2 mm^2 UFSD as a function of the x and y positions. Obtained with the TCT setup.

When the camera is turned on and positioned on the ocular, it can display what it is framing and meanwhile take pictures. Different exposure times can be chosen and the camera can be run both in *normal* mode (the camera is a normal CCD camera) and *gain* mode, in which the light input is multiplied by a certain gain factor that can be chosen by the user. The latter mode is useful when wanting to detect very small light sources. The taken picture can be either displayed in black & white or in 2-/3- colors mode. The light wavelength range in which the camera operates changes dynamically by default, but the user can also set a fixed range.

During the measurements presented in this work, performed at room temperature, the DUT is placed on the probe station's metal chuck, which provides the bias voltage on the backside, while the guard-ring and the pads (the ORCA2 measurements are usually performed on multi-pad sensors) on the front side are contacted with needles and grounded. Then the camera is positioned on the ocular and the microscope is manually focused, with a magnification such that all the region of interest can be framed.

The camera is usually employed to detect the so-called *hot spots*: regions that emit visible photons because of the high current densities flowing through, for which a state-of-the-art camera with very low noise and high signal-to-noise ratio is needed.

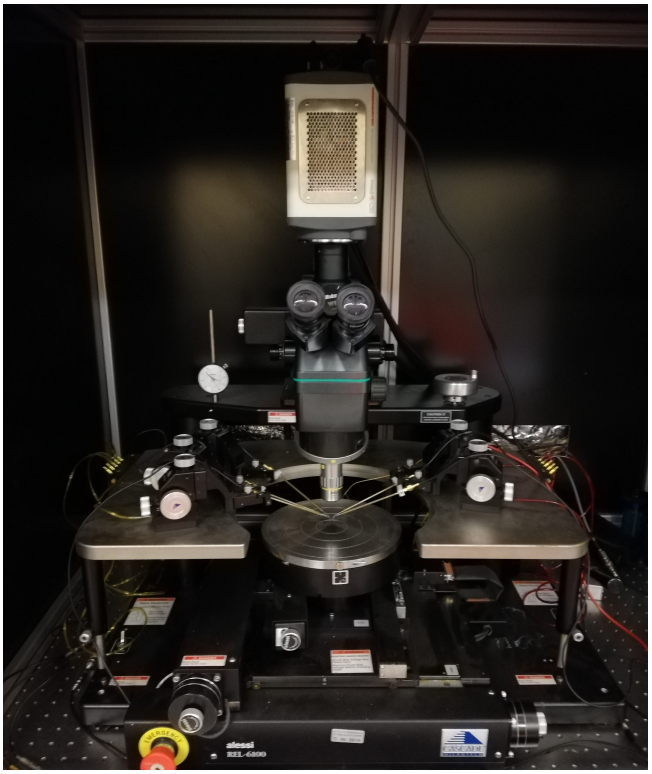


Figure 4.15: The ORCA2 camera mounted on the probe station of the Torino Laboratory of Innovative Silicon Sensors.

Two different pictures of the DUT are taken in the measurement procedure. The first one is a conventional picture taken with an external source of light (probe station's shutter slightly opened, *normal* mode set on the acquisition software), in black & white. A second picture is then taken, with the DUT biased and in complete darkness, this time using the camera *gain* mode, since the intensity of the visible light coming from the DUT is so low that it needs to be multiplied to be detected. The 2-color mode with dynamic frequency range was found to be the most appropriate for this second picture. The second picture is eventually overlapped to the first one so that the region that generates the *hot spot* can be clearly seen (if one looks at the second picture only, he would just see a yellow blob on an almost completely black background).

The ORCA2 camera has been extensively used for the measurements discussed in chapter 5, but it is also a valid diagnostic tool: for instance, a UFSD with a breakdown occurring earlier than expected has been analyzed with the camera. When the sensor was going into a breakdown, an *hot spot* appeared in its periphery, among the guard-rings (figure 4.16); the following optical inspection revealed a scratch that was likely causing the sensor malfunction. Such scratch was tiny, and it would have been impossible to detect it without the ORCA2. This simple measurement proved that the

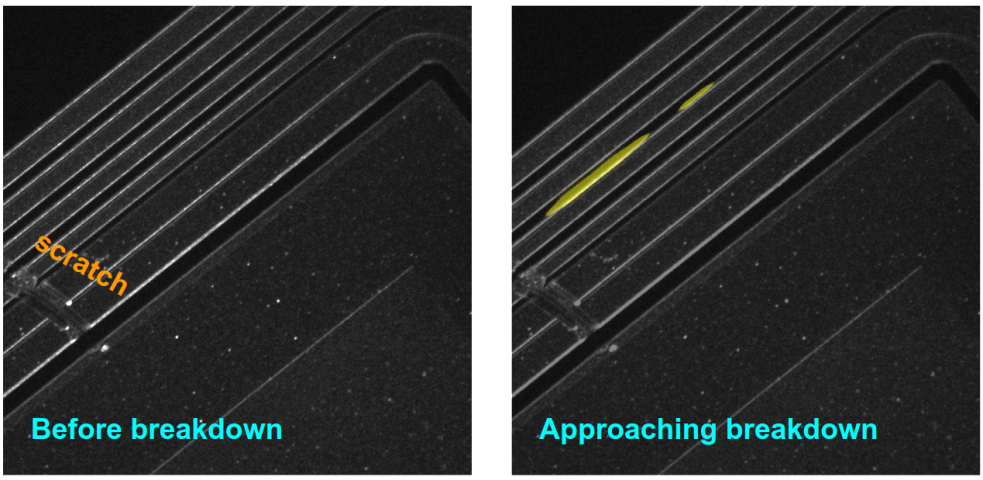


Figure 4.16: Pictures of a UFSD not working properly, taken with the ORCA2 camera. The sensor is framed before breakdown (left) and when approaching breakdown (right), when an *hot spot* appears among the guard-rings, causing the early breakdown. Such issue is likely caused by a scratch, highlighted in the picture on the left.

camera can operate as a diagnostic tool, and it may be potentially used to spot the noisy pads (where the leakage current is usually higher) of large multi-pad arrays, such as those that will instrument ETL.

4.4 The β-source setup

The β-source setup has been key for this thesis work. This experimental setup uses a β-emitting source to study the performances of the DUTs when hit by an ionizing particle. Compared to laboratory tests that use laser sources, this setup is more similar to the conditions that the sensors would experience at the experiments.

The most important figure that is measured is the time resolution [71]: that is the only way to measure the UFSDs total time resolution in the laboratory, otherwise one has to move to beam tests, which are much more complex, expensive, and do not allow measuring the same rate of devices like the β-setup.

The setup described below and shown in figure 4.17 (top) has been built in the Torino Laboratory of Innovative Silicon Sensors. It comprises a two-planes (Trigger plane + DUT plane) telescope, which has been 3D-printed and is needed to align the DUT, the Trigger [1], and the β-emitting source. The β-source is placed on top of the first read-out board by means of a 3D-printed structure specifically designed to align

[1]Hereafter, *Trigger* with the first capital letter will identify the sensor which is used to *trigger* an event.

it with the active areas of DUT and Trigger.

The DUT and the Trigger are glued with a conductive tape to a 10x10 cm^2 read-out board designed at the University of California Santa-Cruz (SC board) [72, 73]. The SC board has a fast inverting amplifier with a trans-impedance of about 470 Ω, followed by a 20 dB broadband amplifier (a 2 GHz Cividec model C1HV), for a total trans-impedance of 4700 Ω.

The first board of the telescope has a hole underneath the active region of the glued sensor, in order not to stop the electrons coming from the source. The diameter is usually $\sim$ 1 mm: given that almost all sensors measured in this work with the β-setup are 1.3×1.3 mm^2 single pads, only the corners of the pad can contact the board. Despite that, the bias voltage is properly brought to the backside of the device, and good mechanical stability is ensured.

A Teledyne-Lecroy HDO9404 oscilloscope with a 20 GS/s sampling rate, corresponding to a 50 ps time discretization, and 10 bit vertical resolution is used to record the signals coming from the DUT and the Trigger. The recorded signals are later analyzed offline.

A ^{90}Sr β-source, an element often used in this kind of setup, has been used. ^{90}Sr produces electrons with two energy branches: one with low energy from ^{90}Sr ($<$ 0.5 MeV), and one with higher energy from ^{90}Y (0.5 - 2 MeV). The low-energy electrons do not manage to traverse hundreds of microns of silicon (the total UFSD thickness, given by the active region plus the support wafer underneath, is in the 300-600 μm range, depending on the producer), and get stopped in the first sensor of the telescope, whereas the high-energy electrons, which are mostly MIPs, reach the second sensor of the telescope. Only events with high-energy electrons should be acquired since UFSDs at the ETL are expected to detect high-energy particles: for this reason, the DUT board is placed on top of the stack, so that only electrons from the high-energy branch can reach the Trigger in the second plane of the telescope.

The source used in the Torino laboratory has an activity of 3.6 kBq, leading to a trigger rate of about 2 Hz. Given the rather low activity, the probability of detecting two electrons so close in time that they cannot be resolved is negligible. The DUT is separated by the Trigger by $\sim$ 1 cm, and less than 0.5 cm separates it from the β-source, so the different paths followed by the electrons when traveling from the source to the Trigger, which differ at most by a few hundreds of micron, have a negligible impact on their time of arrivals and, consequently, on the measured resolution.

The measurements presented in this work have been performed placing the telescope within a climate chamber (FDM T370BX with enhanced temperature system), in order to have stable temperature and humidity. The temperature is usually set to -25°C and

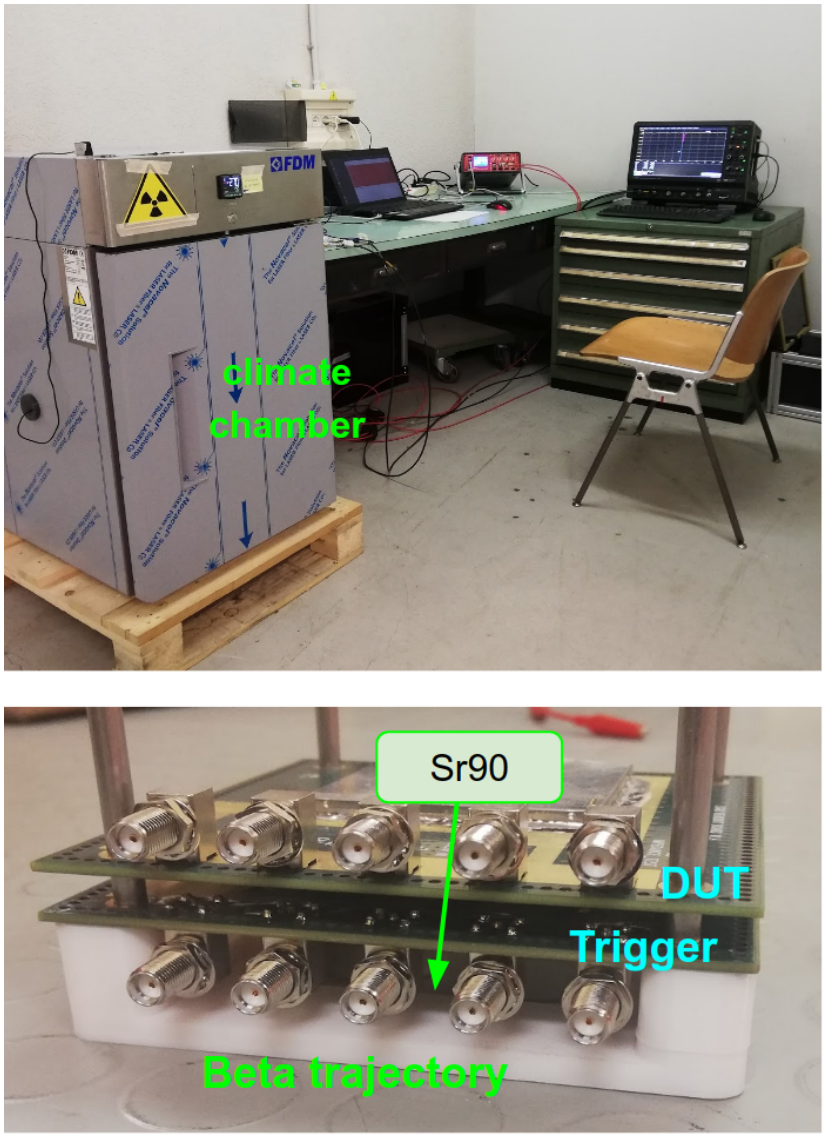


Figure 4.17: The β-source setup of the Torino Laboratory of Innovative Silicon Sensors (top). 3D-printed structure used for the β telescope (bottom).

the relative humidity is below 10%, except during the temperature scan discussed in chapter 6. A thermo-hygrometer is placed close to the DUT to monitor humidity and temperature.

A dry air flux entering the chamber is of primary importance to avoid ice forming on the sensors, possibly breaking them. The procedure followed in this work when cooling the sensors is to flux dry air for a certain amount of time until the minimum relative humidity is reached, then the temperature is slowly reduced, constantly monitoring the humidity. It takes $\sim$ 1 hour to cool down to -25°C the climate chamber, and $\sim$ 20 minutes to heat it up.

The performance of the read-out board does not depend much on the temperature, therefore placing it in the climate chamber is not an issue. The Cividec external amplifier, instead, is always placed outside the climate chamber.

The data acquisition (DAQ) is fully automated: it consists of a python script that, using the pyVISA package [74], is able to control the instruments, oscilloscope included, perform bias scans, and record and save the signals displayed on the oscilloscope for the offline analysis. The raw signals are saved on files manageable by the ROOT framework, the software used in the offline analysis. The details of the analysis are provided in chapter 6.

Both the DAQ and the offline analysis software have been developed in Torino. The DAQ is partially based on a similar software developed at the University of California Santa Cruz.

4.4.1 Trigger Characterization

The Trigger employed in this work is a single pad HPK1 with an area of 1 $\times$ 3 mm^2, read out with the board described in the previous section. It has been used for all the measurements presented in this work, always biased at the same voltage and operated at the same temperature (-25 °C), except for the temperature scan described in chapter 6. The read-out board and the read-out chain instruments never changed, too. Such stable running conditions were crucial to ensure uniform performance throughout the measurement campaign.

The Trigger time resolution has been measured using two identical devices, dividing the measured total resolution by $\sqrt{2}$. A resolution $\sigma_{Trigger} = 31.6 \pm 1.3$ ps was found with the sensor biased at 175 V and operated at -25 °C; such voltage has been found to ensure a safe and reliable operation of the device, while providing a good time resolution and uniform performance.

The characterization has been performed with the telescope previously described, hence, the events recorded by the first sensor are surely due to minimum-ionizing electrons since the Trigger plane is on the second plane of the telescope; however, an electron that crossed the first plane as a MIP may have lost enough energy that it is not a MIP anymore when it hits the second plane. Consequently, although the two sensors used for the Trigger characterization are identical, the signal amplitude distributions are expected to be different, as shown in figure 4.18. The blue distribution represents the second plane of the telescope: it has a larger MPV than the red one because it detects also non-MIP electrons, which lose more energy in the sensor than MIPs. The width is larger as well since the distribution of the second plane is actually given

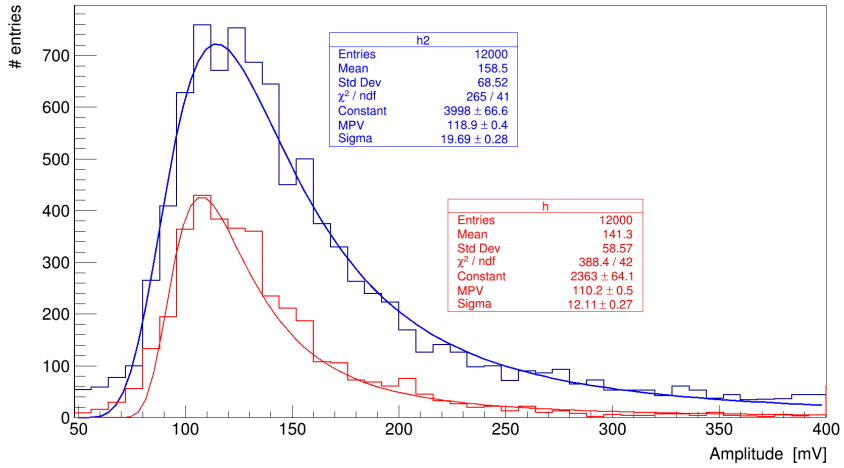


Figure 4.18: Signal amplitude distributions of the two HPK1 UFSD used for the Trigger characterization. The blue distribution represents the second plane of the telescope used in the characterization: it has an higher most probable value (MPV) since it detects also non MIPs, which lose more energy in the sensor.

by the sum of MIP and non-MIP events. Finally, the amplitude distribution in red, representing the first plane of the telescope, has fewer events in the Landau distribution because, in some events, the electrons did not pass through the sensor: in such events, the baseline noise is recorded, resulting in a low signal amplitude (usually less than 3 mV) which is out of range in the plot.

The Trigger offline event selection requires, in order not to trigger on the baseline noise or to saturate the signal, the signal amplitudes to lay in the 80-250 mV range and the collected charge in the 11.5-53 fC one; only events passing such cuts are considered as good and used in the analysis. The cuts are used either in the Trigger characterization and in the analysis discussed in chapter 6.

A temperature scan of a few UFSDs is described in chapter 6, requiring the characterization of the Trigger at different temperatures. The results are shown in figure 4.19: the operating voltage increases with temperature, as expected, and the minimum resolution is $\sim$ 30 ps regardless the temperature. During the UFSDs temperature scan described in chapter 6, the Trigger bias voltage has been changed according to figure 4.19, in order to deliver a time resolution of 30 ps at each temperature step.

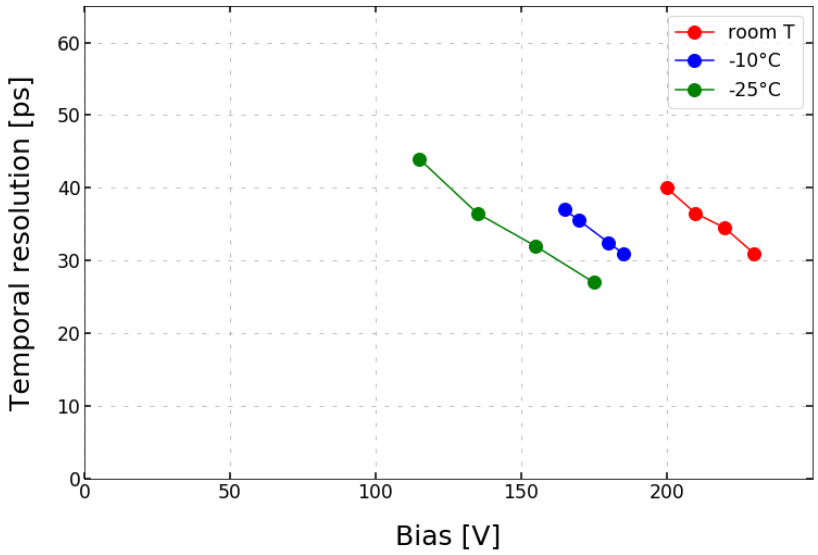


Figure 4.19: Time resolution as a function of the bias voltage, at three different temperatures. The characterization is needed for the temperature scan discussed in chapter 6.

Chapter 5

Development of UFSD pixel arrays

The UFSDs covering the ~ 14 m^2 of the two ETL endcaps will be pixelated arrays, for a total of ~ 8.5 million channels. This chapter will cover the evolution of the UFSD pad isolation and structure isolation designs in the last 3 years, which eventually led to the definition of the optimal design for the 16×16 sensors to be installed in ETL.

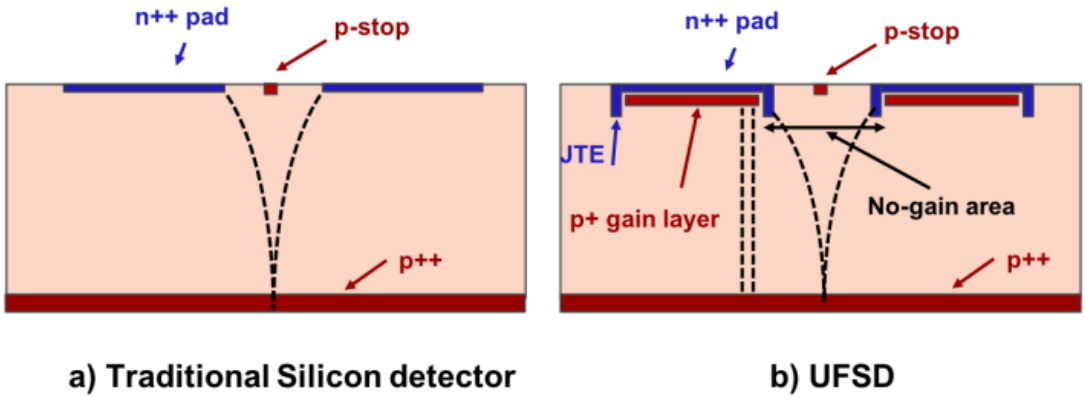


Figure 5.1: Schematic representation of the charge carrier drift lines for a standard silicon sensor (left) and for an UFSD (right)

A key aspect of the ETL sensors is their fill factor, namely the ratio between a device active and total areas. The aim is to produce sensors with a fill factor as high as possible, in order to increase the number of two-hits tracks and, consequently, the performance of the detector.

The main components determining the UFSDs fill factor are the so-called no-gain or inter-pad region and the edge widths, with the inter-pad being the main contributor to the total fill factor.

As shown in figure 5.1, in standard planar silicon sensors the fill factor is almost 100% since the electric field lines are always closing on one of the electrodes. This

remains true in the UFSD design, but not all the electric field lines cross the gain layer: some of them reach the JTE (see chapter 2) where there is no multiplication, as shown in 5.2. As a consequence, the timing of particles passing in the inter-pad region cannot be determined accurately.

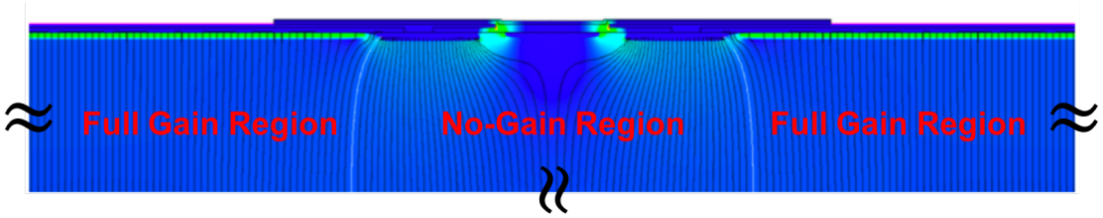


Figure 5.2: TCAD 2D-simulation of the electric field (intensity color map) and drift lines in the inter-pad region of an UFSD sensor.

The overall width of the inter-pad region is determined by: (i) the JTE that surrounds the n^{++} electrode and (ii) the p-stop implant which is placed between each pair of n^{++} electrodes, as described in chapter 2. A narrow width can be achieved by reducing the dimensions of JTE and p-stop and decreasing the distance between them; however, when JTE and p-stop are too close to each other, the risk of premature breakdown (i.e. a breakdown occurring at a voltage lower than expected, not caused by the internal gain) grows, as well as the risk of having non-isolated pads. Finding a trade-off between a high fill factor and the safety of operation of the device is the goal of the measurement campaigns described in the following sections.

ETL requires a fill factor per-disk larger than 85%, which translates into a UFSD array with $\sim$ 90% fill factor and inter-pad width of 65-70 μm, assuming a 16x16 array with 1.3x1.3 mm^2 pixels and 500 μm inactive edges [2]. The UFSD and ETL disk fill factors do not match because, on the ETL disk, there are additional passive elements besides the UFSDs inter-pad and edge widths, such as the gaps between the sensor modules and the disk edges (see chapter A); consequently, to reach the target fill factor, the UFSD fill factor needs to be higher than that of the whole ETL disk.

UFSD producers have developed different inter-pad designs by changing the width of the JTE and the width and doping of the p-stop implant. In the following, the focus will be on the FBK productions, where several innovative designs have been explored in the attempt to find the optimal solution.

The following sections are devoted to: (i) the description of the inter-pad width measurement method, (ii) the measurement campaigns carried on the FBK UFSD3 and UFSD3.1 productions, (iii) the analysis of the micro-discharges effect. Finally, the results of the UFSD3.2 production, which finalizes the inter-pad design for the ETL array, will be presented.

5.1 Measurement of the inter-pad width

The measurement of the inter-pad width has been performed in the laboratory, using the TCT setup described in chapter 4. The sensors tested have a small region without metal traces (optical window) from one pad to the neighboring one, specifically designed for this measurement.

The measurement is made by performing a TCT scan between two adjacent pads and acquiring their collected charges as a function of the laser position. The scan begins with the laser shooting on the active region of one pad, then moving towards the inter-pad region. The charge collected by the first pad progressively decreases and the second pad starts collecting some charge. When the laser shots in the inter-pad, both pads collect the same charge, which is rather low since multiplication does not occur there. For such reason, the laser intensity is usually set to a few MIPs, so that the charge collected in the no-gain region is high enough and the relative uncertainty is not too large. The scan ends with the laser entering the active area of the second pad.

For each sensor to be evaluated, the scan is performed and recorded a hundred times, in order to have an uncertainty of about 2 μm on the inter-pad region width. The laser used for this measurement is the IR laser, focused with a spot of about 10 μm. Such laser has an absorption length in silicon of about 1 mm, therefore it crosses the whole sensor depositing energy along its path, similarly to a minimum ionizing particle [67].

A 2D-map of the region of interest is always produced before moving to the charge scan, to ensure that the device-under-test (DUT in the following) and its optical window over the two pads are aligned with the TCT xy-stage.

The expected charge profile from the acquisition consists of two sigmoidal functions (figure 5.3), one for each read-out pad. The sigmoid is obtained by the convolution of a step function (describing the transition between gain and no gain regions) with a gaussian function, which accounts for the laser beam spot size. The point at which the sigmoid reaches 50% of its height corresponds to the intercept with the step function used in the convolution: hence, the width of the no-gain region is given by the distance between the 50% points of the charge profiles of the two pads. In figure 5.3 such distance is equal to 16.7 μm.

Figure 5.2 shows that the electric field lines reaching the JTE have a lateral bend and expand under the gain layer, therefore the measured inter-pad width can be larger than the nominal distance between the gain implants when the lateral bending is significant.

The lateral bending of the electric field lines depends upon the bias voltage: at

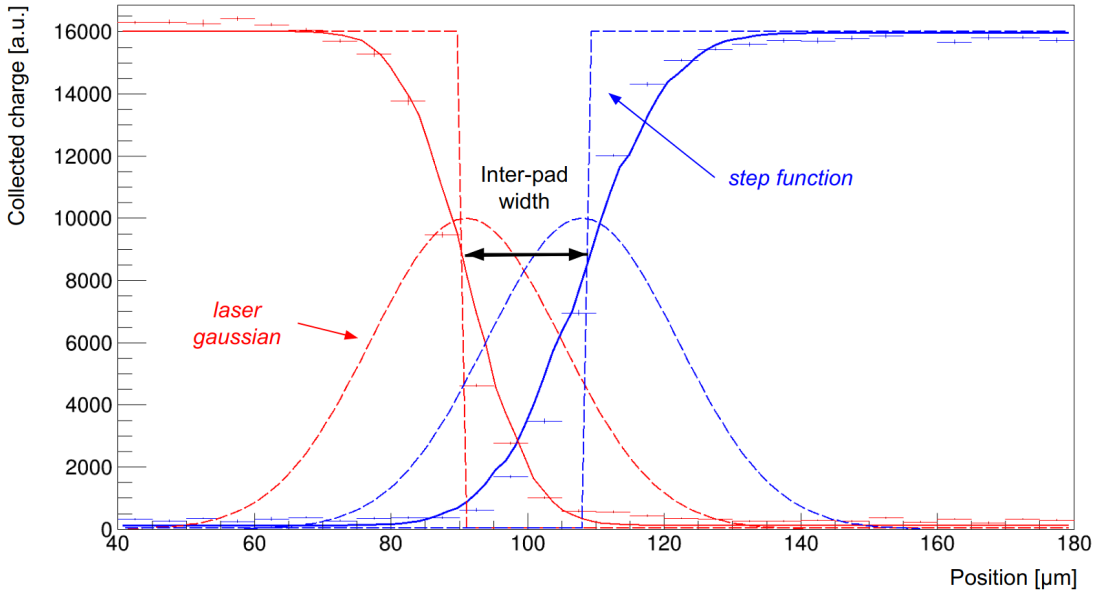


Figure 5.3: Collected charge as a function of position of two neighbouring pads. The step function and the gaussian profile result in a sigmoidal function which is used for the inter-pad width measurement.

low voltages (i.e. low electric field) is strong, while it gets significantly milder at high voltages, with the width changing by 1 μm for a 35-45 V shift [42]. In this work, the inter-pads were measured with the DUT biased at its operating voltage, hence the importance of the bias voltage is negligible since the electric field in the sensor is rather high.

The bias voltage might play an important role in irradiated devices, which have to be operated at higher bias voltage to compensate for the effect of radiation damage. In such devices, not considered for the measurements presented in this chapter, the bias voltage can be sufficiently high to generate a gain at the JTE - p-bulk interface, where the electric field lines get focused. Consequently, the no-gain region is narrowed, and the measured interpad width can be significantly different from that of a pre-irradiation device. This is discussed in detail in [75].

Table 5.1 presents the measured and nominal inter-pad widths of un-irradiated sensors coming from all the FBK productions and from the HPK2 one (see chapter 3). The estimated uncertainty on the measured interpad widths is $\pm$ 2 μm.

Measurements of the inter-pad widths of irradiated FBK UFSD3.2 and HPK2 sensors can be found in [75].

Table 5.1: Inter-pad widths measured with the TCT setup. The estimated uncertainty on the measured widths is ± 2 μm.

Production	design	measured inter-pad [μm]	nominal inter-pad [μm]	Thermal load
UFSD2		67	70	LD
UFSD3	Aggressive	16.5	11	LD
UFSD3	Medium	16.5	20.5	LD
UFSD3	Medium	31	20.5	HD
UFSD3	Safe	30.5	31	LD
UFSD3	Super-Safe	38	41	LD
UFSD3.1	Type 1	32	16	HD
UFSD3.1	Type 10	62	49	HD
UFSD3.1	Type 11	36	20.5	HD
UFSD3.2	Type 4	35	23.5	LD
UFSD3.2	Type 8	38	27.5	LD
UFSD3.2	Type 10	62	49	LD
HPK2	IP3	64	30	
HPK2	IP4	91	40	
HPK2	IP5	102	50	
HPK2	IP7	120	70	

5.2 The inter-pad design of the FBK UFSD3 production

The first production of 50 μm-thick UFSD by FBK (UFSD2, see [42]) featured arrays with an inter-pad width of ~ 70 μm, which is really at the limit of the ETL requirements in terms of maximum width.

The following production, UFSD3 [76], included new inter-pad designs, as detailed in chapter 3. Table 5.1 lists the variety of designs of the production, with their nominal width: the largest width of UFSD3 is almost 30 μm narrower than UFSD2, whereas the smallest one is reduced by a factor 6.

The first step of the characterization was to measure the $I(V)$ curves of arrays with the four different designs. *Super-safe* sensors, namely sensors with the largest inter-pad width, have an exponential curve and go in breakdown after 300 V, see figure 5.4. The exponential trend is due to internal gain multiplication and the electrical breakdown occurs when uncontrolled gain avalanche begins: that is the typical $I(V)$ curve of a UFSD working as expected.

Since all measured sensors have the same gain layer doping, they should all have the same breakdown voltage (V_{BD}), but the *Safe*, *Medium* and *Aggressive* designs

suffer from premature (or early) breakdown, with an abrupt, not exponential, $I(V)$ characteristic. In particular, figure 5.4 shows that the narrower is the inter-pad width of the design, the earlier its breakdown.

The abrupt $I(V)$ curves indicate that breakdown is not caused by an avalanche in the gain region, but it is rather caused by something taking place outside the gain region, which occurs at a certain voltage below the expected V_{BD}.

Since the only difference among tested sensors is the inter-pad design, the early breakdown likely originates there. Such observation lead to a comprehensive study of the UFSD3 no-gain areas, as discussed in the following sections.

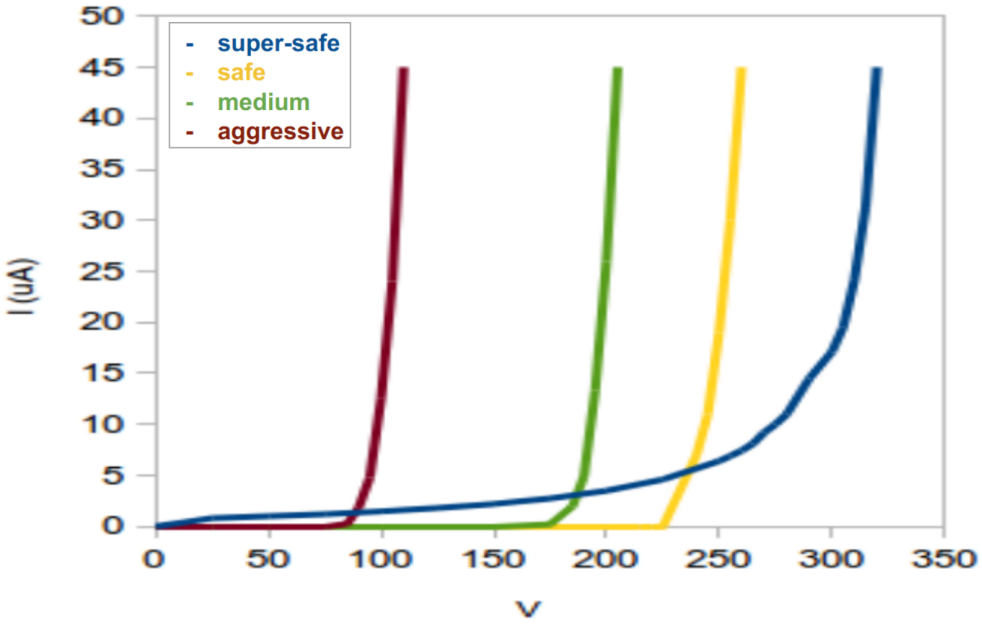


Figure 5.4: $I(V)$ curves of UFSD3 sensors with different inter-pad widths

5.3 TCT Measurement campaign on UFSD3 sensors

The samples used for the study of the inter-pad region are:

- A strip sensor with Super-safe inter-pad design (600 μm pitch, 1 cm length)
- A 2×2 pad matrix with Safe inter-pad design (1×3 mm^2 pads)
- A 2×2 pad matrix with Medium inter-pad design (1×3 mm^2 pads)
- A strip sensor with Medium inter-pad design (300 μm pitch, 1 cm length)
- A 2×2 pad matrix with Aggressive inter-pad design (1×3 mm^2 pads)

All DUTs are covered with a metal layer on the front surface, over the gain region, with the exception of the optical windows used for the inter-pad measurement, therefore only the no-gain and optical window areas could be scanned, as light is absorbed elsewhere.

The measurements have been performed at room temperature, with the sensor bonded to a custom read-out board (figure 5.5). Each read-out channel is connected to a 40-dB external broadband amplifier, whose output is then connected to a fast oscilloscope. The details of the data acquisition with the TCT setup are described in chapter 4.

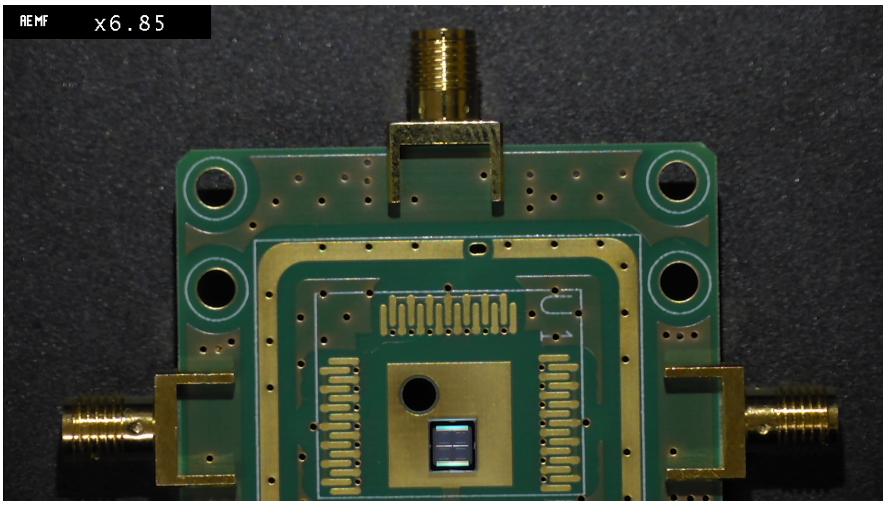


Figure 5.5: A 2×2 UFSD bonded on a custom read-out board used for the TCT measurements.

The DUTs are firstly scanned at a voltage well below breakdown, in order to record the collected charge and produce a x-y map that clearly defines the DUT and, in particular, the inter-pad region. Figure 5.6, for instance, shows the collected charge measured in a 2×2 pad sensor with Safe design. The region displayed in the map is circled in figure 5.7. On the z-axis (color-coded) is the sum of the charges collected by all four pads. The gain regions covered by the metal have a collected charge close to zero (in blue). Figure 5.6 highlights well the four pads and the inter-pad region, and helps the interpretation of the reader in the following plots.

Figures 5.8a, 5.8b, 5.8c show the same area covered by figure 5.6, but, in this case, only the charge collected by Pad 1 is represented in the color-coded z-axis. The measurements are taken at three different bias voltages: while the charge is constant across the inter-pad region at 200 V, it increases a lot around the pad's corner at 250 V and 260 V, when the sensor's breakdown begins.

Figure 5.9 shows the x-projection of these 2D-maps at y=100 μm: the increase of

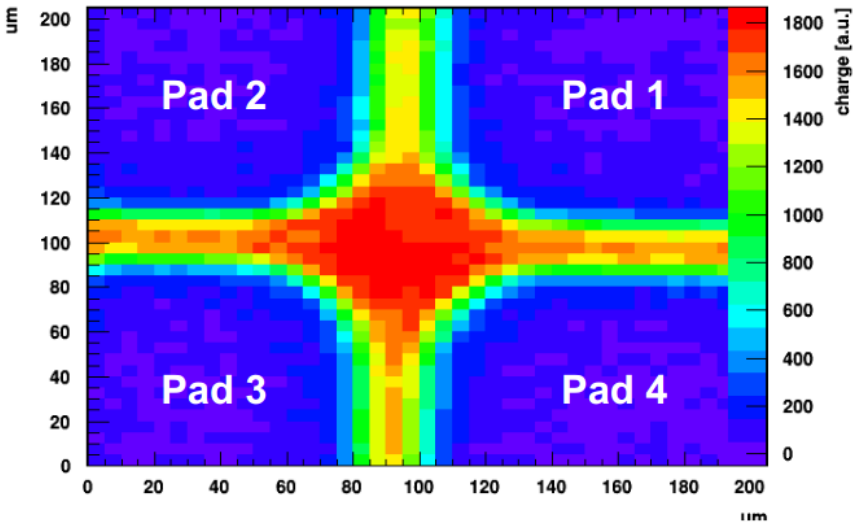


Figure 5.6: 2D-map of the charge collected by a 2×2 sensor with Safe design.

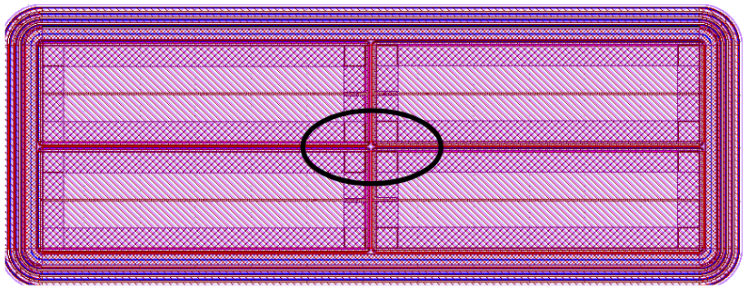


Figure 5.7: A 2×2 sensor, with the scanned inter-pad region circled.

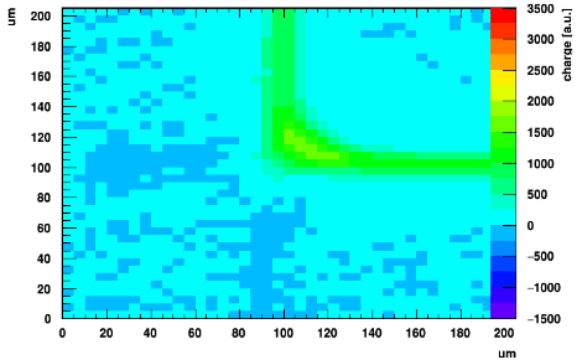
the collected charge while the sensor approaches the breakdown voltage can be noted. Similar observations are obtained measuring the 2×2 sensor with the Aggressive design.

Figure 5.10b shows a similar result for the strip sensor with the Medium design of the inter-pad region: the charge collected is constant at 200 V, then increases with voltage when the sensor is going into breakdown. The effect is particularly significant around the pad's corner.

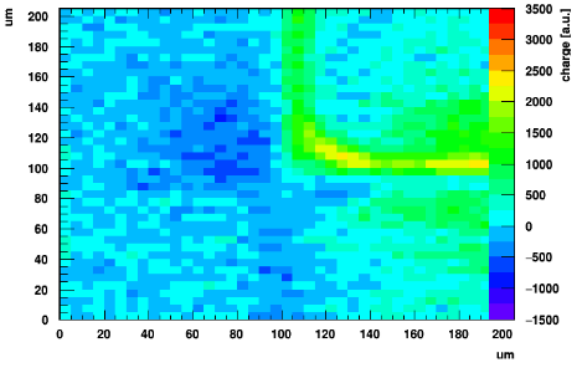
The Super-safe sensor, instead, does not show any increase in collected charge in the inter-pad region, when brought into breakdown, as shown in 5.11: this is the only design able to reach the expected breakdown voltage.

Hence, sensors suffering from premature breakdown show signs of charge multiplication occurring in the inter-pad region, whereas the Super-safe devices have a constant collected charge in the inter-pad up to breakdown.

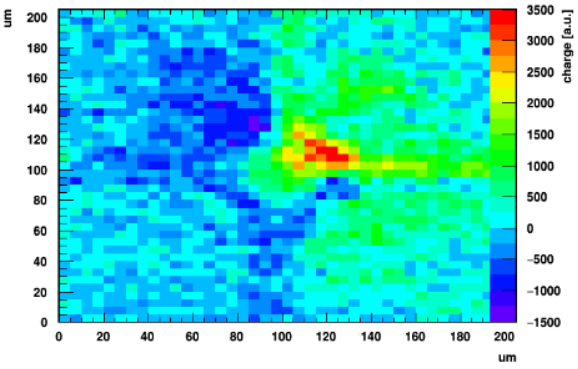
The charge multiplication in the inter-pad region is caused by the onset of a strong electric field, underlying a weakness in the design. The effect of such a field appears



(a) 200 V



(b) 250 V



(c) 260 V

Figure 5.8: TCT 2D-map of the charge collected by Pad1 of the 2×2 sensor with Safe design at 3 different voltages

promptly, as proven by the trend of the $I(V)$ curves in those devices.

The high electric field occurring in the inter-pad region can be explained with an *inversion layer* establishing between the p-stop and the JTE: the positive charges present at the Si-SiO_2 interface (the transition region between silicon and the oxide deposited above), induce a layer of electrons that acts as n-doped silicon, see figure 5.12. The electrons create a pn junction with the p-stop. The value of the p-stop

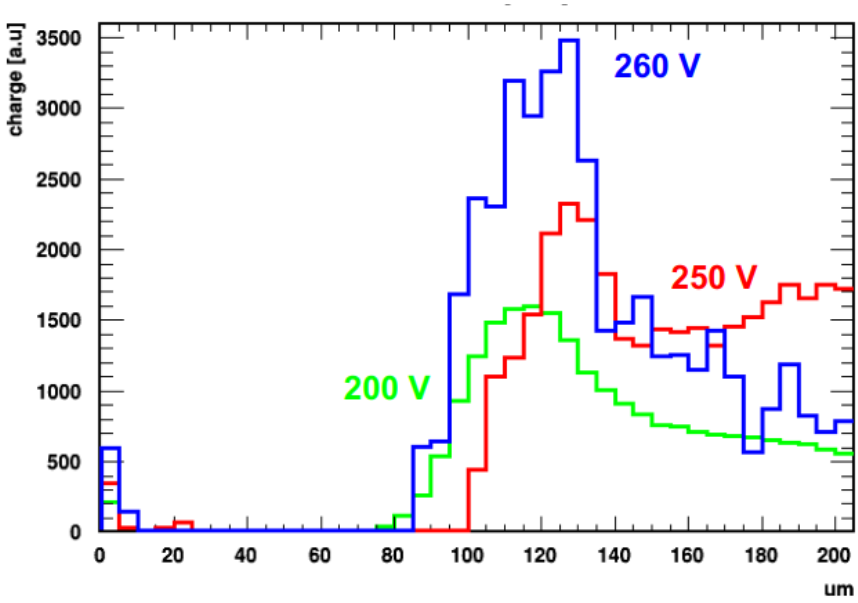


Figure 5.9: Profiles of 5.8a 5.8b 5.8c at y=100 μm

doping determines how abrupt is the *pn* junction and, consequently, how high the electric field is [77]: it follows that sensors with a higher *p*-stop doping would go into a breakdown at a lower bias voltage.

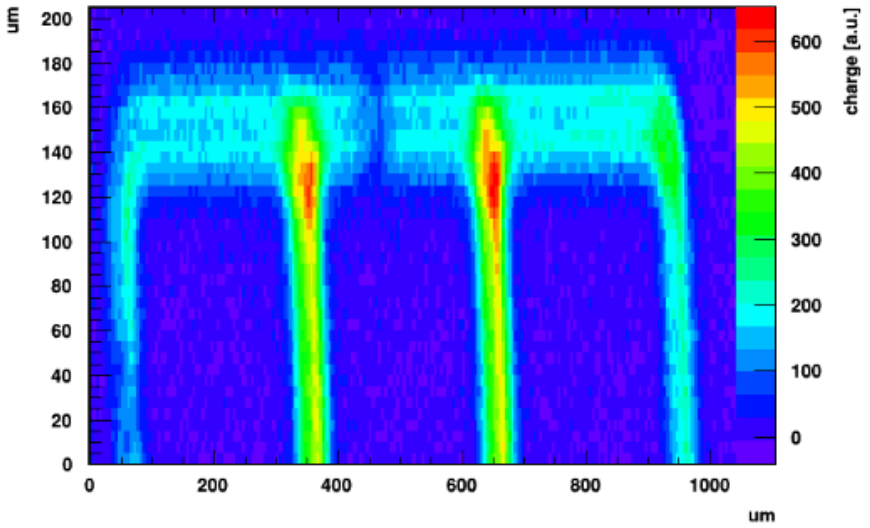
The intensity of the electric field is also affected by the sensor thickness and by the distance between JTE and *p*-stop. The *p*-stop is subject to the bias voltage applied to the sensor back-plane, which is only 50 μm underneath it, therefore it floats to a negative potential, while the JTE is connected to ground. Designs with narrower inter-pad regions are subject to higher fields, explaining their lower breakdown voltage. The electric field generated by a highly-doped *p*-stop is shown in figure 5.13.

Another element which has contributed to the observed early breakdown of some type of sensors, is the usage of the innovative stepper technique to produce UFSD3. The stepper technique creates much sharper images and more defined edges than the mask-aligner technique used for the former productions. As a consequence, the sharp edges locally create a high electric field which can make the breakdown start.

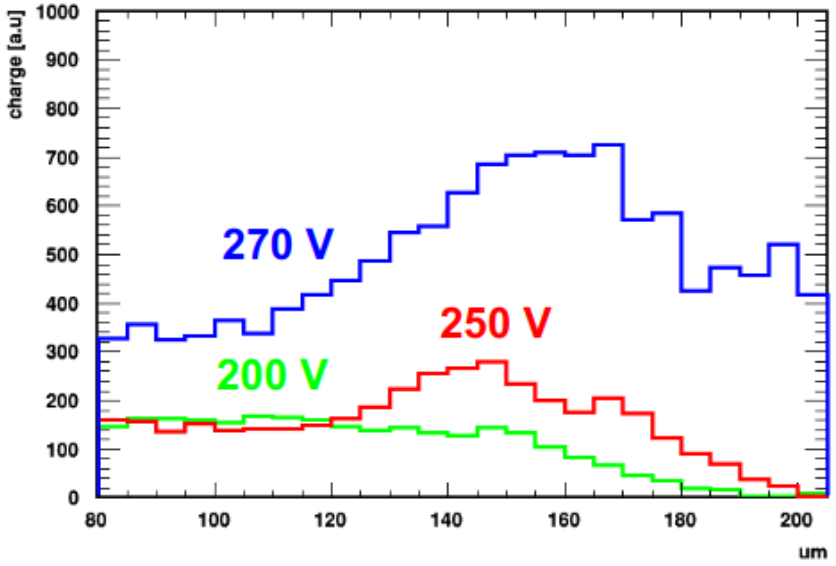
The breakdown usually starts at a location where crystal or structural defects exist and results in a sudden rise in leakage current. The location becomes hot and can be spotted with a state-of-the-art CCD camera, as discussed in the following section [78].

5.3.1 ORCA2 camera measurements

The ORCA2 camera, described in chapter 4, has been used to confirm the results observed with the TCT on UFSD3 devices.



(a) 2D-map obtained with TCT scan of the charge collected by strip sensor with Medium design



(b) Profiles of 5.10a at y=650 μm for 3 different voltages

Figure 5.10: Results of the strip sensor with Medium design

The area of the inter-pad region scanned with the TCT has been framed with the camera at different bias voltages. Figures 5.14 and 5.15 show pictures of a 2×2 array with Safe design, and of the strip device with Super-safe design, taken before and after breakdown .

The *hot spots* (yellow regions) that can be seen in figure 5.14 are regions that emit visible photons due to the high current densities flowing through. Such high densities are tied to the gain avalanche occurring while the sensor is going into breakdown, as

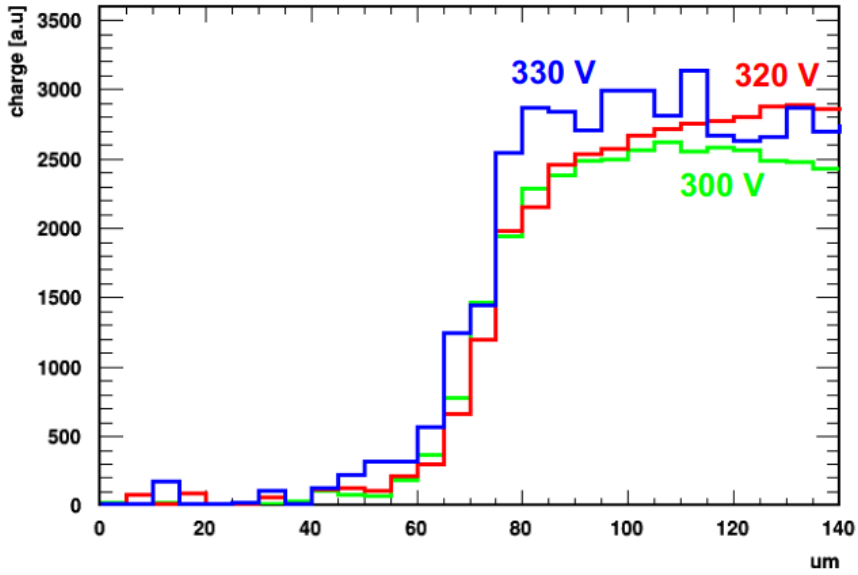


Figure 5.11: Charge collected in the inter-pad region by a strip sensor with Super-safe design for 3 different voltages

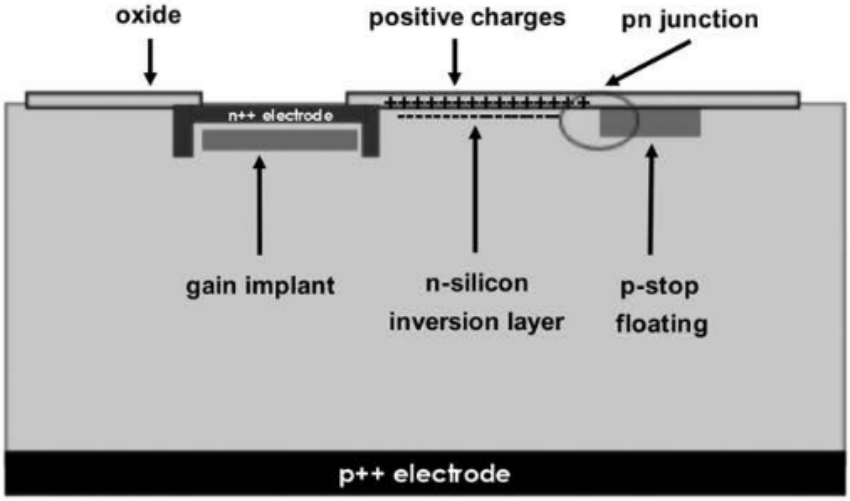


Figure 5.12: Schematic representation of the inversion layer establishing in the inter-pad region.

previously shown with the TCT laser scan.

Hot spots are particularly visible in correspondence to the corners, where the electric field is higher. Similar pictures have been obtained from the devices with Aggressive and Medium designs.

As expected, the Super-safe device does not show signs of *hot spots*, see figure 5.15.

Corners are particularly critical because three or four p-stops join and create large p-doped structures (named *Region A* and *Region B* in figure 5.16), sensitive to the

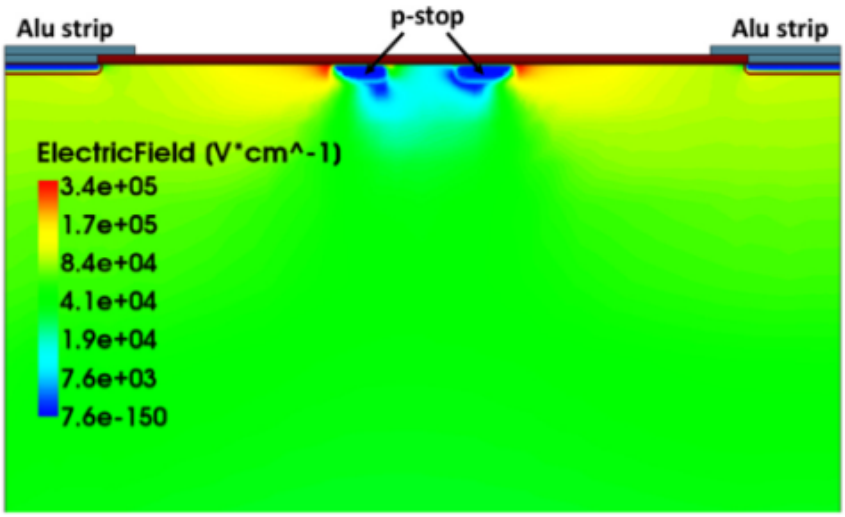


Figure 5.13: Section from T-CAD simulation of a n-in-p silicon diode with a high p-stop doping concentration. The simulation was performed at room temperature at V=-600 V. Picture taken from [77].

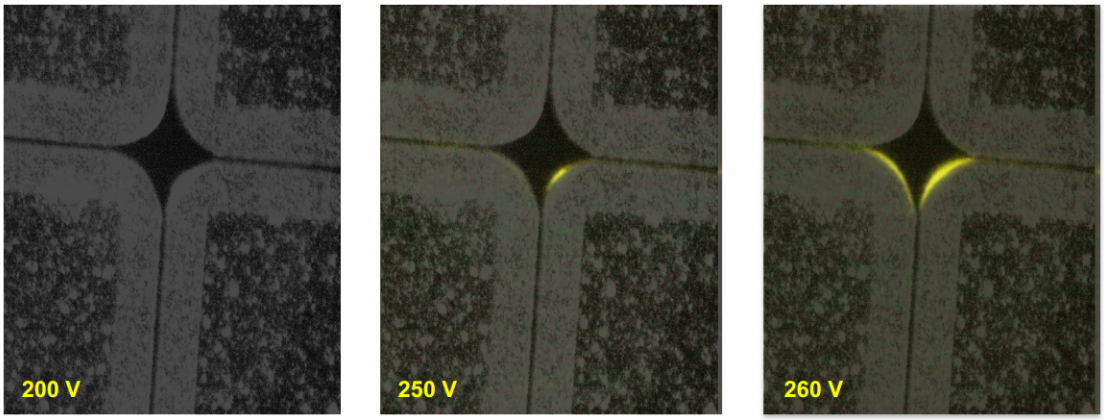


Figure 5.14: Pictures of the inter-pad region of the 2×2 device with Safe design at 3 different voltages

bias voltage applied on the sensor back-plane. As shown in the following, the larger the p-structure in the corners, the earlier the breakdown voltage.

5.4 Micro-discharges in UFSD3 sensors

A second, undesired, effect has been observed on sensors of the UFSD3 production: the appearance of random large spikes in the leakage current, much before the breakdown voltage, which prevents proper operation of the sensor, as the dark count rate increases significantly. Such large spikes have amplitude comparable to that of real signals, but

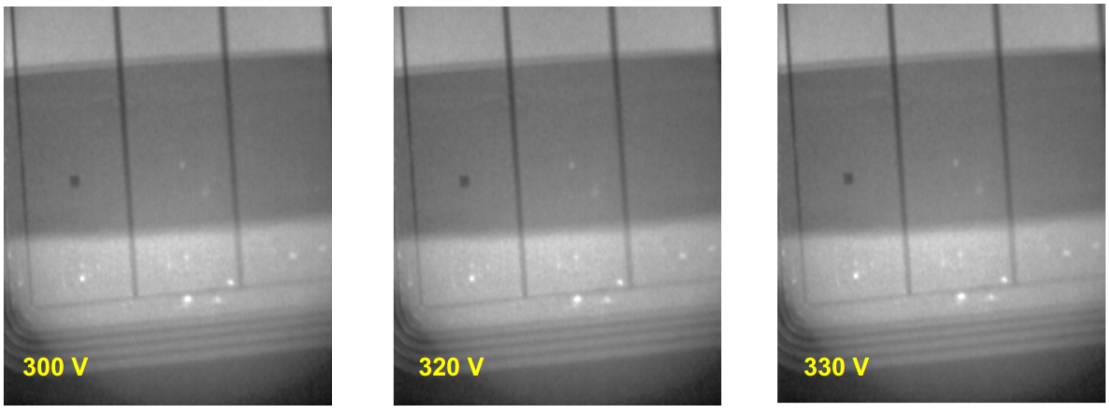


Figure 5.15: Pictures of the inter-pad region of the strip device with Super-safe design at 3 different voltages

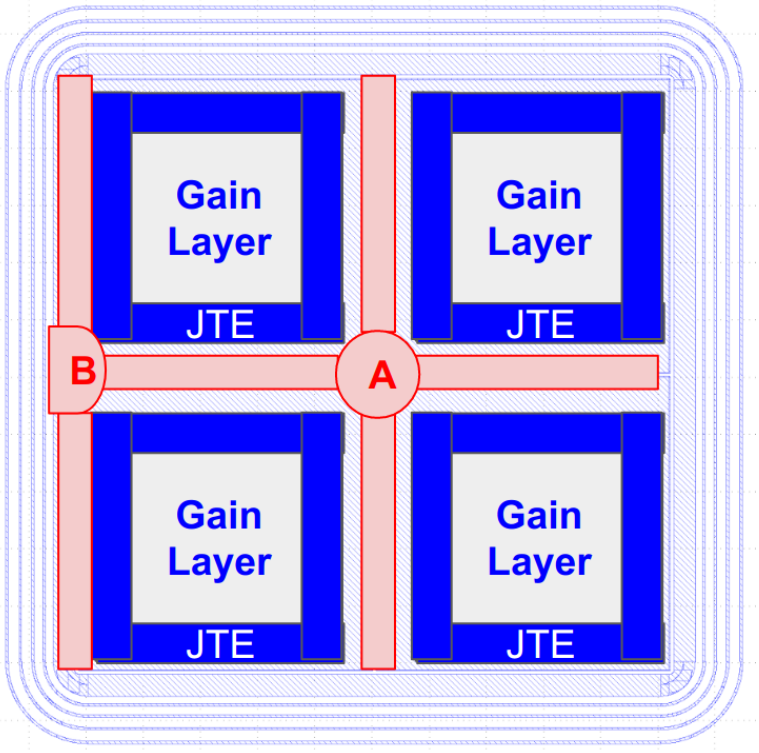


Figure 5.16: Sketch (not to scale) of the structures present in the inter-pad region of the UFSD design: the JTE (blue) and the p-stops (red). The p-stops join to form large p-doped structures (A and B in the picture) which are particularly critical for what concerns the premature breakdown.

randomly distributed in time [78–80]. Figure 5.17 presents a comparison between the normal baseline activity of a UFSD and that of a device affected by micro-discharges.

The micro-discharges appear on both strip and pad sensors and they do not depend on their geometry, neither on their inter-pad designs; they have been observed on both

new and irradiated devices.

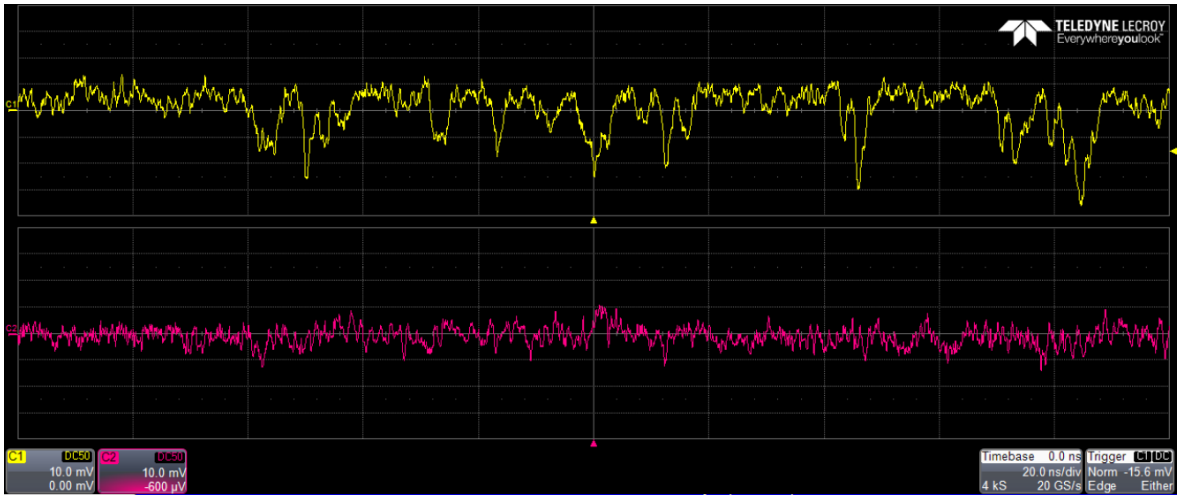


Figure 5.17: Comparison between a sensor with normal baseline activity (pink) and a sensor with micro-discharge (yellow). The vertical scale is 10 mV/division

It is worth pointing out that the increase of the baseline activity naturally happens in all UFSD devices a few volts before breakdown: it is an indication that the gain avalanche is going to start. What distinguishes the micro-discharges in UFSD3 is that they appear at a voltage much lower than the sensor breakdown voltage.

In order to observe the micro-discharges, the DUT has been bonded on a custom read-out board (figure 5.5) with all channels connected to an oscilloscope. The measurements are performed within a climate chamber, at +20 °C, with dry air fluxed. The bias voltage is raised slowly in steps of 5-10 V until the breakdown occurs; meanwhile, the baseline activity is checked on the oscilloscope.

The micro-discharges can be explained similarly to the premature breakdown: an *inversion layer* establishes in the inter-pad region, between the JTE and the *p*-stop; where defects are present, the electric field is high enough to cause a sudden increase of the leakage current, which provokes a spike in the baseline activity.

In sensors with aggressive inter-pad design the micro-discharges appear only a few Volts before breakdown or do not appear at all, simply because they go into very early breakdown (<150 V).

5.4.1 Conclusions of the measurement campaign on UFSD3

Two main issues have been identified in UFSD3 sensor production:

- The sensors with Aggressive, Medium and Safe designs suffer from premature breakdown.

- All designs show micro-discharges much before the breakdown, which prevent the proper operation of the sensors.

Interestingly, the Super-safe design is sufficiently robust to be affected by the micro-discharges without going into premature breakdown, because the JTE and p-stop are relatively far away. Despite that, its proper operation is compromised by the discharges.

The TCT scans and the measurements with the ORCA2 camera demonstrated that the issues of the UFSD3 production are tied to the strong electric fields establishing in the inter-pad region, between the JTE and the p-stop. Such electric fields are caused by:

- Highly doped p-stops
- A short distance between the JTE and the p-stops
- The use of the stepper technique
- Very large p-structures created at the corners of the sensor's pads.

The following section will describe the solutions implemented in the UFSD3.1 production.

5.5 Measurement campaign on UFSD3.1 sensors

The UFSD3.1 production is described in chapter 3. It addresses the issues of the UFSD3 designs.

The measurement campaign started analyzing the $I(V)$ characteristics of several devices: a summary is shown in figures 5.18 and 5.19.

The breakdown voltage is expected to be 360-380 V and constant, since all wafers have the same gain layer dose; sensors going into breakdown earlier are considered to be suffering from premature breakdown.

Figure 5.18 shows that sensors from the same wafer (i.e. same p-stop doping) but with different type (i.e. different inter-pad designs) have different breakdown voltages (V_{BD}), with several of them having $V_{BD} < 350V$. Therefore the premature breakdown is correlated with the sensor type.

Similarly, as shown in 5.19, sensors with the same inter-pad design but coming from different wafers have different V_{BD}, therefore premature breakdown is also correlated with the p-stop doping.

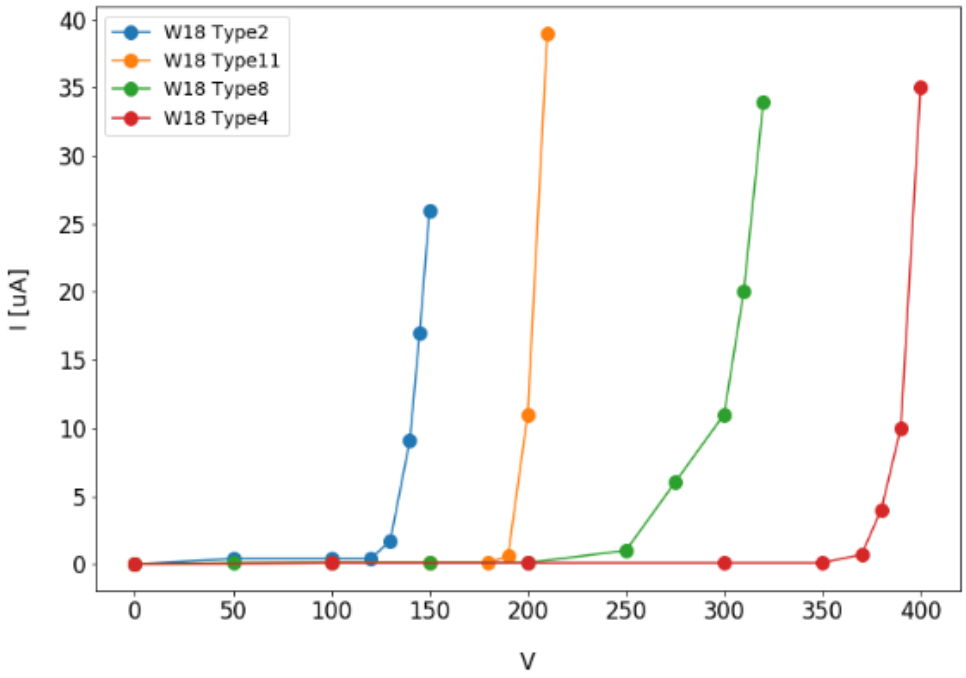


Figure 5.18: $I(V)$ of devices with different inter-pad design having the same p-stop doping

The ORCA2 camera proved to be a reliable tool to detect the *hot spots* highlighting weaknesses in the design. Given the large number of sensors to be measured, the camera has been chosen as a diagnostic tool for this production since it is much faster than the TCT scan.

5.5.1 Measurements with the ORCA2 camera

The measurements have been performed with the same procedure followed with UFSD3 sensors. Some of the pictures taken are shown in 5.20, 5.21.

The obvious feature visible in these pictures is that all sensors with premature breakdown show *hot spots* in the *Regions A/B*, which are particularly critical also in UFSD3.

Regions A/B have been introduced in the previous section and are sketched in 5.16, whereas chapter 3 details their different shapes and areas for the UFSD3.1 production.

The measurements provided some useful observations:

- Premature breakdown occurs in the inter-pad region, as in the UFSD3 production.
- Premature breakdown depends both on the p-stop doping and on the inter-pad design/width.

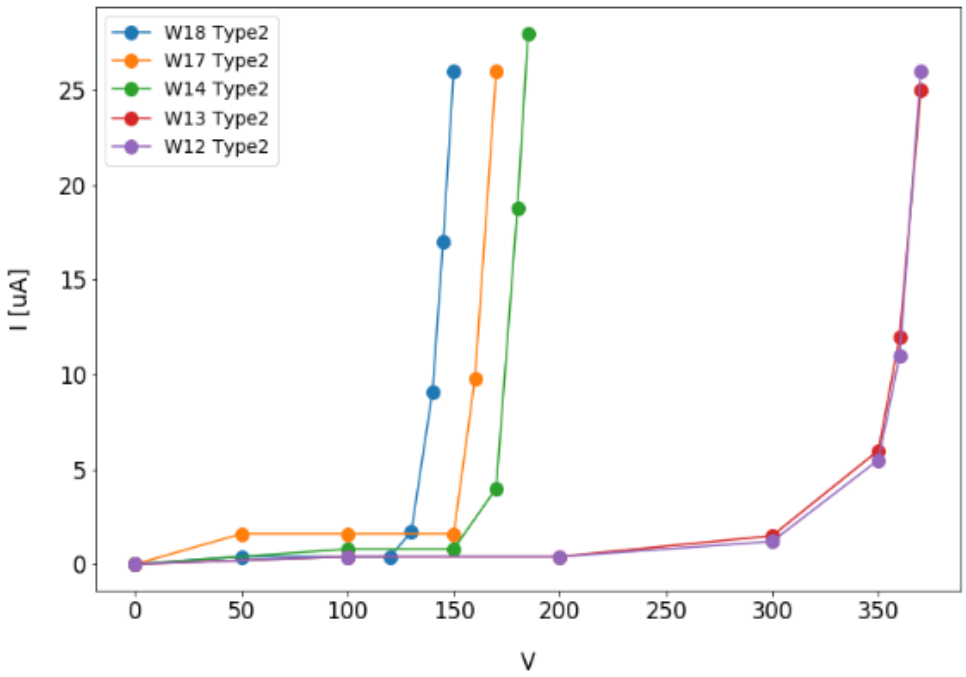


Figure 5.19: $I(V)$ of devices with different p-stop dopings having the same inter-pad design

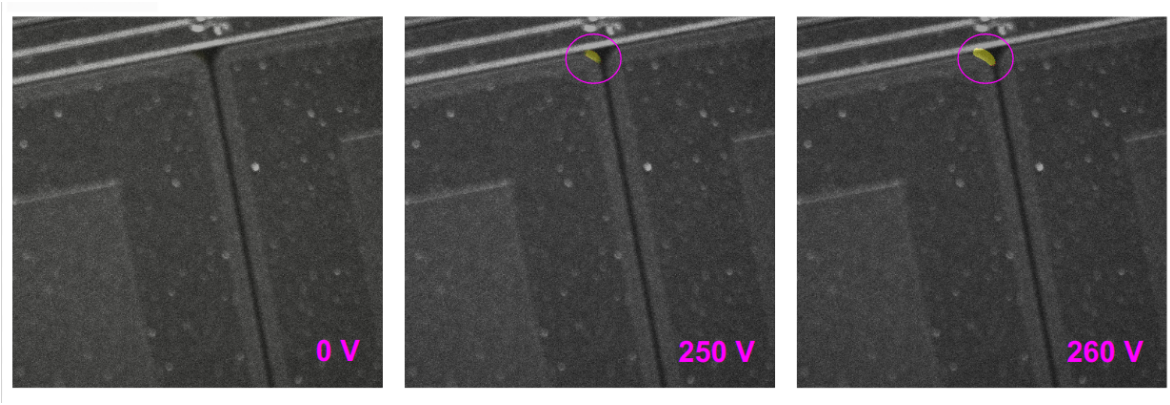


Figure 5.20: *Hot spots* in a Wafer18 type 2 sensor appearing in what has been labelled as *Region A*

- Corners (*Regions A/B*) are the most critical areas.

Figure 5.22 shows the relationship between V_{BD} and the area of *Region A*: larger p-structures lead to earlier breakdown (considering sensors with fixed geometry and gain layer doping), as expected from the measurements on the former production. Sensors with less-doped p-stops are less sensitive to this effect (W13 in the figure). A similar result is obtained when considering V_{BD} as a function of the *Region B* area.

Another important analysis is the study of V_{BD} as a function of the p-stop doping

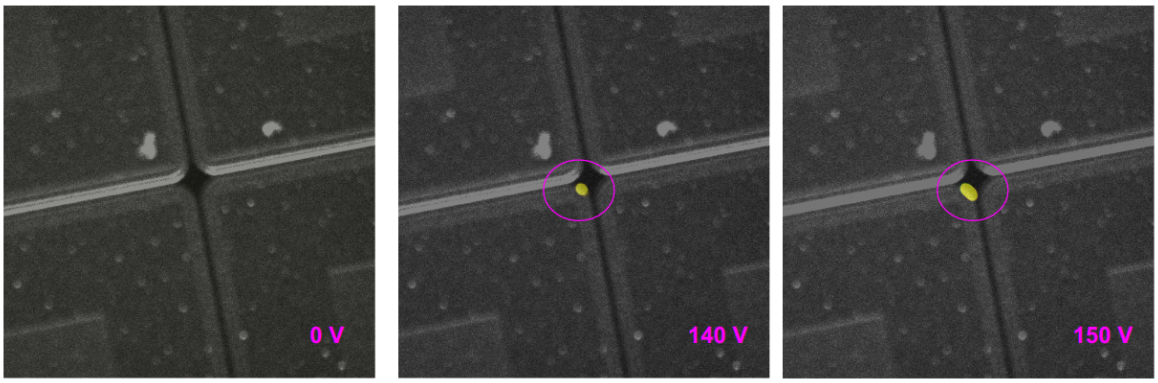


Figure 5.21: *Hot spots* in a Wafer14 type 1 sensor appearing in what has been as *Region B*

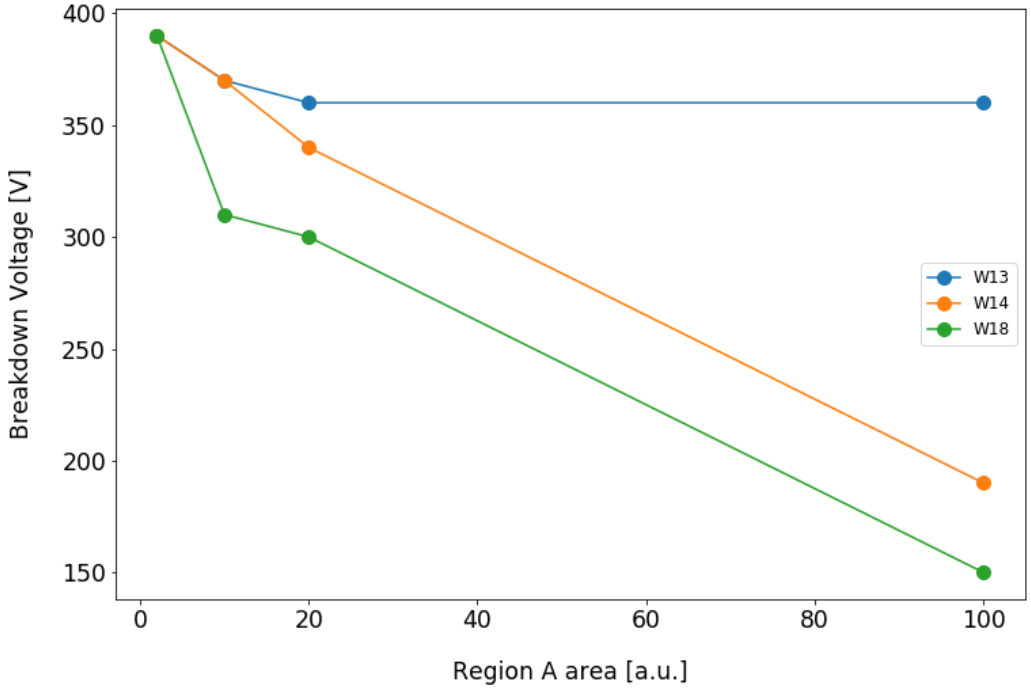


Figure 5.22: V_{BD} as a function of *Region A* area for different Wafers (different p-stop doping)

for sensors with different inter-pad designs. Figure 5.23 highlights that a less-doped p-stop allows for a higher V_{BD}, as expected. Types 3 and 4 have a low dependence on the p-stop doping, because they have small *Region A/B* areas; type 10, instead, is not susceptible to the p-stop doping because of its grid guard-ring design (see chapter 3). Type 1 has small p-structures too, but it is strongly dependent on the p-stop doping,

since it features the most aggressive design (the distance between JTE and p-stop is the shortest among all types).

The figure also shows that sensors from the wafers with lower p-stop doping (W12 and W13) have a V_{BD} that saturates around 360-380 V, regardless of their design, because at that voltage the breakdown is caused by internal gain.

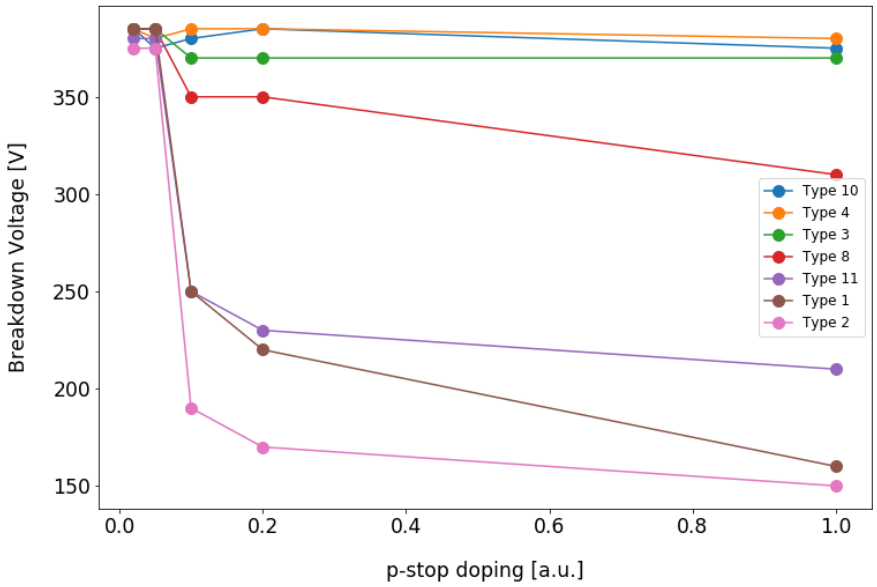


Figure 5.23: V_{BD} as a function of p-stop doping for different sensor types

In summary:

- Large and/or too-doped p-structures lead to premature breakdown, preventing proper operation of sensors.
- The combination of low-doped p-stop and small-area structures in the inter-pad region turned out to be the optimal choice to avoid early breakdown.
- An inter-pad width of $\sim$ 25 μm is achievable without incurring in premature breakdown (UFSD3.1 type 4), adopting low-doped p-stops. Sensors with narrower regions (type 1) look difficult to operate reliably.

Interestingly, the most robust designs are types 4 and 10: while type 10 was expected because it has a wide inter-pad and the grid guard-ring structures, type 4, instead, features a rather aggressive design.

In wafers 12 and 13 the p-stop dose is such to determine breakdown for internal gain, regardless of the inter-pad design. Wafer 14 shows a similar trend, with only the most aggressive designs going into early breakdown.

5.5.2 Measurements of UFSD3.1 sensors with floating pads

The behavior of a sensor with floating pads needs to be studied for sensors to be employed at the ETL, where large arrays are expected to function properly even with a few broken pads. Therefore the breakdown voltage of several 2×2 devices with 0, 1, and 2 floating pads has been measured for different wafers.

Figure 5.24 shows that sensors from W13 are almost insensitive to floating pads, whereas almost all sensor types from W18 have a breakdown voltage decreasing with the number of floating pads (figure 5.25). Hence, a low-doped p-stop looks key to preventing premature breakdown in arrays with one or more floating pads.

Type 10 is very resilient to floating pads, regardless of the p-stop doping, because of its grid guard ring design (see chapter 3), in which all pads are isolated from the neighbouring ones.

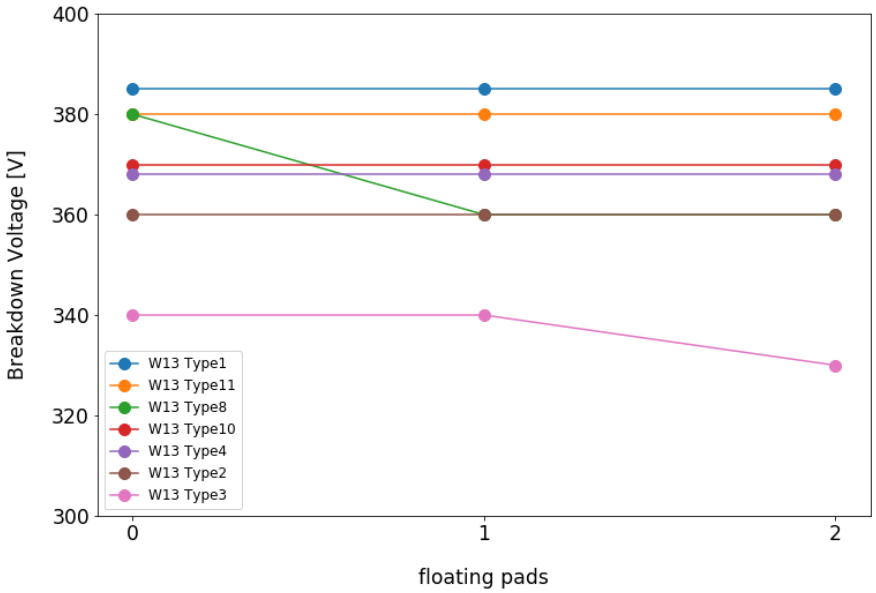


Figure 5.24: V_{BD} as a function of the number of floating pads for different sensors from Wafer 13. Types 4 and 10 have the same V_{BD} (370 V), but the curves have been slightly shifted for illustration purposes.

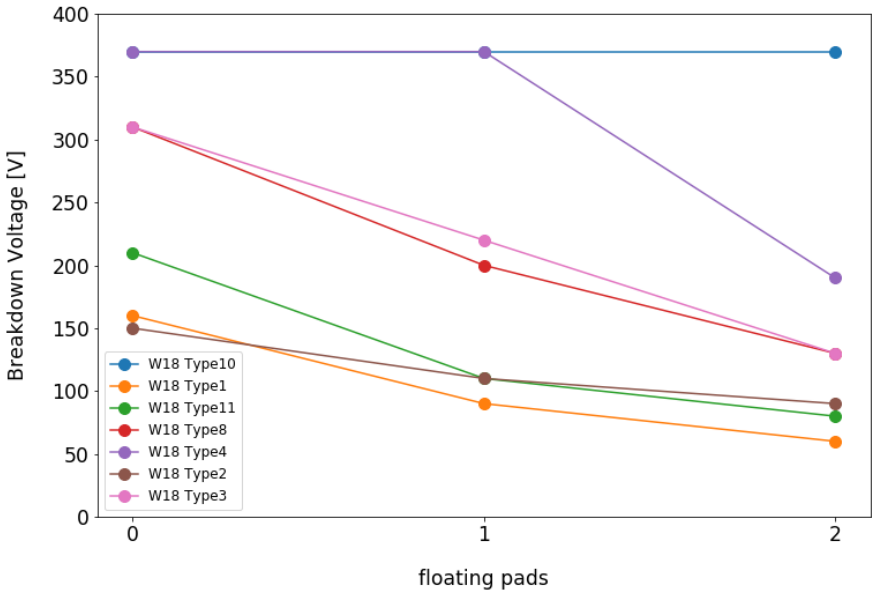


Figure 5.25: V_{BD} as a function of the number of floating pads for different sensors from Wafer 18

5.5.3 Stability of UFSD3.1 sensors with low-doped *p*-stops

The previous results lead to the conclusion that very low-doped p-stops are the most reliable. However, it has been noticed that sensors from wafer 12 (lowest dose), are not stable: they suddenly go into breakdown after being biased for some time (30 to 90 minutes) at a voltage which is lower than their initial V_{BD}. Similarly, if the $I(V)$ characteristics are repeated many times, a progressive downshift of V_{BD} can be seen, as shown in figure 5.26.

Wafer 13 and 14 do not show this behavior, therefore they have been chosen as the best wafers for what concerns the p-stop doping.

5.6 Micro-discharges in UFSD3.1

UFSD3.1 Wafers 13 and 14 have been tested for micro-discharges effect, observed in the UFSD3 production, following the procedure described in section 5.4.

The results are reported in tables 5.2 and 5.3.

Wafer 13 does not show any signs of discharges, even with 1 pad floating; whereas wafer 14 has some, occurring 15-25 V before the breakdown, but only in the type 1 design, which is the most aggressive.

Micro-discharges are therefore not an issue for the two best wafers of the UFSD3.1

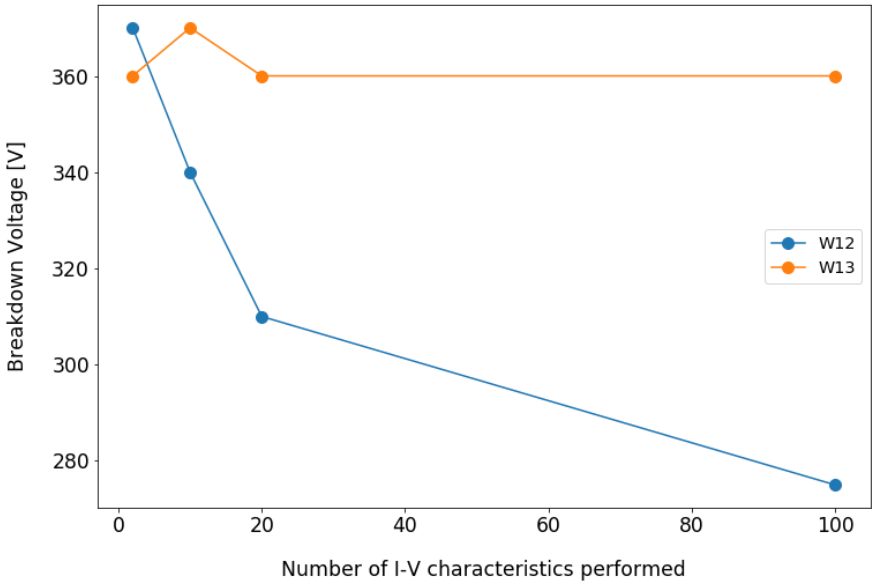


Figure 5.26: V_{BD} as a function of the number of $I(V)$ characteristics performed. The plot shows the instability of sensors with a too low p-stop doping

	0 pad floating			1 pad floating		
Type	V_{BD}	Discharge	$V_{Discharge}$	V_{BD}	Discharge	$V_{Discharge}$
Type 1	385	NO		350	NO	
Type 8	380	NO		380	NO	
Type 9	365	NO		355	NO	
Type 10	370	NO		335	NO	
Type 11	380	NO		335	NO	

Table 5.2: Results on the micro-discharges effect on UFSD3.1 W13.

production, since they show up only in the most aggressive design of W14, and only very close to breakdown, not really affecting the device operations.

5.6.1 Irradiated W14

Irradiation has two main effects that can have an influence on the micro-discharges: (i) the sensor is operated at higher bias voltage to compensate for the loss of collected charge; (ii) the *acceptor removal* mechanism (see chapter 2) decreases the doping of the p-stop structures. The two effects act in opposite directions: (i) leads to an increase in the electric field, whereas (ii) lowers it.

Table 5.4 reports the results obtained on wafer 14, irradiated at $4 \cdot 10^{14}$ n_{eq}/cm^2 and

	0 pad floating			1 pad floating		
Type	V_{BD}	Discharge	$V_{Discharge}$	V_{BD}	Discharge	$V_{Discharge}$
Type 1	260	YES	245	170	YES	145
Type 8	350	NO		350	NO	
Type 9	370	NO		310	NO	
Type 10	380	NO		380	NO	
Type 11	250	NO		185	NO	

Table 5.3: Results on the micro-discharges effect on UFSD3.1 W14.

$8 \cdot 10^{14}$ n_{eq}/cm^2. The sensors have been irradiated, without bias, with neutrons at the JSI TRIGA research reactor in Ljubljana [81].

	$4 \cdot 10^{14}$ n_{eq}/cm^2			$8 \cdot 10^{14}$ n_{eq}/cm^2		
Type	V_{BD}	Discharge	$V_{Discharge}$	V_{BD}	Discharge	$V_{Discharge}$
Type 1	390	YES	270	440	YES	390
Type 8	470	NO		510	NO	
Type 9	470	NO		540	NO	
Type 10	500	NO		590	NO	
Type 11	430	YES	300	530	NO	

Table 5.4: Results on the micro-discharges effect on UFSD3.1 W14 irradiated.

Types 1 and 11 show micro-discharges at $4 \cdot 10^{14}$ n_{eq}/cm^2, whereas only type 1 features this effect at $8 \cdot 10^{14}$ n_{eq}/cm^2. Similar results are obtained in the "1 floating pad" configuration: only types 1 and type 11 show signs of micro-discharges.

Interestingly, at $1.5 \cdot 10^{15}$ n_{eq}/cm^2, all types go into breakdown above 600 V and none has micro-discharges before V_{BD}.

5.6.2 W14 inter-pad resistance

It was previously mentioned that, although low *p*-stop dopings are not affected by premature breakdown, they can impact the pad isolation, since the irradiation decreases the dose value. It is therefore important to measure the pad isolation and its evolution with the irradiation fluences.

The chosen figure of merit that quantifies the pad isolation is the inter-pad resistance, namely the resistance of a pad to ground when all the other pads and the guard-ring are connected to ground.

The measurements have been performed on 2×2 arrays types 8,9 and 10, at -20 °C. Three pads and the guard-ring are grounded, while a voltage sweep between -10 V and +10 V is performed on the pad under test, as shown in figure 5.27. The measured

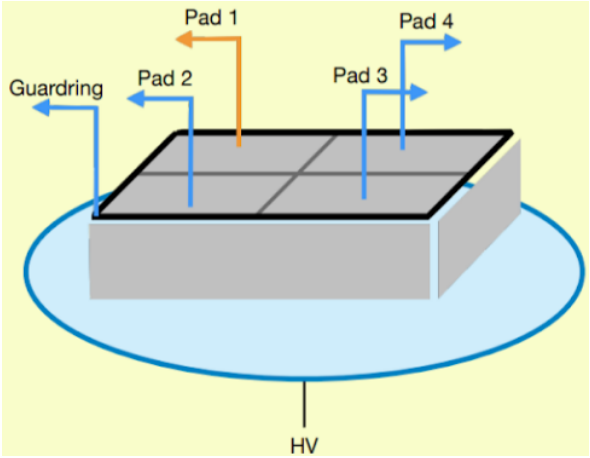


Figure 5.27: Sketch of the setup used to measure the inter-pad resistance.

current is plotted as a function of the bias voltage applied to the pad: the slope of the curve provides the inter-pad resistance. The measurement is performed on a pre-irradiation sensor and repeated on devices irradiated at $4{\cdot}10^{14}$ n_{eq}/cm^2, $8{\cdot}10^{14}$ n_{eq}/cm^2, $1.5 \cdot 10^{15}$ n_{eq}/cm^2, $3 \cdot 10^{15}$ n_{eq}/cm^2.

The results are shown in figure 5.28: the resistance lowers with increasing radiation dose, as expected, but it remains high, about 10 GΩ, even at the highest fluence, proving that the pad is well isolated. As a comparison, ETL requires an inter-pad resistance higher than 0.1 GΩ up to a radiation fluence of $1.5 \cdot 10^{15}$ n_{eq}/cm^2 [82].

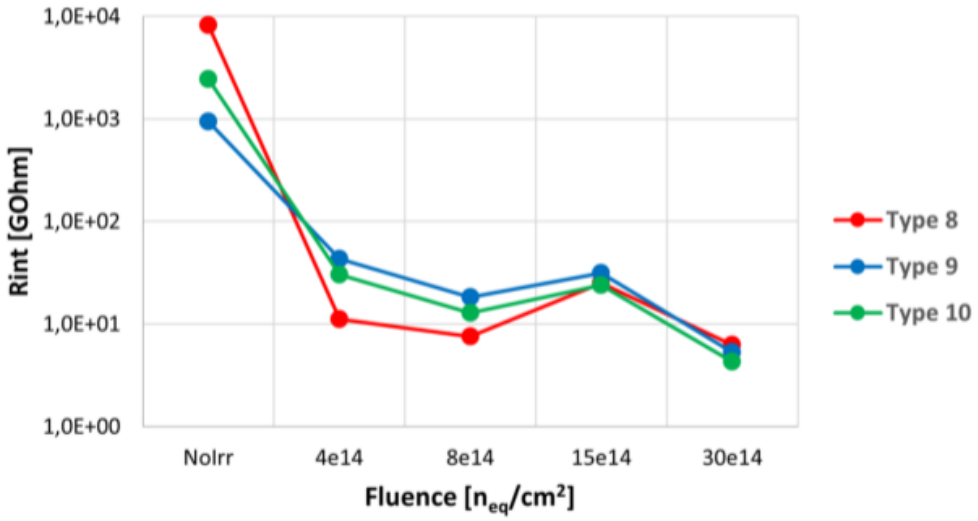


Figure 5.28: Inter-pad resistance of UFSD3.1 W14.

5.7 The optimal inter-pad design and UFSD3.2

The study of the UFSD3 and UFSD3.1 productions provided useful information which helped defining the following FBK production, the UFSD3.2 (see chapter 3).

UFSD3.1 Wafer 13 and 14 demonstrated to have the optimal p-stop doping: low enough to avoid premature breakdown, while being able to reliably hold the bias voltage. In particular, UFSD3.2 has been produced with the p-stop dose equivalent to UFSD3.1 Wafer 14.

Concerning the inter-pad designs, the UFSD3.2 production features almost all the types implemented in UFSD3.1, in the 2×2 arrays. The 5×5 matrices, instead, have been designed with three types of inter-pad, the ones considered as the best candidates from UFSD3.1:

1. Type 9 [1], which is the UFSD3 Super-safe design, but with a low-doped p-stop.

2. Type 10, same as UFSD3.1. The grid guard-ring provides a significant improvement in the inter-pad design. The downside is a rather large width ($\sim$ 62 μm). An improved design with a slightly narrower no-gain region might be possible.

3. Type 12 is an improved version of the UFSD3.1 types 4, with an interpad nominal width of 28 μm. It is the most aggressive design for what concerns the 5×5 matrices.

This production features also the PIN diode version of all 2×2 types, ideal to study the strength of the inter-pad design. Indeed, if the PIN diode goes into breakdown at a voltage higher than its LGAD counterpart, then the LGAD is not suffering from premature breakdown, since early breakdown occurs at the same bias voltage in both the LGAD and the PIN, because it does not depend on the presence of the gain layer.

The lower V_{BD} of the LGAD with respect to the corresponding PIN proves that the device is going into breakdown because of the gain.

At room temperature, all UFSD3.2 LGAD arrays break down below 400 V, as shown in figure 5.29, which presents the $I(V)$ curves of the arrays with the highest V_{BD} in the production. The measurements on the 2×2 PIN diodes are instead reported in figure 5.30, and confirm the UFSD3.1 results: types 4 and 10 are the most robust and go into breakdown between 400 and 500 V, therefore such designs are not suffering from premature breakdown.

[1] Type 9 is not as good as the other two types chosen for the 5×5 matrices, but it is the best known since it has been the reference design since UFSD3.

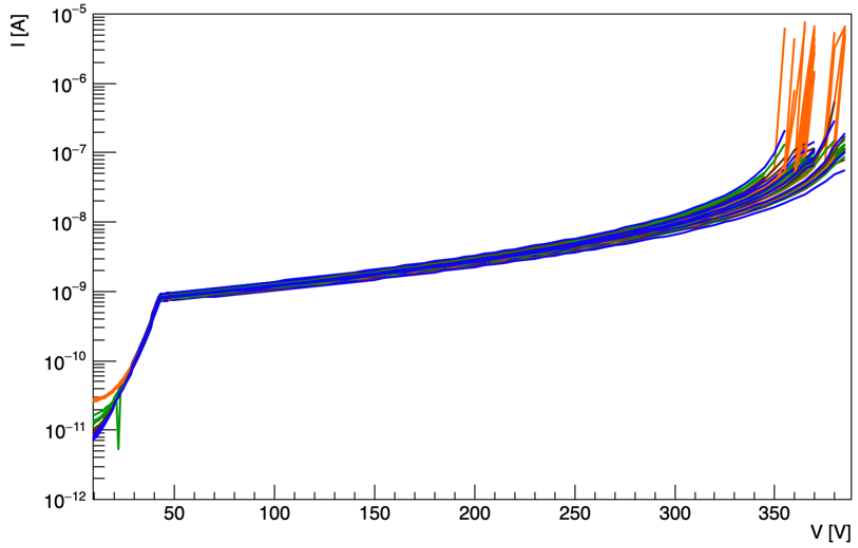


Figure 5.29: $I(V)$ characteristics of LGAD arrays coming from Wafer 10 of the UFSD3.2 production. Measurements performed at room temperature.

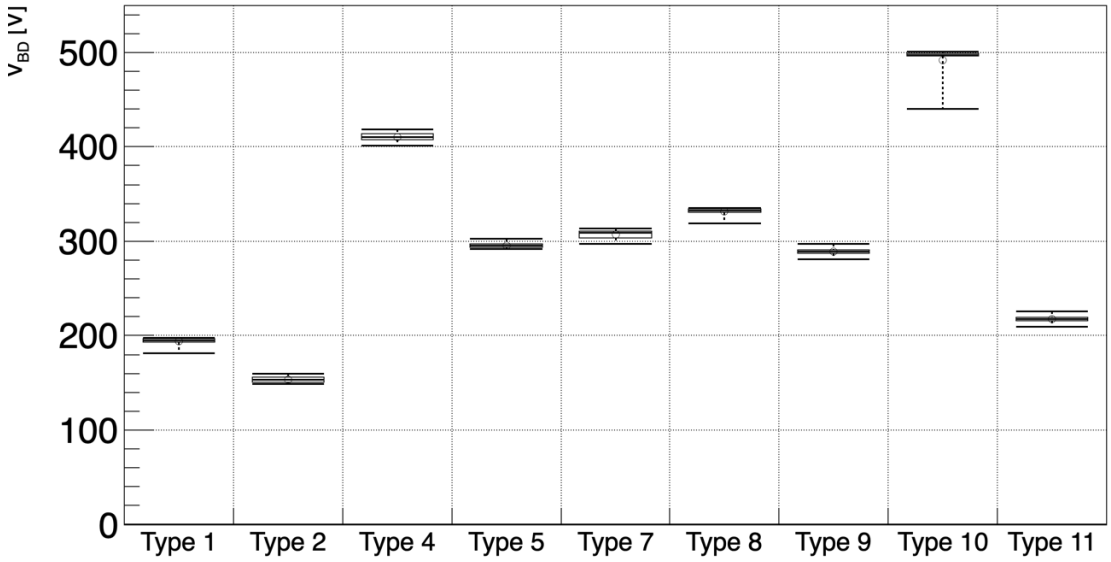


Figure 5.30: Box plot illustrating the breakdown voltages of 2×2 PiN arrays with different inter-pad designs (types). Measurements performed at room temperature.

The same measurements were repeated on the large 5×5 PIN arrays (figure 5.31), this time with 0, 1 and 24 floating pads. Type 10 and the innovative type 12 proved to be the most robust, with a breakdown of $\sim$ 500 V with 0 floating pads, and $\sim$ 450 V with 1 or 24 floating pads.

The micro-discharge effect has not been observed in any of the tested sensors, either

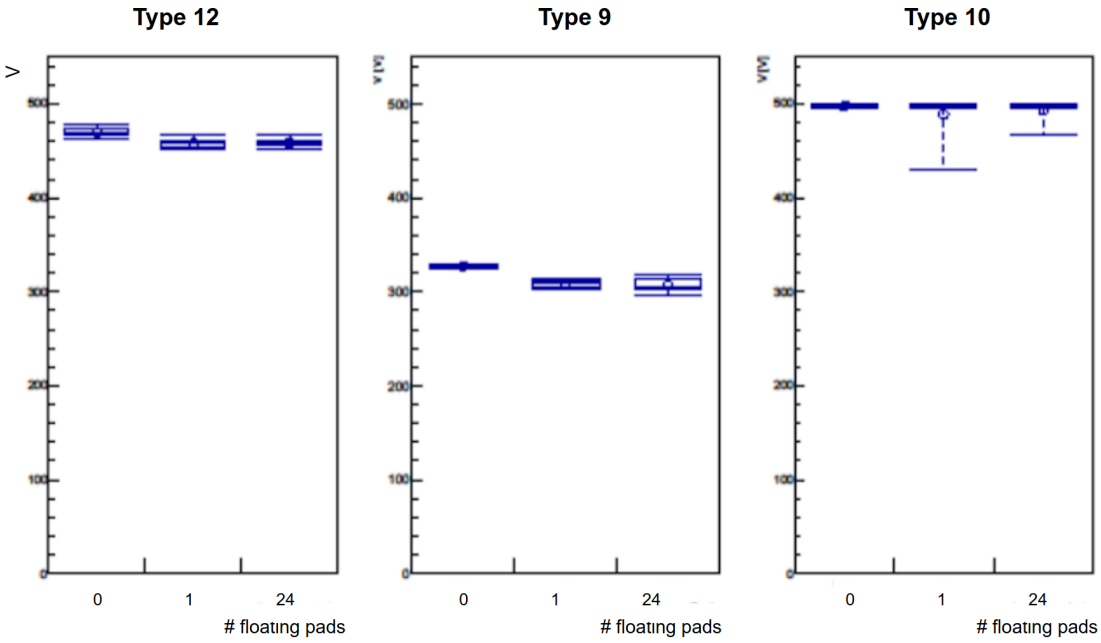


Figure 5.31: Box plot illustrating the breakdown voltages of 5×5 PiN arrays with different inter-pad designs (types) and different configurations of floating pads. Measurements performed at room temperature.

before or after irradiation.

The UFSD3.2 production thus fixes the issues observed in the previous productions, and provides two (or, more precisely, three, since types 4 and 12 are slightly different) robust and reliable inter-pad designs, with high resilience to floating pads and irradiation, for the final UFSD arrays to be installed in ETL.

Chapter 6

Optimization of the UFSD Gain Layer Design

The UFSDs instrumenting the ETL are required to precisely measure the time of passage of charged particles with 30-40 ps resolution up to the end of the HL-LHC lifetime, when they will have received a radiation fluence (Φ), expressed in *1 MeV neutron equivalent*, of $1.6 \cdot 10^{15}$ n_{eq}/cm^2, and absorbed a dose of 450 kGy. The expected maximum radiation fluence further grows to $2.5 \cdot 10^{15}$ n_{eq}/cm^2, when considering also the 1.5x safety factor which accounts for the uncertainties in the predictions from the geometry model, the *pp* inelastic cross-section, and possible sensor-to-sensor variations.

The chapter presents UFSDs with four different types of gain layer and four different active thicknesses, measured using the β-source setup described in chapter 4. The goal of this measurement campaign is to pinpoint the gain layer design that best fits the ETL needs, the one that will be used to produce the final UFSD sensors. The β-source setup is a valuable tool to achieve this goal since it allows measuring fundamental quantities, such as the time resolution, using MIPs, and with a rather high testing rate [71].

Along with this measurement campaign, and exploiting the variety of designs considered, several additional studies on the UFSD working principles have also been performed, whose results will be presented in the following.

Almost all sensors presented in the chapter have been irradiated, without bias, with neutrons at the JSI TRIGA research reactor in Ljubljana [81], up to
$\Phi = 2.5 \cdot 10^{15}$ n_{eq}/cm^2.

6.1 Tested sensors

All tested sensors are single-pad UFSDs with 1.3×1.3 mm^2 active area. A large fraction of the measured devices come from the FBK UFSD3.2 production, presented in chapter 3, which features an unprecedented variety of gain layer designs. Wafers 1, 3, 6, 7, 10, 12, 13, 14, 15, 18, 19 have been tested, whose main characteristics are reported in table 6.1. All sensors have been measured both before and after irradiation, except for Wafer 6, the 35 μm-thick sensor, which has been measured only un-irradiated.

Wafer	Thickness [μm]	GL depth	B dose [a.u.]	C dose [a.u.]	Diffusion scheme
1	45	Shallow	0.98	1	CHBL
3	45		0.98	0.8	
6	35		0.94	1	
7	55		0.98	1	
10	45	Deep	0.70	0.6	CBL
12	45		0.74	1	
13	45		0.74	0.6	
14	45	Deep	0.74	1	CBH
15	55		0.74	1	
18	45		0.78	1	
19	45		0.78	0.6	

Table 6.1: The wafers of the UFSD3.2 production tested in this work. The active thickness, the gain layer (GL) depth, the boron (B) and carbon (C) doses, and the diffusion scheme are reported.

Two different types of sensors manufactured by Hamamatsu Photonics (HPK) are also included in this analysis. The first type, the so-called HPK split 4, belongs to the HPK2 production presented in chapter 3; the second type of sensor here analyzed belongs to the HPK ECX20840 production (see chapter 3), and it is called HPK 80D, featuring a 80 μm-thick active area with a shallow gain implant, whose split dose is called "D", the lowest one in the ECX20840 production. The HPK 80D sensor has been measured un-irradiated only.

The HPK 80D and the FBK UFSD3.2 W6 have been considered only for the additional studies on the UFSD working principles, and not for the measurement campaign to define the best gain layer design for ETL.

6.2 Description of the Data Analysis Methods

Several important parameters related to the sensor performance can be measured using the β-source setup: the collected charge; the electronic noise (sensor-amplifier depen-

dent); the value of the gain; the time of arrival (ToA) and the time resolution.

In the following, the *Trigger*, with a capital t, will identify the sensor used to *trigger* an event.

For all these measurements, the offline event selection requires: (i) the Trigger signal amplitudes to be in the 80-250 mV range; (ii) the Trigger collected charge in the 11.5-53.0 fC range; (iii) the DUT signal amplitude above 10 mV, about 5 times the RMS noise; (iv) the signal not to saturate either the oscilloscope vertical scale or the amplifier. Point (iv) rejects events with DUT signal amplitudes greater than 300 mV.

Point (iii) introduces a small bias in the analysis of the sensors irradiated at a fluence of $2.5 \cdot 10^{15}$ n_{eq}/cm^2 when biased at voltages $\leq$ 500 V. Figure 6.1 shows the Landau distribution of an FBK UFSD3.2 W14 sensor, which is one of the least radiation-hard among tested devices, irradiated at $2.5 \cdot 10^{15}$ n_{eq}/cm^2, and biased at 500 V. In this case, the lower tail of the Landau distribution partially overlaps with the noise.

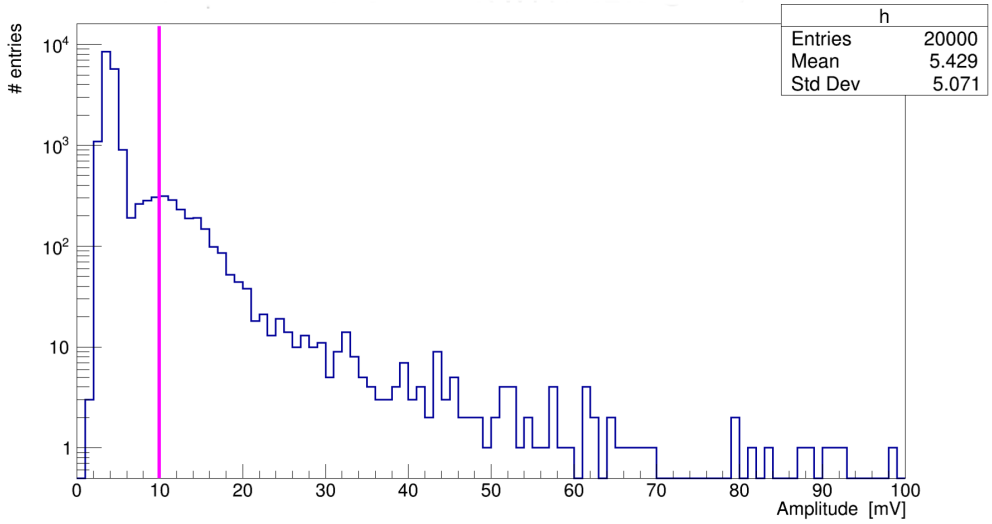


Figure 6.1: Amplitude distribution of the FBK W14 irradiated at $\Phi = 2.5 \cdot 10^{15}$ n_{eq}/cm^2, biased at 500 V. The pink line shows the amplitude cut applied in the event selection. Measurement performed at -25 °C.

6.2.1 Collected Charge, Noise, and Gain

The measurement of the collected charge for each event is derived from the signal area, obtained by integrating the signal over time using the Simpson's method [83], divided by the read-out system trans-impedance, 4700 Ω (see chapter 6). Baseline subtraction is performed on each waveform, using the first 100 sampled points (5 ns), taken before the signal occurs. The same 100 points are also used to measure the sensor-electronics RMS noise.

The gain of the UFSD sensor can be defined as the ratio of the charge collected by the UFSD at a given voltage, $Q(V)$, over that collected by an equivalent PIN diode, Q_{PIN}:

$$Gain = \frac{Q(V)}{Q_{PIN}} \tag{6.1}$$

where $Q(V)$ and Q_{PIN} refer to the most-probable-value (MPV) of the UFSD and PIN collected charge distributions, respectively.

Q_{PIN} is assessed at a voltage below 500 V, so that gain in the bulk cannot happen (see chapter 2 for details on the gain in the UFSD bulk): in this way, Q_{PIN} can be considered a constant, and equation 6.1 takes into account both the UFSD gain from the gain layer and that from the bulk.

Since experimentally the measurement of the charge collected by the PIN diode leads to large uncertainties (given its rather small value, about 0.5 fC), it was decided to use the theoretical value of Q_{PIN}: ~ 0.45 fC in a 45 μm-thick sensor and ~ 0.55 fC in a 55 μm one [69]. The charge computed from theory is affected by the uncertainty on the active sensor thickness, which results in a relative uncertainty of $\sim 4\%$ for a 50 μm-thick sensor, much lower than the one affecting the Q_{PIN} measurement ($\geq 20\%$).

The theoretical Q_{PIN} value has been used to compute the gain of UFSDs before and after irradiation.

6.2.2 Time of Arrival and Time Resolution

The time of arrival (ToA) of a signal is computed using the constant-fraction discriminator (CFD) method, i.e., it is defined as the time at which the signal crosses a certain fraction of the total signal amplitude. A 20% threshold (CFD20) provided the best results for most of the sensors tested. This fraction was increased to 30% (CFD30) for a few heavily irradiated sensors due to their higher noise and a slightly different signal shape caused by charge trapping, see [42] for details.

For the evaluation of the time resolution of a sensor biased at a given value, the distribution of the difference between the DUT ToA and the Trigger ToA is plotted, and then fitted with a Gaussian function whose standard deviation is the squared sum of the DUT and Trigger resolution, hence:

$$\sigma_{DUT} = \sqrt{\sigma^2_{measured} - \sigma^2_{Trigger}} \tag{6.2}$$

6.2.3 Evaluation of the uncertainties on the measured quantities

Due to the large number of sensors tested and the relatively low trigger rate, it would have been very challenging to assess the uncertainties on the measured quantities by repeating the measurement of each sensor several times. However, several steps were taken to keep the uncertainties under control: (i) the same Trigger has been used in each of the 250 measurements of this campaign, (ii) the Trigger plane was always operated at the same temperature, and voltage (except for the temperature scan); and (iii) for a given sensor (pre-irradiation UFSD3.2 W7), a set of repeated measurements was carried out to evaluate all the uncertainties.

Table 6.2 shows the list of uncertainties affecting the quantities measured at the β-source setup.

Table 6.2: The uncertainties on the main quantities measured at the β-source setup.

	uncertainty
time resolution	1.5 ps
collected charge	0.2 fC
gain	0.5
RMS noise	0.2 mV
Slew rate	2.9 mV/ns

6.3 Characterization of a gain layer design for ETL: Q, σ_t, $\mathbf{V}_{10fC}$, $\Delta\mathbf{V}_{10fC}(\Phi)$, k, and noise

The evaluation of a gain layer design for ETL needs to include several aspects. Obviously, the time resolution of the sensor is the first quantity to be considered in its characterization. However, the time resolution achieved in laboratory tests with a state-of-the-art analog read-out board (as the one used in this work, see chapter 4) does not guarantee that the same results can be achieved when the sensor is bonded to the ETL ASIC, ETROC, which is a much more complex, low-power electronics. Additional quantities, such as the collected charge and the noise level, need to be evaluated, as they determine the final achievable performance.

Considering the ETROC requirements [2, 35], a signal charge larger than 5 fC is needed to achieve a time resolution $\sigma_t \lesssim 50$ ps, which is the target resolution of the bump-bonded UFSD+ETROC pair. In general, the higher the charge, the better the

front-end ASIC performs. Likewise, the noise arising from the UFSD gain mechanism needs to be kept as low as possible, below the front-end electronic noise.

When dealing with a 14 m^2 detector, like ETL, to be instrumented with thousands of UFSDs, not only the performance of the single device has to be considered, but also the uniformity and the overall performance of the entire set of sensors. The gain implant non-uniformity, for instance, is a key aspect: different UFSD sensors might have different gain layer doses, with a sensor-to-sensor spread reported to be of about 2% [84]. Hence, ETL will be equipped with UFSDs having slightly different dopings, but biased at the same voltage (it is not possible to bias every single array at its own optimal voltage), leading to non-uniform performance.

Another aspect to be considered, although not as important as the non-uniform doping, is the non-uniform biasing conditions that ETL sensors might encounter: indeed, although a nominal bias voltage will be set, there might be small sensor-to-sensor variations in the actual bias voltage received, leading also in this case to non-uniform performance.

Both aspects depend upon the charge-voltage characteristics, $Q(V)$: the steeper the curve, the less reliable will be the operation of the UFSD arrays. The steepness of the charge curve is quantified by fitting the charge-voltage characteristics with an exponential:

$$Q(V) = Q_{PIN} \cdot e^{k \cdot V} \tag{6.3}$$

with the k [V^{-1}] parameter being a measurement of the curve slope; Q_{PIN} represents the charge collected by a PIN diode in absence of bulk gain. k is proportional to the electric field in the sensor, and, consequently, to the gain layer doping.

It is possible to link the $Q(V)$ steepness to the non-uniform doping. One can consider two UFSDs having different dopings, and, consequently, different slopes: k_1 and $k_2 = k_1 - dk$, with k_2 being smaller because the second UFSD has a slightly lower doping. The corresponding collected charges as a function of the bias voltage for these two UFSDs are:

$$\begin{aligned} Q_1(V) &= Q_{PIN} \cdot e^{k_1 \cdot V}, \\ Q_2(V) &= Q_{PIN} \cdot e^{k_2 \cdot V} = Q_{PIN} \cdot e^{(k_1 - dk) \cdot V} \\ &= Q_{PIN} \cdot \frac{e^{k_1 \cdot V}}{e^{dk \cdot V}} \end{aligned} \tag{6.4}$$

The charge difference, ΔQ, between the two is:

$$\begin{aligned}\Delta Q = Q_1(V) - Q_2(V) &= Q_{PIN} \cdot (e^{k_1 \cdot V} - \frac{e^{k_1 \cdot V}}{e^{dk \cdot V}}) \\ &= Q_{PIN} \cdot e^{k_1 \cdot V}(1 - \frac{1}{e^{dk \cdot V}})\end{aligned} \tag{6.5}$$

Hence, the difference in collected charge is proportional to the steepness of the charge-voltage characteristic: the steeper the curve, the larger ΔQ, the less uniform will be the response of the detector.

With a similar calculation, it is possible to get the ΔQ assuming a fixed k but slightly different bias voltages, with the second UFSD being biased at $V_2 = V_1 - dV$:

$$\Delta Q = Q_{PIN} \cdot e^{k \cdot V_1}(1 - \frac{1}{e^{k \cdot dV}}) \tag{6.6}$$

the equation demonstrates that, for what concerns the non-uniform biasing conditions, the steeper the $Q(V)$, the larger ΔQ, the less uniform the performance.

In the evaluation of the gain layer designs, it is also important to consider at what bias voltage a given charge is reached. In the present analysis, the charge Q = 10 fC was used as a bench point and the value V_{10fC} was measured for every sensor.

Lastly, it is important to consider the evolution of the bias point V_{10fC} with irradiation. For this reason, the quantity $\Delta\mathrm{V}_{10fC}(\Phi)$ is introduced. It describes the voltage increase required to provide 10 fC after a certain irradiation fluence, Φ, with respect to the pre-irradiation condition. Such voltage increase is needed to compensate for the gain decrease, due to the acceptor removal mechanism. $\Delta\mathrm{V}_{10fC}(\Phi)$ is defined as:

$$\Delta V_{10fC}(\Phi) = V_{10fC}(\Phi) - V_{10fC}(\Phi = 0) \tag{6.7}$$

where $\mathrm{V}_{10fC}(\Phi)$ is the bias required to provide 10 fC after an irradiation fluence Φ, and $\mathrm{V}_{10fC}(\Phi = 0)$ is the bias required to provide 10 fC before irradiation.

The importance of $\Delta\mathrm{V}_{10fC}(\Phi)$ for ETL can be explained in this way: one can assume that all ETL sensors must be biased to deliver at least 10 fC (this is just a reference, 8 or 5 fC might be chosen as well). The arrays are 21×21 mm^2, placed in an x-y layout on the ETL disk, as shown in appendix A, therefore each array covers a certain radial interval. The fluence in ETL is a function of the radial position: it decreases with the increasing radius. For instance, at the end of the ETL lifetime, the

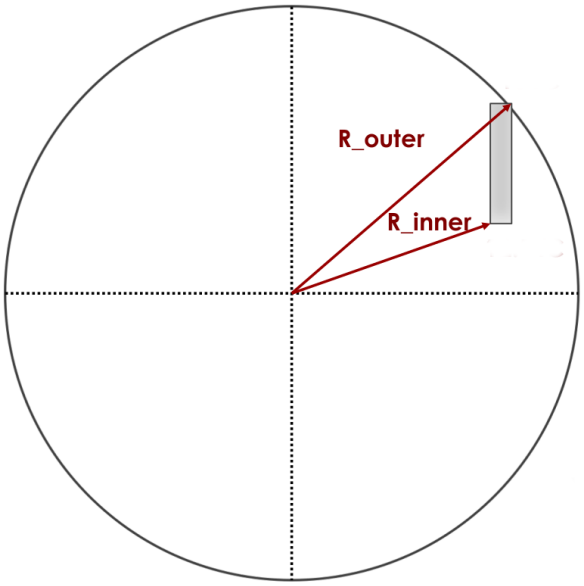


Figure 6.2: Illustration of the non-uniform irradiation of a ETL UFSD array, due to different radial positions. The UFSD is not to scale.

relation between radius and fluence is [2]:

$$\Phi = 1.4 \cdot 10^5 \cdot r^{-1.53} n_{\text{eq}}/cm^2 \tag{6.8}$$

where Φ is measured in 10^{14} $\text{n}_{\text{eq}}/\text{cm}^2$ and r in mm. Hence, all ETL arrays will be subject to non-uniform irradiation: the part of the array at lower radius will be irradiated more than the one at higher radius (figure 6.2). Consequently, not all parts (i.e. all pixels) of the array can be biased at the optimal voltage: the part more irradiated would require higher voltage than the less irradiated one because it is more radiation-damaged. However, the optimal voltage for the most irradiated part cannot be brought to the whole sensor, because the less irradiated part would go into a breakdown since it has a higher gain.

The best solution is to bias the whole sensor at the maximum bias voltage that the less irradiated part can sustain without going into a breakdown and ensuring low-noise, leaving the most irradiated part at a sub-optimal bias voltage. Different parts of the array would thus deliver different performance: how different they will be is determined by $\Delta \text{V}_{10fC}(\Phi)$.

Let $\text{V}_{10fC}(\Phi,\ r)$ be the bias voltage required to deliver 10 fC of charge by the low-irradiation part of a generic ETL UFSD array irradiated at a fluence Φ, whose low-irradiation part is placed, on the ETL disk, at a radial position r. $\text{V}_{10fC}(r)$ can be

written, using equation 6.7, as:

$$V_{10fC}(\Phi, r) = V_{10fC}(\Phi = 0, r) + \Delta V_{10fC}(\Phi) \tag{6.9}$$

The corresponding charge collected in this low-fluence region is:

$$Q_1(\Phi, r) = 10\ fC = Q_{PIN} \cdot e^{k_1 \cdot V_{10fC}(\Phi, r)} = Q_{PIN} \cdot e^{k_1 \cdot (V_{10fC}(\Phi=0, r) + \Delta V_{10fC}(\Phi))} \tag{6.10}$$

The high-irradiation part of array, which is at radial position $r - dr$, is biased at the same voltage, but it delivers lower charge, because it has been exposed to a fluence $\Phi + d\Phi$:

$$Q_2(\Phi + d\Phi, r - dr) = Q_{PIN} \cdot e^{k_2 \cdot (V_{10fC}(\Phi=0, r) + \Delta V_{10fC}(\Phi))} \tag{6.11}$$

with k_2 being slightly lower than k_1, due to the larger radiation damage: $k_2 = k_1 - dk$.

It is now possible to calculate the difference in charge delivered by the low- and high-radiation parts of the array, ΔQ:

$$\begin{aligned} & Q_1(\Phi, r) - Q_2(\Phi + d\Phi, r - dr) = \Delta Q \\ &= Q_{PIN} \cdot e^{k_1 \cdot (V_{10fC}(\Phi=0, r) + \Delta V_{10fC}(\Phi))} - Q_{PIN} \cdot e^{k_2 \cdot (V_{10fC}(\Phi=0, r) + \Delta V_{10fC}(\Phi))} \\ &= Q_{PIN} \cdot (e^{k_1 \cdot (V_{10fC}(\Phi=0, r) + \Delta V_{10fC}(\Phi))} - e^{(k_1 - dk) \cdot (V_{10fC}(\Phi=0, r) + \Delta V_{10fC}(\Phi))}) \\ &= Q_{PIN} \cdot e^{k_1 \cdot (V_{10fC}(\Phi=0, r) + \Delta V_{10fC}(\Phi))} \cdot (1 - e^{-dk \cdot (V_{10fC}(\Phi=0, r) + \Delta V_{10fC}(\Phi))}) \\ &= Q_{PIN} \cdot e^{k_1 \cdot \Delta V_{10fC}(\Phi)} e^{k_1 \cdot V_{10fC}(\Phi=0, r)} \cdot (1 - e^{-dk \cdot (V_{10fC}(\Phi=0, r) + \Delta V_{10fC}(\Phi))}) \end{aligned} \tag{6.12}$$

This equation demonstrates that a higher $\Delta V_{10fC}(\Phi)$ leads to a larger difference in the delivered charge across the ETL UFSD array, and, consequently, to less uniform performance.

6.4 Experimental Results

6.4.1 Collected charge Q

Figure 6.3 shows the collected charge as a function of bias for DUTs from 10 wafers of the FBK UFSD3.2 production and the HPK2 split 4. The color-code of the plot, common to other similar plots in this chapter, identifies different irradiation fluences:

pre-rad sensors are in black, $\Phi = 8 \cdot 10^{14}$ n_{eq}/cm^2 in green, $\Phi = 1.5 \cdot 10^{15}$ n_{eq}/cm^2 in blue, $\Phi = 2.5 \cdot 10^{15}$ n_{eq}/cm^2 in red.

The trend of the collected charge versus irradiation is as expected: as the fluence increases, the bias voltage needs to be raised to compensate for the acceptor removal effect. All tested sensors are able to deliver at least 10 fC up to a fluence of $1.5 \cdot 10^{15}$ n_{eq}/cm^2. At the highest fluence, $2.5 \cdot 10^{15}$ n_{eq}/cm^2, the gain is too low even at high bias values, and none of the sensors deliver 10 fC. FBK UFSD3.2 W19, the most doped design among the FBK UFSD3.2 sensors with deep implant, is the best performer at high fluence, delivering 9 fC at bias V = 600 V.

For each sensor, the point at the highest bias has been taken at the maximum achievable stable biasing condition: at bias voltages above the reported measurements, either the noise is too large, or the sensor goes into a breakdown.

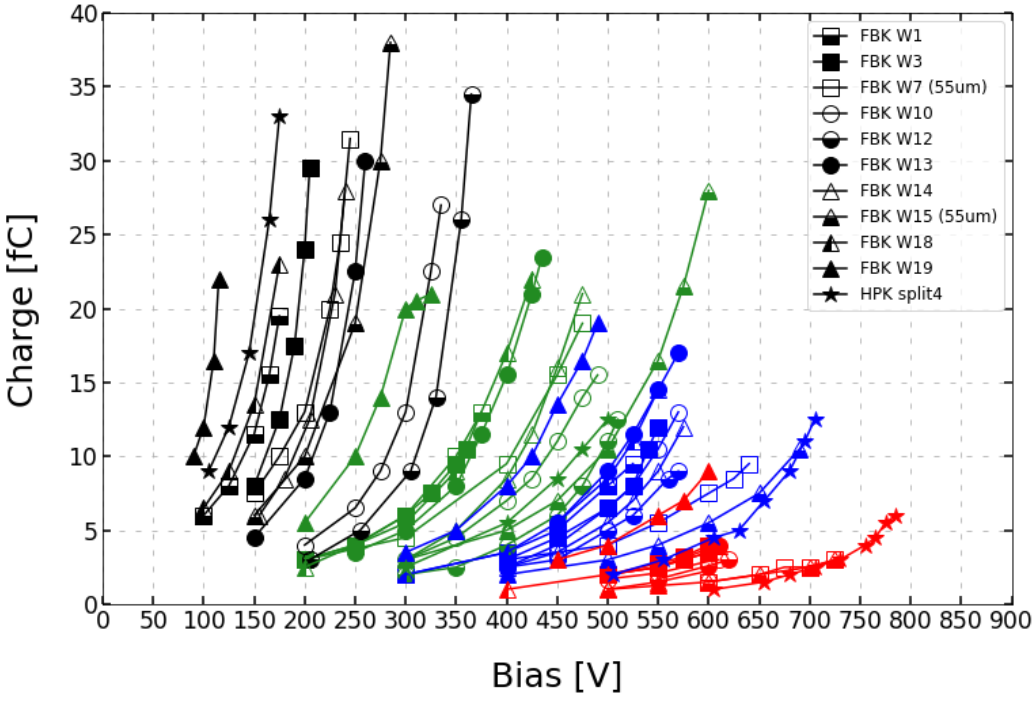


Figure 6.3: Collected charge as a function of bias voltage. Sensors irradiated at different fluences are shown with different colors (black = pre-rad, green = $8 \cdot 10^{14}$ n_{eq}/cm^2, blue = $1.5 \cdot 10^{15}$ n_{eq}/cm^2, red = $2.5 \cdot 10^{15}$ n_{eq}/cm^2). Measurements performed at -25 °C.

6.4.2 Time resolution σ_t

Figure 6.4 shows the time resolution corresponding to the points shown in figure 6.3. All sensors achieve a resolution $\sigma_t < 40$ ps up to $\Phi = 1.5 \cdot 10^{15}$ n_{eq}/cm^2; at $\Phi = 2.5 \cdot$

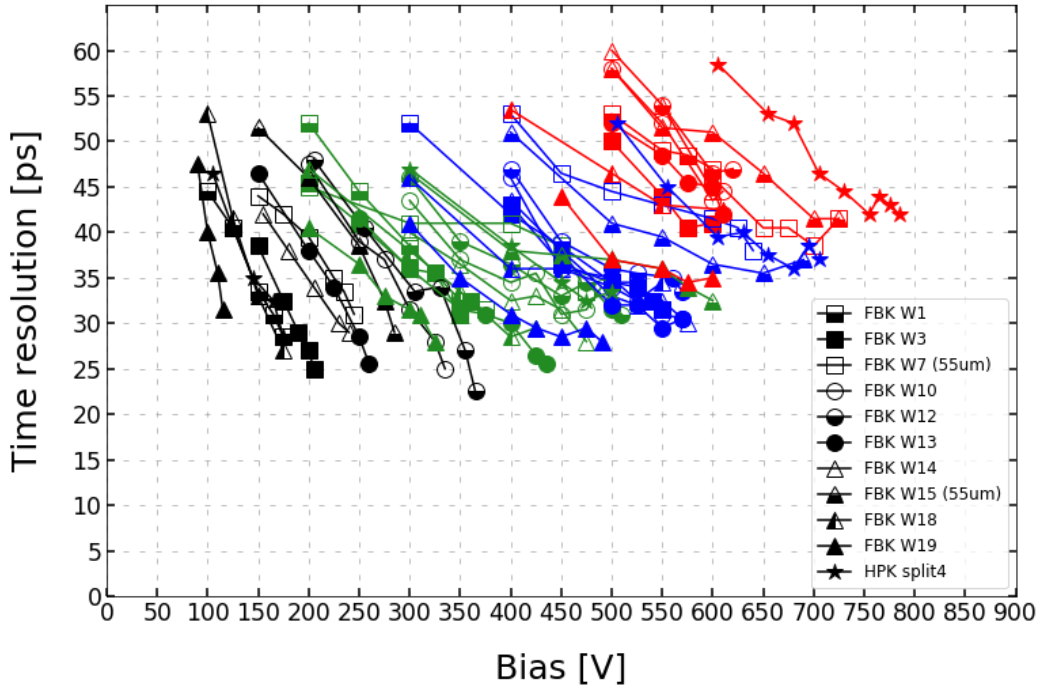


Figure 6.4: Time resolution as a function of bias voltage. Sensors irradiated at different fluences are shown with different colors (black = pre-rad, green = $8 \cdot 10^{14}$ n_{eq}/cm^2, blue = $1.5 \cdot 10^{15}$ n_{eq}/cm^2, red = $2.5 \cdot 10^{15}$ n_{eq}/cm^2). Measurements performed at -25 °C.

10^{15} n_{eq}/cm^2 the resolution degrades up to $\sigma_t \sim 45$ ps. It is worth pointing out that the very good result achieved at the highest fluence is obtained despite the rather small signal charge, $Q < 5\ fC$, thanks to the characteristics of the analog read-out board used in this work.

6.4.3 V_{10fC}, $\Delta V_{10fC}(\Phi)$, and k.

The collected charge as a function of bias ($Q(V)$) for each pre-rad sensor (black points in figure 6.3) has been used to extract the slope parameter k and V_{10fC}. Table 6.3 reports both values; k is the result of the fit of each curve with equation 6.3.

The gain implant doping level controls both parameters: the higher the level, the larger the slope k and the smaller the V_{10fC} value. Sensors with a small V_{10fC} are thus difficult to operate reliably, given their steep gain curve.

Table 6.3: Slopes k [V^{-1}] of the $Q(V)$ curves and V_{10fC}[V] of the tested sensors.

	W1	W3	W7	W10	W12	W13	W14	W15	W18	W19	HPKs4
k	17	16	14	9	9	12	13	12	18	26	20
V_{10fC}	140	165	143	280	310	210	190	164	130	90	110

Figure 6.5 presents the parameter V_{10fC} as a function of the irradiation fluence. The voltages of the two 55 μm-thick sensors (W7 and W15) are scaled by a factor $\frac{45}{55}$, so that they can be compared with the 45 μm-thick ones.

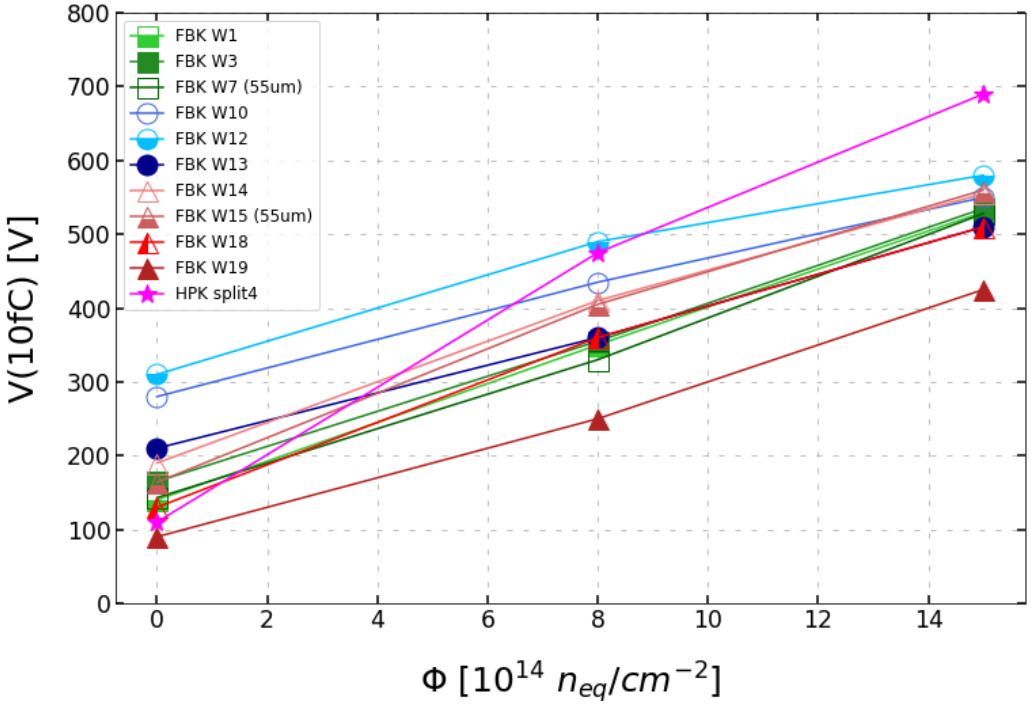


Figure 6.5: The parameter V_{10fC} shown as a function of the irradiation fluence level. Values extracted from measurements performed at -25 °C.

Pre-irradiation sensors show a large spread in V_{10fC}, ranging from $\sim$ 100 V to more than 300 V: this is due to the different doping of the gain implant. The spread is maintained at $\Phi = 8 \cdot 10^{14}$ n_{eq}/cm^2, with sensors requiring between 250 and 500 V to provide 10 fC. At $\Phi = 1.5 \cdot 10^{15}$ n_{eq}/cm^2, the majority of sensors reach the target working point between 500 and 600 V.

The spread in V_{10fC} decreases with increasing fluence since, at high fluences, only a small fraction of the gain implant survives and all the sensors have roughly the same gain at a given voltage, despite their different gain layer designs. Overall, small differences in the gain layer design become gradually less important with irradiation.

Two notable exceptions are shown in figure 6.5: (i) UFSD3.2 W19, due to its high gain layer doping, maintains at high fluences a large fraction of the gain implant, therefore V_{10fC} is reached already at $\sim$ 400 V; (ii) HPK split 4, despite starting with a low value of V_{10fC}, has a low radiation resistance due to the lack of carbon in the gain layer [53], therefore it requires a bias voltage as high as 700 V to provide 10 fC at $\Phi = 1.5 \cdot 10^{15}$ n_{eq}/cm^2.

Table 6.4: The acceptor removal coefficient, c, of the irradiated sensors tested in this study.

	c [10^{-16} cm^2]
W1	1.97 ± 0.31
W3	1.48 ± 0.24
W7	1.91 ± 0.32
W10	2.16 ± 0.28
W12	2.06 ± 0.27
W13	1.63 ± 0.24
W14	2.45 ± 0.30
W15	2.45 ± 0.30
W18	2.05 ± 0.27
W19	1.90 ± 0.26
HPK s4	5.60 ± 0.63

Figure 6.6 shows $\Delta V_{10fC}(\Phi)$ as a function of the acceptor removal coefficient c [cm^2], which describes the exponential removal of the gain implant with fluence, as discussed in chapter 2. Only the 45 μm-thick devices are displayed. The green markers refer to ΔV_{10fC} ($\Phi = 8 \cdot 10^{14}$ n_{eq}/cm^2), the blue ones are for ΔV_{10fC} ($\Phi = 1.5 \cdot 10^{15}$ n_{eq}/cm^2).

Remarkably, the HPK split 4 design is characterized by a c-coefficient almost a factor 3 higher than the average acceptor removal coefficient in FBK sensors, as reported in table 6.4.

6.4.4 Noise

The RMS noise as a function of the bias voltage is presented in figure 6.7. The baseline noise of about 1.2 mV is due to the read-out board. For new sensors (black curves), the noise is always dominated by the read-out board. In irradiated devices, the noise is higher than the baseline noise when the sensor has a large gain and therefore high leakage current. These conditions are met for sensors irradiated at $\Phi = 8 \cdot 10^{14}$ n_{eq}/cm^2 and $\Phi = 1.5 \cdot 10^{15}$ n_{eq}/cm^2; for sensors at $\Phi = 2.5 \cdot 10^{15}$ n_{eq}/cm^2 the noise increase is limited due to the small gain.

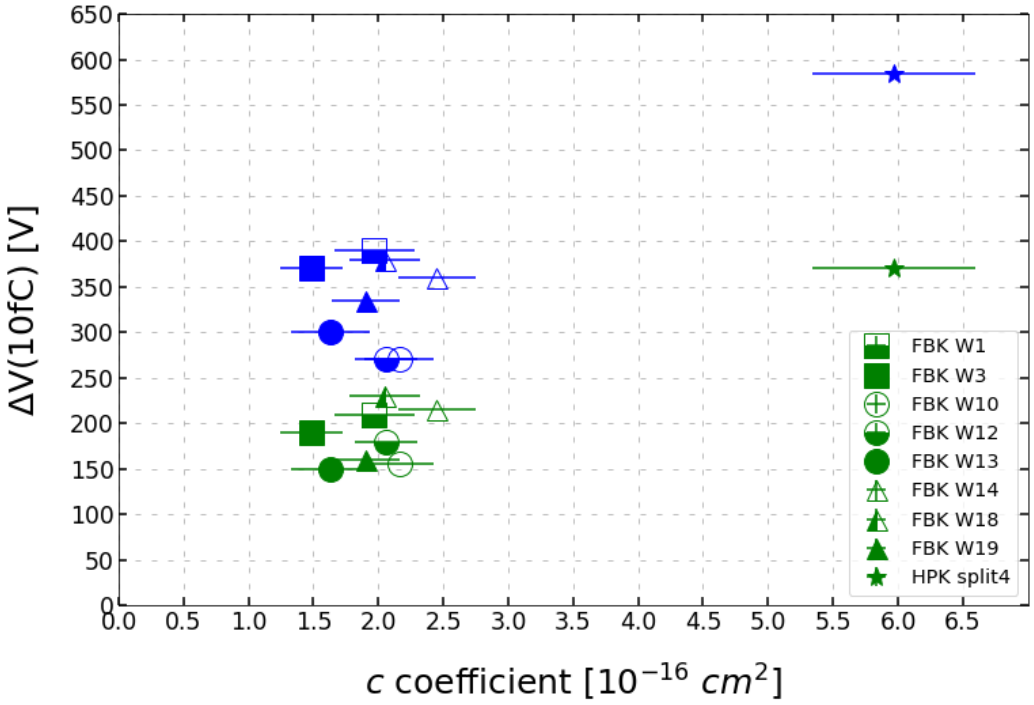


Figure 6.6: $\Delta V_{10fC}(\Phi)$ shown as a function of the acceptor removal coefficient c. Green markers represent ΔV calculated after a fluence of $8 \cdot 10^{14}$ n_{eq}/cm^2, whereas blue ones are calculated at $1.5 \cdot 10^{15}$ n_{eq}/cm^2. This plot reports results only from 45 μm-thick sensors. Values extracted from measurements performed at -25 °C.

Figure 6.8 shows the RMS noise of irradiated sensors, plotted as a function of:

$$\sqrt{Gain \cdot \ln(\Phi/\Phi_0)} \tag{6.13}$$

where Φ_0 is a constant with dimension $[L^{-2}]$.

Interestingly, all tested sensors follow a common trend, with little impact from the specific gain layer design. The underlying physics explanation for such a trend is still unclear and will be investigated by future studies.

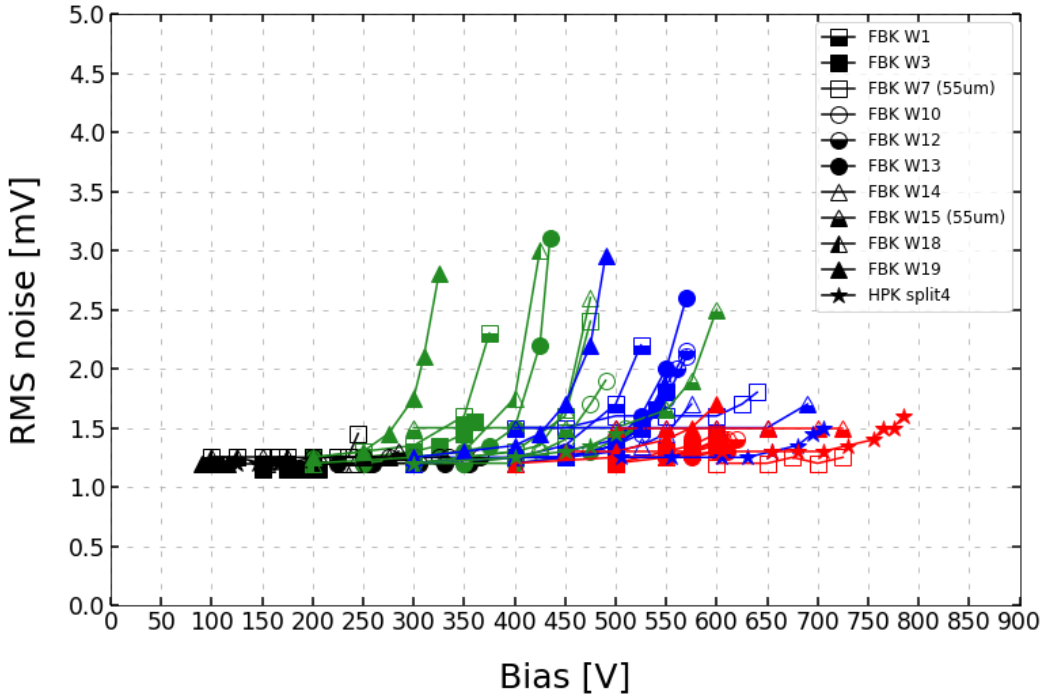


Figure 6.7: RMS noise as a function of bias voltage. Sensors irradiated at different fluences are shown with different colors (black = pre-rad, green = $8 \cdot 10^{14}$ n_{eq}/cm^2, blue = $1.5 \cdot 10^{15}$ n_{eq}/cm^2, red = $2.5 \cdot 10^{15}$ n_{eq}/cm^2). Measurements performed at -25 °C.

A noise level of 2-3 mV in irradiated devices does not spoil the time resolution, which remains below 40 ps. Such noise level is reached when the internal gain of the sensor is larger than 30 (see figure 6.15), therefore the signal-to-noise ratio remains large enough to achieve a good resolution.

6.4.5 Temperature Scan

FBK UFSD3.2 wafers 3, 12, and 14, and the HPK split 4 have been measured at three different temperatures (figure 6.9) to study how their operating points, in terms of the bias voltage, change. The UFSD operating voltage increases with increasing temperature: on average, the bias voltage increase required to compensate for a 1 K temperature increase is 1.4 V (1.1 V) for a UFSD with shallow (deep) gain implant, therefore shallow implants require a higher voltage to compensate for a given temperature increase than deep ones. The uncertainty on these results is ± 0.1 V/K.

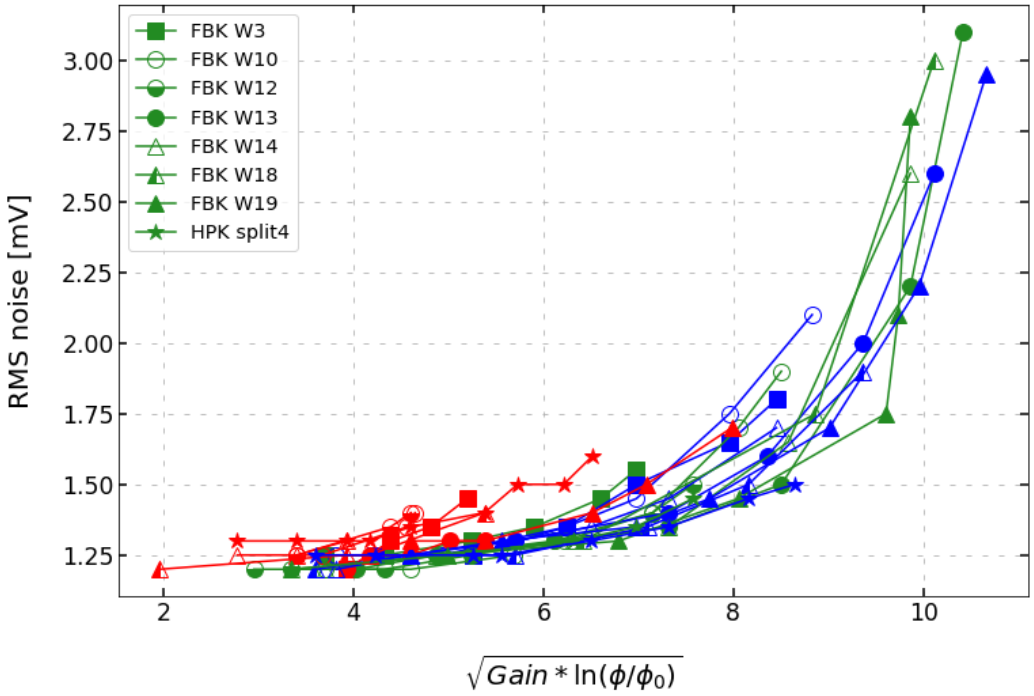


Figure 6.8: RMS noise as a function of $\sqrt{Gain \cdot \ln(\Phi/\Phi_0)}$. A common trend is observed for all irradiated sensors. Sensors irradiated at different fluences are shown with different colors (green = $8 \cdot 10^{14}$ n_{eq}/cm^2, blue = $1.5 \cdot 10^{15}$ n_{eq}/cm^2, red = $2.5 \cdot 10^{15}$ n_{eq}/cm^2). Measurements performed at -25 °C.

6.5 Discussion of results: performance of un-irradiated sensors

6.5.1 Time resolution as a function of gain and electric field

A good time resolution is obtained when both the sensor internal gain and the electric field (or equivalently the bias voltage) at which such gain is reached are sufficiently high [85]. None of these two conditions alone is enough. In fact, the jitter term of the time resolution is minimized when the leading-edge slew rate, dV/dt, is as steep as possible: (i) dV depends upon the signal amplitude (hence the gain); (ii) dt upon the electric field.

For what concerns (ii), in UFSDs the majority of the signal is produced by holes, therefore it is important, from a sensor design point of view, that the holes drift velocity is saturated when the sensor is operated in the gain range 20-30 (the optimal one for UFSDs), so that dt is minimized. The holes drift velocity requires a high electric field

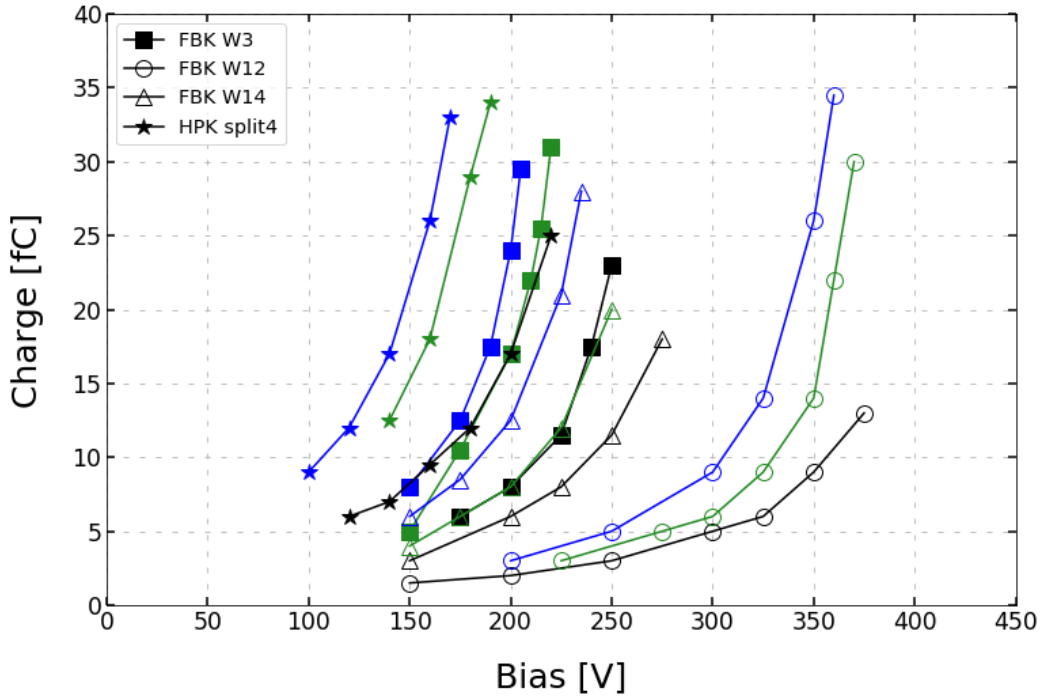


Figure 6.9: Charge-voltage characteristics of UFSDs measured at different temperatures; either FBK and HPK sensors are shown, with both shallow and deep gain implants. A color-code is used to distinguish each temperature: black = room temp. (about 20°C); green = -10 °C; blue = -25 °C.

to approach saturation (50-100 kV/cm [42]): if the gain implant is too doped, the field in the bulk will not be able to reach high enough values before the sensor breakdown.

The pre-irradiated UFSD3.2 W19 is an example of a sensor with a too doped gain layer: large signals are reached at low bias, and the resolution is not as good as less doped sensors.

Figure 6.10 shows the time resolution (blue) of different pre-rad sensors with gain = 20 (each marker represents a different device) as a function of the bias voltage, superposed with the inverse of the computed holes drift velocity. The plot well demonstrates that, for a fixed gain, sensors operated at higher bias voltage reach a better resolution because the holes drift velocity is higher. The blue curve becomes almost flat at about 250-300 V, when the holes drift velocity is approaching saturation, as expected.

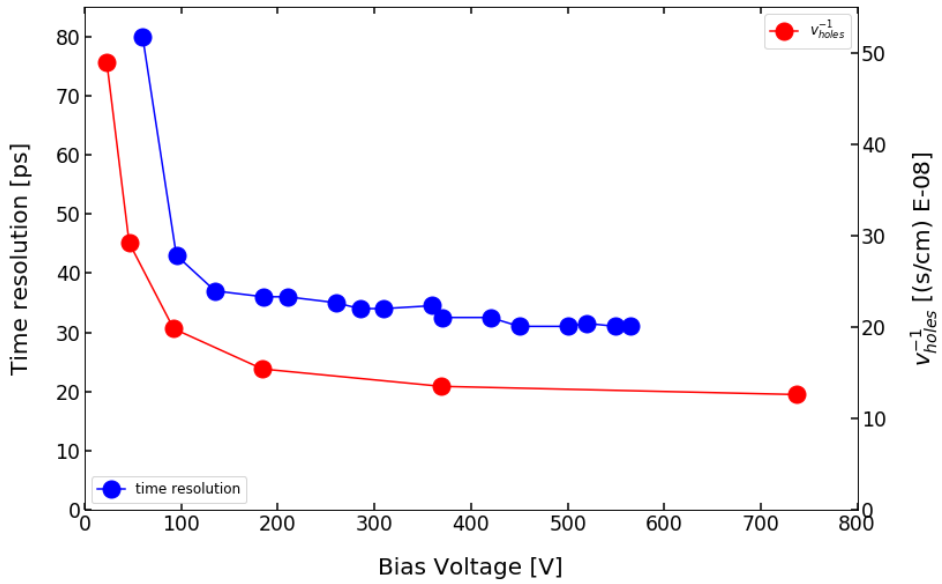


Figure 6.10: Time resolution (blue) as a function of the bias voltage at which the sensor reaches a gain of 20; each blue marker represents a different device. The red curve is the inverse of the holes drift velocity as a function of the bias voltage. Both the experimental results and the computation of the drift velocities are performed at a temperature of -25 °C.

6.5.2 Time resolution as a function of the sensor thickness

Sensors with four different active thicknesses were compared in this study: 35, 45, 55 μm-thick sensors, belonging to the UFSD3.2 production, and 80 μm-thick sensors, manufactured by HPK. All measurements were performed at -25 °C. Figure 6.11 shows the time resolution (dark blue) and the jitter term (light blue) as a function of the collected charge. The jitter term is computed analytically as:

$$\sigma_{jitter} = Noise/(dV/dt) \tag{6.14}$$

Two observations:

- In thicker sensors, a given value of jitter is obtained at higher charges. Indeed, the jitter term can be written as:

$$\sigma_{jitter} = Noise/(dV/dt) \propto Noise/Gain$$

Assuming a constant noise value (the read-out board noise), and a given gain

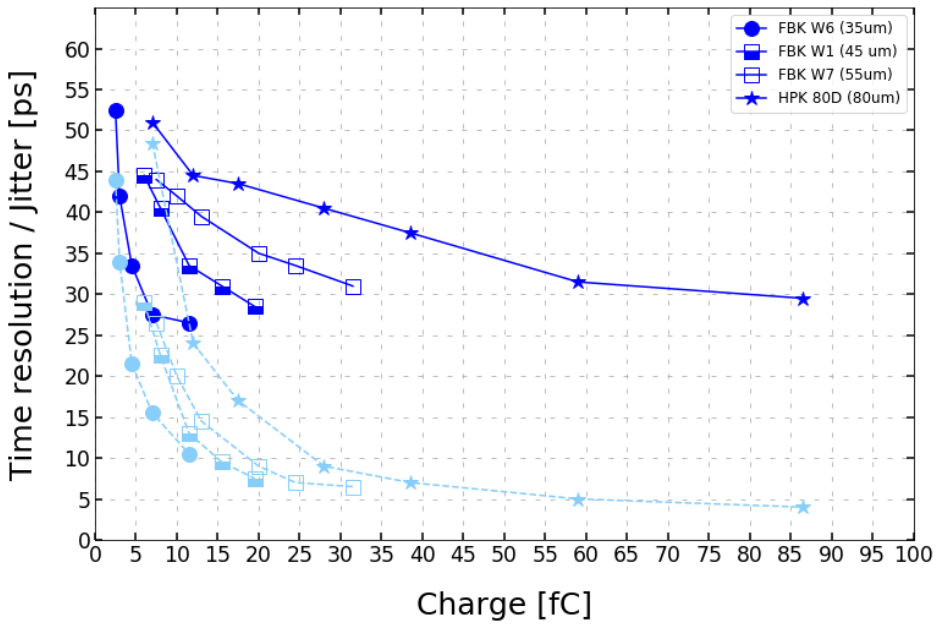


Figure 6.11: Time resolution as a function of the collected charge for sensors with four different active thicknesses. The dark blue curve represents the total resolution, whereas the light blue one is the jitter term only. Measurements performed at -25 °C.

value, the collected charge scales with the sensor active thickness (i.e. with the active volume, since all devices have the same active area), and that explains the spread of the light blue curves in figure 6.11.

- The non-uniform energy deposition generated by an impinging MIP, amplified by the gain, creates variations of the signal shape on an event-to-event basis (the so-called Landau noise [86]). The related uncertainty, σ_{Landau}, which limits the time resolution, can be computed as the difference in quadrature of the total resolution and the jitter [1] :

$$\sigma_{Landau} = \sqrt{\sigma^2_{total} - \sigma^2_{jitter}} \tag{6.15}$$

The Landau term as a function of the active thickness is reported in figure 6.12: it decreases in thinner sensors, as expected [86].

[1] As pointed out in chapter 2, σ_{jitter} and σ_{Landau} are the main contributors to the σ_{total}, whereas the other terms can be neglected.

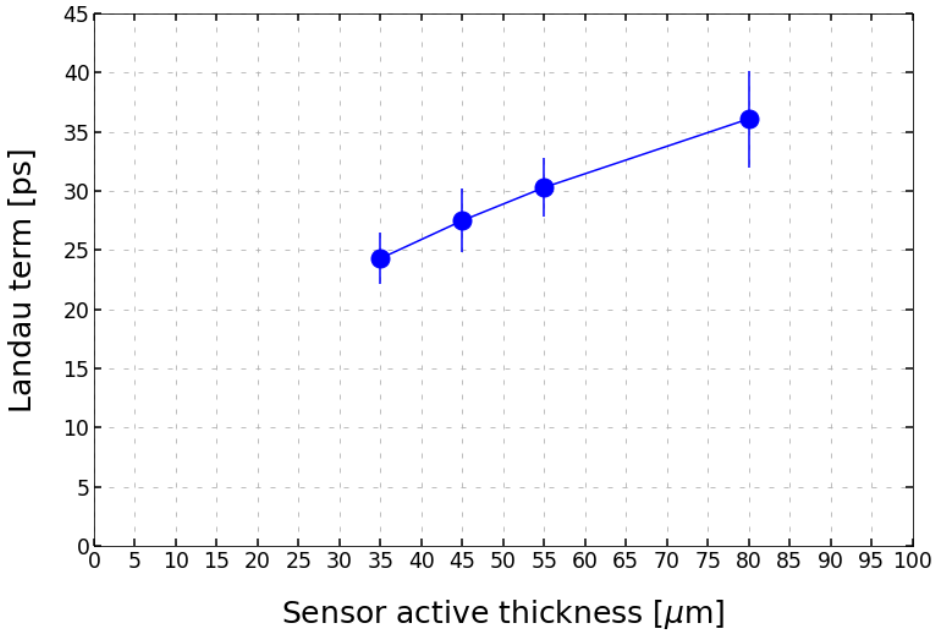


Figure 6.12: Landau term as a function of the sensor active thickness.

6.5.3 Time resolution as a function of the energy deposited in the event.

The energy deposition of MIPs crossing a thin layer of silicon follows a Landau distribution. Events in the lower tail of the distribution tend to have low and uniform energy-per-unit-length deposits, while events in the upper tail have very large localized deposits. Therefore, the position of an event in the Landau distribution is an indicator of the uniformity of the energy deposition: the higher the energy deposited, the higher the non-uniformity.

Figure 6.13 shows the time resolution and the jitter term (measured at -25 °C) in bins of the signal amplitude for a sensor from UFSD3.2 W7 (55 μm-thick). As expected, the jitter contribution decreases with amplitude. The time resolution, instead, first improves as a function of signal amplitude, due to the smaller jitter contribution, then it worsens at high signal amplitudes due to the much larger Landau noise term.

A similar study has also been performed with the UFSD3.2 W6 (35 μm-thick) to investigate the effect of the sensor thickness. Since W6 is thinner, the worsening of the resolution in the upper tail of the distribution should be milder than for UFSD3.2 W7. Figure 6.14 shows the time resolution, normalized to the resolution obtained at the MPV, for UFSD3.2 W6 and W7, in three bins of amplitude (expressed in units of MPV). As expected, the worsening of the Landau term is larger in thicker sensors.

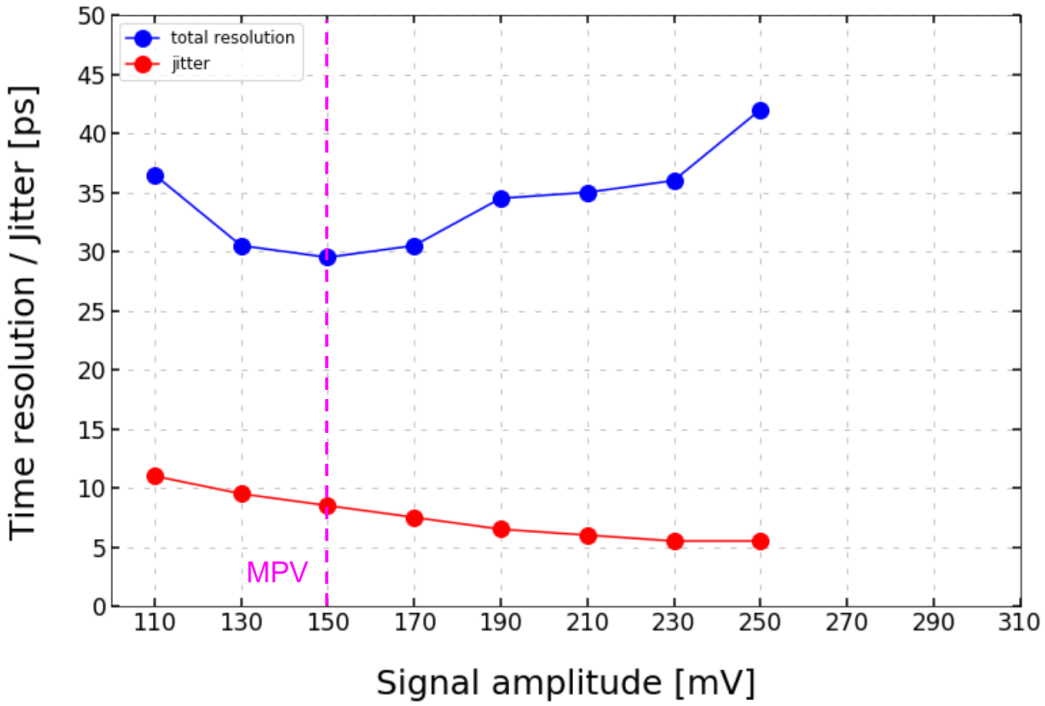


Figure 6.13: Time resolution (blue) and jitter term (red) in bins of signal amplitude. The Landau MPV is $\sim$ 150 mV, corresponding to the best time resolution. Measurements performed at -25 °C.

6.6 Discussion of results: performance of irradiated sensors

Focusing on the time resolution of irradiated devices, it is important to notice that, above a certain bias voltage, once the holes drift velocity is saturated, the time resolution improves with gain in a very similar way for all the wafers, confirming the results of the former section. This common trend can be observed in figure 6.15, representing the time resolution as a function of the gain for sensors irradiated at a fluence of $1.5 \cdot 10^{15}$ n_{eq}/cm^2 and $2.5 \cdot 10^{15}$ n_{eq}/cm^2 (all biased above 300 V).

6.6.1 Time resolution of sensors 45 or 55 μm-thick

The UFSD3.2 production features pairs of wafers with the same gain layer design but different active thicknesses. One pair is formed by W1 (45 μm) and W7 (55 μm) and a second pair by W14 (45 μm) and W15 (55 μm). W1-W7 have a shallow gain implant, while W14-W15 a deep one. Figure 6.16 shows the time resolution of those two pairs, as a function of the electric field in the sensor bulk, calculated as the ratio between V_{bias} and the sensor active thickness.

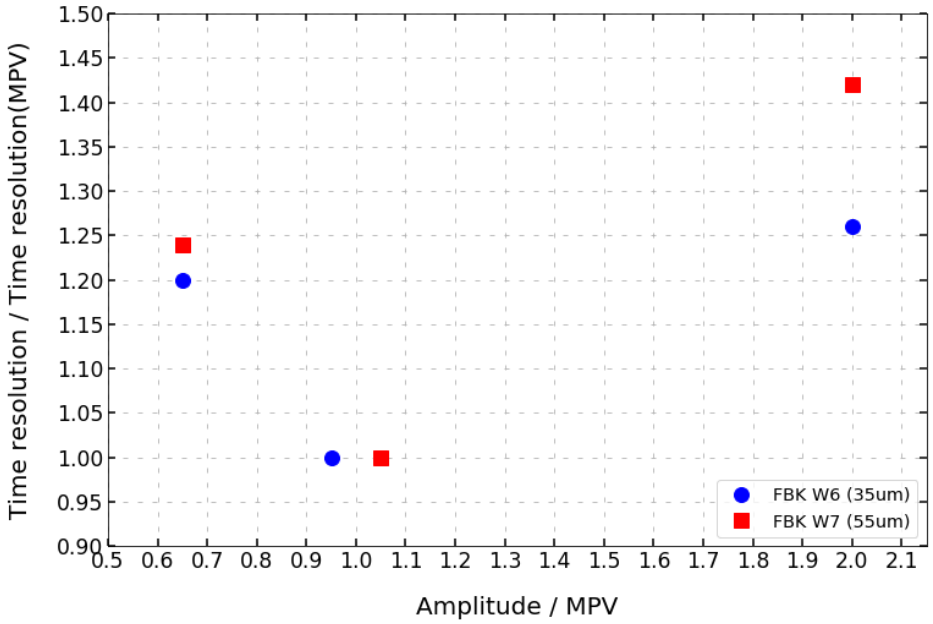


Figure 6.14: Relative time resolution in 3 bins of amplitude (expressed in unit of the MPV) for UFSD3.2 W6 (35μm-thick) and W7 (55μm-thick). The markers in x=1 are slightly shifted only for representation. Measurements performed at -25 °C.

Thin sensors have a better resolution at a given value of electric field (see section 6.5.2), as the Landau noise is larger in thicker devices. At high fluence, $2.5 \cdot 10^{15}$ n_{eq}/cm^2, the sensor gain is close to 1: sensors with thicker bulk have an increased initial charge deposition, leading to a better time resolution.

Hence, thinner sensors, with either shallow or deep gain implants, have better performance than thicker ones, if the sensor gain is high enough. Conversely, thicker sensors are more performing at low or no gain.

6.6.2 Carbon co-implantation

The beneficial effect of the co-implantation of carbon in the boron gain layer can be quantified using the acceptor removal coefficient: the c-coefficient in the HPK split 4 design, which does not feature carbon co-implantation, is almost a factor three larger than the average coefficient of the UFSD3.2 design.

However, a more operational figure can be also used: the voltage increase needed to provide 10 fC after a fluence Φ, $\Delta V_{10fC}(\Phi)$, which highlights the effect of carbon co-implantation in the ETL sensors in a much clearer way than c.

Pre-rad HPK2 split 4 (not carbonated), UFSD3.2 W18, and W19 (both carbon-

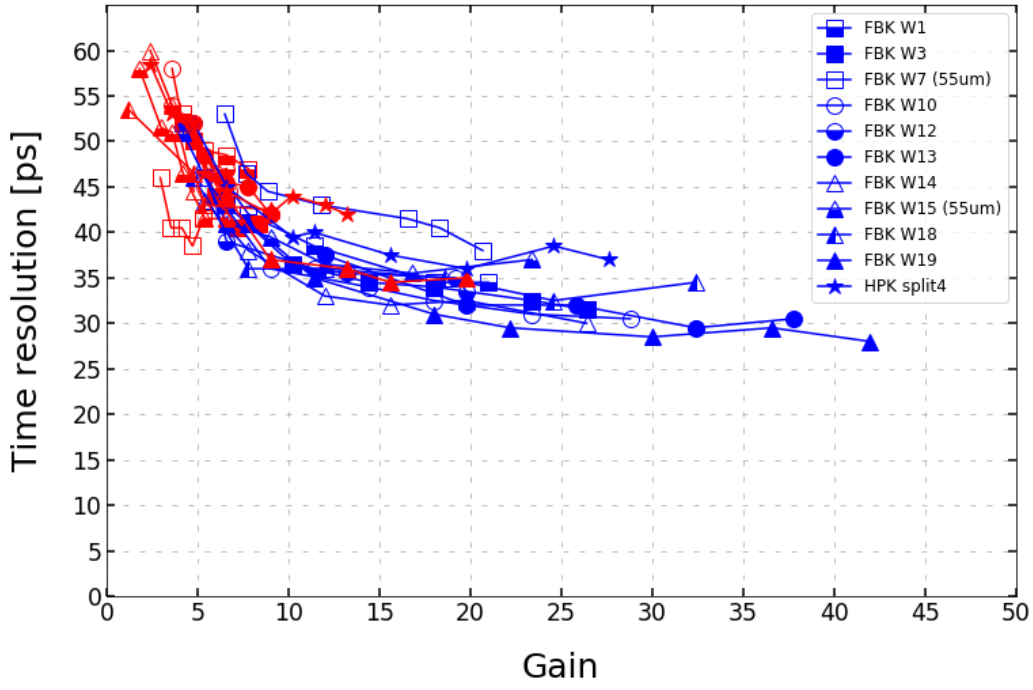


Figure 6.15: Time resolution as a function of Gain. Sensors irradiated at different fluences are shown with different colors (blue = $1.5{\cdot}10^{15}$ n_{eq}/cm^2, red = $2.5{\cdot}10^{15}$ n_{eq}/cm^2). Measurements performed at -25 °C.

ated) sensors behave similarly and have similar gain layer designs. Nevertheless, their behavior differs with irradiation. After a fluence $8 \cdot 10^{14}$ n_{eq}/cm^2, HPK2 split 4 $\Delta V_{10fC}(\Phi = 8 \cdot 10^{14}\ n_{eq}/cm^2) \sim 350$ V, against ~ 200 V for UFSD3.2 W18 and W19 (see figure 6.6). This difference increases further at higher fluences: HPK2 split 4 $\Delta V_{10fC}(\Phi = 1.5{\cdot}10^{15}\ n_{eq}/cm^2) \sim 600$ V while for UFSD3.2 W18, and W19 is ~ 350 V. This difference is rather remarkable, and it has an important impact on the operation of the devices.

One can consider the number of different bias voltages per each line of service hybrids [2] that must be provided to the ETL sensors in order to keep the charge delivered in the 5-10 fC range, where 5 fC is the minimum charge required by the ETROC ASIC to reach the target resolution and 10 fC is a level considered optimal for the operation of the sensors.

When the sensors are new, a single bias voltage is enough for the whole ETL since they all have the same gain, but, with increasing irradiation, the inner sensors, those

[2] As reported in appendix A, the bias voltage is brought to the sensor via service hybrids arranged in lines on the ETL disk.

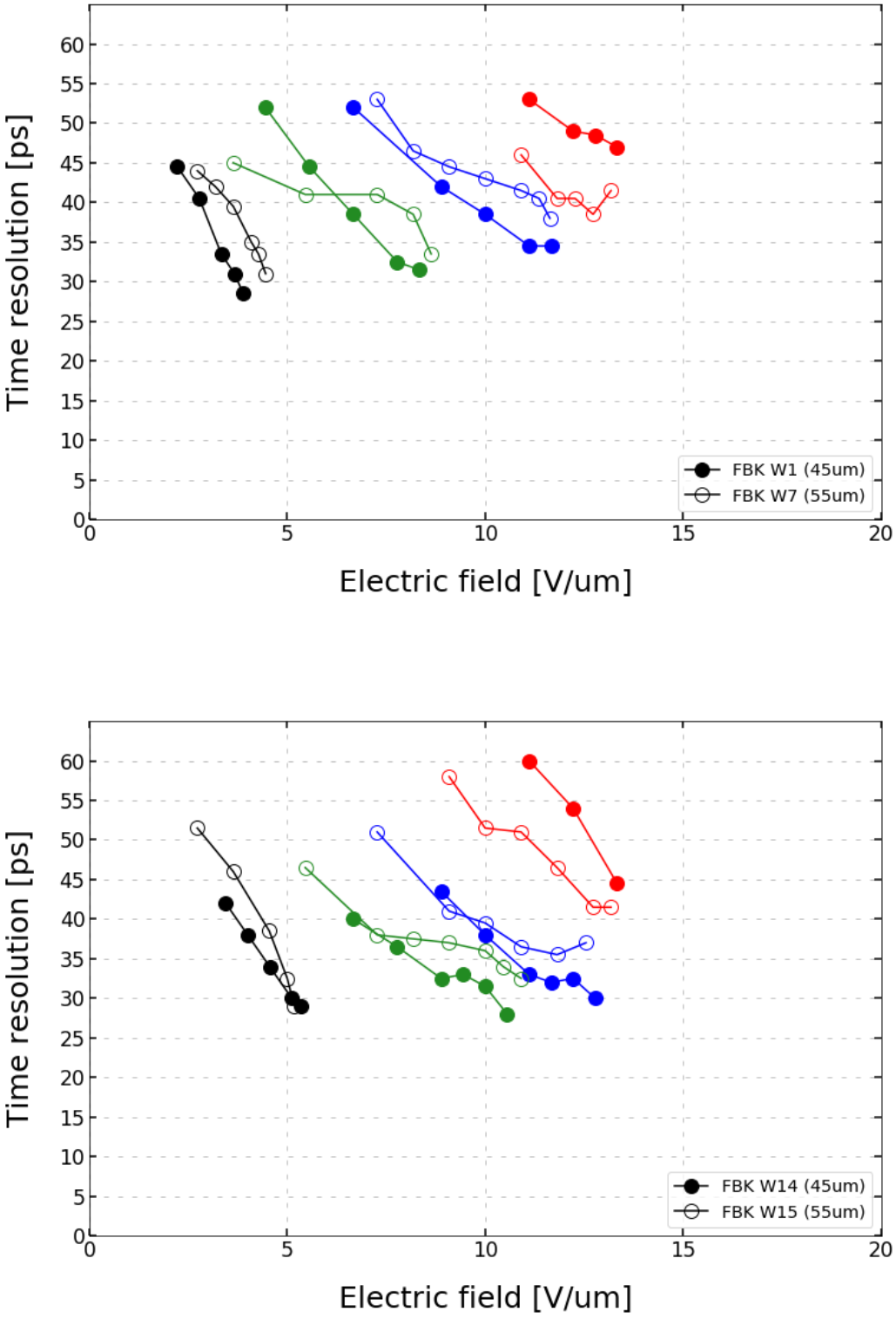


Figure 6.16: Time resolution as a function of the electric field in FBK UFSD3.2 sensors with a shallow (*top*) or deep (*bottom*) gain implant for two different active thickness (45 vs 55μm). Measurements performed at -25 °C.

more exposed to radiations, will need higher bias voltages than the outer ones because of a more severe radiation damage. Because of the radial dependence of the radiation fluence at ETL, the more the irradiation will increase, the more different will be the absorbed dose across ETL, requiring an increasing number of different bias voltages per line to keep all sensors delivering charge in the 5-10 fC range. Such number can be computed using the data shown so far in the chapter.

Equation 6.8 can be used to translate the absorbed radiation fluence at the end of the ETL lifetime (the condition when the largest number of different bias voltages lines will be needed) in the radial position of the sensor that has absorbed such fluence. The interesting radii are those related to the radiation fluences for which experimental data are available: $8 \cdot 10^{14}$ $\mathrm{n_{eq}/cm^2}$, $15 \cdot 10^{14}$ $\mathrm{n_{eq}/cm^2}$, and $25 \cdot 10^{14}$ $\mathrm{n_{eq}/cm^2}$, corresponding to the radial positions 270, 377, and 567 mm, respectively (note that 270 mm is actually smaller than the ETL disk inner radius: 315 mm).

The bias voltages needed to deliver 5 fC or 10 fC can then be plotted as a function of the radial position of the sensor on the ETL disk. Such curves are reported in figure 6.17 for either the FBK W19 and the HPK split 4; the curves are fitted with a power law:

$$V_{5/10fC} = a \cdot R_{in/out}^{-b} \tag{6.16}$$

The inner and outer regions of an UFSD array (see figure 6.2) will be biased at the same voltage in ETL, therefore it is possible to require $V_{5fC} = V_{10fC}$ in the equation, which gives R_{in} as a function of R_{out}. Using this relation, it is possible to fix R_{out}, the 10 fC region, and calculate the corresponding 5 fC position, R_{in}:

- if $R_{out} - R_{in} \geq 21$ mm (the UFSD array length), the sensor whose outer radial position is equal to R_{out} can be biased at a voltage $V_{5fC} = V_{10fC}$ ensuring that all its pixels will provide a charge in the 5-10 fC range.

- if $R_{out} - R_{in} < 21$ mm, it would mean that it is not possible to bias the sensor ensuring that the delivered charge is in the 5-10 fC range for all pixels: fortunately that does not happen in ETL, even at the end of its lifetime and even for the HPK split 4, which features the largest $\Delta \mathrm{V}_{10fC}(\Phi = 1.5 \cdot 10^{15}$ $\mathrm{n_{eq}/cm^2})$ (i.e. it is the most susceptible to non-uniform irradiation).

The number of different voltages to be brought on each line of service hybrids can be computed by considering the R_{out} of each sensor and the corresponding R_{in}. One can consider, for simplicity, a row of arrays arranged along the ETL disk's radial direction and pick the outermost sensor in the disk, with outer radius R_{out}. If

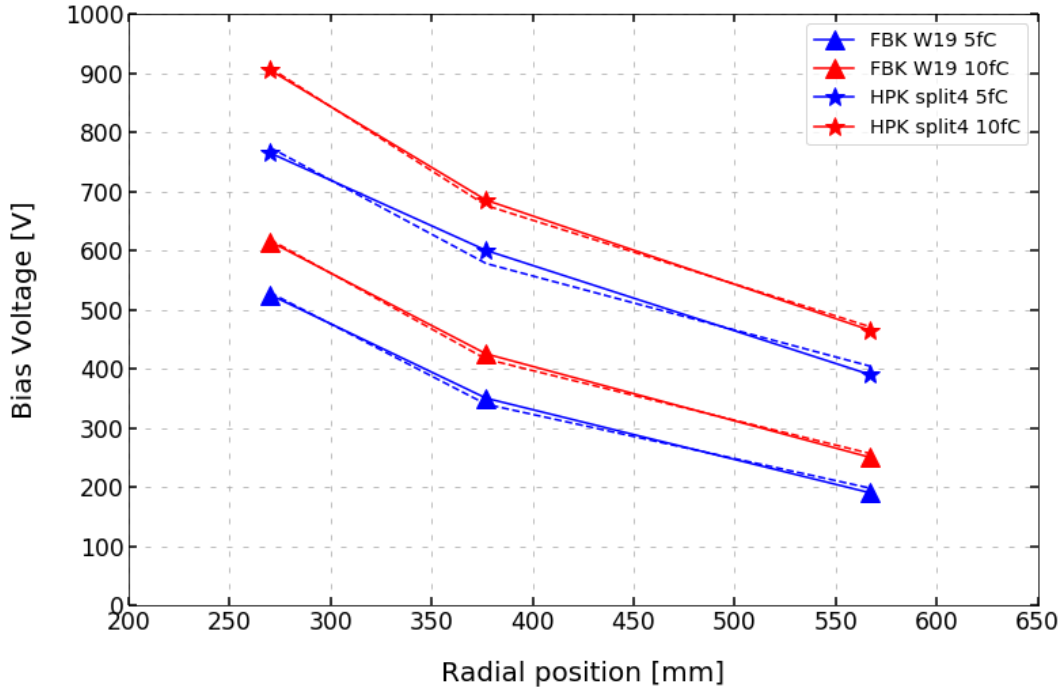


Figure 6.17: Bias voltage required to deliver 5 fC (blue) or 10 fC (red) at the end of the ETL lifetime as a function of the radial position of the sensor on the ETL disk.

$$n \cdot 21\ mm < R_{out} - R_{in} < (n+1) \cdot 21\ mm$$

it would mean that the n outermost sensors in that row can be biased at the same voltage ensuring that the delivered charge is in the 5-10 fC range for all pixels. The same calculation can be then repeated taking the first sensor ruled out by the first iteration and moving inwards up to the innermost sensor. If the row of sensors is not parallel to the radial direction, the calculation is a bit longer, but the outcome is the same. In this way, it is possible to fill in a 2D-map, representing the ETL disk, in which each different bias voltage is represented by a different color: this allows calculating the maximum number of different voltages to be brought to a single line of services. The 2D-maps for either the FBK W19 and the HPK split 4 are shown in figures 6.18 and 6.19, respectively.

The FBK W19 features a smaller $\Delta V_{10fC}(\Phi = 1.5 \cdot 10^{15}\ n_{eq}/cm^2)$ than the HPK split 4, therefore the effect of non-uniform irradiation is milder and 6 different bias voltages are enough, whereas HPK split 4 needs 20 voltages, more than a factor 3 higher, underlining the importance of carbon co-implantation in the gain layer.

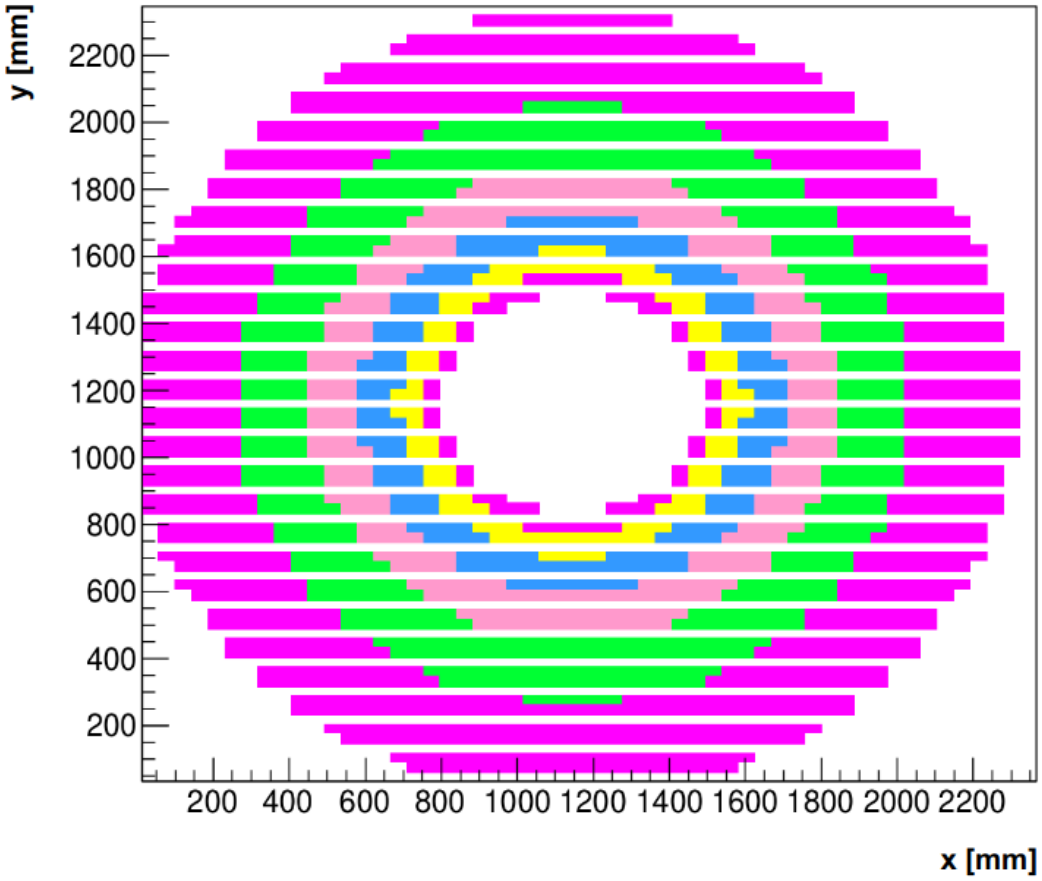


Figure 6.18: 2D-map representing the ETL disk: each color represents a group of FBK W19 sensors (or even a single one) that can be biased at the same voltage ensuring that all pixels deliver a charge in the 5-10 fC range. This is the worst case scenario, since it represents ETL at the end of the operations.

A large number of bias voltages per line leads to higher costs, a more complicated biasing scheme, and, in particular, a large number of different bias voltages to be brought to a single service hybrid (the service module described in appendix A), which is a major constraint, as discussed in detail in appendix B. These results suggest that $\Delta V_{10fC}(\Phi = 1.5 \cdot 10^{15}\ n_{eq}/cm^2)$ is an effective parameter to assess the importance for ETL of carbon co-implantation.

6.6.3 Shallow or deep gain implant

The position of the gain implant has consequences on the radiation hardness and the response to temperature variations of the sensor. To better understand this aspect, it

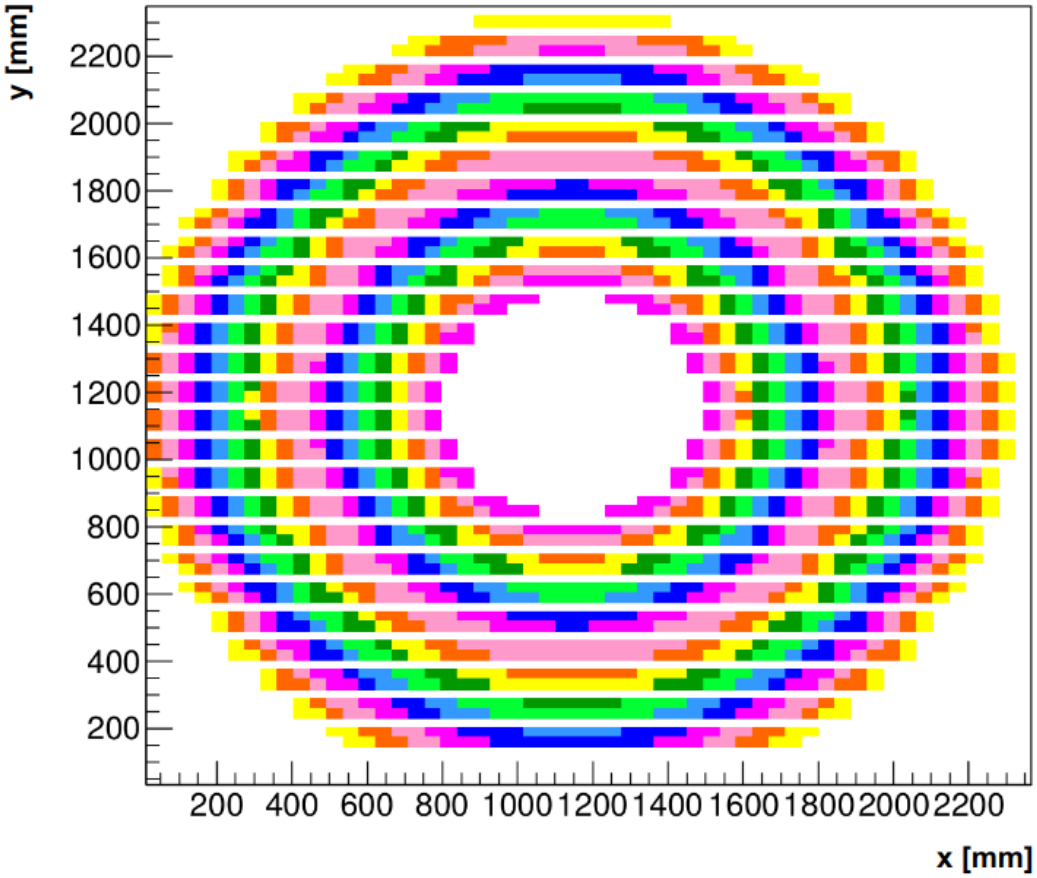


Figure 6.19: 2D-map representing the ETL disk: each color represents a group of HPK split 4 sensors (or even a single one) that can be biased at the same voltage ensuring that all pixels deliver a charge in the 5-10 fC range. This is the worst case scenario, since it represents ETL at the end of the operations.

is useful to introduce the equation of the UFSD gain, which can be written as [57]:

$$G \propto e^{\alpha(E,T,\Phi)\cdot d} = e^{\frac{d}{\lambda(E,T,\Phi)}} \tag{6.17}$$

α is the electrons ionization coefficient. α is a function of electric field, temperature, and radiation fluence, and it is equivalent to λ^{-1}, where λ ([L]) is the mean free path necessary to an electron to acquire enough energy to achieve electron multiplication. d is the length of the high-field region in the sensor bulk, the region where the field is sufficiently high for the electron multiplication to occur.

When the radiation fluence raises, the gain layer doping concentration drops and the mean distance between lattice defects in the sensor is lowered, as pointed out in chapter 2. The decrease in doping concentration causes λ to increase, because of the lower electric field, while the smaller distance between lattice defects calls for a shorter

λ to achieve charge multiplication.

Given that the length of the high-field region is fixed, the only way to maintain a given gain when the radiation dose raises is to decrease λ by increasing the bias voltage: if the electric field in the bulk goes up, the carriers need a shorter path to acquire enough energy to achieve multiplication.

A similar approach can be used to deal with temperature variations [57]: as the temperature raises, the number of phonons in the lattice raises as well, slowing the carriers down in the same way as lattice defects. Hence, when the temperature goes up, the bias voltage needs to be raised as well, to increase the electric field and shorten λ. In this respect, the increase of temperature or radiation fluence have somewhat similar effects on the operation of UFSDs.

A bias voltage increase can recover for the effects of irradiation or temperature increase. The effectiveness of the recovery depends on the value of the electric field and of λ: it is more effective in a deep gain layer, characterized by a lower electric field, since $d\lambda/dE$ is larger. For this reason, as already underlined, deep gain layer designs have a higher gain recovery capability with bias.

The interplay between the intrinsic radiation resistance of the gain layer (acceptor removal coefficient), and the position of the gain implant (gain recovery capability), can be illustrated by analyzing the properties of UFSD3.2 W3 (shallow implant) and W13 (deep implant). Before irradiation, they have roughly the same gain, as shown in figure 6.3, and similar values of the acceptor removal coefficient c (see table 6.4), with W3 being slightly more radiation hard. However, W3 has a higher value of $\Delta V_{10fC}(\Phi = 8 \cdot 10^{14}\ n_{eq}/cm^2)$ and $\Delta V_{10fC}(\Phi = 1.5 \cdot 10^{15}\ n_{eq}/cm^2)$, as shown in figure 6.6, because of the lower gain recovery capability: for this reason, W13 must be preferred for ETL since it provides the same performance of W3 but with lower sensitivity to non-uniform irradiation.

This comparison underlines once more that, from an operational point of view, $\Delta V_{10fC}(\Phi)$ is more effective in assessing the radiation resistance of a UFSD design than the acceptor removal coefficient: the design with the lowest c might not be the most suited for the experiment.

The same result can be obtained when considering temperature variations: a larger bias voltage increase is required to compensate for a given temperature increase for a sensor with shallow implant, as reported in section 6.4.5, due to the limited bias recovery capability with respect to the deep implant.

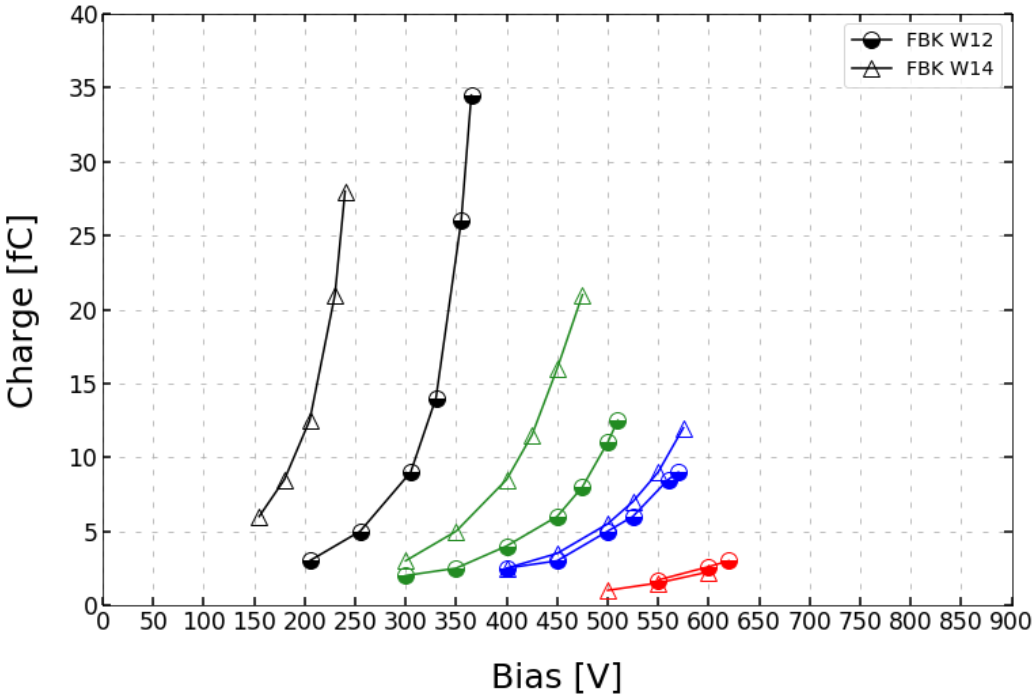


Figure 6.20: UFSD 32 W14 (CBH) requires an higher increase of bias voltage to compensate for the effects of radiations than UFSD3.2 W12 (CBL). Sensors irradiated at different fluences are shown with different colors (black = pre-rad, green = $8 \cdot 10^{14}$ n_{eq}/cm^2, blue = $1.5 \cdot 10^{15}$ n_{eq}/cm^2, red = $2.5 \cdot 10^{15}$ n_{eq}/cm^2). Measurements performed at -25 °C.

6.6.4 Thermal treatments

One of the steps in the sensor production is the activation of the gain implant dopants. In the UFSD3.2 production, this effect has been studied by using two different procedures to activate the carbonated deep gain implants. The two different sequences are called Carbon-Boron Low (CBL, W10,W12 and W13) and Carbon-Boron High (CBH, W14, W15, W18, and W19), where low and high refer to the thermal load. Overall, they have similar performance with one notable difference: CBL is more radiation resistant [42]. Figure 6.20 illustrates this fact: UFSD3.2 W14 (CBH) has a higher pre-rad gain than UFSD3.2 W12, but, with increasing irradiation, the difference gets smaller and, at $\Phi = 1.5 \cdot 10^{15}$ n_{eq}/cm^2, they operate at the same voltage. In other words, UFSD3.2 W12 has a smaller $\Delta V_{10fC}(\Phi)$.

6.6.5 Deep CBL with different dopings

UFSD3.2 W10, W12, and W13 are all deep CBL with an increasing dose of the gain implant doping. They have similar performances and Wafer 10 and 12 have a sligthly better $\Delta V_{10fC}(\Phi = 1.5 \cdot 10^{15}\ n_{eq}/cm^2)$; however, Wafer 13 is operated at a significantly lower bias voltage before irradiation, driving the decision to choose it as the best design among the CBL sensors.

6.7 Summary of the measurement campaign

The most important outcomes of the measurement campaign described in this chapter are:

- The best time resolution is obtained when the electric field in the bulk is large enough: high gain at low bias voltage leads to poor performances.
- $Q(V)$ characteristics not steep (small k values) are to be preferred since they lead to: (i) easier-to-handle sensors, (ii) a better σ_t given the higher V_{10fC}, and (iii) a decrease of the effects of either doping or bias non-uniformity.
- Low values of V_{10fC} before irradiation imply a very steep $Q(V)$ and a worse σ_t before irradiation.
- All sensors were able to reach a time resolution of 30 to 40 ps up to a fluence $\Phi = 1.5 \cdot 10^{15}\ n_{eq}/cm^2$, with at least 10 fC of charge delivered. At $\Phi = 2.5 \cdot 10^{15}\ n_{eq}/cm^2$ the signal charge decreases below 10 fC and the resolution worsens.
- A low value of the parameter $\Delta V_{10fC}(\Phi)$ ensures a better uniformity when the sensor is exposed to non-uniform irradiation. Carbon co-implantation decreases $\Delta V_{10fC}(\Phi)$ by about 50% at every fluence.
- Radiation resistance depends upon two parameters: (i) the acceptor removal coefficient c and (ii) the bias recovery capability. The interplay of these two aspects determines that deep gain implants are more radiation-resistant than shallow gain implants.

This study also analyzes the performance of UFSDs with different active thicknesses (35, 45, 55, and 80 μm-thick):

- Thin devices have a lower Landau noise term.

- At high radiation levels, $\Phi = 2.5 \cdot 10^{15}$ n_{eq}/cm^2 , when the gain value is very small, thicker sensors have better time resolution given their higher initial signal.

- The time resolution has been studied in bins of the Landau distribution. The best results are achieved around the most-probable-value. For smaller amplitudes, the jitter term dominates, while, for higher values, the Landau noise is larger, worsening the time resolution.

6.8 The final gain layer design for the ETL

The measurement campaign underlined important aspects: the optimal UFSD design, considering the ETL needs, should have a deep and carbonated gain layer to ensure good radiation hardness and uniformity in conditions of non-uniform irradiation. The $Q(V)$ should not be too steep to ensure good performance also before irradiation, but, at the same time, V_{10fC} should not be too high for low power consumption.

The FBK UFSD3.2 W13 is the optimal trade-off between all the requests: it features a deep carbonated gain layer with CBL thermal load, its $Q(V)$ is one of the least steep, and V_{10fC} is the lowest among the deep CBL designs.

It is important to stress out here that only the 12% of the ETL surface will receive a radiation fluence higher than $1 \cdot 10^{15}$ n_{eq}/cm^2 (figure 6.21), and, even more importantly, only 2% of the overall dataset will be collected above this point [37]. Therefore, while it is key to ensure excellent performance also in the most irradiated region of ETL throughout its lifetime, it is also important to remember that only a small fraction of the detector will live in really challenging conditions in terms of radiation damage.

The W13 design, compared to more standard ones, is expensive and particularly challenging, since it represents the present state-of-the-art of the UFSD designs. The W7 design, which is a more standard one, costs less and its production would be simpler than with W13 since it does not feature the deep implant.

The performance of W7 below $1 \cdot 10^{15}$ n_{eq}/cm^2 are excellent: it achieves the target resolution delivering more than 10 fC, its $Q(V)$ features a slope parameter (k) similar to W13, and it has one of the lowest V_{10fC} before irradiation. Moreover, the W7 design is a reference: it was already implemented in the UFSD3 production and FBK has a lot of experience with it, which is an additional aspect to account for when producing thousands of sensors.

Hence, in order to limit the costs and facilitate the large-scale production of the sensors, a possible plan would be to instrument the low-fluence area of ETL (fluence $< 1 \cdot 10^{15}$ n_{eq}/cm^2 at the end of operations) with UFSDs having the gain layer

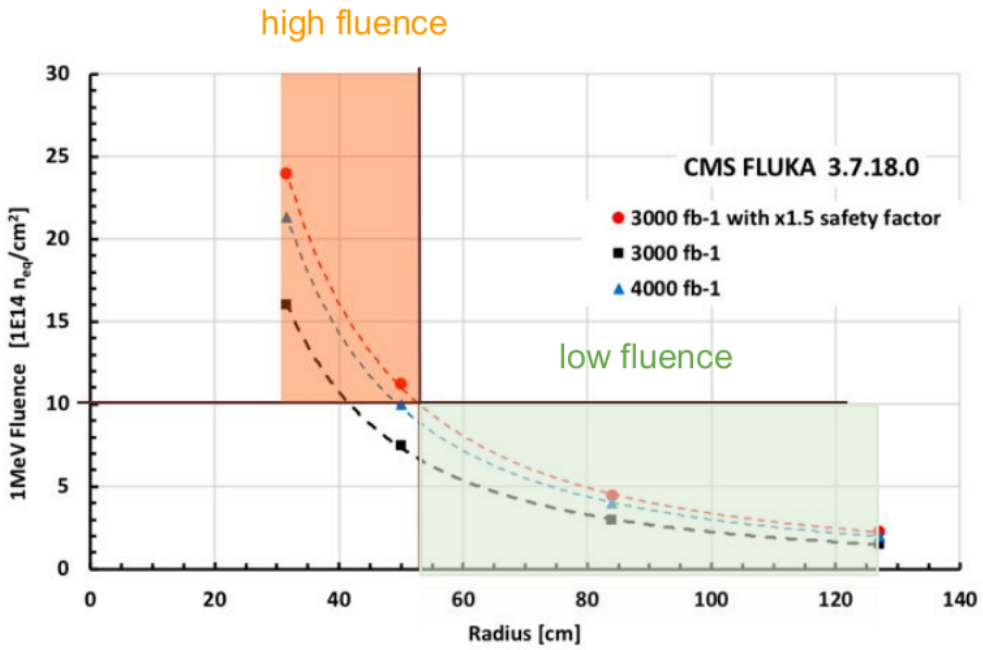


Figure 6.21: Expected received fluence at the end of the HL-LHC lifetime as a function of the radial position on the ETL disk. Projections made in three different scenarios: 3000 fb^{-1} with nominal (black) luminosity or with a 1.5x safety factor (red), and 4000 fb^{-1} (blue).

design of W7, and the high-fluence area with the W13 design. In this way, almost 90% of the UFSDs instrumenting ETL will have a well-known design whose large-scale production should be relatively simple, and only about 10% of the devices will be characterized by the innovative deep gain layer which is expected to be more challenging to produce. This plan, which is further discussed in appendix B, would not affect in any way the target performance of ETL.

Chapter 7

Higgs pair production via vector-boson fusion in the $b\bar{b}b\bar{b}$ final state

This chapter analyzes the production of a pair of Higgs bosons decaying in two bottom, anti-bottom ($b\bar{b}$) quark pairs, in the vector-boson fusion (VBF) production mode, one of the most interesting physics processes to be observed at the HL-LHC, given its unique sensitivity to BSM physics and to the di-vector-boson–di-Higgs-boson ($VVHH$) coupling. The analysis is based on simulated Phase-2 data and focuses mainly on the improvements in signal efficiency due to the addition of the MTD timing in the event reconstruction.

The final state of the process, with two $b\bar{b}$ pairs, is characterized by a rather large branching ratio, $\sim 34\%$, and by the presence of four jets containing the bottom hadrons into which the b- or $\bar{b}$-quark fragment [88], the so-called b-jets, which are key in this analysis.

The $b\bar{b}b\bar{b}$ final state suffers from a large QCD multijet background, but the VBF production mode, despite having a cross section more than an order of magnitude lower than the main production mode, the gluon-gluon fusion, is particularly helpful since it carries a remarkably clear signature: two forward jets, with large invariant mass, separated by a large rapidity gap, and reduced hadronic activity in the central rapidity region. Hence, the combination of the $b\bar{b}b\bar{b}$ final state and the VBF production mode leads to either a high branching fraction and a rather clear signature, making this process of particular interest for CMS Phase-2.

After a first introduction on the simulated data and the definition of the main quantities used in the analysis, the chapter will present the improvements in the primary

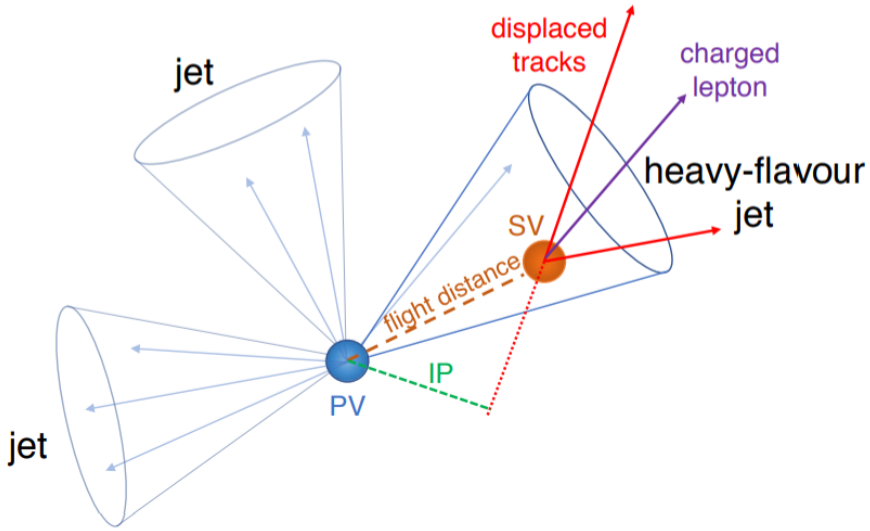


Figure 7.1: Illustration of a heavy-flavour jet with a secondary vertex (SV) from the decay of a B-hadron resulting in tracks that are displaced with respect to the primary interaction vertex (PV), and hence with a large impact parameter (IP) value. Taken from [87].

vertex (PV) tagging and in the pileup rejection due to the introduction of timing in the event reconstruction process, then the results on the signal efficiency with and without timing will be compared. Finally, the result of the Higgs pair search, which heavily relies on previous analyses by CMS [14] and ATLAS [12], will be shown, exploring also the BSM scenario in which the di-vector-boson–di-Higgs-boson coupling differs from the standard model (SM) expectation.

7.1 Simulated events

Monte Carlo (MC) simulated Phase-2 samples, for a total of about 100k events, at 14 TeV center-of-mass energy for the non-resonant VBF HH signal process are generated at LO accuracy in perturbative QCD using MADGRAPH5_aMC@NLO 2.6.5, taking into account the three diagrams shown in figure 1.3 and their interference. A standard model sample is generated with the coupling strength modifiers c_λ, c_V, c_{2V} equal to 1, along with a non-SM one having $c_{2V} = 2$. The latter sample is produced since the VBF production mode has unique sensitivity to BSM physics, as discussed in chapter 1, and a deviation of c_{2V} from its SM value is expected to induce a remarkable increase of the production cross section [13], therefore it is worth studying such scenario with simulated data.

An ultimate HL-LHC scenario with 200 PU is considered for either the SM and

non-SM samples. The response of the CMS detector is simulated using the GEANT 4 package [89]; the updated ETL geometry scenario described in chapter A has been used in the simulation.

7.2 Event reconstruction

Events are reconstructed using the particle-flow (PF) algorithm [22] that aims to reconstruct and identify each individual particle in an event, with an optimized combination of information from the various elements of the CMS detector. The reconstructed particles, which will be simply referred to as "candidates" in the following, are classified into electrons, muons, photons, and charged and neutral hadrons. The candidates are then associated with collision vertices reconstructed from particle tracks.

The vertex reconstruction is based on a two-steps process: first of all, a vertex is identified within a given cluster of tracks, the so-called *vertex finding*; secondly, the tracks used to identify the vertex in the first step are fitted to obtain the best estimate of the vertex position, along with a set of quality parameters, such as the χ^2 or the covariance matrix, which allow accepting or discarding a vertex hypothesis. This second step is named *vertex fitting*. Detailed descriptions of the PV reconstruction in CMS can be found in [90] and [91].

The vertex with the highest $|\vec{p_T}|^2$ is usually taken as primary pp interaction vertex (PV), with p_T being the transverse momentum: $p_T = \sqrt{p_x^2 + p_y^2}$ (see chapter 1 for the reference system used in CMS). The reason to take the highest $|\vec{p_T}|^2$ vertex is that it probes the energy scale from a few GeV to a TeV, which is the interesting one to discover new physics. A different choice, based on the use of the MTD timing, can be made to tag the PV: it will be described in detail in the following.

The z-position of the PV at the MC generator level is available, therefore it is sufficient to compare it with the z-positions of the reconstructed vertices to determine which is the real PV among the $\sim$ 200 ones in the event. This is useful to determine the goodness of the PV tagging, for instance.

Jets, which are key physics objects in this analysis, are reconstructed from three-dimensional topological clusters of energy deposits in the calorimeter with the anti-k_t algorithm [92] implemented in the FastJet package [93] with a radius parameter R = 0.4, where R = $\sqrt{\phi^2 + \eta^2}$ (both ϕ and η angles are defined in chapter 1).

The final state of the process, with two $b\bar{b}$ pairs, is characterized by the presence of four jets containing the bottom hadrons into which the b- or $\bar{b}$-quark fragment [88], the so-called b-jets.

The bottom hadrons have a long lifetime, therefore, when they decay, they are already displaced by the beamline by a few mm to about 1 cm, depending on the momentum, much more than the displacement of light-quark hadrons: consequently, events at the LHC characterized by the presence of b-jets would show tracks with a large impact parameter, namely the distance between the PV and a track at their point of closest approach.

Exploiting the displaced tracks, a secondary vertex (SV), relative to the decay of the B-hadron, can be reconstructed (see figure 7.1). The SVs are extremely important for the improvements achieved in this analysis, as it will be reported in the next sections. A detailed description of SV reconstruction with the CMS detector can be found in [87].

The b-jets are identified using the *deepCSV* tagger, which is an improved version of the $CSVv2$ tagger described in [87]; the *deepCSV* combines the information of displaced tracks with the information on SVs associated with the jet using a deep neural network.

Both vertices and jets are arranged, in each event, in collections which are ordered by decreasing p_T.

The MTD timing is back-propagated to the beamline as described in chapter A, and a 40 ps smearing is added to reproduce the effect of the detector resolution.

The vertex timing is given by the tracks' times used in the vertex fit, whereas the jet timing is the p_T-weighted average of its candidates' times.

7.3 Events selection and main figures of merit

All the events considered in this analysis feature a PV with valid timing information, meaning that it is possible to determine the PV time using the time information of the tracks used to reconstruct the vertex. Usually, the PV, characterized by high-p_T tracks originating from it, is well reconstructed and carries a valid timing information, while PU vertices with very low-p_T tracks have more often a not valid timing information.

Similarly, only the candidates in the reconstructed jets with valid timing information, corresponding to the 52% of the total (neutrals do not have timing information because they are not detected by the MTD), are used in the analysis.

In the analysis dataset, both the reconstructed jets and the generated jets [1], which represent the simulation truth, are present, so that a comparison between them can be performed. In order to achieve that, the reconstructed and generated jets are matched

[1] In order to produce the generated jet, a Monte Carlo simulation generates the particles coming from the quark hadronization, which are then grouped using the anti-k_t algorithm.

in the ϕ-η plane by looking at their angular distance $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$: if $\Delta R \leq 0.2$, the jets are matched, otherwise they are unmatched.

Moreover, in order to further reject PU jets, which usually feature lower momenta than signal jets, a $p_T > 20$ GeV cut is always applied on both the reconstructed and generated jets; sometimes such cut can be raised, as it will be shown in the following. Only jets in the pseudorapidity range $0 \leq |\eta| \leq 5$ are considered.

The p_T of the reconstructed jets is calibrated to account for the effects of the detector response by adding a correction factor which comes from the comparison with the p_T of the generated jets. In order to compute the calibration factor, both jet distributions are divided in p_T and η bins: for each bin, the ratio between the p_T of the reconstructed and the matched generated jets is computed, resulting in a distribution of p_T^{RECO}/p_T^{GEN}, the so-called Jet Energy Scale (JES). The correction factor for a given bin is the mean value of the JES distribution in that bin.

A jet is considered b-tagged if the relative output score of the *deepCSV* algorithm is larger than 0.4184, which is a working point labeled as "medium", meaning that the b-tagging efficiency is 68% and the probability of mistagging a light-flavor jet as a heavy-flavor one is $\sim 1\%$ [94].

Table 7.1: Definitions of efficiency and purity.

	Definition	Selections
Efficiency	$\frac{\#\ matched\ gen\ jets}{\#\ gen\ jets}$	ΔR(reco, gen) < 0.2 Generated jet $p_T > 30$ GeV Matched reconstructed jet $p_T > 20$ GeV
Purity	$\frac{\#\ matched\ reco\ jets}{\#\ reco\ jets}$	ΔR(reco, gen) < 0.2 Reconstructed jet $p_T > 30$ GeV Matched generated jet $p_T > 20$ GeV

The two main figures of merit in this work are efficiency and purity. The efficiency is given by the ratio of generated jets matched to a reconstructed jet over the total number of generated jets; for a generated jet to be considered, its p_T must be larger than 30 GeV, while the corresponding matched reconstructed jet must have $p_T > 20$ GeV. The latter selection sets a lower threshold to account for any not perfect reconstruction of the jet, which can cause the reconstructed jet to have a lower p_T than the corresponding generated one.

The purity, on the contrary, is given by the ratio of reconstructed jets matched to a generated jet over the total number of reconstructed jets; for a reconstructed jet to be considered, its p_T must be larger than 30 GeV, while the corresponding matched generated jet must have $p_T > 20$ GeV. In this case, the p_T threshold is lower for the

generated jet: again this choice is made to account for any contamination of the signal jet by pileup tracks, which can result in a reconstructed jet with a p_T larger than the corresponding generated one.

Both efficiency and purity will be presented in the following in five different η bins, to better map the detector response in different regions; the five intervals are: [0, 1.3); [1.3, 2); [2, 2.5); [2.5, 3); [3, 5].

A high efficiency proves the accuracy of the jet reconstruction since a mis-reconstructed jet can hardly be matched to the corresponding generated jet; whereas, a high purity proves that the jets are well reconstructed, but also that the number of PU jets is low: indeed, a PU jet would not be matched to any generated jets, so the higher the number of PU jets, the larger the denominator in the definition of the purity, the lower the purity.

Other quantities used in the analysis are the Jet-Energy-Scale (JES), already introduced, and the Jet-Energy-Resolution (JER), which is the width of the p_T^{RECO}/p_T^{GEN} distribution. The JER determines the effectiveness of the jet correction factor since the larger the JER, the broader the spread of p_T^{RECO}/p_T^{GEN}, the less effective would be the correction.

7.4 Improved primary vertex tagging: the *pointing method*

The simulated samples used in this work have, on average, 200 pp interactions overlapped with the PV, which is the interesting interaction to discover new physics; it is therefore of primary importance, although challenging, to correctly identify the PV in this dense environment.

In CMS, the PV is usually taken as the vertex with the highest sum of the $|\vec{p_T}|^2$ of the tracks associated with it. This is an effective choice with the pileup level of the present LHC, but the tagging performance is degraded when moving to the HL-LHC scenario. The PV tagging efficiency is defined as the number of events in which the PV is correctly tagged, based on the matching of its position on the beamline with that of the Monte Carlo generated PV, over the total number of events: in this analysis, using the standard procedure for the PV tagging, the efficiency is $88.4 \pm 0.1\%$, meaning that in more than the 10% of the events a PU vertex is wrongly taken as leading, degrading the overall results.

The double Higgs production in the $4b$ final state features a very peculiar topology, with four b-jets and the four corresponding secondary vertices. The *pointing method*

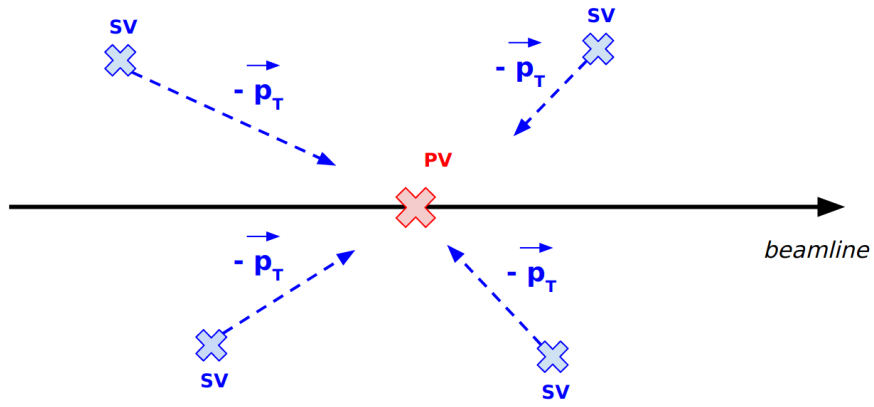


Figure 7.2: Sketch of the beamline with PV and the four SVs relative to the b-jets, whose $\vec{p_T}$ is used to tag the primary vertex.

described in this section aims at exploiting this topology to improve the PV tagging efficiency by using the opposite of the $\vec{p_T}$ vector of the b-jets as a pointer towards the PV, as illustrated in figure 7.2. This cannot be done with light-flavor jets since their tracks are not displaced enough from the beamline to be resolved from the PV, therefore the *pointing method* is suited for the topology considered for this analysis, but it might not be possible to apply it to other processes.

In the first part of the algorithm, a jet is chosen if its p_T is larger than 40 GeV and $|\eta|< 2$: this is justified since the signal b-jets are produced in the central rapidity region and are expected to have a large momentum; moreover, the chosen jets are required to have a b-tagging score higher than 0.4184, which is the typical working point for b-jets.

The angular distance (ΔR) between the selected jet and each SV present in the event is computed: the SV with the smallest ΔR is chosen and, if ΔR(jet, SV) < 0.4, the momentum of the jet and the position of the corresponding SV are recorded for the following step. This procedure is repeated for all reconstructed jets in the event, until four jets matched to a SV, or less, depending on the event, are found.

The procedure stops, at most, after the fourth matched jet, even if a larger number satisfies the selection, to avoid picking a PU jet: since the signal b-jets are four and expected to have high momentum, and given that jets are ordered by decreasing p_T in each event, it is reasonable to take the first four matched jets in the event collection, although it does not grant to choose signal jets only.

The $-\vec{p_T}$ of a jet chosen in the first step of the algorithm and the corresponding SV define a line, named r. The z-position corresponding to the distance of closest approach of r to the PV is searched: the plane α orthogonal to r and passing through the PV is computed, then the intersection between r and α, H, is found. H have coordinates

(x_H, y_H, z_H). The distance H-PV is the distance of closest approach, and z_H is taken as the best estimate of the vertex position on the beamline for what concerns the jet represented by r.

This calculation is repeated for each b-jet chosen in the first step, and the p_T-weighted average of the different z_H positions found ($z_{pointing}$) is taken as the reconstructed z-position of the PV. Similarly, the reconstructed PV time ($t_{pointing}$) is calculated as average of the b-jet times [2].

This second step is repeated for all vertices in the event: the vertex chosen as PV is that with either the smallest distance from $z_{pointing}$ and the smallest temporal distance from the reconstructed vertex time, $t_{pointing}$. Moreover, the PV candidate needs to have a normalized χ^2, given by the fit of the tracks originating from it, in the 0.53 - 0.93 interval.

If none of the vertices in the event satisfy these requirements, the one with the highest $|\vec{p_T}|^2$ is selected as PV, but this situation is negligible, occurring in less than 1% of the events.

The tagging efficiency of the *pointing method* can be assessed by computing the fraction of events in which the PV found with the *pointing method* is the real PV, based on the MC truth (the z-position of the PV at the generator level is available). The efficiency is $89.9 \pm 0.1\%$, $\sim 1.5\%$ larger than simply taking as leading the vertex with the highest $|\vec{p_T}|^2$. The effect of introducing the *pointing method* on the reconstruction efficiency is discussed in the following.

The *pointing method* is not only based on the peculiar topology of the $4b$ final state, but it also importantly relies on the MTD timing: indeed, using only the spatial selection, the tagging efficiency drops to only 77%, while the additional temporal requirement raises the efficiency to about 85%. The importance of the temporal selection can be explained by the spatial overlap of the vertices, already discussed in chapter 1: because of the 200 PU, several vertices are overlapped in space, but not in time. The spatial selection of the *pointing method* can mistag a pileup vertex as leading because it overlaps with the real PV: the temporal cut can correct that by adding the temporal dimension, allowing to resolve the two vertices in the $t - z$ plane.

However, the 85% is still lower than the tagging efficiency given by the standard reconstruction method: the ultimate improvement that determines that the *pointing method* is more efficient than the standard one is the selection on the normalized χ^2 of the vertices, obtained from the *vertex fitting* process.

Figure 7.3 illustrates the distributions of the normalized χ^2 in the events where the

[2] That is not the standard way of determining the PV time, it is just the solution chosen for the *pointing method*.

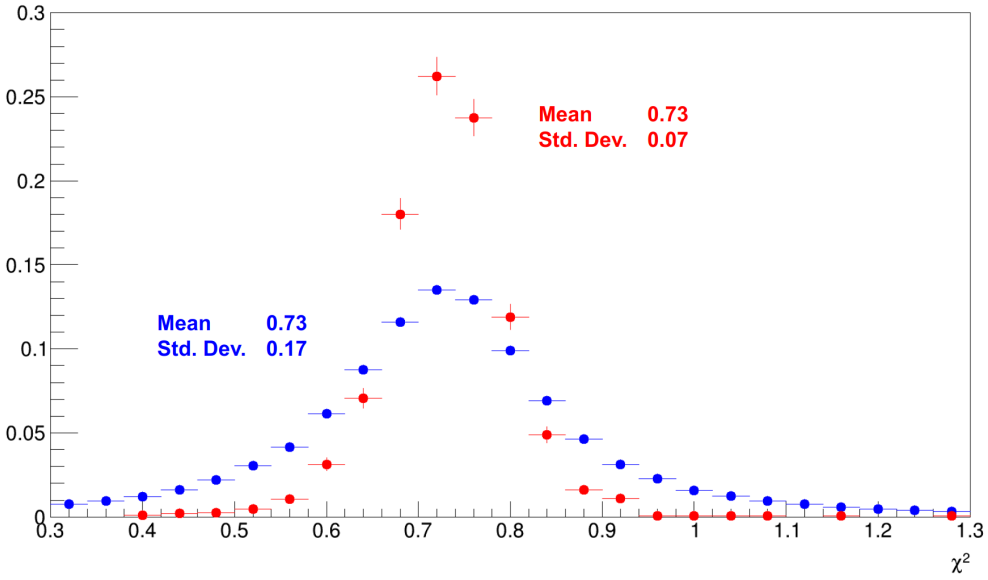


Figure 7.3: Distributions of the normalized χ^2 in the events where the PV is correctly tagged: the red distribution represents the χ^2 of the PV, the blue one the PU vertices. The PV features a narrower distribution, demonstrating that the selection on the normalized χ^2 can improve the PV tagging efficiency. Both distributions are normalized to one.

pointing method (with the χ^2 selection not implemented yet) correctly tags the PV, based on the MC truth: the red distribution represents the χ^2 of the PV, the blue one the PU vertices. The primary vertices have a narrower distribution with a σ a factor x2.5 smaller than in the distribution of the PU vertices, therefore the additional selection on the χ^2 in the 3σ interval (0.53 - 0.93) has a remarkable rejection power. Indeed it improves the tagging efficiency, raising it to $\sim$ 90% and establishing the *pointing method* as an effective PV tagging tool for the $HH \to b\bar{b}b\bar{b}$ channel.

Finally, it is worth reminding that, as stated in chapter 1, the number of spurious reconstructed secondary vertices is reduced by 30% with the MTD, with a non-negligible impact on the performance of the *pointing method*, which will be quantified in future studies.

7.5 Improved PUPPI algorithm

Pileup interactions have been always present at the LHC, requiring solutions to disentangle them from the hard-scatter interaction. Pileup mitigation is already needed in the PU condition of the present LHC, with $\sim$ 40 overlapping interactions, and will thus be even more important at the HL-LHC, where the PU will be at the 140-200

level.

CMS uses two techniques to remove PU: the Charged-Hadron Subtraction (CHS) [95], which has been the standard technique for years, and the newer Pileup Per Particle Identification (PUPPI) [23], which is the focus of the remaining part of this section.

Differently from other techniques, the PUPPI algorithm aims at removing PU tracks from jets, rather than just correcting the measured jet quantities: in doing so, it moves towards a local reconstruction, where each particle composing the jet is individually assessed and given a weight w. Such individual assessment of each particle candidate is performed prior to the jet clustering: a technique that follows the so-called "jets without jets" paradigm [96]. The weight w is then used to rescale the four-momentum of each particle; ideally, particles coming from pileup would get a weight of zero, and particles coming from the hard scatter would get a weight of one. This leads to a pileup-corrected event, where one can then proceed with jet finding without the need for further PU correction.

All charged particles used in the PV fit are assigned a weight of 1, while all charged particles associated with a PU vertex have their weight set to 0. Charged particles not assigned to any vertices are given a weight of 1 if their distance of closest approach to the PV along the beamline is smaller than 3 mm.

The PUPPI algorithm then attempts to assign a weight to neutral particles by defining a discriminating variable α, which describes the local shape of the jet. The underlying idea of the α parameter is that a particle from a shower is expected to be close to other particles from the same shower, when observed in the laboratory frame, because of their high Lorenzt boost (the hard-scatter interaction is highly energetic), whereas PU particles, coming from different interactions and without a large boost, are usually distributed more homogeneously.

The local shape α_i for a given particle i is defined as:

$$\alpha_i = log \sum_{j \neq i, \Delta R_{ij} < R_0} (\frac{p_{T,j}}{\Delta R_{ij}}))^2 \tag{7.1}$$

where $p_{T,j}$ is the transverse momentum of particle j, and it is used as discriminating variable, given that the p_T spectrum of PU particles falls much faster than the PV one. The sum is performed on particles j whose angular distance from the particle i, ΔR_{ij}, is smaller than R_0, i.e. by all particles in the cone of radius R_0 defined by the particle i. For $|\eta_i| < 4$, where tracking is available, the sum considers only the particles j which are charged and originate from the PV, while, at larger rapidity, it runs over all kind of reconstructed particles. R_0 is usually set to 0.4 [22]. In the case where no particles are found within the cone defined by particle i, α_i is equal to 0. This parameter is

designed such that the i particle gets a large value of α if it is close to PV particles, or, where tracking is not present, to highly energetic particles, whereas α would be small if i is a PU particle [23].

In order to translate the local shape α into a weight that can be used to rescale the neutral particles four-momenta, the distribution of the α_i values relative to the charged particles assigned to a PU vertex (which were found in the first step of the algorithm) is computed, and a median and RMS are extracted. Each neutral particle is then compared to the distribution, and a signed χ^2 is calculated:

$$\chi_i^2 = \Theta(\alpha_i - \overline{\alpha}_{PU}) \times \frac{(\alpha_i - \overline{\alpha}_{PU})^2}{\alpha_{PU}^{RMS}} \tag{7.2}$$

where Θ is the Heaviside step function; $\overline{\alpha}_{PU}$ the median of the pileup distribution relative to charged PU particles, and α_{PU}^{RMS} the corresponding RMS. The definition is such that, if a neutral particle has $\alpha_i < \overline{\alpha}_{PU}$, it is assigned a χ_i^2 of zero.

The p_T rescaling weight is eventually given by the cumulative distribution function of the χ_i^2 distribution, F_{χ^2}:

$$w_i = F_{\chi^2}(signed\ \chi_i^2) \tag{7.3}$$

the definition is such that all particles with $\alpha_i < \overline{\alpha}_{PU}$, having a χ_i^2 of zero, would receive a weight of zero as well; whereas all neutral particles with large departures of their α_i from $\overline{\alpha}_{PU}$ would get a weight close to one.

The PUPPI algorithm was originally developed for a PU lower than 200, therefore it is important to test it at such level, using the simulated VBF HH sample previously described.

The purpose of the test is to check whether particle candidates in signal and PU jets are compatible with originating from a primary or PU vertex, respectively. Such a check can be performed using the candidates' timing information, which is not used in the standard PUPPI algorithm and, therefore, constitutes an independent observable to test the signal-pileup hypothesis.

All candidates' times are computed with respect to the same reference point, which is the simulated time t_0 at which the hard-scatter interaction occurs, and then are smeared according to the 180 ps beamspot spread. Hence, the distribution of the candidates' times in signal or PU jets are equivalent: they are gaussian, centered in zero, and with a σ given by the convolution of the beamspot spread and the simulated MTD timing resolution (which is simulated by adding another smearing to the candidate's time), for a total σ of about 210 ps, see figure 7.4. The candidates' times alone are

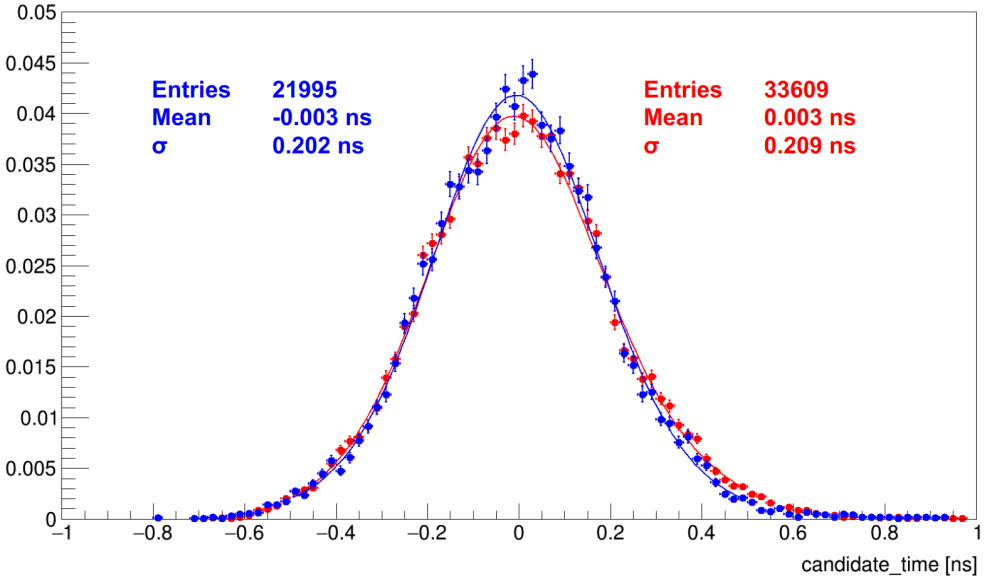


Figure 7.4: *Candidate time* distribution concerning particle candidates in signal (red) and pileup (blue) jets. Each distribution is normalized to one.

therefore not effective in distinguishing signal from PU candidates.

A more effective variable is the temporal difference between the PV time and the times of the particle candidates (available in the analysis framework) composing a jet. Candidates coming from a signal jet are correlated (in an ideal world they would have the exact same time) with the PV time, therefore the resulting *candidate time* - *PV time* (Δt hereafter) distribution should be a gaussian centered in zero, with a σ representing the detector resolution, about 40 ps. Conversely, candidates from a PU jet are not correlated with the PV time, therefore the Δt distribution should be a gaussian centered in zero, but with a σ given by the squared sum of the PV temporal spread, 180 ps, and the candidate time uncertainty, 210 ps, given by the beamspot spread convoluted with the detector resolution, for a total of about 275 ps.

The Δt can be a very effective discriminating variable: the distribution of candidates in signal jets should have a σ of about 40 ps, much narrower than the distribution of candidates in PU jets, $\sim$ 275 ps. Unfortunately, that would be the case if a perfect pileup rejection algorithm could be developed: actually, PUPPI (and any other rejection algorithm) is not 100% perfect, meaning that some PU candidates can be included in a signal jet, as well as a signal candidate can be rejected and fall into a PU jet. Hence, the distributions of the candidates' times, relative to either signal or PU jets, have a fraction of signal-like candidates, arranged on a gaussian with a σ of 40 ps , and a second part, representing PU candidates, with a σ of 275 ps; in other words,

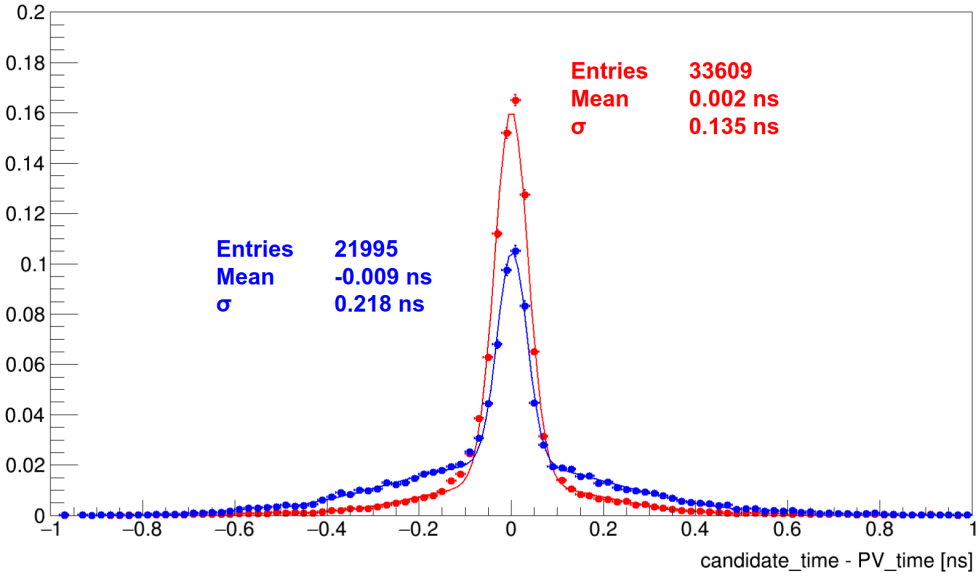


Figure 7.5: *Candidate time - PV time* distributions (normalized to one) concerning particle candidates in signal (red) and pileup (blue) jets. Both distributions are characterized by the sum of two gaussians, with different widths, relative to signal-like and pileup-like candidates.

the resulting distribution is always given by the sum of two different distributions.

This is illustrated in figure 7.5, where the red distribution represents the Δt in signal jets, while the blue one is relative to PU jets; both distributions are normalized to one. The figure shows rather clearly that both curves are given by the sum of two different distributions, but the overall σ is $\sim$ 135 ps for the signal jets (red), rising to more than 215 ps for PU jets (blue), proving that timing can be an important discriminating variable to deal with high levels of pileup.

The following step is to decide how to use the Δt variable to improve the PU rejection. The simplest idea would be to reconstruct the jets with PUPPI and then rule out those candidates in the signal jets for which $|\Delta t| > 3\sigma_{Signal}$, where σ_{Signal} is the 40 ps detector resolution. However, it must be noted that PUPPI, when applied to a 200 PU scenario, does not fail in rejecting pileup-like candidates, but, conversely, it is very performing in doing that, even too performing, as it ends up removing also signal-like candidates from the signal jets: this is demonstrated by the large signal-like peak of the blue distribution in 7.5. Hence, adding a timing cut after PUPPI to further reject candidates is useless since the algorithm already did the job.

In this context, the purpose of the MTD timing within PUPPI is to avoid discarding signal-like candidates rather than enhance the PU rejection.

An alternative, and more effective, way of exploiting the MTD timing is to introduce

it directly into PUPPI, instead of using it afterward. Among the many ways this can be done, the solution that gave the best performance, based on the present CMS reconstruction code for Phase-2 [3], is a modified version of PUPPI, which does not consider the tracks used in the PV fit, but only looks at their spatial (dz) and temporal (dt) distances from the PV to decide whether they can be assigned to it.

Several selections in dz and dt have been tested, with the optimal one featuring dz= 3 cm, dz/σ_{dz}= 10 and dt/σ_{dt}= 3, where σ_{dz} (σ_{dt}) is given by the squared sum of the track and vertex spatial (temporal) uncertainties [4]. The selection implies that all tracks within 3 cm or $10\sigma_{dz}$ can be associated to the primary vertex, provided that they are within $3\sigma_{dt}$ to it. As a comparison, the default PUPPI version only requires a spatial selection on dz, but an order of magnitude tighter: dz= 0.3 cm.

The comparison proves that, with the present CMS reconstruction code for Phase-2, a tight spatial cut is not very effective in the rejection of pileup, instead it is preferable to use loose dz selections and set an additional timing cut [5].

The change in the PUPPI code can be made on charged candidates only, while neutrals, which are not detected by the tracker, are subject to the same treatment of the standard PUPPI version, based on the local discriminating variable α: nevertheless, also neutrals are affected since their α values depend upon the distribution of the local shapes of the charged candidates, which change moving from the standard algorithm to the new one.

Figure 7.6 shows the Δt distribution for candidates in signal jets only. Both curves are normalized to one. The black distribution is relative to the standard PUPPI code, whereas the blue one presents the updated version using the MTD timing; with the introduction of timing, the standard deviation of the distribution is lowered by 25%.

The improvement is likely tied to the higher number of candidates assigned to the signal jets by the version of PUPPI with timing (+3%): since those candidates are mainly signal-like, as demonstrated by the higher peak of the blue distribution, the σ is decreased.

The PUPPI with timing version attempts to avoid rejecting signal-like candidates, the main issue with the standard PUPPI configuration, and it partially achieves that:

[3]The CMS reconstruction code for Phase-2 is being developed and it is still far from its ultimate version, therefore it is likely that the best configuration for the pileup rejection algorithm will change again in the future.

[4]Either σ_{dz} and σ_{dt} depend upon many parameters such as the track momentum, its η, the number of hits the particle left in the tracker/MTD.

[5]It is worth stressing that this will likely change with more advanced versions of the reconstruction code for CMS Phase-2, and that spatial selections will be as important as timing ones. Nevertheless, this study wants to underline the possible strong impact of the MTD timing on such an important aspect as the pileup rejection.

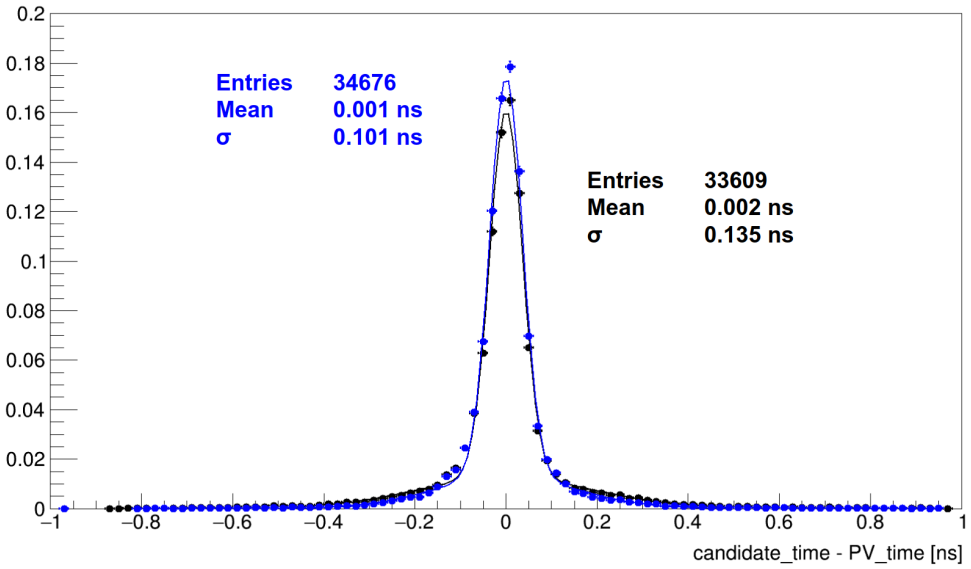


Figure 7.6: *Candidate time - PV time* distributions (normalized to one) for candidates in signal jets only. The black distribution is relative to the standard PUPPI code, whereas the blue one presents the updated version using the MTD timing. A better resolution is achieved with the latter version.

indeed, more signal-like candidates are present in signal jets, although a significant fraction of signal-like candidates is still present in the PU jets, as shown in figure 7.7 (blue distribution), underlining the need for further updates of the algorithm.

7.6 Improvements in the jet reconstruction

Two main updates that have an impact on the VBF $HH \to b\bar{b}b\bar{b}$ channel are presented in this chapter: (i) the *pointing method*, which improves the PV tagging; (ii) the introduction of the MTD timing in the PUPPI algorithm. It is therefore worth comparing the main figures of merit previously introduced, the signal efficiency, the purity, and the JER, before and after the implementation of the updates.

The comparison is presented in figures 7.8a, 7.8b, 7.8c with the efficiency, the purity, and the JER before and after the implementation in the reconstruction code of the two main updates. The uncertainties are $\leq 0.1\%$.

The implementation of the updates gives a clear advantage, as it improves the efficiency in the central region, in particular, while leaving the purity and the JER almost unchanged with respect to the default configuration. It is worth pointing out that either the *pointing method* and the updated version of the pileup rejection algorithm

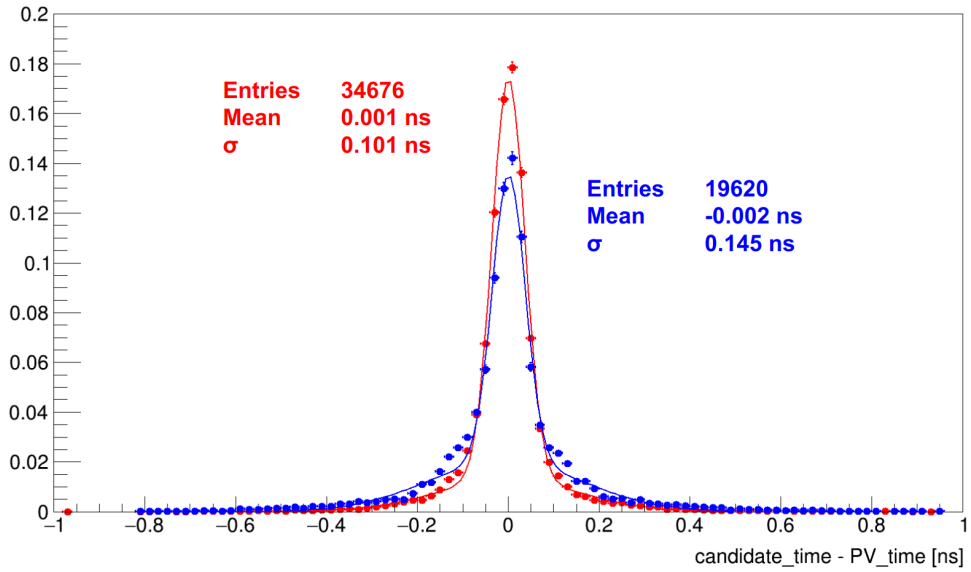


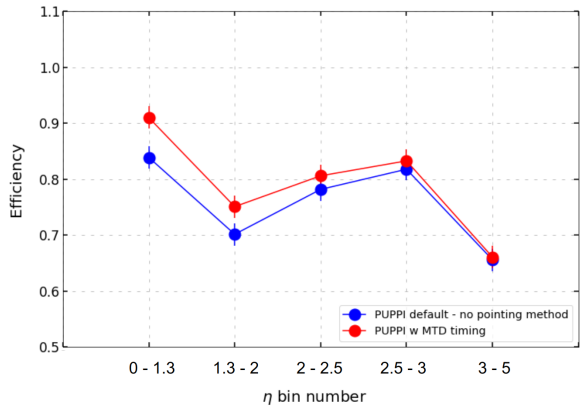
Figure 7.7: *Candidate time - PV time* distributions (normalized to one) concerning particle candidates in signal (red) and pileup (blue) jets. Both distributions are characterized by the sum of two gaussians, with different widths, relative to signal-like and pileup-like candidates. The jet assignment is based on the PUPPI rejection algorithm using the MTD timing.

rely on the MTD timing, whose contribution is fundamental. Indeed, in the fifth η bin, both efficiency and purity are similar in the two configurations, given that the MTD is not present in the 3-5 η interval.

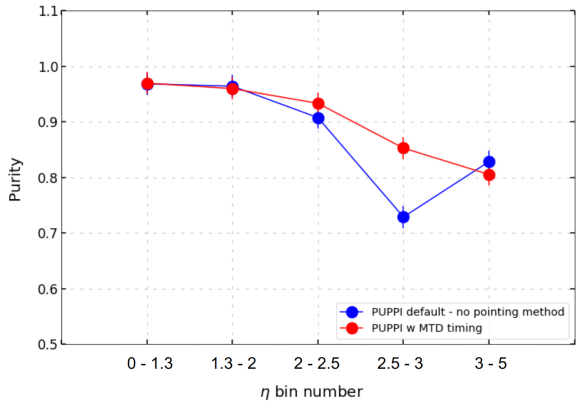
The drop in purity in the fourth η bin of the blue curve can be explained with figure 7.9, which presents the distributions, as a function of η, of the jets with $p_T > 30$ GeV (blue) and of the jets with $p_T > 30$ GeV matched to a generated jets (red); (a) is relative to the default PUPPI code without the use of the *pointing method*, (b), instead, uses the *pointing method* and the version of PUPPI with timing. The figure shows that both (a) and (b) have almost all jets matched over the entire η range, with the notable exception of the 2.5-3 η bin, where the number of total jets is much larger than the matched ones, meaning that, in that bin, the number of PU jets is significantly larger than in the other bins. Consequently, the purity is lowered in both the (a) and (b) scenarios, but the drop in (a) (corresponding to fourth bin of the blue curve in figure 7.8b) is much more evident because of the larger number of PU jets.

Figure 7.10 clarifies that the excess of jets in the 2.5-3 η bin in (a) and (b) is mainly due to PU jets, since they are concentrated at low p_T values; this is particularly evident in the scenario (a).

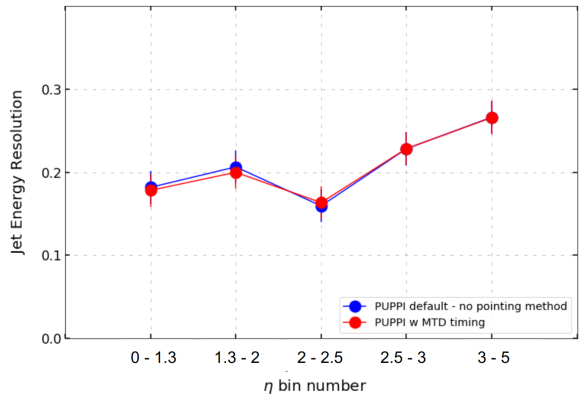
The interesting figures of merit can be also calculated for b-jets only, as shown in



(a) Efficiency



(b) Purity



(c) Jet Energy Resolution (JER)

Figure 7.8: Efficiency, Purity, and JER in bins of η, relative to the $HH \to b\bar{b}b\bar{b}$ channel, before (blue) and after (red) the *pointing method* for PV tagging and the new PUPPI algorithm using the MTD timing are applied.

figures 7.11a, 7.11b, 7.11c, with similar results.

Table 7.2 summarizes the results presented in this section, comparing the default scenario with the standard PUPPI code with the newest configuration using the *point-*

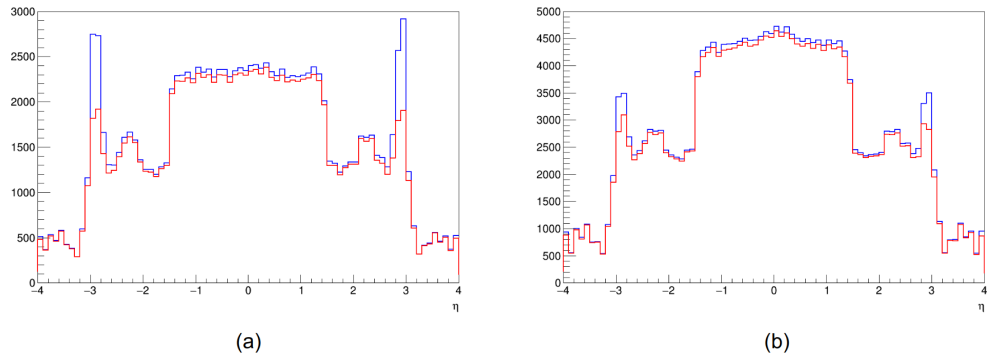


Figure 7.9: Number of total (blue) and matched-only (red) jets as a function of η. (a) represents the scenario where the default PUPPI version is used and the *pointing method* is not implemented; whereas (b) uses the PUPPI version with timing and the *pointing method*.

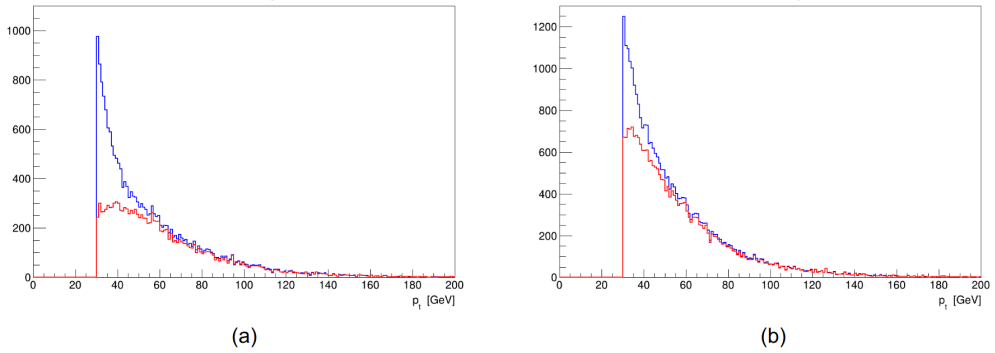
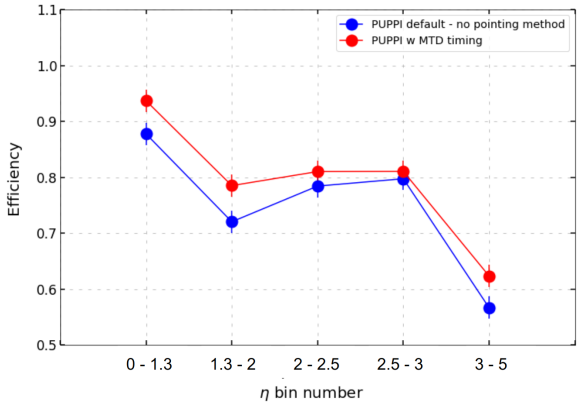


Figure 7.10: p_T distributions relative to the total (blue) and matched-only (red) jets in the 2.5-3 η bin.(a) represents the scenario where the default PUPPI version is used and the *pointing method* is not implemented; whereas (b) uses the PUPPI version with timing and the *pointing method*.

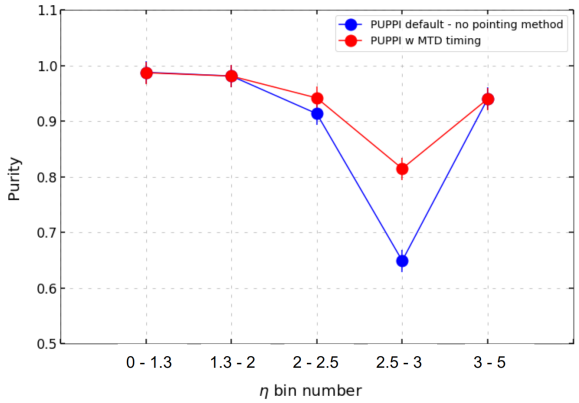
ing method and the MTD timing. The efficiency and purity reported in the table are the overall values, not divided in η or p_T bins.

7.7 Double Higgs search

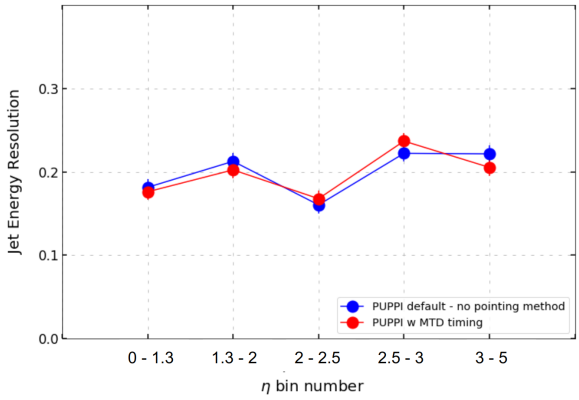
In this section, the same analysis presented in [12] is repeated, but using simulated data relative to CMS Phase-2, and employing the updated PUPPI with timing and the *pointing method*. This is a simplified analysis, which preliminary attempts to demonstrate the improvements related to the addition of the MTD timing in the reconstruction process; for this reason, the analyzed data are signal only, with no background added: additional studies with more realistic conditions will be carried out in the near future.



(a) Efficiency



(b) Purity



(c) Jet Energy Resolution (JER)

Figure 7.11: Efficiency, Purity, and JER in bins of η, concerning b-tagged jets only, before (blue) and after (red) the *pointing method* for PV tagging and the PUPPI algorithm using the MTD timing are applied.

A total of $\sim 8{\cdot}10^4$ signal events have been analysed.

Only events where the vertex tagged as primary has valid timing information are used.

	Default	MTD timing && *pointing method*
PV tagging efficiency	88.4 ± 0.1%	89.9 ± 0.1%
Efficiency	78.7 ± 0.1%	83.6 ± 0.1%
Purity	92.8 ± 0.1%	94.6 ± 0.1%

Table 7.2: PV tagging efficiency, efficiency, and purity computed in the default scenario and the new one using the *pointing method* and the MTD timing.

The reconstructed jets have their energy and p_T corrected using the MC truth, as explained at the beginning of this chapter.

The reconstructed jets are analyzed and can be tagged as b-jets or VBF candidates:

- A reconstructed jet is b-tagged if it features: (i) $|\eta|< 2$; (ii) $p_T > 40$ GeV; (iii) a b-tagging score larger than 0.4184, corresponding to a "medium" working point.
- A reconstructed jet is tagged as coming from the vector-boson fusion if it features: (i) $|\eta|> 2$; (ii) $p_T > 30$ GeV.

The selection continues if exactly 4 b-tagged and at least 2 VBF-tagged jets are found in the event, otherwise the event is discarded.

The two highest-p_T jets with opposite η are then chosen: if their η difference is larger than 4 and their invariant mass is larger than 500 GeV [6], the selection continues, otherwise the event is discarded.

An additional requirement that could be added to the simplified analysis shown here is that VBF jets candidates have compatible timings. Indeed, in the case of more than two jets passing the rapidity gap and invariant mass cuts, the request of compatible timings helps choosing the correct pair, given that the VBF jets originate from the same interaction. In this simplified analysis, however, due to the low statistics and the lack of a background, the requirement could not effectively be used for the VBF tagging, since it never occurs to have more than two jets passing the rapidity gap and the invariant mass selections.

Nevertheless, the *jet time - PV time* distributions (normalized to one) for signal (red) and pileup (blue) VBF jets have been produced, as shown in figure 7.12. The matching with the generated jets is based on the ΔR angular distance, while the requirements for a jet to be a VBF candidate are those presented above.

Since the jet time is given by the p_T-weighted average of its candidates' times, the distributions in 7.12 are similar to those presented in figure 7.5: signal jets have a

[6] The selection in [12] is tighter, requiring an η difference larger than 5 and an invariant mass larger than 1000 GeV. The selection in this analysis is a bit more relaxed and it is the same presented in the CMS paper on VBF HH, see [14].

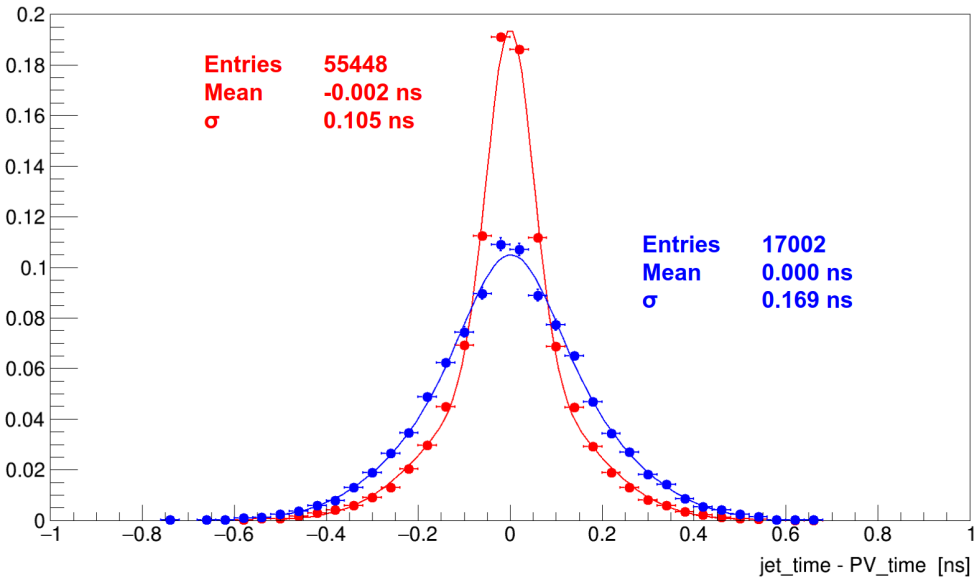


Figure 7.12: *Jet time - PV time* distributions (normalized to one) concerning signal (red) and pileup (blue) VBF jet candidates. The different widths of the two distributions prove that jet timing can be used to discriminate between signal and pileup VBF jets.

narrower width since their timings are correlated with the PV, whereas pileup jets have a significantly broader distribution, given that their timings are unrelated with that of the PV. Hence, a tight requirement on *jet time - PV time* can improve the accuracy of the VBF candidates selection by testing their temporal compatibility. Such an additional step will be introduced once the background is added in the analysis process.

The four b-jets are considered at this point, in three different combinations of two-jets pairings. Their invariant mass, m_{4b}, is used to define criteria to select signal-like events. Each pairing is characterized by a leading and a sub-leading pair, based on their p_T. Depending on the value of m_{4b}, the angular distances between the jets in both the leading and sub-leading pairs, ΔR_{bb}^{lead} and $\Delta R_{bb}^{sublead}$, determines whether the event is signal-like.

If $m_{4b} \geq 1250$ GeV, both ΔR_{bb}^{lead} and $\Delta R_{bb}^{sublead}$ are required to be lower than 1, for the event to be considered signal-like.

If $m_{4b} < 1250$ GeV, instead, the requirements are:

- $\frac{360\ GeV}{m_{4b}} - 0.5 < \Delta R_{bb}^{lead} < \frac{653\ GeV}{m_{4b}} + 0.475$
- $\frac{235\ GeV}{m_{4b}} < \Delta R_{bb}^{sublead} < \frac{875\ GeV}{m_{4b}} + 0.35$

Such requirements, the same ones used in [12], reflect the correlation between m_{4b},

	Default	MTD timing && *pointing method*
$Acceptance \times efficiency$	5.9 ± 0.1%	6.4 ± 0.1%
Purity	61.5 ± 0.1%	67.6 ± 0.1%

Table 7.3: $Acceptance \times efficiency$ and purity computed in the default scenario and in the new one using the *pointing method* and the MTD timing.

the Lorentz boost of the Higgs bosons, and the angle between their decay products in the laboratory frame.

Among the possible pairings fulfilling the above selection, the one that gives a dijet mass closest to that of the SM Higgs boson should be chosen. However, due to semileptonic B-hadrons decays, which lead to missing energy, the criterion is slightly relaxed, with the optimal leading Higgs boson mass set to 123.7 GeV and the subleading one set to 116.5 GeV; such choices maximize the signal significance, as reported in [12].

The selection of the leading and subleading pairs is not simply based on the two dijet masses, but on the D_{HH} parameter, namely the distance of the leading and subleading Higgs boson candidate masses, in the (m_{bb}^{lead} , $m_{bb}^{sublead}$) plane, from the line connecting (0 GeV, 0 GeV) and (123.7 GeV, 116.5 GeV). D_{HH} can be computed as:

$$D_{HH} = \sqrt{(m_{bb}^{lead})^2 + (m_{bb}^{sublead})^2} \left| \sin(\tan^{-1}(\frac{m_{bb}^{sublead}}{m_{bb}^{lead}}) - \tan^{-1}(\frac{116.5\ GeV}{123.7\ GeV})) \right| \quad (7.4)$$

The pairing leading to the smallest D_{HH} is chosen, and the corresponding distributions of m_{bb}^{lead} and $m_{bb}^{sublead}$ are plotted on each event and eventually fitted with a gaussian, the mean values providing the masses of the leading and subleading Higgs boson candidates.

The leading and subleading mass distributions, resulting from the $\sim 8{\cdot}10^4$ simulated data which have been analyzed, are presented in figure 7.13: the Higgs bosons masses peak at $\sim$ 124 GeV and $\sim$ 112 GeV, with a rather large $\sigma \sim$ 25 GeV, due to a very low statistics.

Figures 7.14 and 7.15 present the p_T distributions of the b-tagged and VBF jet candidates.

The signal $acceptance \times efficiency$ (ε in the following) has been defined in this work as the number of events over the total in which exactly 4 b-tagged jets and at least 2 VBF jets pass the selections illustrated above. The ε is 6.4 ± 0.1 %, to be compared with the 5.9 ± 0.1 % efficiency obtained by not using the *pointing method* and the MTD timing in PUPPI. The value of ε is in line with the result obtained from simulated data reported in [12].

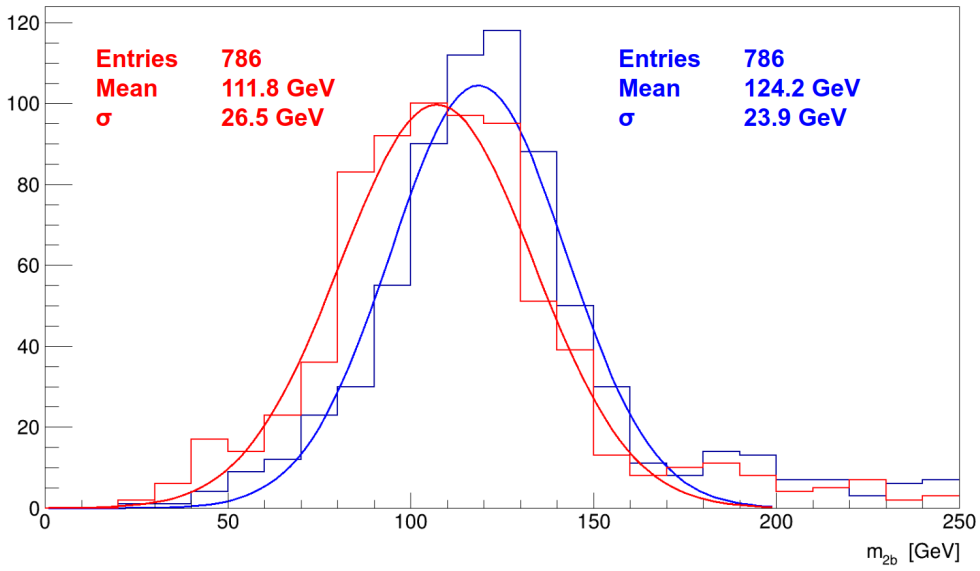


Figure 7.13: Leading (blue) and sub-leading (red) Higgs boson candidates mass distributions, reconstructed from the dijet mass, m_{2b}.

The signal purity (defined in a different way than the purity previously discussed) is another important figure of merit, which describes the fraction of events that passed the b-tagging and VBF selections in which six or more reconstructed jets are matched to a generated one: in this simplified analysis, it is found to be 67.6 ± 0.1 %, a 5% higher than not using the *pointing method* and the MTD timing in PUPPI. Such remarkable improvement underlines once more the impact of the updates in the reconstruction code described in the previous section.

As reported in section 7.1, a non-SM sample with the di-vector-boson–di-Higgs-boson coupling modifier, c_{2V}, set to 2 has been also produced and analyzed, given the unique sensitivity of the VBF production mode to BSM physics. Indeed, as previously discussed, any departure of the coupling modifiers from their SM value leads to an important increase of the VBF production cross section, therefore it is very interesting to simulate and analyze also this BSM scenario. Moreover, the VBF production mode is the only one to include the $VVHH$ channel, therefore only this channel is sensitive to a modification of the c_{2V} coupling modifier.

The increase of the production cross section in the BSM sample leads to an increase of the signal sensitivity. In particular, since the growth of the VBF production cross section in a BSM scenario is proportional to the partonic center-of-mass energy, as reported in chapter 1, the increase in signal sensitivity is more significant at high values of m_{HH}, the invariant di-Higgs mass, causing the m_{HH} spectrum to harden, and

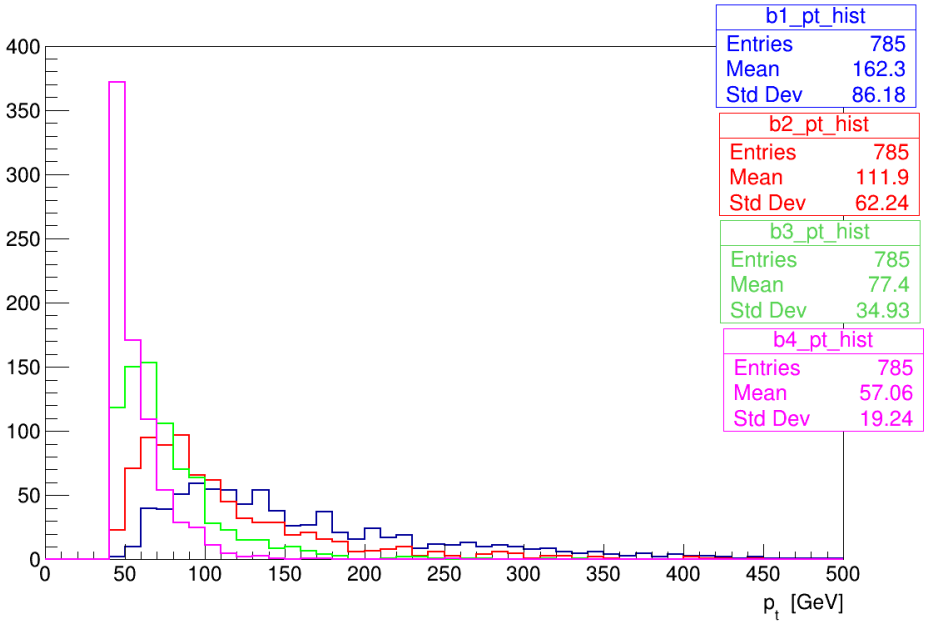


Figure 7.14: Transverse momentum distribution of the b-jet candidates.

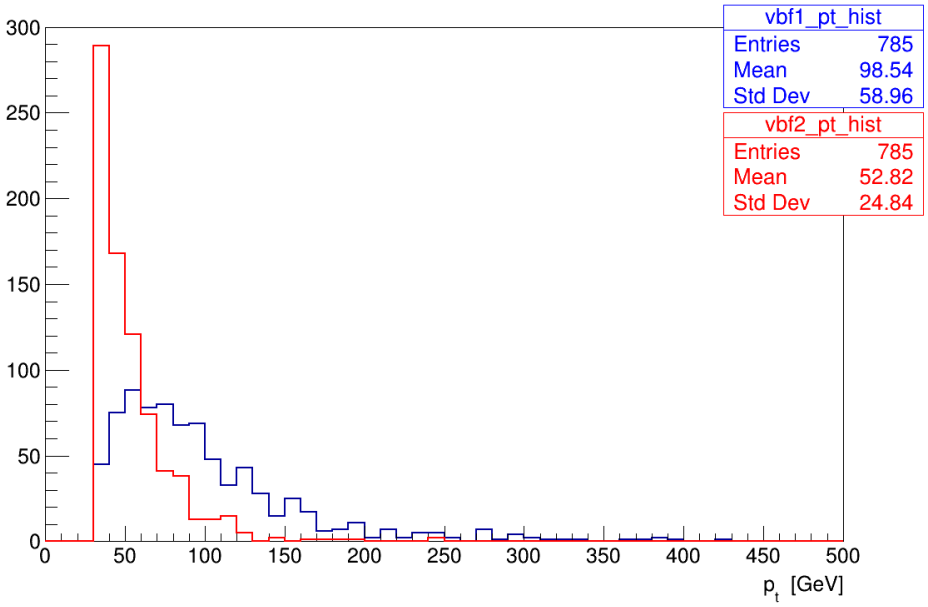


Figure 7.15: Transverse momentum distribution of the VBF jet candidates.

	Default	MTD timing && *pointing method*
$Acceptance \times efficiency$	15.0 ± 0.1%	16.7 ± 0.1%
Purity	62.1 ± 0.1%	68.3 ± 0.1%

Table 7.4: $Acceptance \times efficiency$ and purity computed for the BSM sample in the default scenario and in the new one using the *pointing method* and the MTD timing.

the p_T spectra of the b-jets to do the same [12, 13]. The p_T spectra of the b-jets in the BSM sample are shown in figure 7.16, to be compared with figure 7.14.

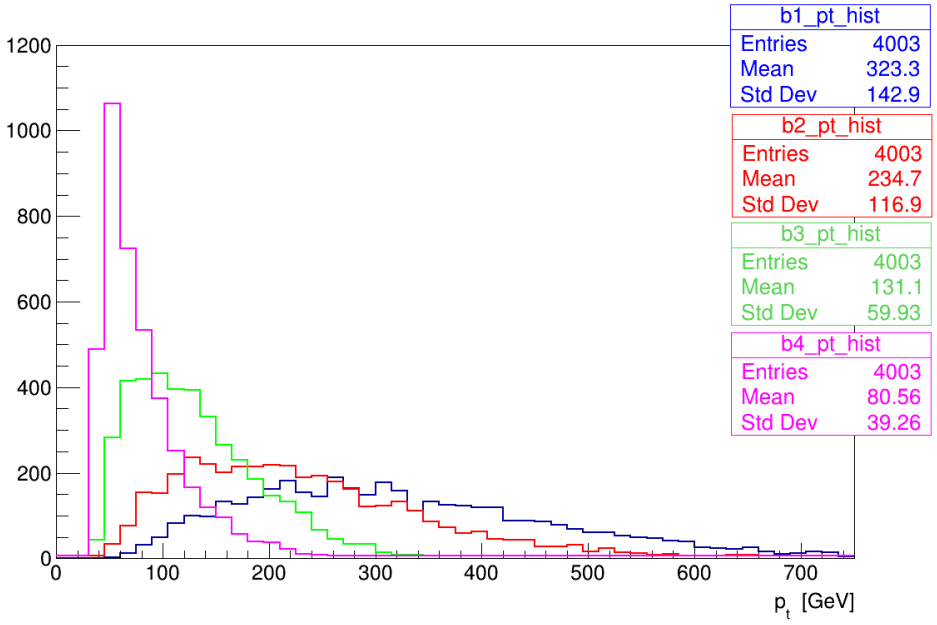


Figure 7.16: Transverse momentum distribution of the b-jet candidates, in the BSM scenario with $c_{2V} = 2$. The spectra are harder than in the SM sample, as expected.

The harder p_T spectra of the simulated BSM sample causes ε to grow significantly, as shown in table 7.4, while the purity is almost unchanged with respect to the SM scenario. A similar growth of ε with respect to its SM value can be seen in the ATLAS analysis [12], when setting $c_{2V} = 2$.

Also in the BSM scenario, as illustrated in the table, the improvement caused by the *pointing method* and the new PUPPI version is significant on both figures, particularly on the signal purity.

The mass distributions of both leading and sub-leading Higgs candidates are shown in figure 7.17: the σ of both distributions is slightly lowered with respect to the SM sample, likely because of the larger statistics (a consequence of the larger ε). The leading Higgs boson mass is further away from the expected value than in the SM

scenario, whereas the sub-leading candidate masses are very similar.

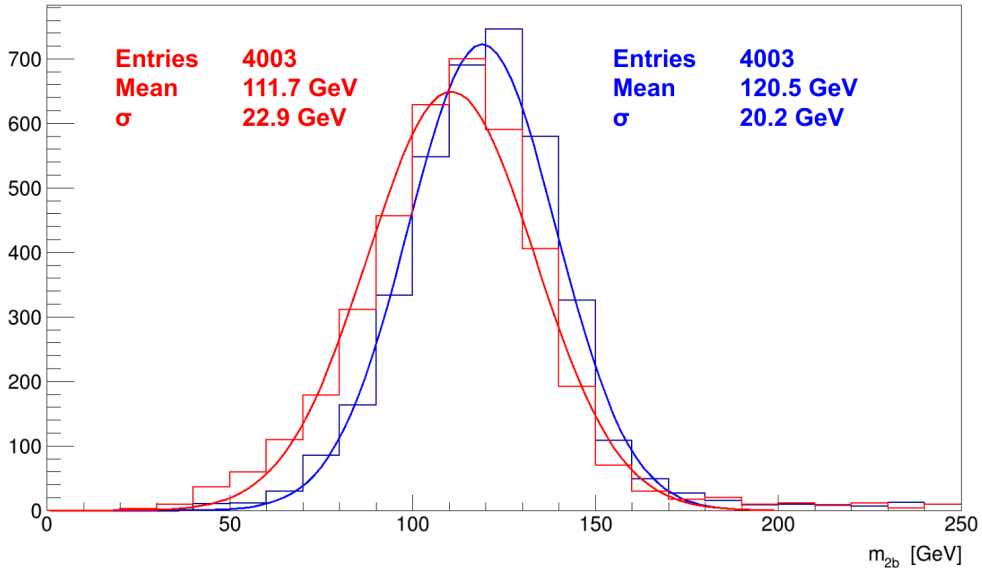


Figure 7.17: Leading (blue) and sub-leading (red) Higgs boson candidates mass distributions, reconstructed from the dijet mass, m_{2b}, in the BSM scenario with $c_{2V} = 2$.

Conclusions

This work focuses on the development of the CMS ETL towards the HL-LHC. The main results achieved concern the definition of the optimal inter-pad design for the UFSD arrays, the optimization of the gain layer, and the study of the impact of the MTD timing on the HH $\rightarrow b\bar{b}b\bar{b}$ channel in the VBF production mode.

The first two results led to the definition of an almost final UFSD design for ETL, the reference for all future productions. The analysis of the Higgs pair production proved the impact of MTD on the CMS reconstruction code for Phase-2: in particular, an increase in the reconstruction efficiency has been achieved, thanks to the timing information.

The measurements on the UFSD inter-pad demonstrated that a robust design features a low-doped p-stop with a grid structure, so that premature breakdown and micro-discharges cannot occur. A low-doped p-stop also ensures low sensitivity to floating pads. The FBK UFSD3.2 types 10 and 12 are the best designs: type 10 is slightly more robust because of the grid guard-ring structure, although it features a broader inter-pad width (62 μm against 28 μm of type 12). The behaviour of such sensor types has been also studied after irradiation, with particular attention to pads isolation, finding no issues.

The measurement campaign on the optimization of the gain layer design identified the FBK UFSD3.2 carbonated deep implant with CBL thermal diffusion scheme (W13) as the most suited for ETL, able to achieve 40 ps up to a radiation fluence of $2.5 \cdot 10^{15}$ $\mathrm{n_{eq}/cm^2}$ delivering at least 5 fC of charge. This design features a rather shallow charge-bias characteristic, ensuring a low sensitivity to non-uniform biasing conditions.

A new figure of merit has been introduced to assess the radiation hardness of a gain layer design: $\Delta V_{10fC}(\Phi)$, which describes the voltage increase required to provide 10 fC after a certain radiation fluence, Φ, with respect to the pre-irradiation condition. The lower $\Delta V_{10fC}(\Phi)$, the higher the overall radiation resistance of the sensor. The UFSD3.2 CBL sensors with deep implant feature the lowest $\Delta V_{10fC}(\Phi)$. $\Delta V_{10fC}(\Phi)$ is complementary to the radiation resistance coefficient (c): for instance, it has been

used to demonstrate that, despite having a slightly larger c-coefficient, the UFSD3.2 W13 is more suited for ETL than the standard UFSD3.2 W3 with shallow implant, given its lower sensitivity to non-uniform irradiation.

The effect on the operation of the ETL of having carbonated UFSDs has been highlighted: without carbon, the number of bias voltages lines per ETL disk grows more than a factor 3, increasing the complexity of the biasing scheme, the sensitivity to non-uniform irradiation, and the power consumption. Carbon co-implantation decreases $\Delta V_{10fC}(\Phi)$ by 50% at every fluence.

An important outcome of this work is the study of the time resolution as a function of the bias voltage at which the sensor is operated: assuming a fixed gain, sensors operated at higher voltages achieve a better resolution because of their sharper and larger signals. The resolution has been also studied as a function of the energy deposited by an impinging particle: low-energy deposits yield a good resolution, because the energy deposition is uniform along the particle path, while high-energy events, being characterized by large localized deposits, tend to have a poorer resolution because of the larger Landau noise.

The last part of this work is dedicated to the analysis of a Higgs bosons pair decaying in two $b\bar{b}$ pairs, in the vector-boson-fusion (VBF) production mode, using Phase-2 simulated data. The results have been compared with similar ATLAS and CMS analyses based on Run 2 data [12, 14]. A first improvement in the analysis is given by the *pointing method*, which exploits the peculiar topology of the process and the MTD timing to enhance the primary vertex tagging. The efficiency increase brought by this method in tagging the primary vertex is $\sim 2\%$. Secondly, a pileup rejection algorithm aware of the MTD timing has been developed: it is the first version of the algorithm to implement timing, and it represents the starting point for the rejection techniques to be developed for the Phase-2 CMS. Thanks to the updated rejection algorithm, the jet reconstruction efficiency and purity grow by 0.5% and 6%, respectively.

Using the improved reconstruction code, the ATLAS analysis has been repeated, in order to determine the Higgs bosons mass distributions, which respectively peak at 124 and 112 GeV, with $\sigma \sim 25$ GeV.

The same analysis has been repeated with a BSM dataset, finding results consistent with expectations. In this scenario, the efficiency and purity increase by 1.5% and 6%, respectively.

Appendix A

The update of the ETL geometry in CMSSW

This appendix presents the description of the ETL within the CMS software framework, CMSSW [97]. In particular, it describes the update of the ETL geometry in CMSSW from the old version present in the MTD Technical Proposal [98] to the present, post-TDR scenario. The impact of such updates on the CMS event reconstruction is also discussed.

In CMSSW, the event reconstruction and the study of final state observables are based on a complete GEANT [89] simulation of the Phase-2 CMS detector. The event reconstruction relies on a Particle Flow algorithm [22,23] that provides the most global description of an event, and the track-time information from MTD is added to improve the reconstruction, with the time information from charged tracks which is used to better reconstruct, both in space and time, the vertices (see chapter 7). The final state particles and observables are defined using vertices and track collections that are cleaned from pileup tracks using space and time compatibility requirements.

The time and the position of the particles crossing the MTD are reconstructed from the energy deposited in the active detector elements: if it is larger than 0.1 MIP-equivalent, which is the read-out threshold, the active element is considered hit by the particle. The effect of the detector resolution in the time measurements is reproduced by adding a gaussian smearing to the simulated times of tracks.

The track time at the collision vertex is obtained from the association of a time measurement in the MTD to a track and the subsequent time-of-flight correction to account for the track path length and the particle velocity. In the first step of the MTD track reconstruction, a simple topological clustering is performed to associate adjacent MTD hits above the read-out threshold. The barycenter, weighted by the single hit

energy, is used as an estimate of the cluster position and time. In the following step, tracks that have been reconstructed using the Tracker are propagated to the MTD and spatially matched with compatible clusters [1]. The track is then propagated from the point of closest approach to the beamline to the MTD cluster position, in order to compute the total path length. This path length is used, together with the particle velocity based on its momentum and the pion mass hypothesis, to estimate the time-of-flight (TOF) from the vertex. The MTD cluster time corrected for the TOF provides the track time at the point of closest approach to the beamline [2].

Vertex reconstruction in time and position along the beamline is performed using a time-aware extension of the deterministic annealing technique used for the CMS vertex reconstruction [2, 99]. At momenta below a few GeV, the difference in TOF between pions, kaons, and protons becomes significant with respect to the time resolution of the detector. In order to address this, the 4D vertex reconstruction is carried out in two stages. In the first stage, the pion mass hypothesis is used to compute the time of each track at the beamline, but the uncertainty assigned to this measurement is inflated by adding in quadrature the difference in time-of-flight between the pion and proton mass hypothesis. After this initial reconstruction, the compatibility of tracks with the reconstructed vertices is tested under the kaon and proton mass hypotheses in addition to the nominal pion hypotheses. Tracks identified as pions, kaons or protons have the additional contribution from the mass hypothesis ambiguity removed from the uncertainty assigned to their time measurement, and those identified as kaons or protons have their time measurement at the beamline recomputed as appropriate. These recomputed timestamps and uncertainties are then used to run the 4D vertex reconstruction a second time.

The simulation of the BTL geometry is accurate and has not been updated in this work. The trays are built within the Outer Tracker support tube, with a radial position of the crystals at 1174.5 mm from the beam axis. The BTL geometry model in CMSSW is based on 48 rows of modules per tray, each including 3 modules of 16×1 crystals of transverse dimensions 57.6×3.15 mm^2 with gaps between them corresponding to the space needed by SiPMs. The crystal bars have their long side orthogonal to the z dimension of the CMS detector. An aluminum plate with a surface of 92×52.2 mm^2 and 3 mm thick is placed below the crystals in front of the Tracker volume to simulate the support plate, while a PCB board of equivalent transverse size and 1.6 mm thick is placed 3 mm above the crystals to approximately simulate the readout electronics

[1] In a future update of the software, an additional tracking step will be present, in which the track reconstructed by the tracker is refitted considering also the additional MTD hit, to get a more accurate description of the particle trajectory. This is not yet present in the software version used for this work.

material budget.

Differently from BTL, the Technical Proposal version of the ETL geometry scenario was very simplified: it included a single disk per endcap, with eleven overlapping concentric rings of UFSD arrays, hermetic in ϕ and covering the pseudorapidity interval $1.5 < |\eta| < 3.0$. The scenario did not include any passive structure: electronics, services, cables, support structures were not present, only active UFSD elements covered the single ETL disk. Furthermore, a simplified approach was adopted to describe the timing performance of such design (regarding both BTL and ETL), with a constant single-cluster resolution of 25 ps: this choice is now outdated and a new, more realistic parameterisation has been developed and will be illustrated in the following. Multi-hit clusters have a resolution given by the energy-weighted average of the single-cluster uncertainty.

The implementation of a new ETL geometry scenario (see figure A.9) within CMSSW began at the end of 2019, following the geometry updates present in the TDR and the development of the ETL engineering drawings. The version presented in this chapter has been developed and used for the studies discussed in chapter 7 and is an update of the TDR geometry.

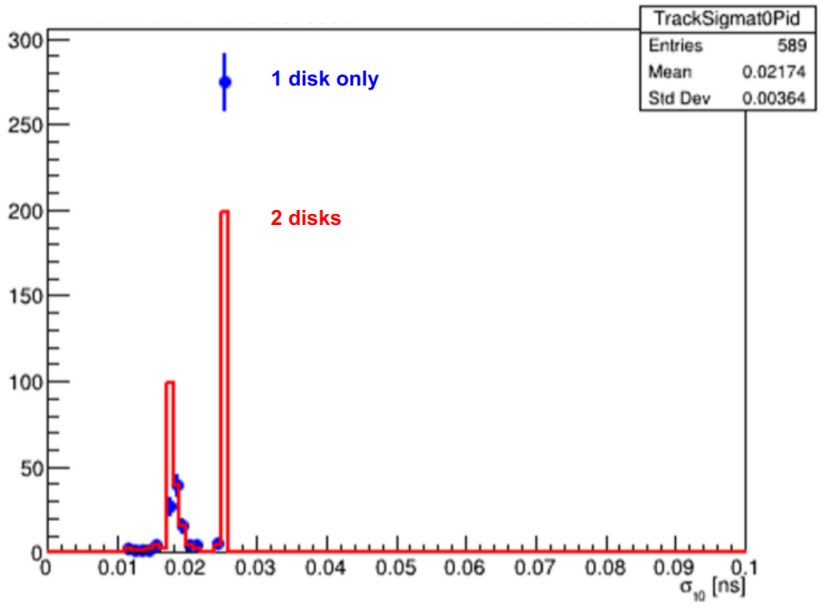


Figure A.1: Time uncertainty distribution in the Technical Proposal geometry scenario (blue) and in the 2-disks scenario used in this work (red). In the 2-disks scenario, the average resolution is lowered due to the higher number of multi-clusters in ETL. The uncertainties used for this plot do not consider the updated parameterisation that will be illustrated in the following.

The ETL detector is assembled from disks, with each disk serving as mechanical

support as well as a cooling plate. The disks are made of $\frac{1}{4}$-inch MIC6 aluminum, in which a small diameter stainless steel tube carries dual-phase CO_2 for cooling purposes. A 2-disks endcap, instead of the old version with just a single disk, is one of the main updates in the new geometry.

The presence of two disks allows measuring the time of passage of a large fraction of the tracks twice (the average number of hits in the 2-disks ETL is 1.7), improving the resolution. During the reconstruction process, the track time measured by Disk 2 (the one further away from the interaction point) is projected onto Disk 1, assuming a pion time-of-flight (TOF) between the two disks. The uncertainty of this projection is obtained as the difference between the proton and the pion TOFs between the two disks. The final time is given by the average of the two timestamps, weighted on their timing uncertainties; while the final uncertainty is given by the weighted average of Disk 1 and Disk 2 uncertainties, to which the TOF uncertainty is added.

Figure A.1 illustrates a study on the effect of having a double-disk instead of a single one; the study still considers the old description of the timing uncertainties, where all MTD single-clusters are given an uncertainty of 25 ps. The blue distribution in the figure represents the MTD uncertainties considering the single-disk scenario: a major fraction of the tracks has a single-hit cluster with 25 ps resolution, while a smaller part has a lower resolution, peaking at about 18 ps, mostly from BTL tracks, since the probability of having two UFSD hits [2] in the single-disk ETL is very low. With the two-disks scenario (in red) the number of multiple hits in ETL increases, consequently the peak at 18 ps is doubled, while the single-cluster peak is lowered. Therefore, the use of two disks improves the overall track time resolution, as expected.

In the new geometry, the sensor modules are placed on both faces of the two disks in the x-y layout presented in figure A.2. The sensor module is the building block of the active region of the ETL geometry, built from sub-assemblies containing an UFSD array, with 1.3×1.3 mm^2 pads, bump-bonded to the ETROC ASIC. The UFSD bonded to the ETROC is glued on an Aluminium Nitride (AlN) substrate which provides a cooling path. A thermally conductive film is glued to the bottom side of the AlN baseplate. A second AlN cover is placed on top of the module for protection purposes. Figure A.3 presents the exploded view of the module, in the design that is implemented in CMSSW.

The latest ETL module design, presented in chapter 1, is based on a 16×16 UFSD array, for a total 21×21 mm^2 surface, to be read out by a single ETROC of 20×20 mm^2

[2] The resolution of a track time measured n times is given by: $\sigma_{tracks} = \sigma_{hit}/\sqrt{n}$, with σ_{hit} being the resolution of the single hit. Therefore, the resolution of a track measured by two UFSDs with 25 ps single-hit resolution is 25 ps$/\sqrt{2} \sim 18$ ps.

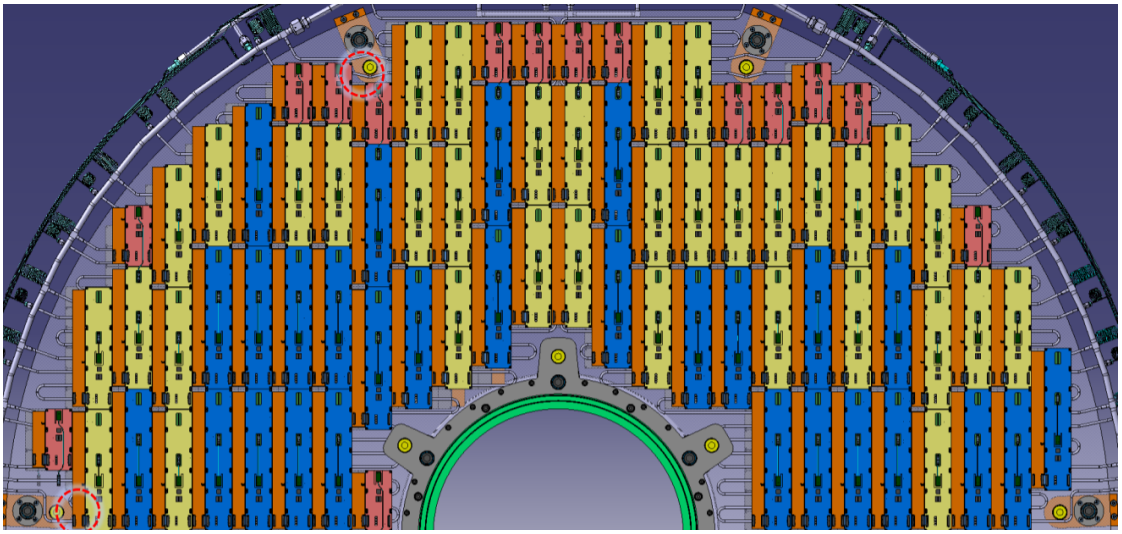


Figure A.2: A section with the most recent arrangement of sensor modules and services on a $D-shaped$ section.

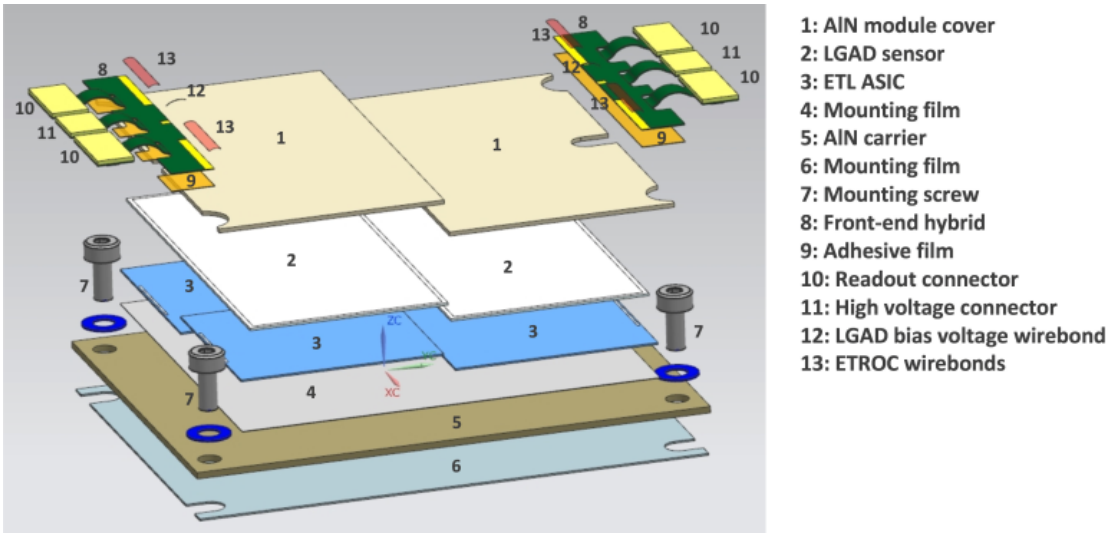


Figure A.3: The sensor module as implemented in the MTD TDR.

size. This is not yet implemented in CMSSW, and, for the simulation studies presented in this work, the UFSD array is 32×16 with $\sim$ 42×21 mm^2 surface, read out by two ETROC ASICs, as in the TDR design. In addition, the UFSD arrays do not implement, in the CMSSW version used in this work, the inactive area between pixels, therefore the sensors feature 100% occupancy (or fill factor), namely the ratio between the array's active and total areas.

The real UFSD arrays to be installed at the ETL will have $\sim$ 85% occupancy, and that is already accounted for in the most recent developments of the CMSSW code: a dedicated method takes four input parameters (the x/y sizes of the inter-pad gap

and of the inactive sensor edge) and the position of the simulated hit and returns the corresponding (n, m) coordinate of the hit pixel together with a boolean (true = hit; false = not hit, the particle crossed the inactive region). Simulation studies aware of such upgrade will be repeated in the near future.

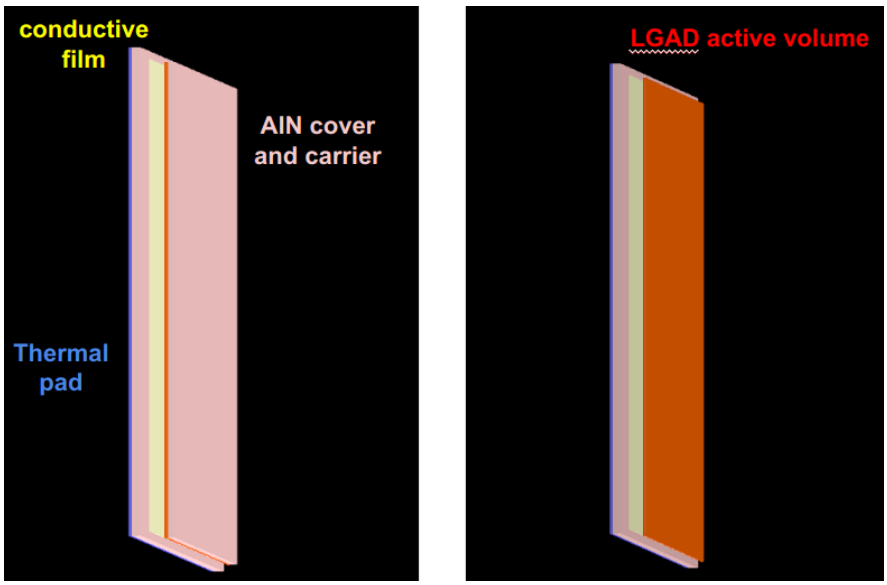


Figure A.4: The sensor module as implemented in CMSSW.

The sensor modules are placed on the disk surface in a staggered way such that the services arranged in channels along each line of sensors on one face are covered by the sensors on the opposite face; each line of sensors is made of two sensor modules, arranged so that the long sides [3] of the modules are adjacent, as illustrated in figure A.3. Each row of sensor modules has two adjacent rows of service structures, named service hybrids: one module is wire-bonded to the service hybrid on the left, one is bonded to the service hybrid on the right, this is shown in figure A.5.

The service hybrids, which are implemented in CMSSW with a simplified design (figure A.11), provide power and read-out services to the modules via flex circuit connectors; in particular, the boards (i) deliver power to the ETROCs and the bias voltage to the sensors, (ii) control and monitor signals and the clock to the ETROCs, (iii) transfer data from the ETROCs to the DAQ. In the simplified version used in this work, the service hybrids are just box-shaped volumes of a homogeneous passive material and come in three different sizes: 3-, 6-, and 7-sensor-modules-long, so that the whole disk surface can be covered in the most efficient way (see figure A.8).

[3] In the CMSSW version, the UFSD are 32×16 arrays, so they have sides of different length. That will not be the case when the 16×16 UFSDs of the most updated design will be implemented in the simulation software.

Each disk is split down the center vertically, forming a *clam shell* around the beam pipe, each *shell* being populated with modules on both faces (figure A.9). The whole set of modules and services on one face is called a $D-shaped$ section (figure A.2). A thermal screen, which isolates the ETL volume, is also implemented in the new scenario, along with a significant amount of passive volumes that were not considered in the previous geometry: the patch panels, the cables attached, and their support plates, the CO_2 pipes used for cooling of the sensors (see figure A.10).

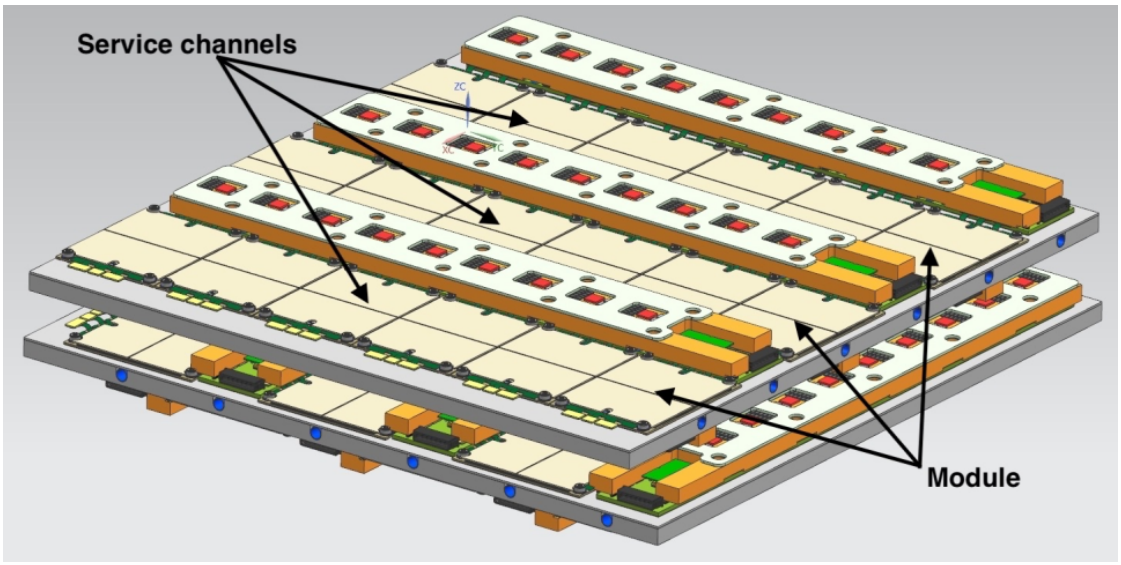


Figure A.5: Arrangement of sensor modules and service hybrids on the ETL disks.

In the new scenario, not only the ETL geometry changed, but also the parameterisation of the MTD time resolution, which is now more accurate (prior to this parameterisation, a constant 25 ps resolution was given to all MTD single-clusters). For ETL, the resolution per single-hit is flat at $\sim$ 39 ps both as a function of η and p_T; whereas, for BTL, it is flat in η and parameterized as a function of the charge generated by the impinging particle (figure A.6):

$$\sigma_{BTL}(Q) = p_0 \cdot Q^{p_1} \tag{A.1}$$

with Q being the charge, and p_0, p_1 parameters. Q is proportional to the energy released by the particle.

The parameterisation has been obtained by using simulated data and comparing the simulated hit time (t_{Sim}) with the reconstructed one (t_{Reco}). The $t_{Sim} - t_{Reco}$ difference is then plotted both in bins of η_{hit} and E_{hit} (the energy released by the particle in the hit) and the σ of the distribution is taken as the detector resolution. The study has been performed with minimum bias events, single muons and pions, finding compatible

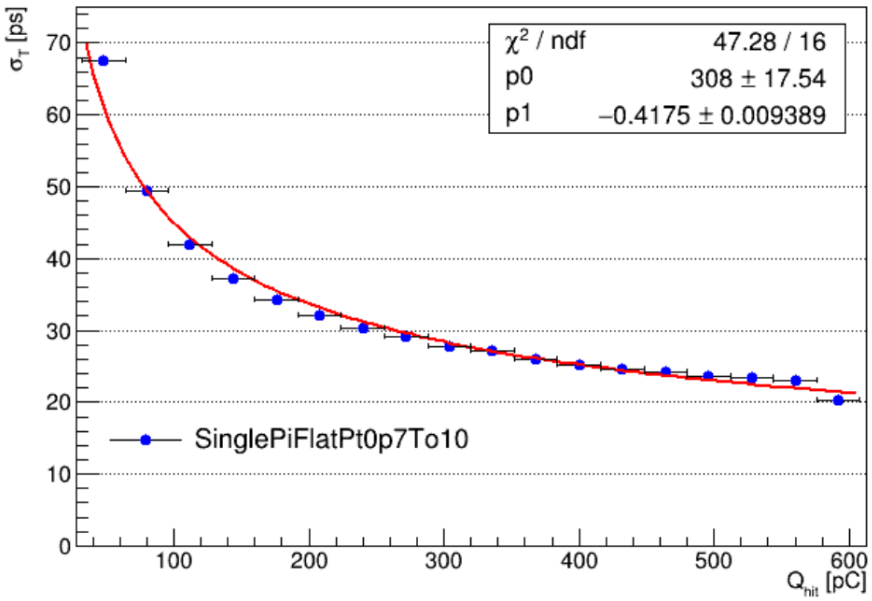


Figure A.6: Parameterisation of the BTL resolution as a function of charge generated by the incoming particle.

results.

The distribution of the track timing uncertainties with the new parameterisation is shown in figure A.7 for both ETL (red) and BTL (blue). The ETL distribution have almost all tracks peaking at 39 ps and 27 ps, which are the single-hit and double-hit (39 ps/$\sqrt{2}$ $\sim$ 27 ps, assuming two hits with the same uncertainty) uncertainties for ETL; a smaller fraction of tracks, those producing multi-clusters with more than two hits in ETL, have a better resolution. BTL, instead, have a broader distribution, due to the parametrized resolution which spans across a larger range of values. Both distributions have peaks at zero, due to tracks without valid timing information, i.e. tracks where the timing information could not be properly reconstructed (this happens, for instance, with low-energy tracks).

This new track uncertainty parameterisation does not provide a smaller resolution compared to the previous scenario, but it does set a much more realistic distribution.

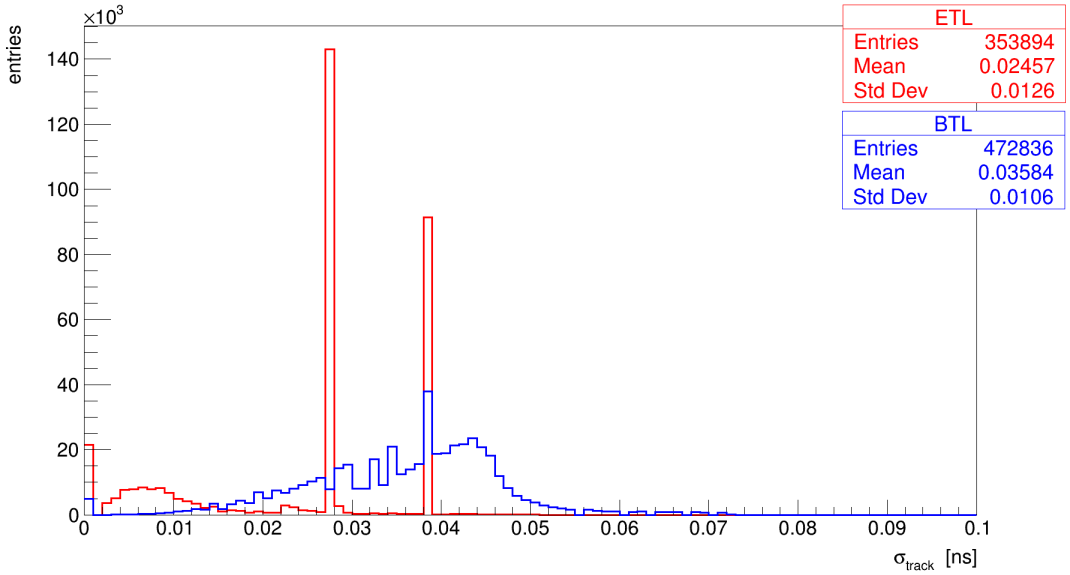


Figure A.7: Distribution of the timing uncertainties using the updated parameterisation of the BTL (blue) and ETL (red) resolutions.

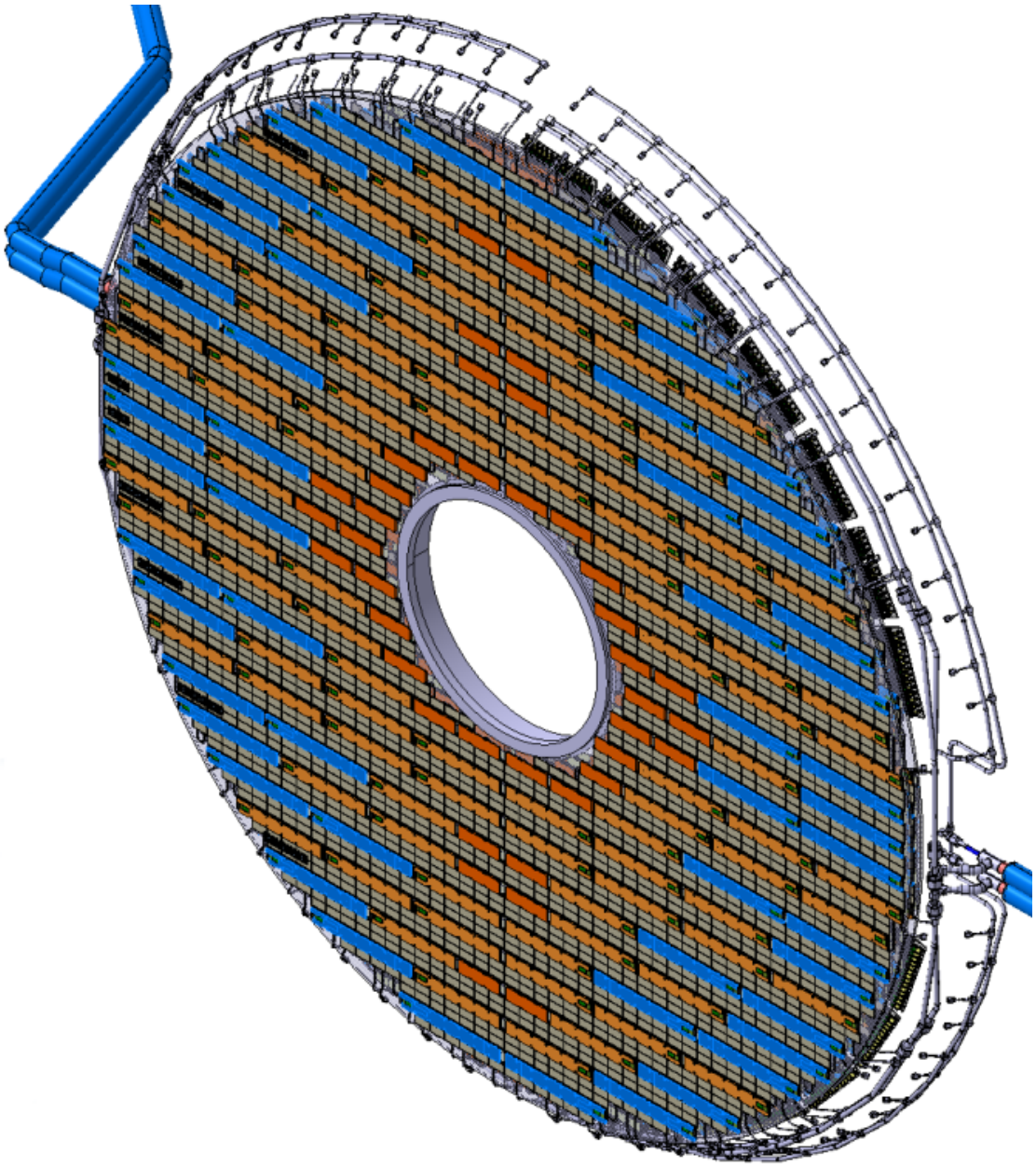


Figure A.8: Two $D-$ *shaped* sections forming one disk's face. The circular outer structure (in grey) represents the CO_2 pipes.

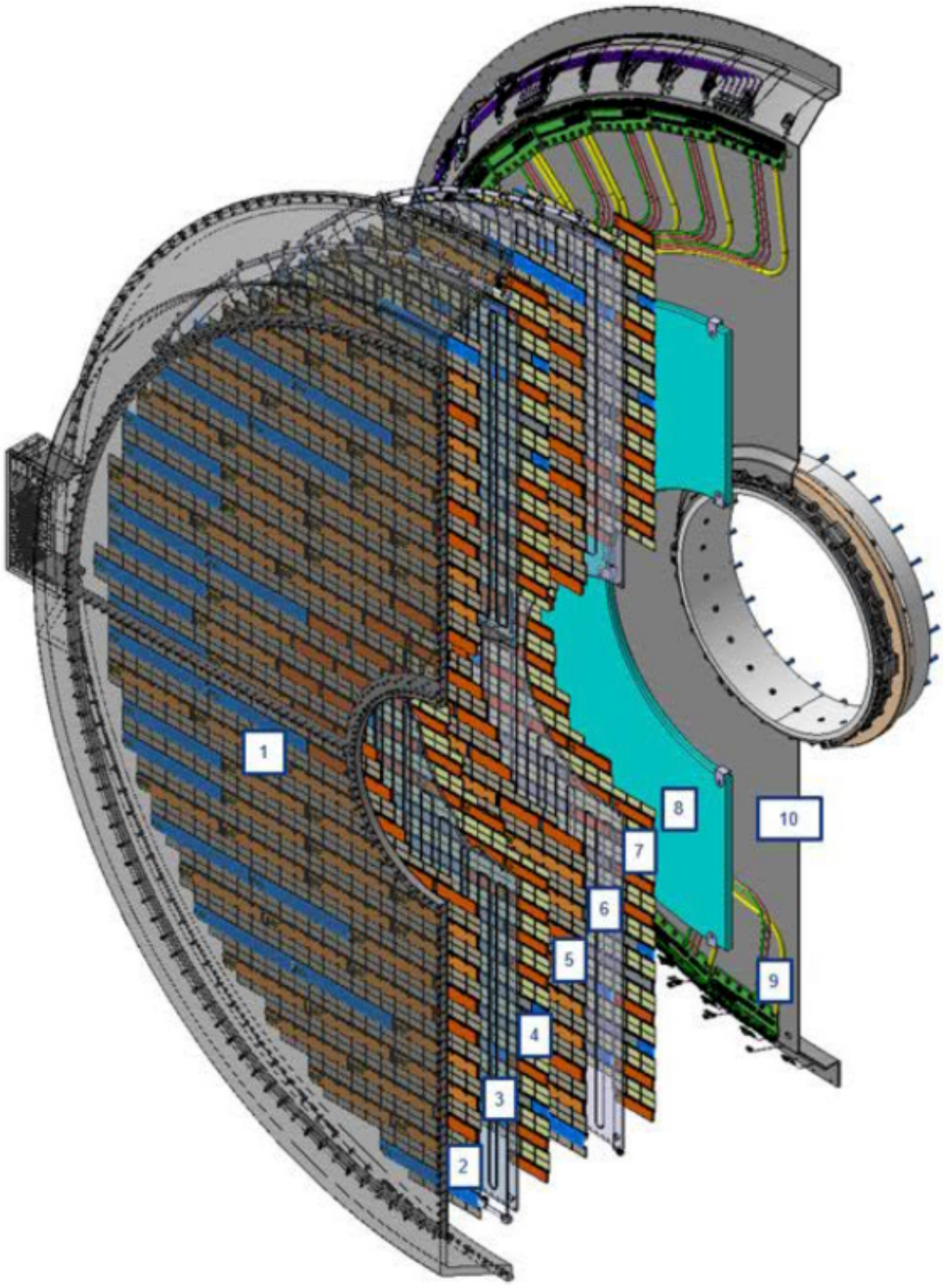


Figure A.9: Cross-sectional view of ETL along the beam axis. The interaction point is on the left. Only one half of the endcap is shown, with the different components numbered: (1) is the thermal screen; (2,4) front and rear faces of Disk 1 with active elements, electronics, and services shown; (5,7) front and rear faces of Disk 2 with active elements, electronics, and services shown; (3,6) support structures of Disks 1 and 2; (8,9) patch panels, cables; (10) back neutron moderator.

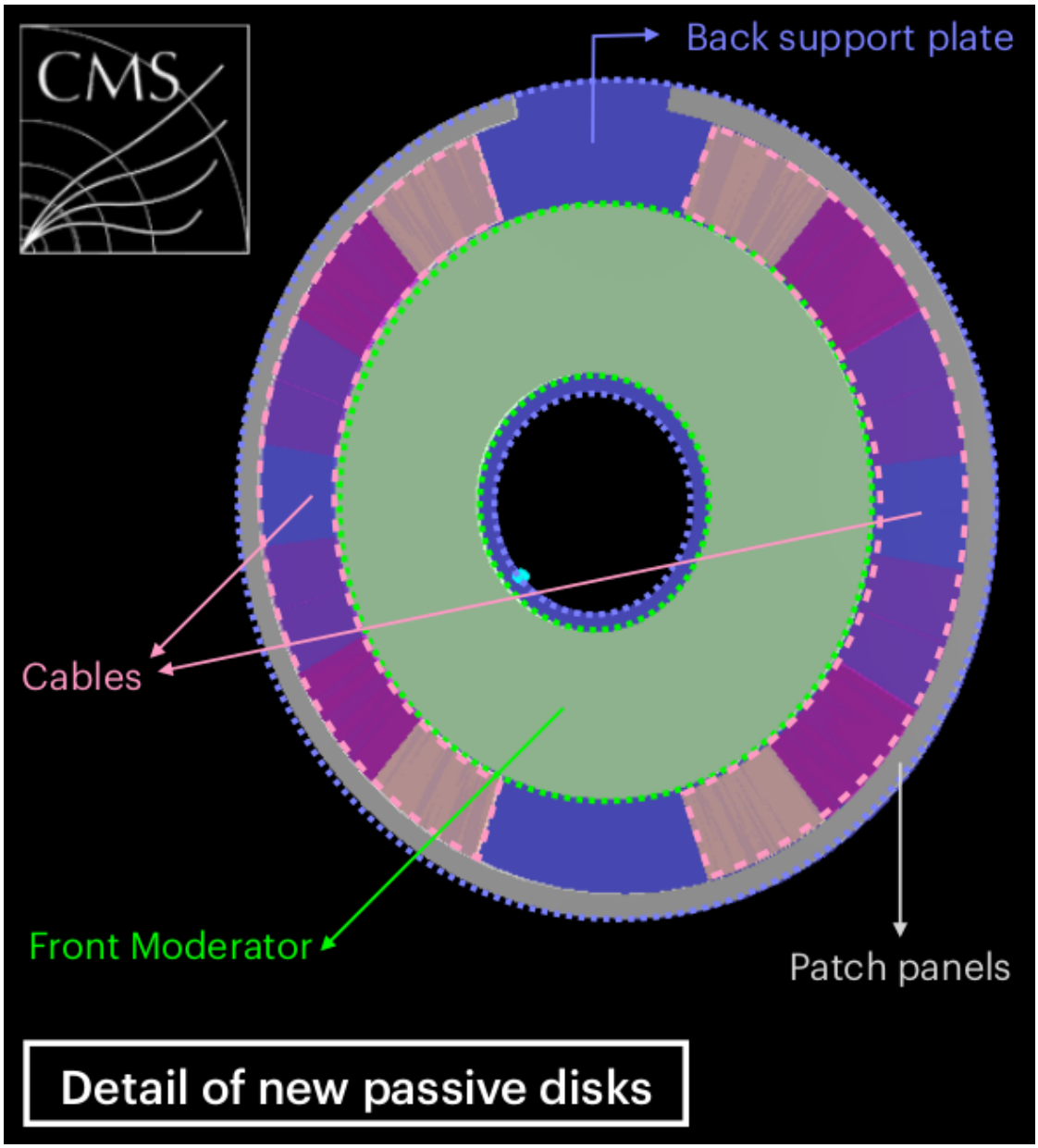


Figure A.10: The ETL disk in the latest CMSSW version, with the passive materials shown.

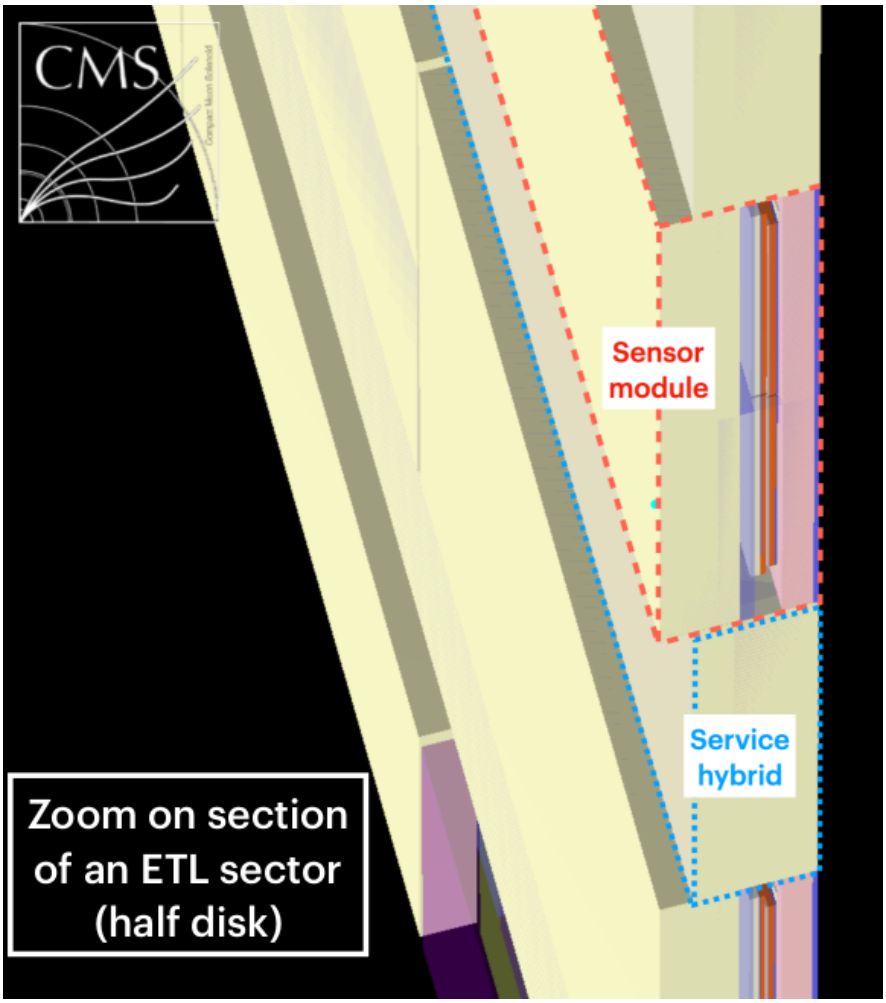


Figure A.11: Detail of the ETL disk with the sensor modules and service hybrids highlighted.

Appendix B

The biasing scheme of ETL

In this appendix, the present status of the ETL biasing scheme is reported. UFSD sensors in ETL will be subject to different radiation levels, thus requiring different bias voltages to reach the target time resolution; in particular, the innermost sensors, heavily irradiated, will require a significantly higher voltage than those in the outer part of the ETL disks.

The most challenging scenario when studying the biasing scheme is the end of the ETL lifetime, when the difference in the radiation fluences received by the UFSDs across an ETL disk will be at its maximum, requiring the largest number of different bias voltages to deliver uniform performance. Bringing a large number of bias voltage lines on an ETL disk is not trivial and require accurate studies.

In ETL, the high-voltage (i.e. bias voltage) lines are brought to the outer edge of the disks (figures B.1 and B.2), and from there they are connected to a patch-panel (the so-called PP0 module, figure B.3) which distributes the voltages to the sensor modules, via the service hybrids (described in appendix A). According to the current studies, a maximum of three different bias voltages can be delivered by a single service hybrid: this is a major constraint to be considered when selecting the UFSD design for ETL.

With the HPK split 4 design (not carbonated), six different voltages would be needed already at medium radius (i.e. not even in the most irradiated part) at the ETL end of lifetime, therefore such design cannot be considered for equipping the whole ETL disk. On the contrary, the FBK deep carbonated designs (W13, but also W19) would require a maximum of three different bias voltages per service hybrid even in the most irradiated part at the end of lifetime, therefore, according to the current studies, they are fully qualified in terms of bias voltage granularity. The reason for this significantly better performance is the smaller $\Delta V_{10fC}(\Phi)$ of the UFSD3.2 W13 and

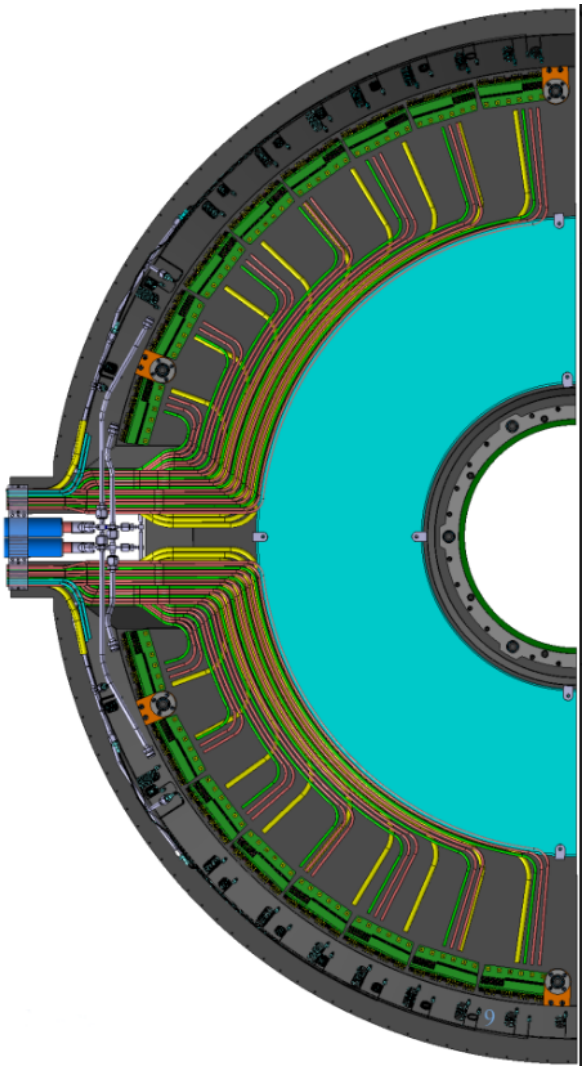


Figure B.1: Overview of the high- and low-voltage lines (yellow and red) in the outer edge of an ETL half-disk.

W19 designs compared to the HPK split 4, as already discussed in chapter 6.

This study further highlights that a deep carbonated gain implant is the optimal choice for ETL sensors: however, as previously mentioned, equipping the ETL is also a matter of costs, and the FBK UFSD3.2 W13 is too expensive for equipping the whole detector. The current plan is to instrument the largest possible area with sensors with a shallow gain implant (less expensive and well known by foundries), ensuring that the sensors can deliver the target resolution up to the end of the lifetime, and that they will not require more than three different bias voltages lines per service hybrid; the remaining, high-fluence part (fluence $> 1 \cdot 10^{15}$ n_{eq}/cm^2 at the end of operations) will be covered with UFSDs with deep carbonated gain layers, such as the UFSD3.2 W13.

The optimal UFSD design to operate in the ETL low-fluence part is the UFSD3.2

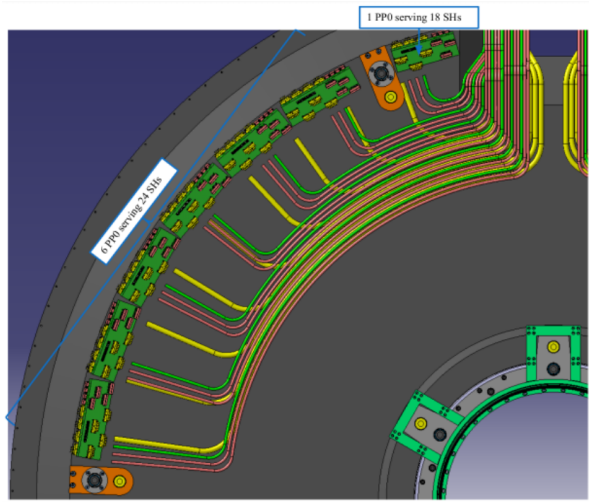


Figure B.2: High- and low-voltage lines in the outer edge of an ETL quarter-disk. PP0 modules (green) are well visible.

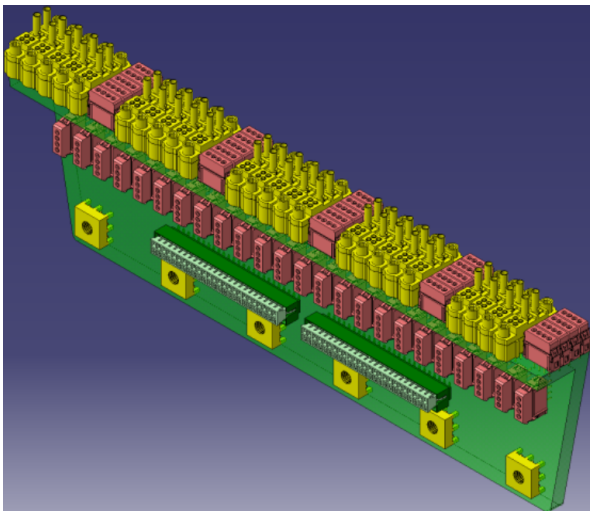


Figure B.3: The PP0 module.

W7, as already pointed out in chapter 6: in order to further confirm this, a study on the maximum number of bias voltages per service hybrid, similar to the ones performed on the HPK split 4 and UFSD3.2 W13, will be performed in the near future.

Acknowledgements

We thank Fondazione Bruno Kessle, INFN - Gruppo V and RD50 collaborations for the development of UFSD sensors. The work was supported by the United States Department of Energy, U.S. Department of Energy grant number DE-SC0010107. Part of this work has been financed by the European Union Horizon 2020 Research and Innovation funding program, under Grant Agreement no. 654168 (AIDA-2020) and Grant Agreement no. 669529 (ERC UFSD669529), and by the Italian Ministero degli Affari Esteri and INFN Gruppo V. Thanks to Dipartimento di Eccellenza, University of Torino (ex L. 232/2016, art. 1, cc. 314, 337). Thanks to Ministero della Ricerca, Italia, PRIN 2017, Grant 2017L2XKTJ – 4DinSiDe; and Ministero della Ricerca, Italia, FARE, Grant R165xr8frtfare.

Research in physics is team working, so, first of all, thank you to all my colleagues: Marta, Valentina, Marco, Luca. We spent together infinite hours, from the lab to the conference dinners... even after the dinners, when Pino Mugo pours in; you have often been more than just colleagues. Thank you to Marco Costa, my tutor, for his continuous support: you have advised me many times, and I am very grateful for that. Roberta, you have reviewed all my talks, every single slide, and, above all, you have always believed in me, what else should I say? Thanks! Nicolo, thanks for all the hours of discussion: you have inspired me, as only great teachers do, and I have learnt a lot. Thank you Roberto for helping me a lot with the analysis, and for all the suggestions. Thanks also to all the other colleagues from both INFN Torino and the University. Thanks Fabio Cossutti: CMSSW is a wild place to go without guidance, you have supported me with endless patience. Thanks to Barbara e Flori, our bond girls. Thank you Anna Benecke for all the explanations about PUPPI. Thanks for their help to all the students that worked with us: Giulia, Leonardo, Matteo, Filippo.

Thank you David, Gregor for reviewing my thesis, and for your helpful comments, and thank you also for coming to Torino!

A Claudia: questo lavoro è anche un po' tuo, che ci sei sempre stata. Vorrei anche ringraziare le persone che mi hanno accompagnato al mio primo giorno di scuola: Mamma, Papà, pensate quanta strada abbiamo fatto per arrivare alla fine di questo lungo ciclo, che ora si chiude. Grazie teo: per tutte quelle che abbiamo fatto e che ci racconteremo quando saremo vecchi. A tutti i miei amici, che rendono la mia vita migliore. Ringrazio anche tutta la mia famiglia: nonna Cristina, nonna Lina, Umbi, gli zii e le zie.

A Sansone, perchè dopo 29 anni non si vergogna a dormire ancora con me.

List of Figures

List of Tables

Bibliography

[1] L. Rossi and et al., "High Luminosity Large Hadron Collider (HL-LHC)," *CERN Yellow report 2015-005*, pp. 1–19, 2017.

[2] CMS Collaboration, "A MIP Timing Detector for the CMS Phase-2 Upgrade," tech. rep., CERN, Geneva, Mar 2019.

[3] https://hilumilhc.web.cern.ch/content/hl-lhc-project.

[4] B. Schmidt, "The High-Luminosity upgrade of the LHC:Physics and Technology Challenges for the Accelerator and the Experiments," *Journal of Physics: Conference Series*, 2016.

[5] https://hilumilhc.web.cern.ch/article/ls3-schedule-change.

[6] http://www.fluka.org/fluka.php.

[7] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Physics Letters B*, vol. 716, no. 1, pp. 30–61, 2012.

[8] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Physics Letters B*, vol. 716, no. 1, pp. 1–29, 2012.

[9] M. Cepeda, S. Gori, P. Ilten, M. Kado, F. Riva, and et al., "Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC," *CERN Yellow Rep. Monogr.*, vol. 7, pp. 221–584, 2018.

[10] M. Gouzevitch and A. Carvalho, "A review of Higgs boson pair production," *Reviews in Physics*, vol. 5, p. 100039, 2020.

[11] S. Di Vita, G. Durieux, C. Grojean, J. Gu, Z. Liu, G. Panico, M. Riembau, and T. Vantalon, "A global view on the higgs self-coupling at lepton colliders," *JHEP*, vol. 2018, no. 2, 2018.

[12] ATLAS Collaboration, "Search for the HH $\rightarrow b\bar{b}b\bar{b}$ process via vector-boson fusion production using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," *JHEP*, vol. 2021, no. 1, 2021.

[13] F. Bishara, R. Contino, and J. Rojo, "Higgs pair production in vector-boson fusion at the LHC and beyond," *The European Physical Journal C*, vol. 77, no. 7, 2017.

[14] CMS Collaboration, "Search for Higgs boson pair production via vector boson fusion with highly Lorentz-boosted Higgs bosons in the four b quark final state at $\sqrt{s} = 13$ TeV," tech. rep., CERN, 2021.

[15] CMS Collaboration, "Search for resonant pair production of Higgs bosons decaying to bottom quark-antiquark pairs in proton-proton collisions at 13 TeV," *JHEP*, vol. 2018, no. 8, 2018.

[16] CMS Collaboration, "Search for a massive resonance decaying to a pair of Higgs bosons in the four b quark final state in proton–proton collisions at s=13TeV," *Phys. Lett. B*, vol. 781, p. 244–269, 2018.

[17] CMS Collaboration, "Search for production of Higgs boson pairs in the four b quark final state using large-area jets in proton-proton collisions at $\sqrt{s} = 13$ TeV," *JHEP*, vol. 2019, no. 1, 2019.

[18] CMS Collaboration, "Search for nonresonant Higgs boson pair production in the $\mathrm{b\bar{b}b\bar{b}}$ final state at $\sqrt{s} = 13$ TeV," *JHEP*, vol. 2019, no. 4, 2019.

[19] ATLAS Collaboration, "Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," *JHEP*, vol. 2019, no. 1, 2019.

[20] CMS Collaboration, "The CMS experiment at the CERN LHC," *J. Instrum.*, vol. 3, no. 08, pp. S08004–S08004, 2008.

[21] D. Contardo, M. Klute, J. Mans, L. Silvestris, and J. Butler, "Technical Proposal for the Phase-II Upgrade of the CMS Detector," tech. rep., CERN, Geneva, 2015.

[22] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector," *J. Instrum.*, vol. 12, no. 10, pp. P10003–P10003, 2017.

[23] D. Bertolini, P. Harris, M. Low, and N. Tran, "Pileup per particle identification," *JHEP*, vol. 2014, Oct 2014.

[24] CMS Collaboration, "The Phase-2 Upgrade of the CMS Barrel Calorimeters," tech. rep., CERN, 2017.

[25] CMS Collaboration, "The Phase-2 Upgrade of the CMS Endcap Calorimeter," tech. rep., CERN, 2017.

[26] D. Anderson, A. Apreysan, A. Bornheim, J. Duarte, H. Newman, C. Pena, A. Ronzhin, M. Spiropulu, J. Trevor, S. Xie, and R.-Y. Zhu, "Precision timing measurements for high energy photons," *Nucl. Inst. Meth. A*, vol. 787, pp. 94–97, 2015.

[27] S. Gundacker, E. Auffray, B. Frisch, P. Jarron, A. Knapitsch, T. Meyer, M. Pizzichemi, and P. Lecoq, "Time of flight positron emission tomography towards 100ps resolution with L(Y)SO: an experimental and theoretical analysis," *J. Instrum.*, vol. 8, no. 07, pp. P07014–P07014, 2013.

[28] G. Pellegrini and et al., "Technology developments and first measurements of Low Gain Avalanche Detectors (LGAD) for high energy physics applications," *Nucl. Inst. Meth. A*, vol. 765, no. 2, pp. 12–16, 2014.

[29] W. Vasilescu and G. Lindstrom, "Displacement damage in silicon." https://rd50.web.cern.ch/niel/, 2000.

[30] M. Huhtinen, "Simulation of non-ionising energy loss and defect formation in silicon," *Nucl. Inst. Meth. A*, vol. 491, no. 1, pp. 194 – 215, 2002.

[31] G. Kramberger and et al., "Radiation effects in low gain avalanche detectors after hadron irradiations," *J. Instrum.*, vol. 10, no. 07, p. P07006, 2015.

[32] J. Varela, "The TOFHIR2 readout ASIC of the CMS Barrel MIP Timing Detector," tech. rep., CERN, Jan 2021.

[33] R. Abbott, A. Abreu, F. Addesa, M. Alhusseini, T. Anderson, Y. Andreev, A. Apresyan, R. Arcidiacono, M. Arenton, E. Auffray, and et al., "Test beam characterization of sensor prototypes for the CMS Barrel MIP Timing Detector," *J. Instrum.*, vol. 16, no. 07, p. P07023, 2021.

[34] ATLAS Collaboration, "Technical Design Report: A High-Granularity Timing Detector for the ATLAS Phase-II Upgrade," tech. rep., CERN, Geneva, Jun 2020.

[35] H. Sun, D. Gong, W. Zhang, C. Edwards, G. Huang, X. Huang, C. Liu, T. Liu, J. Olsen, and et al., "Characterization of the CMS Endcap Timing Layer readout

chip prototype with charge injection," *J. Instrum.*, vol. 16, no. 06, p. P06038, 2021.

[36] H. Sun, Q. Sun, S. Biereigel, R. Francisco, D. Gong, G. Huang, X. Huang, S. Kulis, P. Leroux, C. Liu, T. Liu, P. Moreira, J. Prinzie, J. Wu, J. Ye, L. Zhang, W. Zhang, and T. Liu, "A radiation tolerant clock generator for the CMS Endcap Timing Layer readout chip," tech. rep., CERN, 2021.

[37] F. Siviero, "Precision Timing with Low-Gain Avalanche Diode Sensors with the CMS Endcap Timing Layer for HL-LHC," *PoS Proc. Sci.*, vol. EPS-HEP2021, p. 815, 2022.

[38] G. Oh, "ETROC project for the CMS MTD Endcap Timing Layer (ETL)," in *2021 Meeting of the Division of Particles and Fields of the American Physical Society (DPF21)*, 2021.

[39] S. Ramo, "Currents Induced by Electron Motion," *Proceedings of the IRE*, vol. 27, no. 9, pp. 584–585, 1939.

[40] W. Shockley, "Currents to conductors induced by a moving point charge," *J. Appl. Phys.*, vol. 9, no. 10, pp. 635–636, 1938.

[41] W. Maes, K. De Meyer, and R. Van Overstraeten, "Impact ionization in silicon: A review and update," *Solid-State Electronics*, vol. 33, pp. 705–718, 1999.

[42] M. Ferrero and et al., *Ultra-fast silicon detectors : design, tests, and performances*. Boca Raton: CRC Press, 2021.

[43] F. Cenna and et al., "Weightfield2: A fast simulator for silicon and diamond solid state detector," *Nucl. Inst. Meth. A*, vol. 796, pp. 149 – 153, 2015. Proceedings of the 10th International Conference on Radiation Effects on Semiconductor Materials Detectors and Devices.

[44] N. Cartiglia and et al., "Issues in the design of Ultrafast silicon detectors." indico.cern.ch/event/351695/contributions/828366, 2015.

[45] N. Cartiglia and et al., "Design optimization of ultra-fast silicon detectors," *Nucl. Inst. Meth. A*, vol. 796, pp. 141–148, 2015.

[46] V. Sola and et al., "Ultra-Fast Silicon Detectors for 4D tracking," *J. Instrum.*, vol. 12, p. C02072, 2017.

[47] N. Cartiglia, "Beam test studies of the LGAD sensors at FNAL," in *30th RD50 Workshop, Krakow, Poland*, 2017.

[48] M. Moll and et al., "Effects of displacement damage involving the shallow acceptor doping in p-type silicon devices," in *Proceedings of Vertex 2019: 28th International Workshop on Vertex Detectors*, 2019.

[49] H. F.-W. Sadrozinski and et al., "In-depth experimental Study of Acceptor Removal in Low-Gain Avalanche Detectors," in *Proceedings of the 12th International "Hiroshima" Symposium (HSTD12)*, 2019.

[50] S. Terada and et al., "Proton irradiation on p-bulk silicon strip detectors using 12 GeV PS at KEK," *Nucl. Inst. Meth. A*, vol. 383, no. 1, pp. 159 – 165, 1996.

[51] G. Kramberger and et al., "Radiation effects in Low Gain Avalanche Detectors after hadron irradiations," *J. Instrum.*, vol. 10, p. P07006, 2015.

[52] N. Cartiglia, "A naive parametrization of initial acceptor removal," in *32th RD50 Workshop, Hamburg, Germany*, 2018.

[53] M. Ferrero and et al., "Radiation resistant LGAD design," *Nucl. Inst. Meth. A*, vol. 919, pp. 16 – 26, 2019.

[54] Y. Shimizu and et al., "Impact of carbon co-implantation on boron distribution and activation in silicon studied by atom probe tomography and spreading resistance measurements," *Jpn. J. Appl. Phys.*, vol. 55, no. 2, p. 026501, 2016.

[55] R. Wunstorf and et al., "Investigations of donor and acceptor removal and long term annealing in silicon with different boron/phosphorus ratios," *Nucl. Inst. Meth. A*, vol. 377, no. 2, pp. 228 – 233, 1996. Proceedings of the Seventh European Symposium on Semiconductor.

[56] G. Paternoster, "New developments in Ultra Fast Silicon Detectors at FBK," in *31th RD50 Workshop, CERN, Geneva*, 2017.

[57] N. Cartiglia and et al., "The effect of temperature and irradiation on the LGAD gain mechanism," in *34th RD50 Workshop*, 2019. https://indico.cern.ch/event/812761/contributions/3459057/.

[58] D. Massey, J. David, and G. Rees, "Temperature dependence of impact ionization in submicrometer Silicon devices," *IEEE T. Electron Dev.*, vol. 53, no. 9, p. 2328, 2006.

[59] http://www.imb-cnm.csic.es/index.php/en/.

[60] https://www.hamamatsu.com/eu/en/index.html.

[61] M. Ferrero and et al., "Study and characterization of low gain avalanche diode," in *Proceedings of the 12th "Trento" Workshop on Advanced Silicon Radiation Detectors*, 2017.

[62] https://www.fbk.eu/en/.

[63] G. Paternoster and et al., "Developments and first measurements of ultra-fast silicon detectors produced at fbk," *J. Instrum.*, vol. 12, p. C02077–C02077, 2017.

[64] Keysight Technologies, "Keysight technologies b1505a power device analyzer/curve tracer."

[65] http://particulars.si.

[66] M. Green and M. Keevers, "Optical properties of intrinsic silicon at 300 K," *Progress in Photovoltaics*, vol. 3, p. 189, 1995.

[67] G. Kramberger, "Advanced Transient Current Technique Systems," in *Proceedings of the 23rd International Workshop on Vertex Detectors*, 2015.

[68] http://www.particulars.si/TCTAnalyse/.

[69] S. Meroli, D. Passeri, and L. Servoli, "Energy loss measurement for charged particles in very thin silicon layers," *J. Instrum.*, vol. 6, no. 06, pp. P06013–P06013, 2011.

[70] https://www.hamamatsu.com/eu/en/product/cameras/cmos-cameras/index.html.

[71] F. Siviero and et al., "Optimization of the gain layer design of ultra-fast silicon detectors," *Nucl. Inst. Meth. A*, p. 166739, 2022.

[72] N. Cartiglia and et al., "Beam test results of a 16ps timing system based on ultra-fast silicon detectors," *Nucl. Inst. Meth. A*, vol. 850, pp. 83–88, 2017.

[73] Z. Galloway and et al., "Properties of HPK UFSD after neutron irradiation up to 6e15 n/cm^2," *Nucl. Inst. Meth. A*, vol. 940, 2019.

[74] https://pypi.org/project/PyVISA/.

[75] P. Skomina, B. Hiti, V. Cindro, A. Howard, I. Mandić, M. Mikuž, and G. Kramberger, "Studies of inter-pad distance in low gain avalanche detectors," *Nucl. Inst. Meth. A*, p. 166158, 2021.

[76] M. T. et al., "Performances of the third UFSD production at FBK," in *33rd RD50 Workshop*, 2018. https://indico.cern.ch/event/754063/contributions/3222642/.

[77] M. Printz, "P-stop isolation study of irradiated n-in-p type silicon strip sensors for harsh radiation environments," *Nucl. Inst. Meth. A*, vol. 831, pp. 38–43, 1984.

[78] Y. Unno and et al., "Optimization of surface structures in n-in-p silicon sensors using TCAD simulation," *Nucl. Inst. Meth. A*, vol. 636, pp. 118–124, 2011.

[79] F. Siviero, "Innovative TCT studies on the breakdown of UFSD3 sensors by FBK," in *33rd RD50 workshop*, 2018.

[80] F. Hartmann, "Evolution of Silicon Sensor Technology in Particle Physics," *Springer Tracts Mod. Phys.*, vol. 275, pp. pp.1–372, 2017.

[81] L. Snoj, G. Žerovnik, and A. Trkov, "Computational analysis of irradiation facilities at the JSI TRIGA reactor," *Appl. Radiat. Isot.*, vol. 70, no. 3, pp. 483–488, 2012.

[82] "Supply of LGAD sensors for the CMS MIP Timing Detector Endcap: 07010202M." http://cds.cern.ch/record/2777890.

[83] Atkinson, Kendall, and et al., *An Introduction to Numerical Analysis*. John Wiley & Sons., 1989.

[84] M. Ferrero and et al., "Recent studies and characterization on UFSD sensors," in *34th RD50 Workshop*, 2019.

[85] F. S. et al., "Characterization with a β-source setup of the FBK UFSD 3.2 & HPK2 LGAD productions," in *37th RD50 Workshop*, 2020. https://indico.cern.ch/event/896954/contributions/4106335/.

[86] H. F. W. Sadrozinski, A. Seiden, and N. Cartiglia, "4D tracking with ultra-fast silicon detectors," *Rep. Prog. Phys.*, vol. 81, no. 2, p. 026101, 2018.

[87] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV," *J. Instrum.*, vol. 13, p. P05011, 2017.

[88] CMS Collaboration, "Identification of b-quark jets with the CMS experiment," *J. Instrum.*, vol. 8, no. 04, p. P04013, 2013.

[89] S. Agostinelli and et al., "Geant4—a simulation toolkit," *Nucl. Inst. Meth. A*, vol. 506, no. 3, pp. 250–303, 2003.

[90] W. Erdmann, "Vertex reconstruction at the CMS experiment," *Journal of Physics: Conference Series*, vol. 110, no. 9, p. 092009, 2008.

[91] T. Speer, K. Prokofiev, R. Fruhwirth, W. Waltenberger, and P. Vanlaer, "Vertex fitting in the CMS tracker." CERN-CMS-NOTE-2006-032, 2006.

[92] M. Cacciari, G. Salam, and G. Soyez, "The anti-ktjet clustering algorithm," *JHEP*, vol. 2008, no. 04, pp. 063–063, 2008.

[93] M. Cacciari, G. Salam, and G. Soyez. FastJet user manual, 2012. 10.1140/epjc/s10052-012-1896-2.

[94] De Iorio, A., *Search for new physics contributions entering the tWb electroweak vertex at 13 TeV with the CMS experiment.* PhD thesis, Universita degli Studi di Napoli Federico II, 2021.

[95] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector," *J. Instrum.*, vol. 12, no. 10, pp. P10003–P10003, 2017.

[96] D. Bertolini, T. Chan, and J. Thaler, "Jet Observables Without Jet Algorithms," *JHEP*, vol. 04, p. 013, 2014.

[97] https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSSWFramework.

[98] CMS Collaboration, "Technical proposal for a MIP timing detector in the CMS experiment Phase 2 upgrade," tech. rep., CERN, 2017.

[99] E. Chabanat, N. Estre, and P. Vanlaer, "Deterministic annealing for vertex finding at CMS," 2003.