Design and characterization of pixel IC electronics and sensors for new pixel detector generations

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von
Piotr Rymaszewski
aus
Ostróda, Polen

Bonn, 01.2022

Contents

iv

Introduction

Physics is one of the oldest scientific disciplines and its origins come from humans' fascination of the universe and the laws that govern it. It is difficult to say when the beginning of physics was, but if we narrow the search to particle physics only (understood as study of matter and its composition), than one could argue that the atomism proposed by Democritus over 2000 years ago could be the starting point. Atomism proposes that the physical world is composed out of fundamental, indivisible components knows as atoms. It took until the 20th century to advance technology and experimental methods sufficiently to start investigating this concept. Based on the new data obtained from experiments, the theories describing particles evolved and the one which presently is considered as the best description of particle physics is the Standard Model. It manages to unify electromagnetic, strong and weak interactions, leaving out only the fourth fundamental interaction - gravity. Its predictions agree very well with many experimental results, but the model does not explain everything. In order to advance further, new and more elaborate experiments are needed. An example of such an endeavour is the Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN). The LHC is the biggest and most complex particle collider to date, it collides protons at a center of mass energy of 14 TeV. The LHC is operating since 2008 and it allowed observing the Higgs boson with a mass of 125 GeV, which was in accordance with the Standard Model. The collider is expected to still run for many years, though with upgrades along the way. The preparations for the biggest upgrade are currently going on, leading to a significant increase of the luminosity and so-called High Luminosity LHC (HL-LHC) era.

Particle collider experiments require the simultaneous detection of particle's tracks with high spatial resolution and precise timing, especially close to the interaction point. The technologies used for achieving this changed a lot during the last 50 years, but presently the state-of-the-art instruments rely on pixel detectors. In such device the sensitive area is composed out of large amounts of small sensing elements (pixels), which can process the observed signals and can be read out individually. The detectors are built out of several layers of pixels surrounding the interaction point. The most common pixel design is the hybrid approach, where the sensing element is one entity, the signal processing electronics is the second one and they are connected together with metal bumps. Presently it is possible to construct detectors with very large pixel density (thousands per cm^2) thanks to the usage of modern CMOS (Complementary Metal Oxide Semiconductor) electronics. At the same time the sub-micron CMOS processes allow integrating complex electronic circuits in the pixels, enabling high data rate processing. The readout chips for the pixels have to be custom designed for each

experiment, and the designs are becoming more and more complex as the luminosity, particle hit rate and amount of information expected from every interaction grows. Additionally, integrated circuits (ICs) for High Energy Physics (HEP) experiment have to operate in very harsh radiation environment, which leads to circuit performance degradation. All this makes the development of integrated circuits for HEP experiments very difficult, while at the same time being mandatory for the correct operation of detectors. This thesis is dedicated to two IC development topics.

As mentioned before, currently most pixel detectors follow the hybrid approach, where the sensing element and the readout electronics are two separate devices. While this allows achieving the best performance and radiation hardness, hybrid devices are complicated and expensive to make. An alternative approach would be to combine the sensor and electronics into one device. Such devices, called Monolithic Active Pixel Detectors (MAPS), have been in use for a long time, but mostly in low radiation and low rate applications due to slow charge collection through drift. In recent years the HEP community tried to take advantage of the advances of CMOS manufacturing technology in order to improve the performance of MAPS. This lead to Depleted MAPS (DMAPS), where the bulk of silicon substrate is depleted and is used to quickly collect charge through diffusion. This development of DMAPS design capable of meeting the requirements of outer layer of HL-LHC ATLAS Pixel Detector, both in terms of electrical performance and radiation hardness, is the first topic of this thesis. If those requirements could be met, a DMAPS device would be an interesting alternative to hybrid pixels, as it would make production of the detector cheaper, simpler and faster.

The second topic is the design of a Clock Data Recovery (CDR) circuit for future hybrid pixel readout chips. The performance requirement for the pixel readout chip in HL-LHC era for ATLAS and CMS are very similar, therefore the groups decided to work together on the chip design. This collaboration, named RD53, started in 2013 is one of the biggest chip design groups in HEP – over the years it included members from over 24 institutes world-wide for chip design work and from many more institutes for chips testing. RD53 has successfully proven the feasibility of producing the required ICs and is currently working on the final production versions of the chips for both ATLAS and CMS. The CDR circuit is an important part of the communication interface of the RD53 chips and it also produces clock signals with different frequency ranging from 40 MHz up to 1.28 GHz needed for data processing inside the chip. Usage of a CDR instead of a Phase Locked Loop (common in older generations of HEP ICs) also allows reducing the number of cables needed for operating the chip, thus helping to reduce the detector's material budget.

All the work presented in this thesis was done in the context of LHC and developing integrated circuits needed for the upgrade of pixel detectors planned for the High Luminosity LHC upgrade. Chapter 2 presents a short overview of the LHC, focusing on the areas and devices related to this thesis. The next two chapters present the main projects developed in the framework of this thesis. Chapter 3 describes work on monolithic silicon radiation sensors designed in 150 nm CMOS technology, while Chapter 4 shows the development and measurement results of Clock-Data Recovery circuits designed in 65 nm CMOS technology. Finally, the conclusions are presented in Chapter 5.

Experimental environment

The aim of this chapter is to provide a short overview of the LHC and the ATLAS experiment as the context for the work done in this thesis. Section 2.1 introduces the LHC itself and its upgrade. Section 2.2 provides some details on the ATLAS detector, changes expected for the HL-LHC and simulation predictions regarding the radiation environment, which is a very important constraint for the designs which will be presented in the next chapters.

2.1 The Large Hadron Collider

The Large Hadron Collider [1] is the world largest particle accelerator and largest particle accelerator constructed by mankind to date. It was built at CERN near Geneva, just at the border of Switzerland and France, as shown in Fig. 2.1(a). The LHC is a circular accelerator with 27 km circumference placed 45 m - 170 m below ground, designed for proton - proton collisions at a center of mass energy of 14 TeV. The two beams (composed of bunches of particles) circulate in the ring in opposite directions and are collided in four points, as indicated in Fig. 2.1(b), where the main particle detectors are placed:

- ATLAS (A Toroidal LHC ApparatuS) [2] - a general purpose detector designed to measure wide range of physics phenomena, especially the prediction of Higgs boson by the Standard Model

- CMS (Compact Muon Solenoid) [3] - a second general purpose detector, aimed at very similar types of measurements as ATLAS, but designed and constructed completely independently

- LHCb (Large Hadron Collider Beauty) [4] - experiment dedicated to study of B-hadrons and CP violation

- ALICE (A Large Ion Collider Experiment) [5] - an ion collision experiment (in addition to protons LHC is designed to also collide heavy ions) aiming to study strongly interacting matter at high energy collisions

In addition to those four, there are three smaller experiments in the LHC: TOTEM [6], MoEDAL [7] and LHCf [8], each of them designed for a very specialized and narrower research purpose.
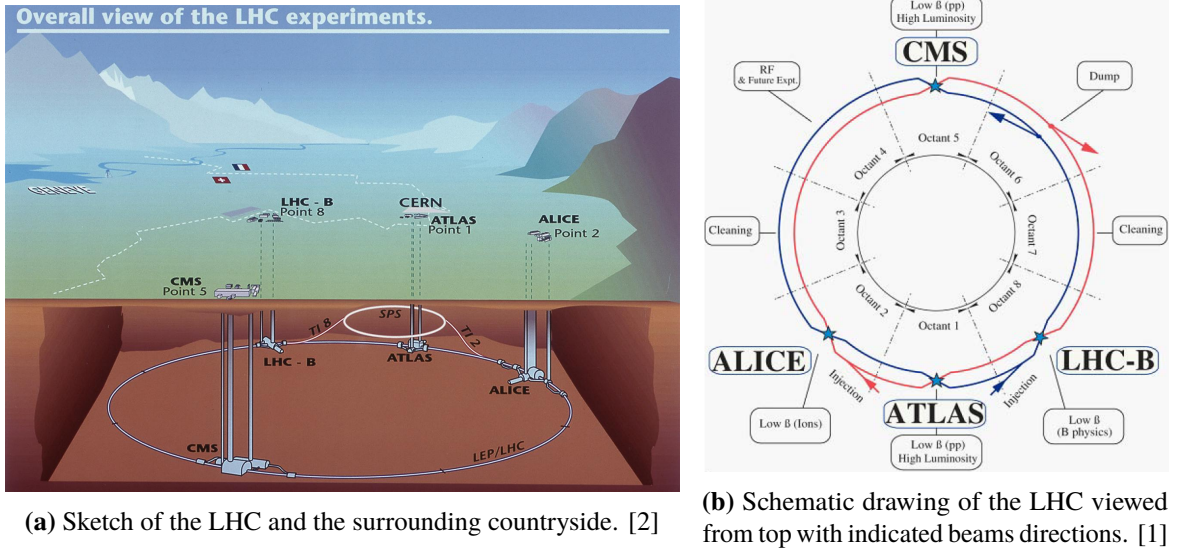
(a) Sketch of the LHC and the surrounding countryside. [2]

(b) Schematic drawing of the LHC viewed from top with indicated beams directions. [1]

Figure 2.1: Drawings of the LHC with indicated position of the main detectors.

Every time the particles collide a vast amount of different types of interactions can occur due to the statistical nature of the phenomenon. The frequency at which a given event type happens f_{event} can be calculated as:

$$f_{\text{event}} = L \cdot \sigma_{\text{event}} \tag{2.1}$$

where σ_{event} is the events cross-section and L is the beam luminosity given by:

$$L = \frac{n_{\text{bunch}} \cdot N_1 \cdot N_2 \cdot f_{\text{col}}}{A} \tag{2.2}$$

for a ring collider. Here n_{bunch} denotes number the of bunches per beam, N_i is the number of particles in the bunch of the i-th beam, f_{col} is the frequency of collisions and A is the cross-sectional area. The amount of statistics collected over a period ΔT can be expressed by the integrated luminosity $\mathcal{L} \equiv \int_{\Delta T} L dt$. While the event cross-section σ_{event} is a physical property and cannot be changed in an experiment, the luminosity clearly depends on the collider design and therefore its increase will ease the observation of rare events.

The LHC was originally designed for a peak luminosity of $10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$, but this number is being gradually increased as the accelerator is upgraded over the years, as indicated in Fig. 2.2. This allowed to collect approx. $190\,\text{fb}^{-1}$ of data, which already lead to observation of the existence of Higgs-like boson at the mass of $125\,\text{GeV}$. By then end of 2023 LHC is expected to collect $\mathcal{L} = 350\,\text{fb}^{-1}$, which will further increase the certainty of that discovery. Afterwards a very major upgrade is planned, called the High Luminosity upgrade, where the luminosity is expected to increase to approx. $7 \times 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$. The HL-LHC installation phase is currently scheduled for 2024 - 2026 and will involve major changes not only to the accelerator but also to the detectors, resulting in the end-of-life $\mathcal{L} \approx 4\,000\,\text{fb}^{-1}$.
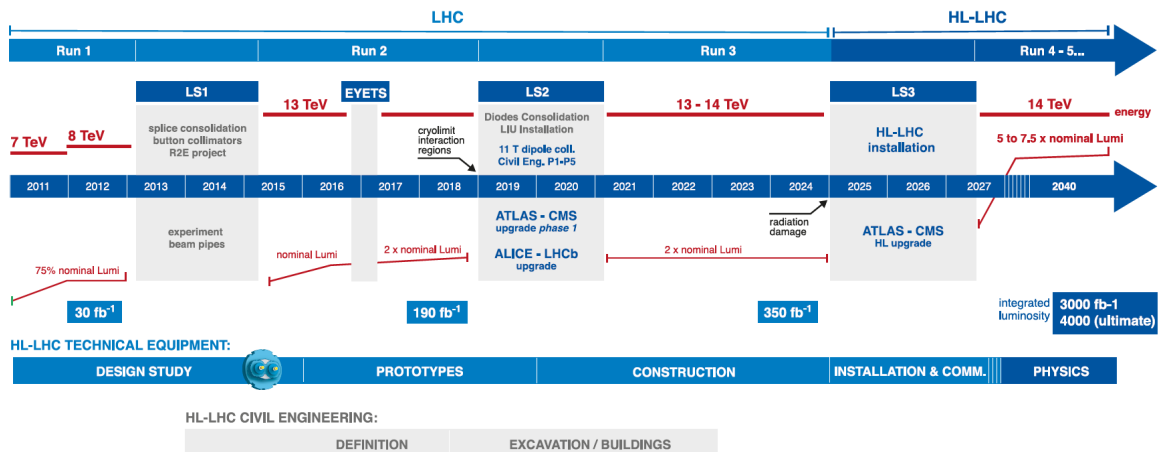
Figure 2.2: Roadmap of the LHC upgrade plan with indicated expected centre of mass collision energy, luminosity and integrated luminosity. [9]

.

2.2 The ATLAS detector

The particles created in the proton - proton collisions can scatter at any angle, therefore the detectors in High Energy Physics (HEP) experiments are often built as concentric devices around the beam pipe, with the collision point in the middle of the detector. ATLAS follows this principle and with its length of 46 m along the beam pipe and 26 m diameter, it has the largest detection volume of all the detectors installed in the LHC. As can be seen in Fig. 2.3 ATLAS is composed of several sub-detectors. Each of them allows identifying and characterizing different types of particles as illustrated in Fig. 2.4. Particles produced in the collision first traverse the tracking system, where charged particles leave electron-hole pairs along their path. The tracking system is composed of several thin detector layers in order to disturb the original particles as little as possible, while at the same time getting enough information to be able to precisely reconstruct the particle's path. The charge of the particle can be found by analysing the curvature of the reconstructed track, since there is a strong magnetic field inside the tracking system produced by the solenoid magnet. After passing through the tracker, the particles enter calorimeters, where most of them are stopped and their energy can be reconstructed based on the particle showers created inside the calorimeters. The outermost ATLAS detector is designed to capture the paths of muons, which pass essentially undisturbed though all previous layers. Neutrinos are invisible to the entire detector and are reconstructed based on undetected energy in the total energy of the observed events.

Since the amount of collisions seen by the detector is extremely high (the design luminosity is achieved by colliding 10^{11} protons every 25 ns) all the subsystems produce a tremendous amount of data, most of which is considered "not interesting" since the physical phenomena with high cross-sections were already extensively measured by other experiments in the past. For this reason the ATLAS detector is operated in triggered mode, where partial data produced after each collision is analysed in real time and a decision is taken whether a given event is potentially interesting or not. Only the promising ones are fully readout from all sub-systems and stored for offline analysis, while for all the rest the data is discarded. On average 1000 out of 1.7×10^9 events are saved per second [2],

which is a very significant benefit of the triggered mode of operation, but it comes at the cost of a further increase of complexity of the detector's sub-systems and of the off-detector data acquisition system (DAQ).
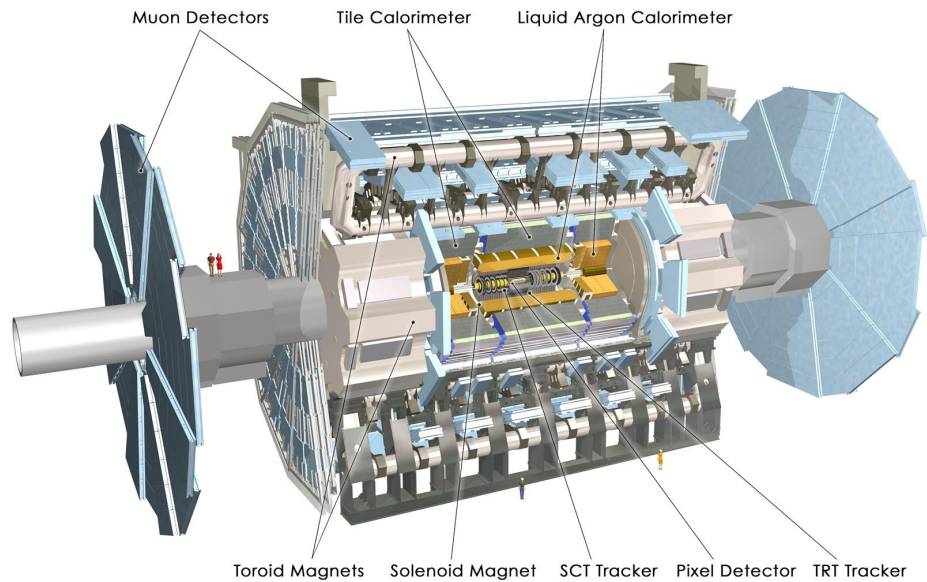

Figure 2.3: An image of the inside of the current ATLAS detector. [2]
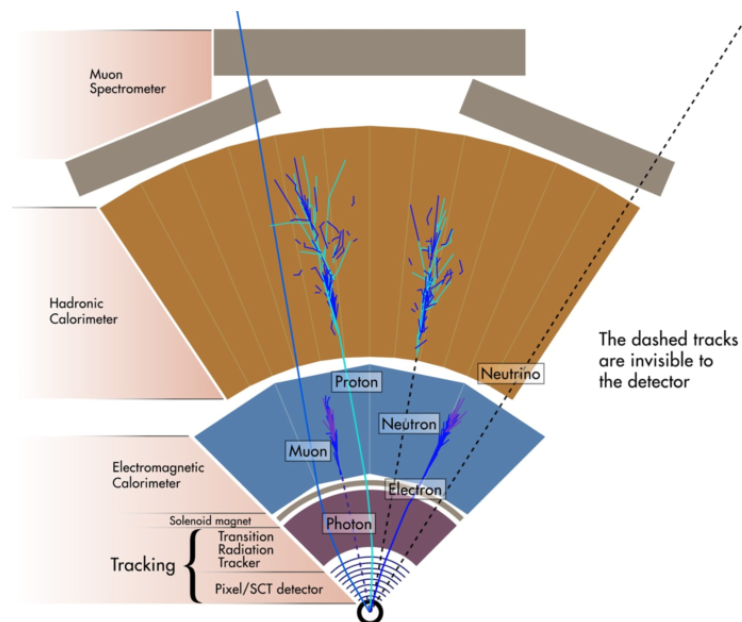

Figure 2.4: Transverse view of an octant of the present ATLAS detector with examples of particle interactions with different sub-systems. Adapted from [2].

2.2.1 ATLAS upgrade for the HL-LHC era

Data acquisition with ATLAS started in 2009 and the detector performed very well, by 2012 together with CMS confirming the observation of the Higgs boson. During Long Shutdown 1 of LHC most of ATLAS remained unaffected, only the innermost part (the pixel detector) was upgraded by installing an additional layer (Insertable B-Layer) in order to improve tracking performance and efficiency at luminosity twice higher than the LHC original design goal. However, in order to prepare for the operation in HL-LHC conditions a major upgrade of ATLAS will be carried out during the Long Shutdown 3 (referred to as "Phase-II upgrade" within ATLAS collaboration). This modernization effort will affect all sub-systems and DAQ in order to ensure that the detector can cope with the effects of increase of luminosity. Description of all the planned changes is far beyond the scope of this work, therefore focus will be given only to the part of the detector most relevant to this work which is the Inner Detector. Presently the Inner Detector (marked as "Tracking" in Fig. 2.4) consists of a silicon pixel detector (closest to the beam pipe), followed by a silicon microstrip detector (Semiconductor Tracker, SCT), which is in turn surrounded by a straw tube detector (Transition Radiation Tracker, TRT). For the Phase-II upgrade the whole Inner Detector will be replaced by so called Inner Tracker (ITk) built solely using silicon detectors. Some aspects of the ITk design are discussed below.

Coverage

Because the products of collisions can scatter in any direction an ideal detector would be a sphere around the proton - proton interaction point. This is impossible to realize due to geometrical and mechanical constrains, instead the detector is build as several concentric cylindrical layers ("barrel region") with discs on the sides ("endcaps") as visualized in Fig. 2.5. Layers of barrel and endcaps are build out of large number of relatively small (approx. $4\,\text{cm}^2$) sensors and readout chips.
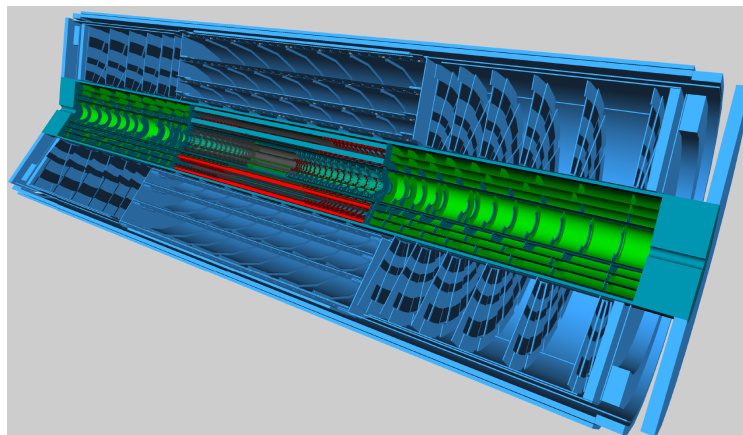

Figure 2.5: Visualization of the latest ATLAS ITk layout. The green and red sections represent the Pixel Detector, while the blue parts show the Strip Detector. [10]

Adjusting the arrangement of the sensors and spacing between layers allow changing the detector's effective coverage of the collision. A variable often used to describe the capability of a detector in

terms of coverage is pseudorapidity η defined as:

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \tag{2.3}$$

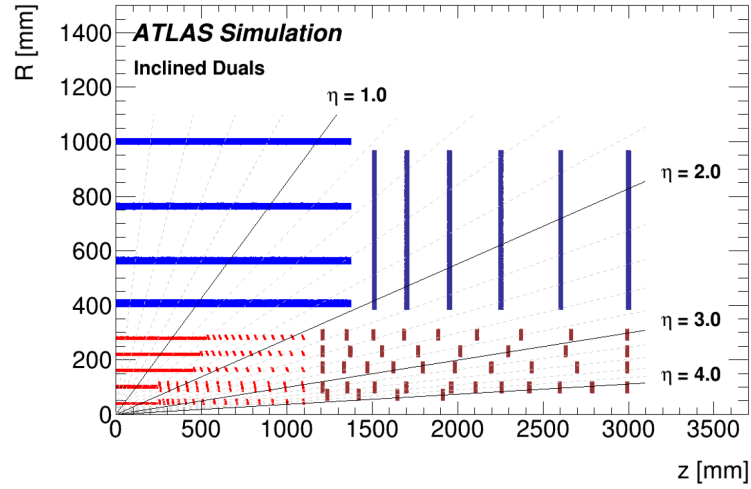where θ the angle with respect to the beam (parallel to the beam being at $0°$).

Figure 2.6: Schematic of the layout of the ATLAS IT, only a quarter of the detector is shown. The active elements are coloured blue for strip detectors and red for pixel detectors, with lighter colours indicating the barrel region and darker colours for the end-caps. The horizontal axis is the axis along the beam line with zero being the interaction point. The vertical axis is the radius measured from the interaction region. [10]

The ATLAS Inner Detector currently is located within $|\eta| < 2.5$ and the ITk upgrade aims to increase that to $|\eta| < 4$, i.e. much more forward. The detector design is not finalized yet, but the most up-to-date layout is presented in Fig. 2.6. In order to realize this layout the strip detectors will cover an area of $165\,\mathrm{m}^2$ and the pixel detectors will cover $13\,\mathrm{m}^2$, resulting in one of the biggest silicon sensor based detector in HEP.

Material budget

The main task of the tracking subsystem of ATLAS is the precise registration of the position where ionising particles pass through the detection layers. The obtained information is later on combined with data from other sub-systems in order to reconstruct the full paths of particles and determine their primary vertex. This requires that the whole tracking detector influences the particles' paths as little as possible, leading to much effort being put to minimize the amount of material in the particles' way. One commonly used metric to quantify such effects is expressing the material thickness normalized to the radiation length X_0. The radiation length is a material property and can be given as [11]:

$$X_0 \approx \frac{716.408 \left[\frac{g}{cm^2}\right]}{\rho} \cdot \frac{A}{Z(Z+1) \cdot \log\left(\frac{287}{\sqrt{Z}}\right)} \tag{2.4}$$

where ρ is the density, A is the mass number and Z is the atomic number. The radiation length corresponds to the mean distance over which a high-energy electron loses all but $\frac{1}{e} \approx 37\%$ of its energy by bremsstrahlung or alternatively it is $\frac{7}{9}$ of the mean free path for $e^- e^+$ pair production by a high-energy photon through photon conversion [11]. In addition to energy loss, a charged particle traversing material is subject to multiple scattering, where the interaction with Coulomb field of nuclei causes numerous small angle deviations of the path. Radiation length also plays a role, since the standard deviation of the scattering angle is proportional to $\sqrt{\frac{1}{X_0}}$ [12].

A comparison of material budgets of current and future tracking detectors is presented in Fig. 2.7, where it is clearly visible that a significant reduction is expected when the entire Inner Detector is upgraded to the ITk. This is of course beneficial from the point of view of measuring the effects of beam collisions, but it comes at a cost of very large amount of engineering work on all aspects of the detector. Few examples of modification needed in order to achieve this goal are:

- building lightweight and rigid support structures out of all-carbon foam

- reducing the size of the titanium cooling pipes and the amount of coolant needed by using a more efficient CO_2 cooling gas instead of C_3F_8

- significantly reducing the amount of copper power cables going to the pixel readout chip by using a new powering scheme ("serial powering")

- reducing the amount of cables going to the pixel readout chip by removing the need for dedicated input clock lines

The last point is especially important for this thesis, since it gives the motivation to the design of the clock-data recovery circuits described in Chapter 4.

(a) Material budget of ATLAS Inner Detector. [10]

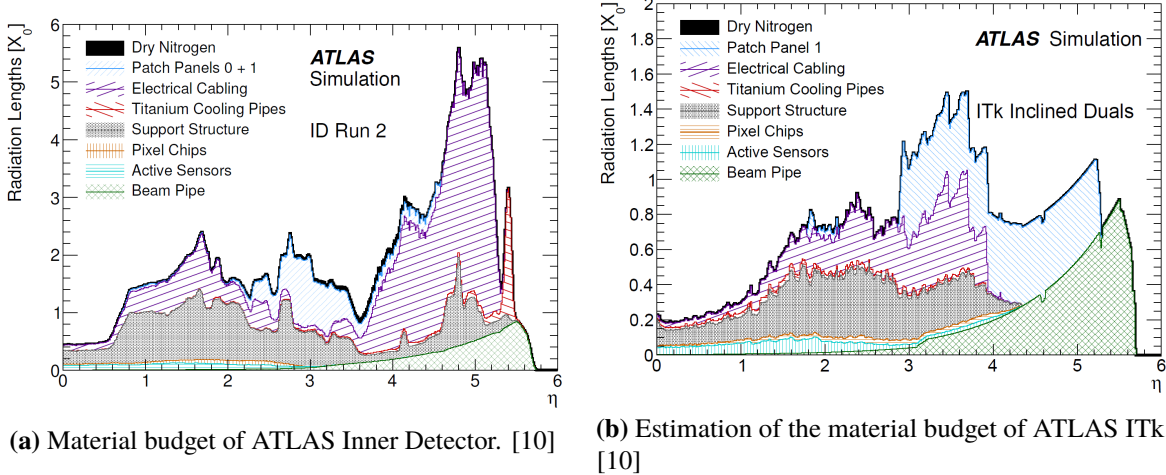**(b)** Estimation of the material budget of ATLAS ITk. [10]

Figure 2.7: Comparison of estimated material budgets of ATLAS' current Inner Detector and the planned Inner Tracker (ITk) expressed as radiation length X_0 versus pseudorapidity η. Only positive η is shown, since the values for negative η is identical. Note the different vertical scales on the plots.

Hit rate and data bandwidth

Hit rate is the number of elementary sensors (e.g. pixels for pixel detector) responding to particles passing through them per area and per time. It is one of the most important parameters guiding the design of the detector, because if the readout cannot cope with the hit rate during the collision a data loss will occur and in turn event reconstruction will get corrupted. The increase of luminosity in the HL-LHC era will lead to much higher hit rates, but also many other factors influence this number e.g. charge sharing among sensors, the area of the individual sensors, the capabilities of the readout or the amount of interactions of primary particles with detector materials. Therefore, the estimation of the expected hit rates are carried out based on system-level numerical simulations. Presently, the inner most layers of the ATLAS ID experience a hit rate of 50 kHz per pixel ($50{\times}250\,\mu\text{m}^2$) and this is expected to increase to 75 kHz per pixel ($50{\times}50\,\mu\text{m}^2$) for the innermost layer of the ITk.

One of the consequences of a higher hit rate is an increase in the amount of data generated by the sensor readout chips (this is also influenced by the volume of data produced by the chip for every hit, which will also increase for ITk with respect to ID). This leads to many technical complications in the design of such chips e.g. the amount of on-chip memory needed or the complexity of data processing logic, but of special interest for this thesis is the output bandwidth of the readout chips. This value will grow from 160 Mbps used currently to 5.12 Gbps (realized as four physically separate lanes running at 1.28 Gbps) for the ATLAS ITk. Because the downlink (connection from the control room of experiment to the readout chips) will run at 160 Mbps, the higher frequency clock needed to produce the output data stream will have to be created inside the chip itself. This is one of the tasks of the clock-data recovery circuits described in Chapter 4.

Radiation levels

By the very nature of the collider experiment there are high levels of radiation created inside the detector volume, which lead to a hostile environment for all the exposed devices. The damage caused by radiation-related effects is a significant concern for the design of all detector systems. Therefore, a lot of effort is put into predicting the radiation dose. Such simulations, however, require a precise modelling of the entire detector, since the secondary interactions with the device material account for a significant portion of particle fluences [10].

Radiation damage can be classified in several ways with three being most relevant to this work (the effects of radiation on silicon devices will be described in more detail in Chapter 3):

- Total Ionizing Dose (TID) - the measure of ionization of a material by charged particles or photons. The damage done is proportional to the energy absorbed during the ionization process. The unit used to quantify this effect is "Gy" (1 Gy is defined as the absorption of 1 J of energy by 1 kg of matter), however, in the field of radiation hard electronics it is still quite common to express TID in "rad" (1 rad = 0.01 Gy). The expected TID levels inside ITk detector are shown in Fig. 2.8(a).

- Single Event Effects (SEE) - a general term encompassing many effects in electronics circuits (transistor damage, bit flips, etc.) caused by charge created in ionization processes being collected by a sensitive node of a circuit. The ionization process leading to SEE is usually not caused directly by the particle passing the material, but rather by secondary effects connected to its interaction with the atoms. In the LHC environment it is expected that most SEE will be

caused by hadrons with energies above 20 MeV, therefore their fluences are carefully simulated as shown in Fig. 2.8(b).

- Non-Ionizing Energy Loss (NIEL) - the measure of damage caused by displacement of atoms inside the crystal lattice. The severity of this effect depends on the type and energy of the impinging particle, therefore NIEL is commonly expressed as equivalent damage caused by 1 MeV neutrons with the unit of n_{eq}/cm^2, as shown in the predictions for the ATLAS ITk in Fig. 2.8(c).
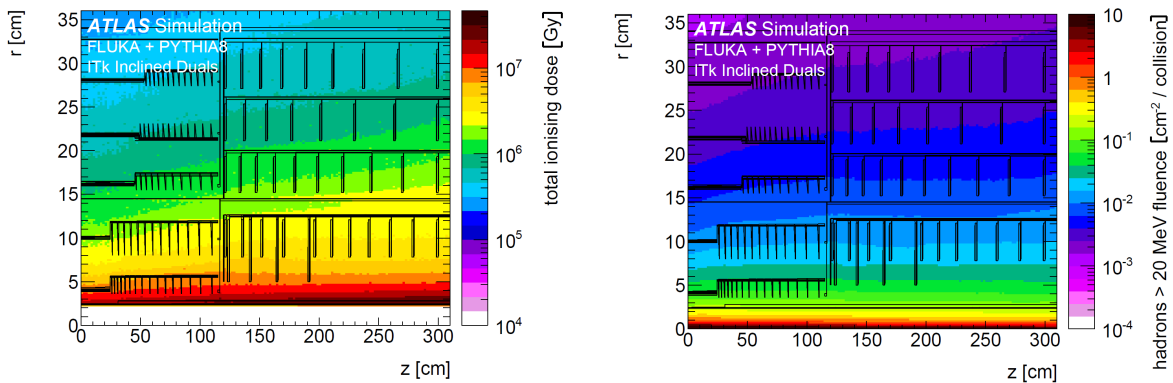
(a) Total ionizing dose normalized to $4\,000\,fb^{-1}$.

(b) Hadron fluence for energies above 20 MeV.
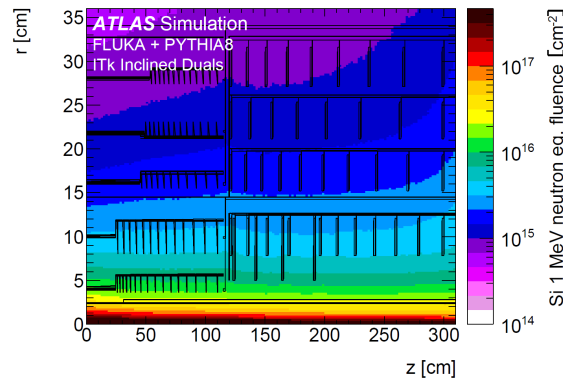
(c) 1 MeV neutron equivalent fluence normalized to $4\,000\,fb^{-1}$.

Figure 2.8: Simulation of the expected radiation environment of ATLAS ITk Pixel Detector. All figures visualize one quadrant of the detector, with the detector structures drawn with black lines. The results are based on the layout shown in Fig. 2.6 and as such should be considered preliminary, no safety factors are added. [10]

Comparison to other detectors

All detectors used in HEP experiments are tailor made for a specific task and for the particular environment created by their accelerators. This makes it difficult to compare them directly, however

one can juxtapose them by looking at a few universal characteristics as shown in Table 2.1. The electron - positron collisions measured by Belle II produce much fewer secondary particles compared to collisions done at LHC. Additionally the proton - proton collisions inside ATLAS have a much higher repetition rate than collisions at other experiments. Those two facts result in very harsh radiation environment of ATLAS. The high repetition rate also poses great challenges for the readout electronics, since fast data processing becomes mandatory. The end results is that achieving the ATLAS ITk upgrade specifications poses a difficult challenge, even when compared to other detectors.

Also worth noting is that the outer layer of ATLAS ITk has similar expected specification to the performance of current ATLAS ID. Based on this, one can assume that any new technology that can be shown to perform at least as good as the devices used presently in ATLAS ID could also be used for ITk. Such validation method is very valuable when developing new technical concepts, as for example the monolithic sensors described in Chapter 3.

| Specification | Belle II | ALICE (LHC) | ATLAS ID | ATLAS ITk | |
				Outer	Inner
Colliding particles	e^+e^-	heavy ions	p^+p^+	p^+p^+	p^+p^+
Time between collision [ns]	20 000	20 000	25	25	25
Hit rate [kHz/mm^2]	400	10	4×10^3	4×10^3	3×10^4
TID [Mrad]	1	0.7	80	50	> 500 [1]
NIEL [n$_{eq}$/cm^2]	10^{12}	>10^{13}	10^{15}	10^{15}	2×10^{16}

Table 2.1: Comparison of selected characteristics of silicon-based pixel detectors operating in HEP experiments.

---

[1] The TID at the end of life is expected to be approx. 1 Grad, but the two most inner layers of the ITk are designed to be replaced in the middle of detector's lifespan.

Depleted Monolithic Active Pixel Sensor

This chapter presents work done on Depleted Monolithic Active Pixel Sensor (DMAPS) in LFoundry 150 nm technology. Before getting into the design details selected general concepts related to pixel detectors will be presented in Section 3.1. This section also includes some details on radiation related effects in silicon. Next Section 3.2 will discuss the concept of monolithic detector and possible approaches to sensor implementation. Afterwards, in Section 3.3, the design prototype chips will described in detail together with measurement results. Finally, this chapter concludes with a short summary and outlook in Section 3.4.

3.1 Fundamentals of silicon pixel detectors

The task of the silicon pixel detector is to sense a particle passing through it and provide information about such an event (particle's hit position, deposited energy, time of hit, etc.) to the DAQ. As described in Section 2.2.1 ATLAS Pixel detectors consists of several layers, with each layer being constructed out of long staves build out of smaller (few cm^2) sensitive elements called modules, as shown in Fig. 3.1. For the rest of this chapter, whenever a pixel detector is mentioned it is referring to the modules.

A commonly used choice for a pixel design in HEP is the so-called hybrid architecture, where the particle detection part (sensor) and the data processing part (readout chip) are two separate entities connected e.g. with bump bonds, as shown in Fig. 3.2. In this work only the case where both parts are made out of silicon will be considered, but many different configurations are possible e.g. creating sensor out of diamond or using GaAs-based readout chips. An alternative approach to hybrid would be to combine the a sensor and a readout chip into one physical device - a realization of this this idea will be discussed further in Section 3.2, but the concepts described in this section apply to both types of designs.
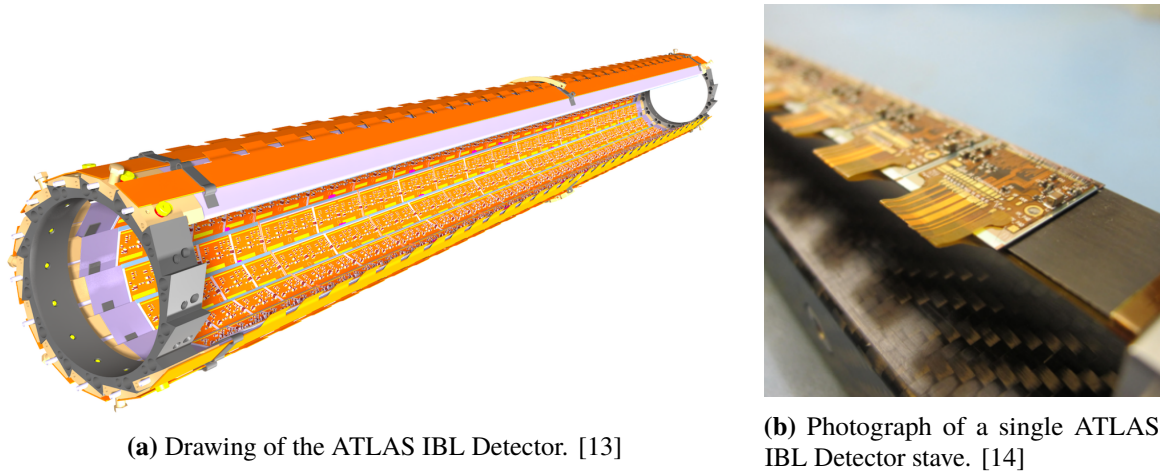
(a) Drawing of the ATLAS IBL Detector. [13]

(b) Photograph of a single ATLAS IBL Detector stave. [14]

Figure 3.1: ATLAS IBL Detector. The individual detector staves are mounted on carbon fibre support structures.

(a) Isometric view. As indicated here it is possible to connect several readout chips to one larger sensor. [15]

(b) Cross-sectional view of a single pixel showing the sensor connected to a pixel of the readout chip. [16]
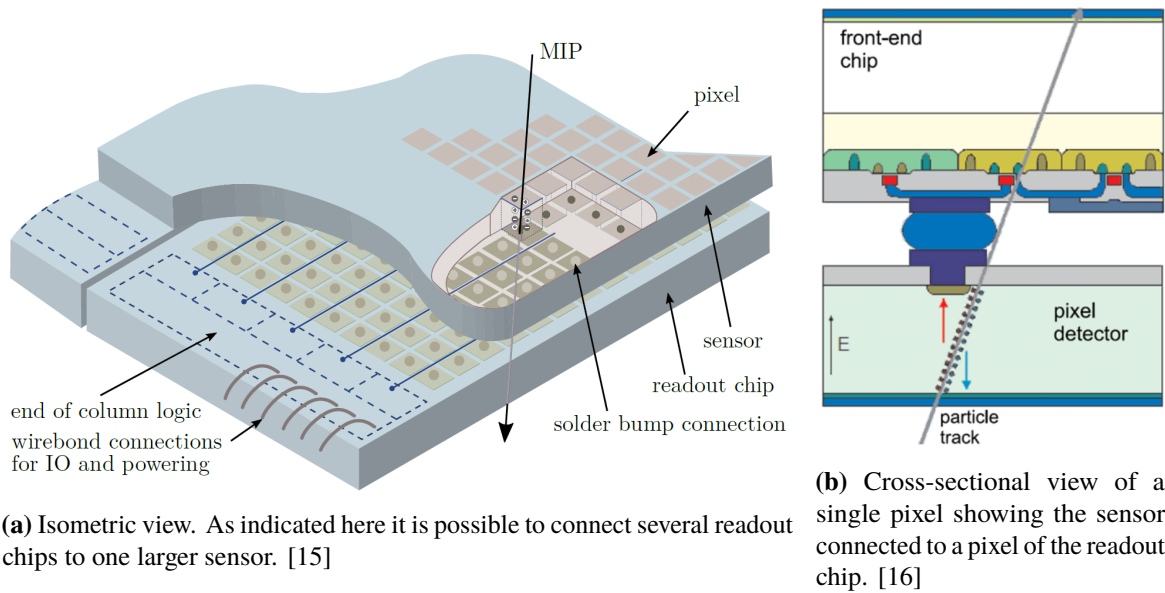
Figure 3.2: A drawing of a hybrid pixel detector hit by a minimum-ionizing particle (MIP).

3.1.1 Charge generation in a pixel detector

3.1.1.1 Charged particles

A particle passing through a material interacts with its atoms and in this process looses energy along the path. The amount of energy lost depends on the charge of the particle, its mass, energy and the traversed material in a rather complicated way (details can be found e.g. in [11]). For particles in the energy range of interest to this work the energy loss is mostly caused by inelastic scattering off the shell electrons of atoms. The average energy loss per unit of length and density of the traversed material $\left\langle -\frac{dE}{dx} \right\rangle$, also referred to as mass stopping power, is analytically described in this energy range by Bethe-Bloch formula [11] and is plotted in Fig. 3.3. A particle with $\beta\gamma \gtrsim 3$ is referred to as Minimum Ionizing Particle (MIP) as it reaches the minimum value of the mass stopping power.
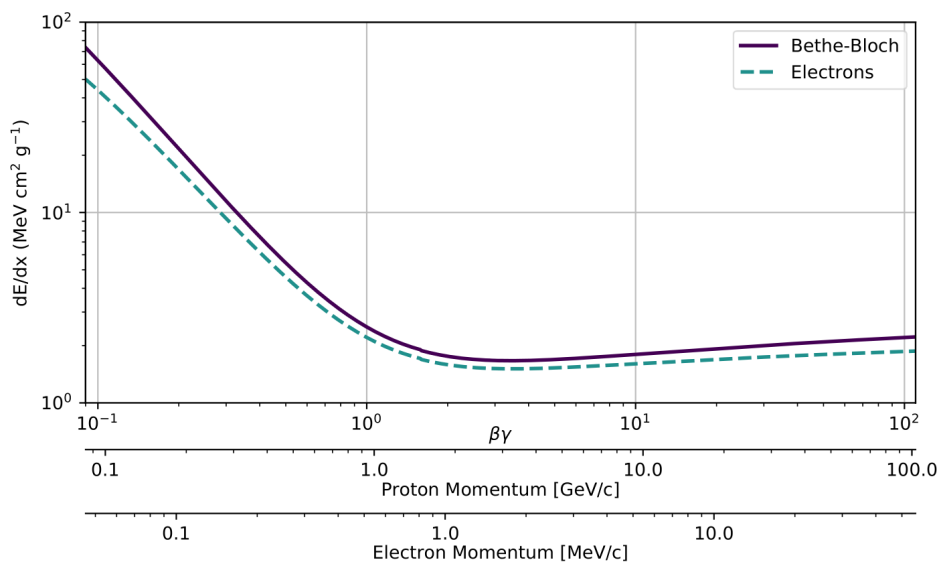

Figure 3.3: Mass stopping power in silicon as a function of $\beta\gamma = p/Mc$. Solid line is for particles heavier than electrons, dashed line indicate electrons (please note different horizontal momentum scales). [15]

The total energy loss due to the collisions along the particle's path is a sum of a series of discrete energy transfers, where the energy loss is subject to statistical fluctuations. For thin sensors of few hundreds of μm (such as used in ATLAS) an appropriate description is provided by a Langau function (a convolution of Landau and Gaussian distributions)[1]. Examples of such distribution for several widths of silicon sensor are presented in Fig. 3.4, with the Most Probable Value (MPV, the peak of the distribution) noted as Δ_p. The MPV is preferred over the mean value when describing energy deposition, since determining the average requires detailed sampling of the tail of the energy-loss distribution, which is not practical [15].

Each interaction causing the energy loss results in ionization or excitation of sensor atoms leading

---

[1] A Gaussian distribution would not provide a good description due to the distribution's long tail towards higher energy-losses caused by delta electrons produced in rare inelastic scattering events.
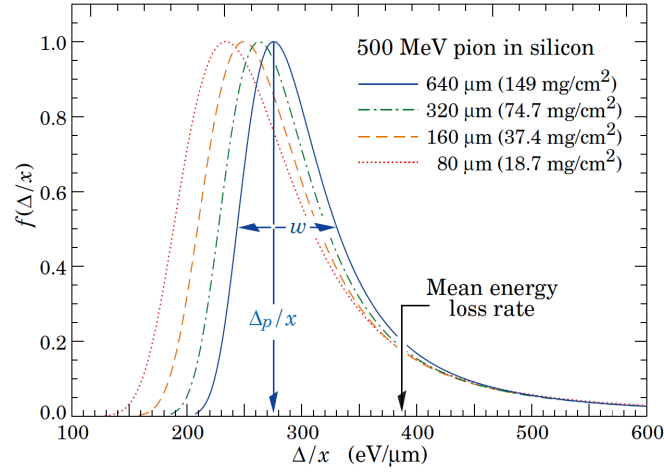
Figure 3.4: Energy loss distribution in silicon for 500 MeV pions, normalized to unity at the most probable value Δ_p/x. The w is the full width at half maximum. The horizontal axis is energy loss normalized by the mass thickness (product of the density of the material and its thickness)). [11]

to an electron-hole pair creation. Then number of electron-hole pairs created N_{eh} can be calculated as:

$$N_{eh} = \frac{\Delta_p}{E_{eh}} \qquad (3.1)$$

where E_{eh} is the energy needed to create one pair. For silicon $E_{eh} = 3.65$ eV, which combined with $\Delta_p = 250\,\mu\mathrm{m}$ (from Fig. 3.4 for $160\,\mu\mathrm{m}$ sensor) leads to an estimate of $N_{eh} \approx 70 \cdot d$, where d is sensor thickness in μm.

Since the higher amount of charge created is the easier it is to detect, as will be explained in Section 3.1.3. Eq. (3.1) could suggest that sensor should be made very thick. Such approach, however, has several disadvantages. One of them is the increase of the material budget which, among other issues for the detector system, would increase the multiple scattering effect as mentioned in Section 2.2.1. Secondly, the sensors used in an experiment are biased in order to deplete them of free charge carriers (more details are give in Section 3.1.2) and the voltage needed to fully deplete the sensor grows with its thickness, leading to practical issues with high voltage distribution. Lastly, after radiation induced damage the sensor will generate unwanted leakage current proportional to the volume of the sensor (more details are discussed in Section 3.1.5). All this means that the thickness of the silicon sensor is a very sensitive parameter, which has to be chosen as a compromise between the amount of charge created, capabilities of the readout electronics and the problems arising from the increase of material budget.

3.1.1.2 Photons

In contrast to charged particles, the energy of photons passing through material does not change, but due to the absorption of individual photons the intensity of the photon beam diminishes with the distance x travelled in the sensor:

$$I(x) = I_0 e^{-\frac{x}{\lambda}} \qquad (3.2)$$

where I_0 is initial intensity and λ is the mean free path. The result of photon interaction is mainly creation of low-energetic electrons, which then behave as described above. The attenuation coefficient, defined as $\frac{1}{\lambda}$, is dependent on the energy of the photon as illustrated in Fig. 3.5. As visible in the plot, the effective attenuation coefficient curve is defined by three processes, each dominant in different photon energy ranges (depending on the material, here values for silicon are given):

- Below 60 keV the photoelectric effect is dominant, where the photon is fully absorbed by an atom, leading to electron ejection (provided that the absorbed energy is greater than the work function). As the absorption coefficient value is high in this range, the photons get absorbed quite close to the surface (few tens of μm for energies below 10 keV [17]), which combined with precisely defined energy deposition makes sources of low energy gamma radiation very useful in testing silicon sensors.

- In the range of 60 keV - 10 MeV mostly the Compton scattering process is dominant, which leads to a partial transfer of the photon's energy to an electron. The amount of energy depends on the scattering angle, therefore energy deposition has a continuous spectrum.

- Above 10 MeV mainly pair production occurs. In this process the photon energy is converted in creation of an electron-hole pair in addition to causing recoil of a nearby nucleus. This phenomenon is likely to happen during the events observed by ATLAS, but it is considered undesirable since the photons are supposed to be stopped in the electromagnetic calorimeter, not in the tracking detector, thus causing difficulties for event reconstruction. Since the pair production cross-section is inversely proportional to the radiation length X_0 (as mentioned in Section 2.2.1), one way to reduce the frequency of this process is to limit as much as possible the material budget of pixel detector.
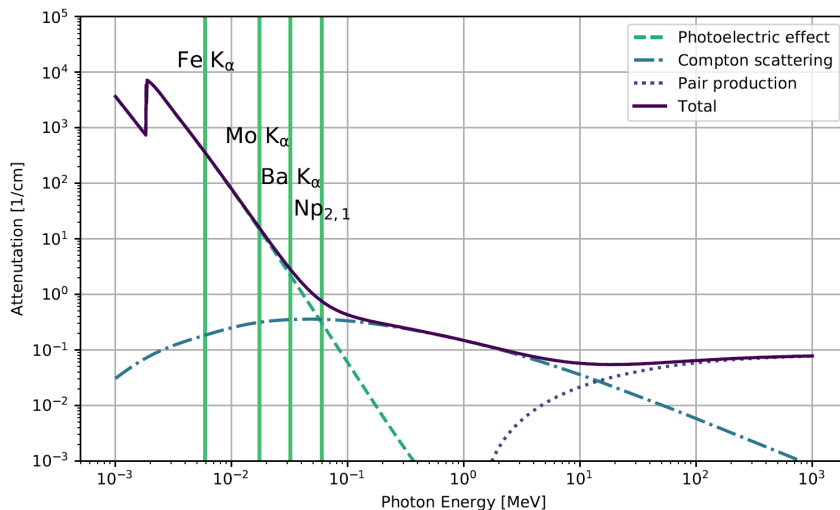

Figure 3.5: Attenuation factor for photons interacting with silicon plotted versus energy of the photon. Most probable photon energies of common gamma sources used for calibration of silicon sensors are depicted as vertical lines. [15]

3.1.2 The p-n junction as a detector

In order to adjust electrical properties of silicon impurities are introduced to the lattice in a process called doping. If elements from the third group of the periodic tables are used (acceptors) the end result is p-doped silicon with positive charge carriers (holes) present in the lattice. Alternatively, elements from fifth group can be used (donors) to create n-doped silicon with unbound electrons in the crystal. The concentration of the dopants is chosen depending on the intended application, but usually it ranges from 10^{12} cm^{-3} (for cases where high resistivity after doping is desired) to 10^{18} cm^{-3} (if high conductivity is important).

When n-type and p-type silicon are put together a p-n junction is formed. On the interface of the two zones the free charge carriers recombine and once thermal equilibrium is reached a zone devoid of free charge carriers, called depletion zone, is established. The now ionized dopant atoms present in the depletion zone form space charge, thus an electric field is created, which stops the diffusion of electrons and holes through the depletion zone. The potential difference over the space charge is called the built-in voltage V_{bi} and for silicon its typical value is approx. 0.6 V. An example of a p-n junction with p-doping slightly higher than n-doping, resulting in asymmetric depletion zone, is shown in Fig. 3.6.

The depletion zone in configuration described so far is quite narrow, but it can be widened by applying an external reverse bias voltage V_{ext} (positive at n-doped side, negative at p-doped side), resulting in the width of depletion zone W [19]:

$$W = \sqrt{\frac{2\epsilon_0 \epsilon_{Si}}{q_e} \left(\frac{1}{N_A} + \frac{1}{N_D} \right) (V_{bi} + V_{ext})} \qquad (3.3)$$

where ϵ_0 is the vacuum permittivity, ϵ_{Si} is the relative silicon permittivity, q_e is the elementary charge and N_A (N_D) is the doping concentration of acceptors (donors). For p-type silicon bulk and $V_{ext} \gg V_{bi}$ Eq. (3.3) can be approximated as:

$$W \approx 0.3 \cdot \sqrt{\rho \cdot V_{ext}} \qquad (3.4)$$

where ρ is the bulk resistivity. In a realistic case the concentration level asymmetry between n- and p-doping might be several orders of magnitude e.g. the structures described in Section 3.3 use lightly doped p-type bulk and highly doped n-type implants (readout electrodes) resulting in depletion zone extending mostly into the sensor bulk. The electric field inside the depletion zone has a triangular shape, as shown in Fig. 3.6, and is described as [20]:

$$E(x) = \begin{cases} \frac{2(V_{ext}+V_{bi})}{W} \left(1 - \frac{x}{W} \right) & , V_{ext} \leqslant V_{dep} \\ \frac{2(V_{ext}+V_{bi})}{d} \left(1 - \frac{x}{d} \right) + \frac{V_{ext}-V_{dep}-V_{bi}}{d} & , V_{ext} > V_{dep} \end{cases} \qquad (3.5)$$

where the full depletion voltage $V_{dep} = \frac{q_e N_A W^2}{2\epsilon_0 \epsilon_{Si}} - V_{bi}$ (voltage required to deplete the entire detector volume).
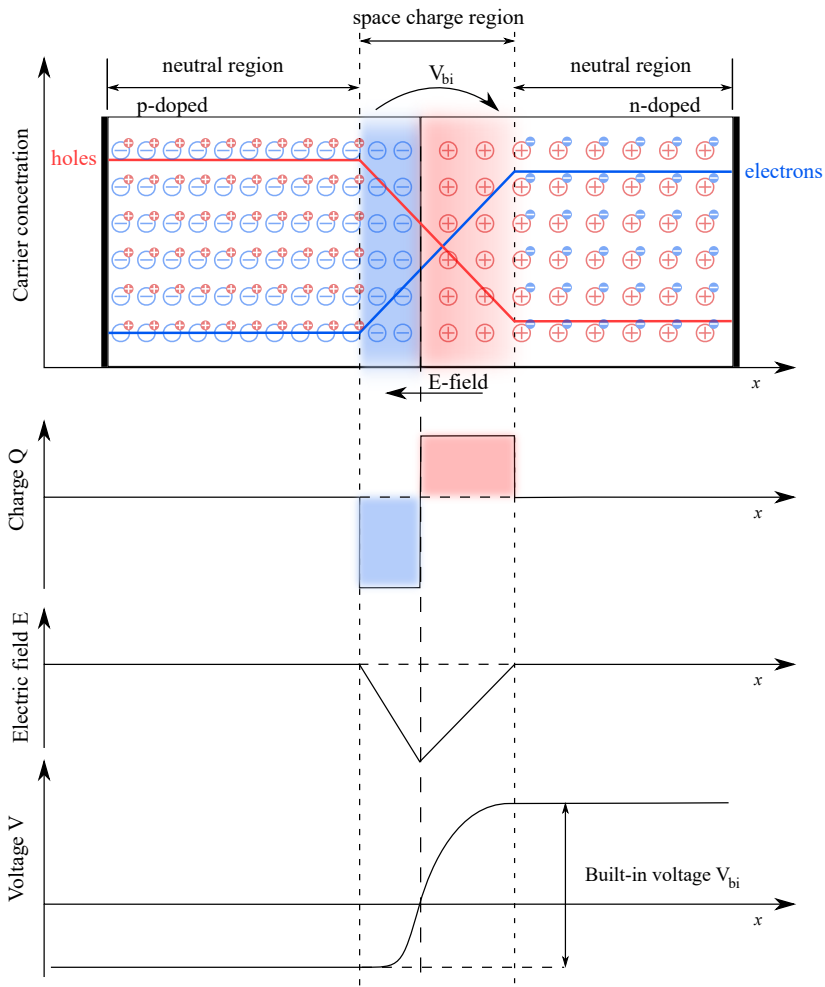
Figure 3.6: A p–n junction in thermal equilibrium with zero-bias voltage applied. Based on [18]
.

As described in Section 3.1.1 a charged particle traversing silicon sensor generates electron-hole pairs along its path. The electric field present in the depletion zone causes those charge carriers to drift to the electrodes, as shown in Fig. 3.7, inducing in them an electric current i_{ind} described by the Shockley - Ramo theorem [21, 22]:

$$i_{ind} = \sum_i^N q_i \overrightarrow{v_i} \cdot \overrightarrow{E_w} \tag{3.6}$$

where N is the number of charge carriers, q_i is the charge of a carrier, $\overrightarrow{v}$ is its velocity and $\overrightarrow{E_w}$ is the weighting field. The velocity is defined as:

$$\overrightarrow{v} = \mu(E)\overrightarrow{E} \tag{3.7}$$

where $\mu(E)$ is the mobility of a carrier, which is dependant on the strength of electric field $\overrightarrow{E}$, as illustrated in Fig. 3.8 (doping concentration also has an effect on the mobility, but for bulk resistivities above $10\,\Omega\,cm$ which are of concern for this work, this effect is negligible [23]). The $\overrightarrow{E_w}$ is not the same as the electric field $\overrightarrow{E}$ inside the sensor, it rather describes the coupling of the created charge inside the sensor to a given electrode, following the Gauss law. The weighting field is calculated by applying a unit potential to the electrode under consideration and zero potential to all others. In the simplest case of a parallel plate configuration with large electrodes placed d apart, the weighting field is $\frac{1}{d}$ and perpendicular to the electrodes, but it gets more complicated for realistic configurations.

A wider depletion zone translates into more deposited charge (typically an average of $80\,e^-/\mu m$ is assumed) and therefore a larger signal induced in the readout electrode. From Eq. (3.4) it is clear that in order to increase the depletion zone width both the resistivity of the bulk and the bias voltage should be increased. Additionally, Eq. (3.5) shows that overdepleting the sensor (using bias voltage higher than needed to deplete it fully) results in uniform increase of the electric field inside the bulk, which according to Eq. (3.7) leads to faster charge collection. While the resistivity is defined based only on the materials available from the vendors and cannot be improved by designers, the maximal applicable bias voltage is defined by the breakdown voltage of the sensor, which can be enhanced by design. The breakdown voltage of the p-n junction is affected by the sensor implant geometry and the design of the guardrings surrounding the pixel matrix. More details can be found e.g. in [23, 25].

The total charge Q_{tot} received by the readout electrode is:

$$Q_{tot} = \int_0^{t_{col}} i_{ind} dt \tag{3.8}$$

where t_{col} is the time needed to collect all the charge carriers. Since pixel detectors are by definition segmented and the size of pixels relevant to this work is small (few tens to few hundreds of μm in both directions), the charge carriers created by ionizing particle can be collected by more than one electrode, depending on the position where the particle hit the sensor and the tilt of its trajectory. This phenomena is called "charge sharing" and results in smaller signal induced in the readout electronics of neighbour pixels leading to many pixels responding to the same particle (increase in hit rate). While this effect is unavoidable, it can be partially mitigated by adjusting the readout electronics parameters (discussed in Section 3.1.3) or reducing the sensor thickness.
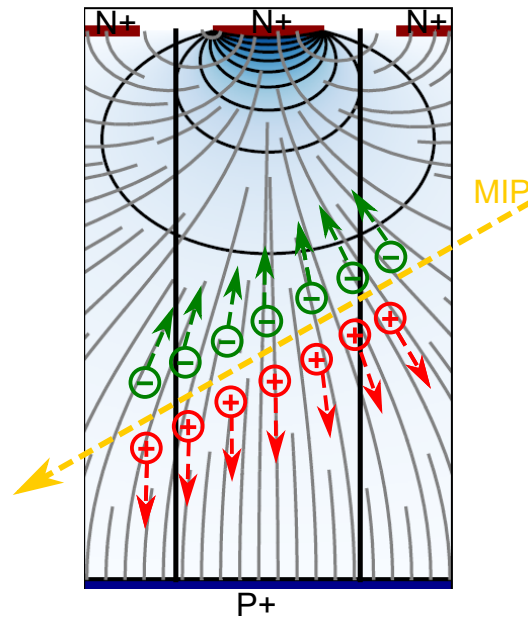
Figure 3.7: Cross-section view of a silicon pixel detector with a passing ionizing particle creating electron-hole pairs. Small n+ implants on the top side represent readout electrodes of individual pixels. Grey lines visualize the electric field, black elliptical lines are equipotential surfaces of the weighting potential. Vertical black lines indicate the size of a single pixel. Adapted from [15].
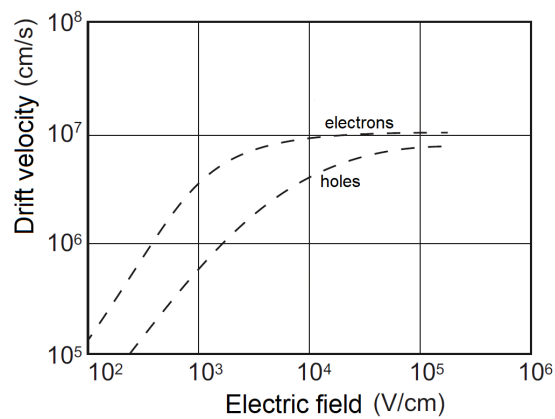
Figure 3.8: Velocity saturation of charge carriers versus electric field strength in silicon. Adapted from [24]

3.1.3 Basics of pixel readout electronics

The current induced in the readout electrode has to be received by a readout circuit in order to shape the signal according to the needs of a given application. Such circuits can vary from very simple ones (e.g using just three transistors [26]) to very complex ones with multi-stage analogue shaping and filtering combined with extensive digital processing (examples from HEP applications can be found in [27] or [28]). The choice of the readout circuit depends on many factors such as available area for electronics, technology used, allowable power consumption, required signal processing speed, etc. In order to fulfil those requirement and HEP specific needs related to the radiation environment, it is common that the readout circuit is developed from scratch for each experiment, rather than using commercially available chips. Such custom designs are referred to as Application Specific Integrated Circuits (ASIC).

This section focuses on a readout chain architecture which is often used for hybrid pixel detectors and is also a basis for the devices described in Section 3.3. Descriptions provided here will be kept rather general, since the goal of this section is to establish terminology and metrics which will be used later on in this thesis when discussing circuit implementation details and measurement results. This selection of circuit aspects is by no means exhaustive, rather only few, most relevant, aspects of the readout chain will be discussed, as this topic is extremely broad.

A generic model of a readout chain is presented in Fig. 3.9 and an illustration of the most important signal waveforms as a response to traversing particles is depicted in Fig. 3.10. In the following paragraphs each sub-circuit will be discussed in more detail, but the overall circuit behaviour can be described as follows: the sensor (depicted as a biased diode with a parallel capacitor C_d) produces a current pulse as a result of the interaction with the traversing particles. A Charge Sensitive Amplifier (CSA) converts the current pulse into a voltage output corresponding to the integral of the current, the charge. The voltage signal is shaped and amplified as needed, in order to best fit application requirements. The output of the CSA is then compared to a pre-defined threshold voltage V_{th} by a comparator circuit (commonly also referred to as a discriminator) in order to determine if the observed signal from the sensor is large enough to be read out. The digital output signal of the comparator is then processed by the digital logic in order to the extract required information e.g. the pulse width. It is important to note that not all elements of such a readout chain have to be placed inside the pixel or in the chip periphery, e.g. the chip described in Section 3.3.1 has no digital logic inside the pixel, instead relies on a second chip for digitalization.
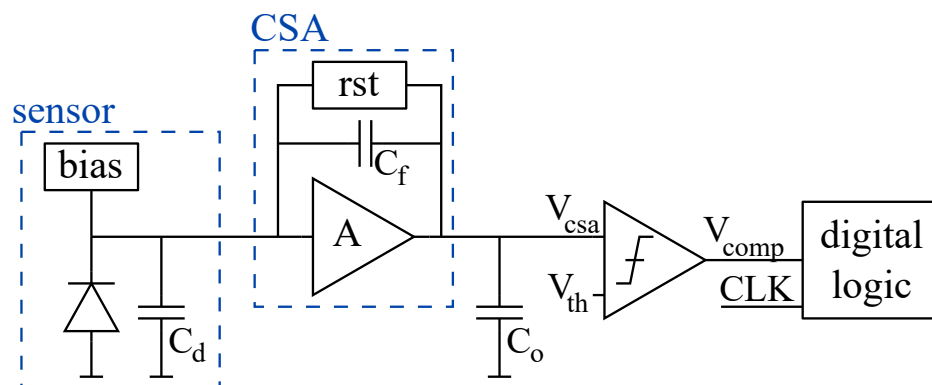

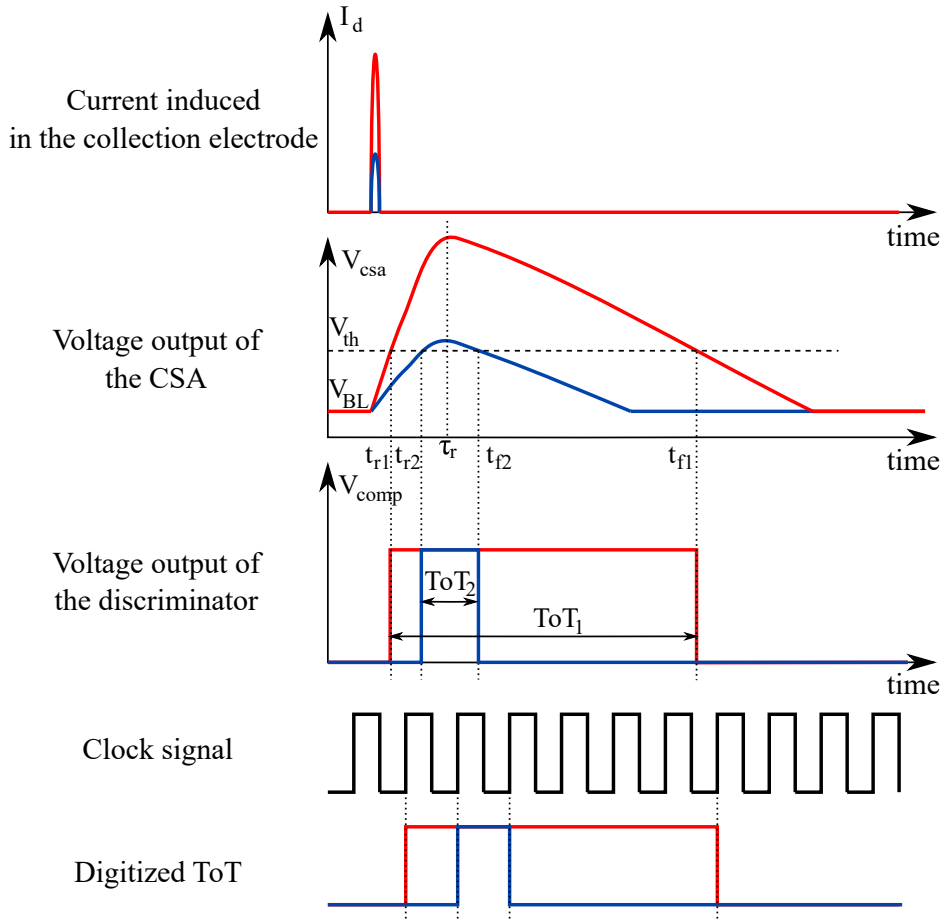**Figure 3.9:** A model of a readout electronics chain for a single sensor.

Figure 3.10: Drawings of signal waveforms of an in-pixel readout as in Fig. 3.9 for larger (red traces) and smaller (blue traces) energy depositions. In this example the reset circuit of the CSA is a constant current source.

3.1.3.1 Charge Sensitive Amplifier

The CSA is composed of a voltage amplifier with a feedback loop, made out of a capacitor C_f and a reset circuit. The capacitor C_o in Fig. 3.9 depicts the total output capacitance of the CSA. The CSA integrates the charge Q coming from the sensor and outputs a voltage V_{csa} following, in an ideal case, Eq. (3.9). The term $\frac{1}{C_f}$ is therefore the ideal gain value A_{ideal} and depends only on the value of the capacitor C_f in the feedback loop of the amplifier.

$$V_{csa} = \frac{Q}{C_f} \tag{3.9}$$

In a more realistic case the gain is influenced also by the detector capacitance C_d and a finite open

loop gain of the voltage amplifier A_{csa}, resulting in the CSA output V_{csa} being:

$$V_{csa} = \frac{Q}{C_f + \frac{C_d + C_f}{A_{csa}}} \tag{3.10}$$

Eq. (3.10) indicates that in order to have a large (easier to further processing) signal a high CSA open loop gain and low detector capacitance are desired.

The rise time τ_r of the CSA's output can be approximated as shown in Eq. (3.11) (g_m is the transconductance of the input transistor of the voltage amplifier) [29]. From this equation several interesting properties of the CSA's rise time can be noticed. Firstly, the rise time depends only on the circuit properties, not on the amount of charge collected by the sensor. Secondly, in order to achieve fast rise time the detector capacitance should be minimized, while the input transistor should have a large transconductance. The feedback capacitance has to be chosen appropriately, depending on the size of C_o (dependant on e.g. the comparator architecture and parasitic capacitances of the metal lines) and its impact on τ_r. All those parameters have to be carefully optimized in order to strike a good balance between circuit performance and other specification like physical size of circuit or the power consumption.

$$\tau_r = \begin{cases} \frac{C_o C_d}{g_m C_f} & : C_f \ll C_o \\ \frac{C_d}{g_m} & : C_f \gg C_o \end{cases} \tag{3.11}$$

The reset circuit of the CSA is necessary in order to avoid saturation after detecting few particles and the type of reset circuit used defines the shape of the CSA output waveform. The fall time of the CSA output (or its return to the baseline voltage V_{BL}) is yet another parameter, which has to be carefully optimized for a given application in order to cope with the expected particle hit rate. A few typically used feedback implementations are presented in Fig. 3.11 together with sketches of expected waveforms. In case of the resistive feedback (Fig. 3.11(a)) the signal after reaching a maximum falls exponentially with the time constant of $\tau_f = R_f C_f$. Since the feedback capacitance C_f has a strong influence on gain and rise time of the CSA, the possible values are strongly restricted. Because the typical requirement for the fall time is in the order of microseconds or below, and the feedback capacitance is often small (~fF), this would result in a very large R_f (many MΩ). Implementing very large resistance values in a modern CMOS technology is possible, but usually requires a large area and the absolute resistance value is prone to variations due to process non-idealities. Therefore this approach is problematic. Instead of a resistor the reset circuit can be implemented by a switch, as shown in Fig. 3.11(b). This can be an attractive option, since it allows to obtain very fast rise and fall times, however, it comes at the cost of more complex circuitry required to properly time the reset action. The third option is to use a constant current source in the feedback loop as shown in Fig. 3.11(c). Implementation of a current source is relatively easy and the constant current discharges the feedback capacitor linearly, therefore enabling the usage of the Time over Threshold (ToT) technique (discussed in Section 3.1.3.2) to measure the CSA signal amplitude. Each of the three mentioned architectures are viable approaches and the choice between them depends on the requirements of the whole readout chain. For the designs described in this thesis the constant current reset mechanism was chosen, as will be explained in more detail in Section 3.3.

For the reset circuit implementation with either a resistor or a current source it is important to consider the ballistic deficit effect. This effect is caused by the fact that already during the rise of the
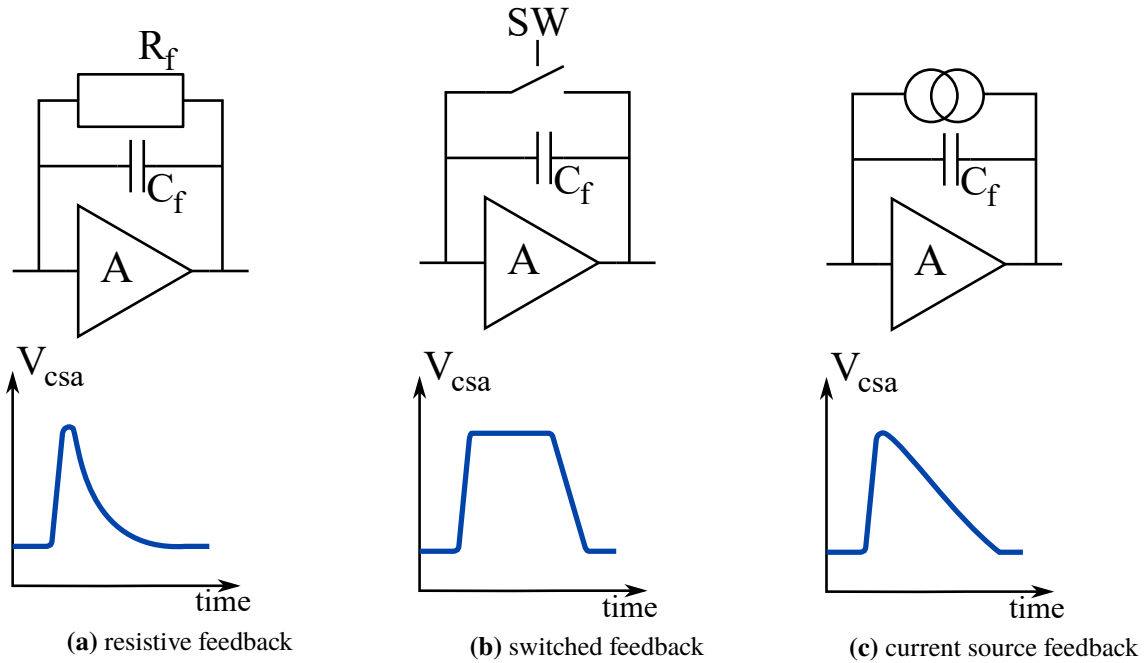
(a) resistive feedback
      (b) switched feedback
      (c) current source feedback

Figure 3.11: Illustration of three possible implementations of the reset circuit of CSA.

CSA output signal the reset mechanism works and discharges the feedback capacitor, thus reducing the overall gain of the CSA. For this reason the ratio of τ_r to τ_f has to be chosen such that the ballistic deficit is acceptable, usually leading to $\tau_r \ll \tau_f$.

For all electronic circuits, especially for sensitive analogue devices such as a CSA, noise is a very significant issue. Noise is an unwanted component added on top of all electrical signals. It can have many sources, some of which are due to physical phenomena inside elementary electronic devices such as transistors and resistors, some are due to "system-level" issue e.g. variation of power supply voltage, crosstalk between metal lines due to parasitic coupling, coupling between wells through substrate etc. Noise can be minimised through careful design of the circuit and its testbench, however it can never be completely eliminated. Several types of noise are distinguished and their statistical behaviour in time and frequency domains are known, which allows to model them and optimize circuits in order to minimize the unwanted effects. Common way to quantify noise is to express it in Equivalent Noise Charge (ENC). The ENC is equal to the charge fluctuation at the input of a CSA which results in the noise voltage at the output of CSA equal to noise of the real circuit. In the case of a CSA the total noise ENC_{CSA} is a quadratic sum of three main noise components:

$$ENC_{CSA} = \sqrt{ENC_{therm}^2 + ENC_{shot}^2 + ENC_{1/f}^2} \tag{3.12}$$

Where the contributions are [19]:

- ENC_{therm} - thermal noise: caused by interactions of the electrical current with atoms in the crystal lattice. Those interaction lead to random movements of the electrons, which in turn changes the voltages at terminals of transistors and resistors. For transistor channel the thermal

noise is (k_B - Boltzman constant, T - temperature):

$$ENC_{therm} = \sqrt{\frac{2}{3} \frac{k_B T}{q_e^2} \frac{C_d C_f}{C_o}} \tag{3.13}$$

- ENC_{shot} - shot noise: result of statistical charge carrier fluctuation occurring when they are emitted independently of each other over a potential barrier, e.g. pn junction. For pixel detectors the source of this noise is the leakage current of a sensor. While the sensor leakage current I_{leak} is usually rather small for undamaged devices, the harsh radiation environment inside HEP detectors damages the sensor's lattice structure over time, potentially leading to a significant increase of the shot noise.

$$ENC_{shot} = \sqrt{\frac{I_{leak}\tau_f}{2q_e}} \tag{3.14}$$

- $ENC_{1/f}$ - $\frac{1}{f}$ noise (also refereed to as flicker noise): it is cause by charge carriers being trapped and released randomly at the channel-gate boundary of the transistor. The number of traps depends on the technology and therefore cannot be analytically predicted for a general case. The noise level can be approximated as (K_f - technology dependent constant, C_{ox} - gate oxide capacitance, W - width of the transistor, L - length of the transistor):

$$ENC_{\frac{1}{f}} \approx \frac{C_d}{q_e} \sqrt{\frac{K_f}{C_{ox}WL}} \sqrt{\ln\left(\tau_f g_m \frac{C_f}{C_o C_d}\right)} \tag{3.15}$$

From the Eq. (3.15) it is clear that the area of transistor influence the flicker noise and by increasing the transistor size the noise can be reduced. Additionally, if we consider that the flicker noise of a transistor is modelled by a voltage source connected in series to the transistor gate, where the root mean square value of the voltage source $< V_{\frac{1}{f}}^2 >$ is:

$$\frac{d}{df}\langle V_{\frac{1}{f}}^2 \rangle = \frac{K_f}{C_{ox}WL} \frac{1}{f} \tag{3.16}$$

it becomes clear that the main contribution from this noise originates from low frequencies.

For pixel detectors before irradiation the main contributor to ENC_{CSA} is the thermal noise from the transistor channel, however, after heavy irradiation the sensor leakage might increase to levels high enough, such that shot noise dominates.

3.1.3.2 Comparator

As mentioned in the beginning of Section 3.1.3 the output voltage produced by the CSA is compared to a threshold voltage V_{th} at the comparator. The comparator produces a digital signal, as visible in Fig. 3.10, and if the CSA uses constant current feedback, the width of the comparator's output is directly proportional to the CSA maximal amplitude and therefore also to the charge seen at the sensor electrode due to the traversing particle. This method of measuring the produced charge this way is called Time over Threshold (ToT). The threshold voltage has to be set high enough that no random hits

(caused e.g. by noise) trigger the comparator, but not so high that interesting events are missed. The minimum threshold should be set to fit the needs of a given application, e.g. the noise occupancy and the used sensor. Since the charge produced by the sensor has a distribution as shown in Fig. 3.4 and since the charge is often shared between several pixels (e.g. if the particle crosses between pixels), the threshold voltage should be set much below the MPV value of a given sensor. An additional difficulty with finding a proper threshold for a pixel chip is the fact that due to manufacturing non-idealities in the CMOS process each comparator inside a matrix will have a slightly different threshold. A method of compensating for such dispersion will be described in Section 3.3. When comparing the waveform for small and large charge deposit in Fig. 3.10 it is clearly visible that the discriminator responds faster to the larger CSA signal ($t_{r1} < t_{r2}$). The fluctuation of the moment in time when the charge injection is recognized is called "time walk" and it is an undesirable phenomenon caused by the rise time of the CSA, comparator trigger delay and slower response of the discriminator to charges that are barely crossing the threshold. The time walk value can be decreased, e.g. by decreasing the response time of both CSA and comparator (often leading to a power consumption increase) or by circuit architecture changes. Therefore, the time walk becomes yet another parameter which has to be optimized for in the readout chain.

3.1.3.3 Digital processing logic

The comparator output signal is processed by the digital logic, as depicted in Fig. 3.9. The extent and complexity of this processing strongly depends on the application needs such as the type of information required from each event, requisite readout speed or physical size limitation of the circuitry inside pixel. A very simple approach is to store only the information about an event occurring (comparator output going high) at some point in time since the last readout of a given pixel. An example of an implementation of such a binary (hit/no hit) readout system is presented in Section 3.3.1 (stand-alone operation mode). Many HEP experiments, including ATLAS, require additional information about each event in order to correctly identify and interpret results of particle collisions. Such information might be e.g. energy deposited in the sensor or a time stamp assigning the event to a particular bunch collision (so called "bunch crossing ID"). The circuitry required for those functions can be placed either in the pixel itself (as in the design described in Section 3.3.3), in the chip periphery outside the pixel matrix (as implemented in few column flavours in Section 3.3.3) or even outside of the chip in another ASIC (design in Section 3.3.1 in the coupled operation mode) or FPGA.

Once the comparator output signal processing is done, the obtained information is extracted from the chip. Many readout architectures exist and the choice of an optimal one is guided by criteria such as the expected hit occupancy (number of hits per time and area), the bunch crossing scheme or the acceptable event loss. In the LHC environment, due to the 40 MHz bunch crossing frequency and the high hit occupancy (order of several MHz/mm^2 for LHC and up to few GHz/mm^2 for HL-LHC) simple architectures like rolling shutter (reading all pixels in the matrix row by row) or shift register based readout (Section 3.3.1) would not be able to cope with the demands, therefore more sophisticated readout methods are employed e.g. "column drain" [30] or "distributed latency counters" [27]. A common concept in those architectures is not to read all pixels, but only those which were hit. This allows for significant reduction of the amount of data needed to be sent out of a chip, which can be further improved by performing data compression. The current state-of-the-art in the HEP community in regard to on-chip data processing is the RD53B chip [31].

3.1.4 CMOS technology

All chips presented in the following chapters were designed in CMOS technologies (LFoundry 150 nm for Section 3.3 and TSMC 65 nm for Chapter 4). The technological process steps needed to produce CMOS chips are very complicated, varying between vendors and process nodes, and will therefore not be described here (details can be found in [32]). In extreme oversimplification the CMOS process starts with a silicon wafer into which acceptor and donor atoms are implanted and defused in patterns corresponding to desired shapes and sizes of transistors. In the next steps the poly silicon gates are deposited in the appropriate places between sources and drains of the transistors. Once the transistors are made, their terminals are connected in the desired way by metal layers (horizontal lines and vertical vias) as shown in Fig. 3.12(b). As visible in the structure of devices in Fig. 3.12(a), even within a given technology several types of transistors are available, usually differing in the maximal gate voltage they can sustain (regulated by the gate oxide thickness), the threshold voltage (regulated by doping levels within the channel region) and optional isolation for the wafer bulk (obtained by creating doped regions around and under transistors, so called wells).
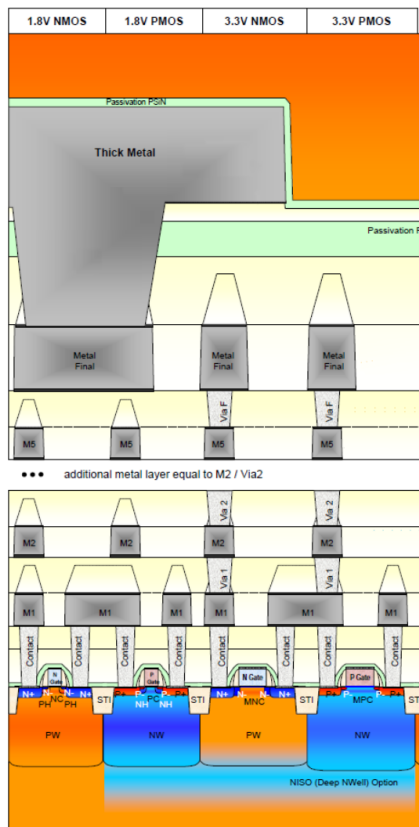
(a) Cross-section of LFoundry 150 nm (LF15A) CMOS technology. [33]
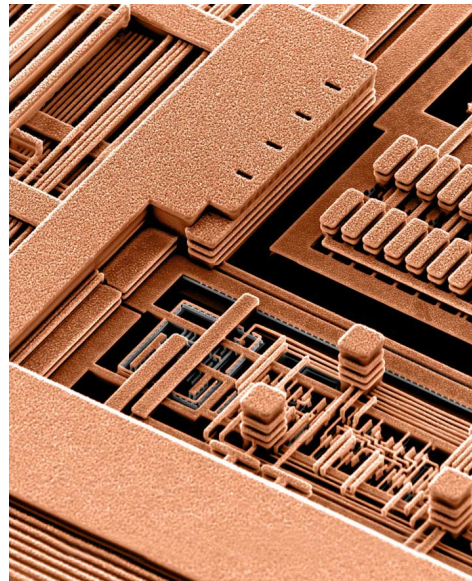
(b) Visualization of copper wires inside an IBM chip. [34]

Figure 3.12: CMOS technology process examples.

When describing different CMOS process generations, often referred to as "nodes", one of the most distinguished features is the minimum gate length L (Fig. 3.13(a)) of a transistor possible to

manufacture. While shrinking L is by far not the only change between nodes, it is a convenient value to refer to, since it always reduces in a newer generation of the process. The number of available metal layers is a characteristic feature of a given technology node and varies from a few layers for older processes to ten or more for modern ones.

A big appeal of the smaller technology nodes is the possibility to implement the same functionality in a smaller area. Fig. 3.13(b) visualizes this with an example of a D Flip-Flop realized in several technologies from 250 nm down to 65 nm. The size reduction is achievable not only thanks to smaller minimal L, but also due to narrower metal wires. It is also clear from this figure that an architecture change can also have a significant impact on the occupied area, even when using the same technology - the first Flip-Flop (DFF_skt) is 12 times smaller than E_dff thanks to using standard linear transistors instead of enclosed layout transistors (ELT).

(a) Simplified drawing of an nMOS transistor with indicated width W and length L.

(b) Comparison of the size of a circuit with the same functionality in different technologies (node indicated in bold). [35]
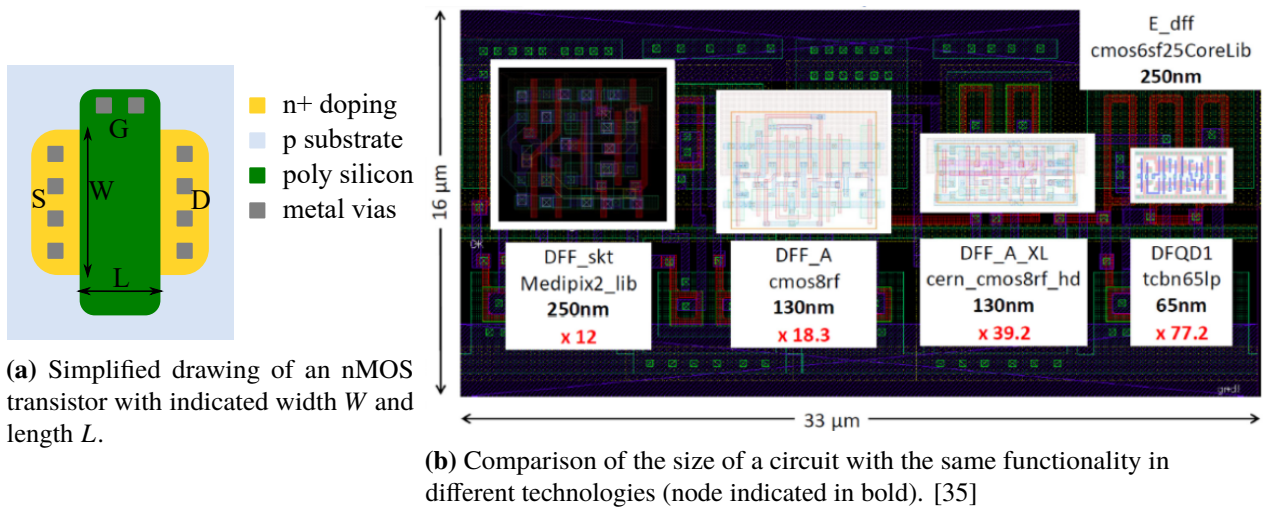
Figure 3.13: Transistor dimensions and their impact on the area occupied by a circuit.

Smaller technology nodes often reduce the nominal power supply voltage V_{DD} needed to operate the devices. A 350 nm technology usually requires a power supply voltage of 3.3 V, while a 65 nm process needs 1.2 V. The V_{DD} reduction combined with smaller parasitic capacitances (thanks to narrower wires) is especially beneficial for digital logic, since the dynamic switching power P_{dyn} is defined as:

$$P_{dyn} = C_L f V_{sw}^2 \tag{3.17}$$

where C_L is the load capacitance, f is the switching frequency and V_{sw} is the swing voltage of the digital gate output, in most cases equal to V_{DD}. Therefore, the same circuit realized in a smaller technology node can occupy a smaller area and consume less power or run at a higher frequency with unchanged power consumption. On the other hand the gate thickness is often reduced in the new generations of the CMOS processes, which leads to a very strong electric field developing across the gate's oxide, thus enabling electrons from the silicon bulk to pass through the oxide to the gate by a quantum tunnelling effect. This gate leakage current can become quite significant for modern technologies as shown in Fig. 3.14, leading to a noticeable static power increase for large chips with millions of transistors. Additionally, as the transistor's dimensions shrink, the distance between source and drain gets small enough that the channel cannot be completely closed by controlling gate voltage,

leading to off-state current flow. This effect is relevant only in modern, deep submicron nodes.
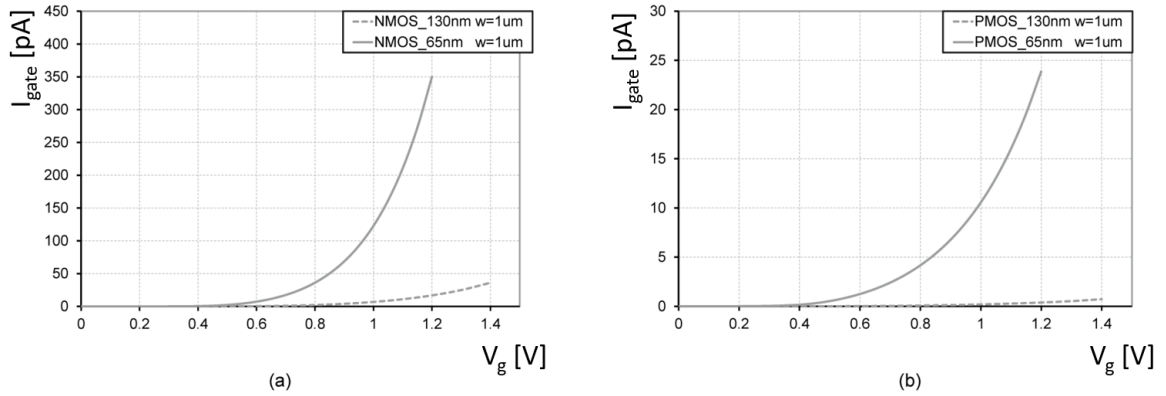

Figure 3.14: Comparison of the gate leakage current for nMOS (a) and pMOS (b) transistors as a function of gate voltage V_g in two different technologies. Dimensions of the transistor are kept constant at $1 \times 1\ \mu m^2$. [36]

For analogue circuits the situation is more complicated. While the influence of the gate leakage current is also noticeable, the lower power supply voltage does not provide benefits. It reduces the available voltage swing and can make it more problematic to bias stacked transistors. At the same time the transconductance g_m of a transistor of a given length can be higher in a smaller technology node, as shown in Fig. 3.15, leading to a potential better performance of a circuit (e.g. rise time of amplifier as shown in Eq. (3.11)). Overall the benefits of using a smaller technology node for analogue design are not straightforward to predict and have to be evaluated on a case-by-case basis.
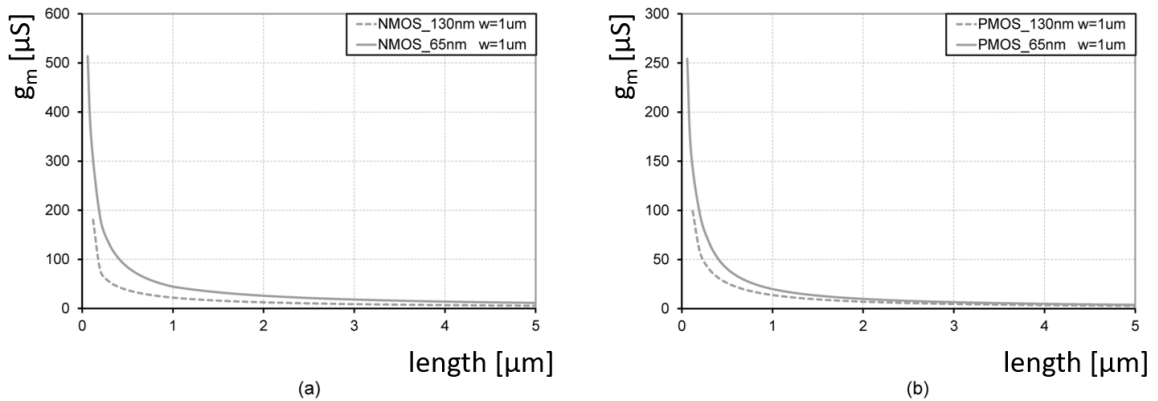
Figure 3.15: Comparison of the transconductance of an nMOS (a) and a pMOS (b) transistor as a function of the gate length in two different technologies. [36]

The electrical performance, power and area trade-offs described shortly in the previous paragraphs are not the only things to be considered when choosing a technology for implementing an ASIC. Of very high importance also is the price of manufacturing. While for technologies older than a few years the cost per mm^2 of IC's area scales roughly linearly with the reduction of the node size, the prices for very latest and most complicated processes grows much faster and become prohibitive for

everyone except the biggest companies. Additionally, when choosing a technology, the long term availability and stability of manufacturing has to be considered, especially when planning production of chips for long running HEP experiments. On top of that, the effects of high radiation levels on the chip performance (discussed in the next section) have to be evaluated, which is a very HEP-specific requirement, usually not considered by the CMOS foundries. All those factors result in the selection process for technologies for particle detectors being lengthy and complicated.

3.1.5 Effects of radiation on silicon and electronics

Particles passing through the detector not only create charge in the sensor as described in Section 3.1.2, but can also have negative effects on the device. This is especially true for the silicon tracking detector, which is the subsystem closest to the interaction point in ATLAS, thus receiving the highest dose of radiation. For this reason a lot of effort is put into understanding the mechanisms of detector degradation due to radiation and implementing techniques to prevent or mitigate it.

The effects of radiation on silicon sensors and CMOS electronics can be classified into three categories:

- Bulk damage from Non-Ionizing Energy Loss (NIEL)

- Surface damage, scaling with the Total Ionizing Dose (TID) received

- Single Event Effects (SEE)

The qualitative description of those three phenomena is presented in the following sections, while a quantitative one can be found in [15] or [37].

3.1.5.1 Bulk damage in silicon

Bulk damage is caused by the displacement of silicon atoms from their usual positions in the lattice by a high energy particle. The empty lattice site (vacancy) and the dislocated Si atom (interstitial) can for example form a Frenkel pair, as shown in Fig. 3.16, but other configurations can take place [37]. The minimum energy required to remove an atom from its natural position depends on the binding energy, which is a lattice property, as well as the type of particle interacting with material. In case of silicon the binding energy is 25 eV, and an electron needs at least 225 keV for the dislocation to occur, while a proton or a neutron would require only 185 eV. If the impinging particle has higher energy, the dislocated primary-knock-on-atom (PKA) is accelerated and can cause additional damage to the lattice leading to defect clusters, as illustrated in Fig. 3.17. The defects created in the silicon bulk are not permanent, as they can change the configuration, diffuse or recombine over time, especially in high temperature. This process is called annealing and it can be used to reduce negative effects of radiation on the sensors in the experiment.
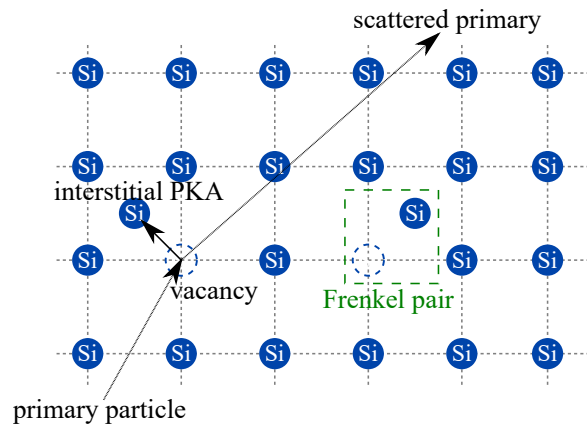
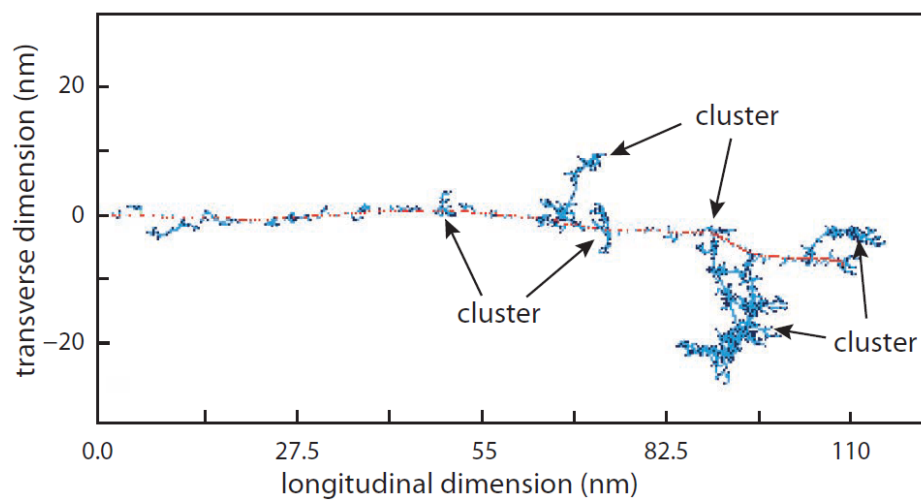**Figure 3.16:** Example of a displacement of silicon atoms in the lattice caused by high energy particle.

Figure 3.17: Visualization of lattice damage in silicon caused by primary lattice atom kicked off at $(x, y) = (0, 0)$ with a kinetic energy of $50\,\mathrm{keV}$ [24].
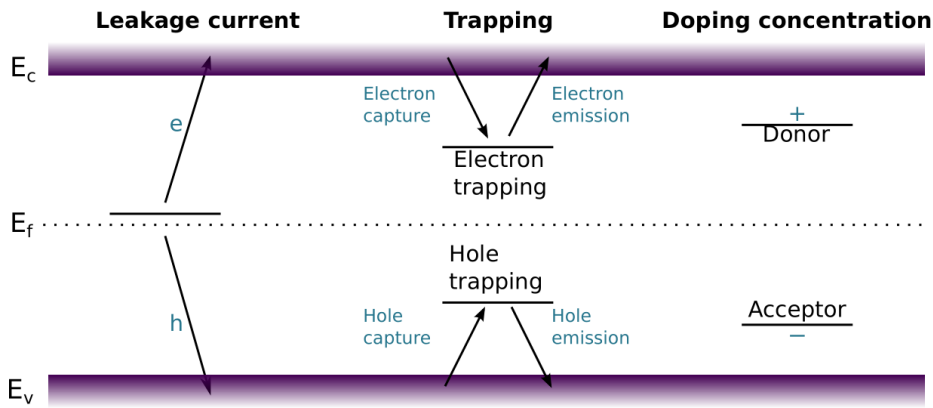
Figure 3.18: Visualization of additional energy levels caused by bulk damage. Each of the three shown types leads to different effects, which in turn affects the detector operation in various ways. An increase in the leakage current can lead to an increase of the device's temperature, higher noise levels and potential issues to analogue front-ends. Creation of the trapping centres decreases the signal seen by the readout electronics. Alteration of the doping concentration modifies the electric field, impacts the collection time and breakdown voltage. E_c - conduction band, E_v - valence band, E_f - Fermi level of intrinsic silicon. [15]

Particles passing through the detector create defects in the whole volume of the silicon bulk, thus leading to creation of additional energy levels in the bandgap. If the irradiation is intensive and long enough the density of defects can become sufficient to meaningfully impact the characteristics of the silicon. Depending on the position of the created energy levels in the bandgap, different sensor properties will be impacted (Fig. 3.18):

- Leakage current - creation of energy levels near the middle of the bandgap reduces by half the energy barrier which an electron has to pass in order to move from the valence band to the conduction band, though two subsequent jumps are needed. This means that more electrons can be lifted to conduction band by thermal fluctuations. For detector operation this leads to an increase of the leakage current I_{leak} of the detector (current flowing in the p-n diode with reverse bias voltage applied) [38]:

$$I_{leak} = V q_{e^-} \sum_{\text{defects}} G_t \tag{3.18}$$

where V is the volume of the depleted damaged silicon, q_{e^-} is the elementary charge and G_t is the generation rate of a defect level. Because $G_t \propto T^2$, the temperature of the silicon T has to be well controlled during the usage of the sensor in order to avoid "thermal runaway" - a situation where increase of leakage current causes an increase of temperature, which in turn increases leakage current. Such a self amplifying process could lead to destruction of the sensor. Additional negative effect of the increased leakage current is an increase of the shot noise of analogue electronics (Eq. (3.14)). This problem can be mitigated by e.g. AC-coupling the analogue readout to the sensor (as done in Section 3.2) or by adding a leakage compensation circuit [39]. It was experimentally shown (Fig. 3.19) that the increase of leakage current ΔI_{leak}

is linearly proportional to the particle fluence ϕ_{eq} (for many cases) and independent from the type of material used for sensor manufacturing. While exceptions from this exist (e.g. non-linear behaviour was observed for irradiations producing mainly point defects [38]), this is a useful feature of silicon, since it allows to predict the leakage current for any particle fluence.

- Trapping - energy levels created close to valence or conduction bands can trap charge carriers created in the sensor by a traversing particle. Such trapped carriers are released after some time constant characteristic for given energy level. If the release time is longer than the integration time of used analogue front-end electronics the measured total charge will be reduced for the given event. The effective trapping time τ_{eff} can be used to describe this effect assuming that the loss of charge depends mostly on the transport time of the charge carriers inside the sensor [38]:

$$\frac{1}{\tau_{\text{eff}}} = \frac{1}{\tau_{\text{eff},0}} + \beta\phi_{eq} \tag{3.19}$$

where β is the proportionality factor (different for electrons and holes) and $\tau_{\text{eff},0}$ is the effective carrier lifetime before irradiation. Fig. 3.20 shows an example of a simulated charge collection efficiency for a 18 μm thick epitaxial sensor biased at 1 V after irradiation to levels relevant for ATLAS detector in the HL-LHC era, where it is clearly visible that charge collection efficiency is strongly impacted after high irradiation. This is a crucial information for the design of the analogue front-end electronics, since the noise and threshold levels have to optimised such, that reliable particle detection is guaranteed at the end of life of the detector.

- The doping concentration - defects can interact with dopants resulting in creation of complexes or removal of dopants from their usual position in the lattice. This changes the effective doping concentration N_{eff} (difference of donor-like and acceptor-like state concentration, equivalent to effective space charge). N_{eff} can be defined as:

$$N_{\text{eff}} = \frac{2\epsilon_0\epsilon_{Si}V_{FD}}{q_e d^2} \tag{3.20}$$

where $\epsilon_0\epsilon_{Si}$ is the permittivity of silicon, d is diode thickness, V_{FD} is full depletion voltage (can be measured as the reverse bias voltage at which the diode reaches a minimum capacitance) and q_e is the elementary charge. From Eq. (3.20) it is clear that changes of N_{eff} have to be followed by changes of V_{FD} if full depletion of bulk volume is desired. Fig. 3.21 presents measured V_{FD} as a function of the particle fluence and it reveals that the bias voltage has to change as the sensor gets irradiated.
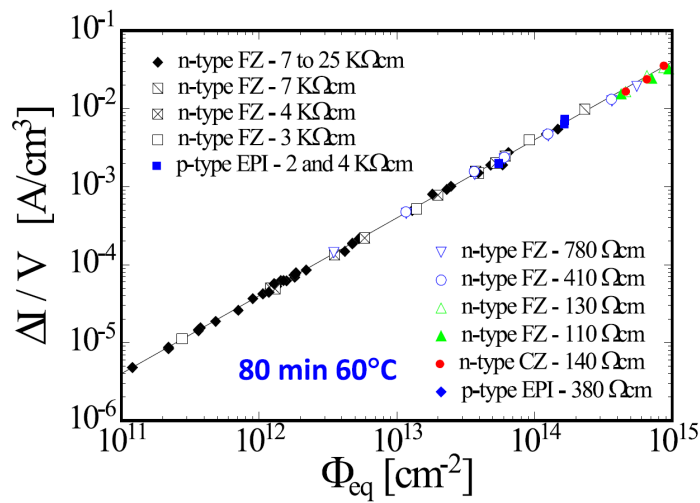
Figure 3.19: Increase of the leakage current (normalized to sensor volume) as a function of particle fluence for varius silicion detectors producesed in few different technologies. The current is measured after annealing for 80 min in 60 °C. [37]
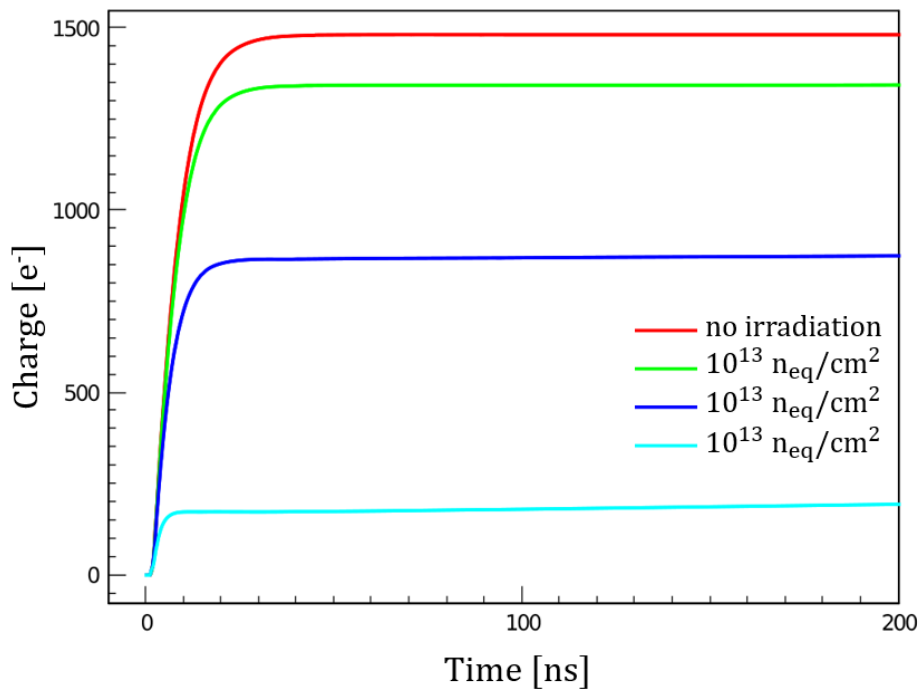

Figure 3.20: Charge collection in a 20 μm thick silicon sensor biased at 50 V after different NIEL fluences of radiation [16].
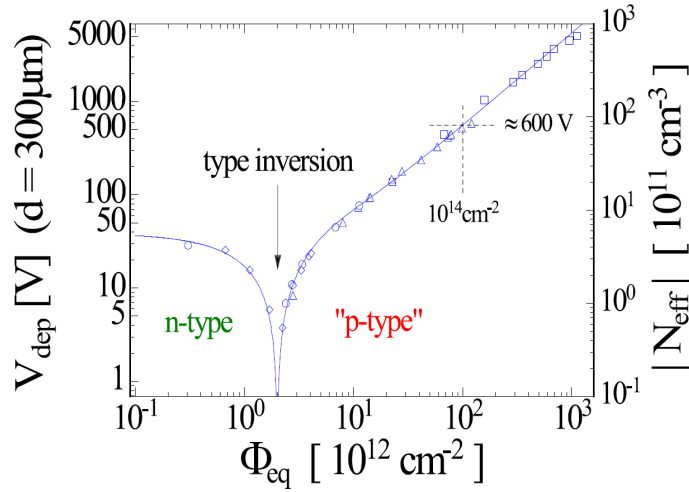
Figure 3.21: Change of full depletion voltage and effective doping concentration as a function of radiation fluence [38].

The amount of bulk damage caused by a particle depends on its type and energy. This means that the silicon sensor will degrade differently in different radiation environments. Since it is not possible to replicate the conditions inside ATLAS detector before running the experiment, the sensors have to be irradiated for evaluation using available particle beams and the damage must be scaled to levels expected inside ATLAS. For such purposes a NIEL (Non-Ionizing Energy Loss) hypothesis has been proposed, which states that fluence dependence of lattice radiation damage in silicon scales linearly with NIEL. The effective cause of the damage are primary defects (point defects and clusters) and the difference in their distribution due to different origins (homogenous over large volume when caused by low-energy protons or gamma rays, or dense clusters when induced by neutrons) is negligible [24]. The lattice radiation damage in silicon can be related to a displacement damage function $D(E)$ defined as:

$$D(E) = \frac{M}{N_A} \frac{dE}{dx}\bigg|_{\text{NIEL}} \tag{3.21}$$

where M is the molar mass of the target material, N_A is the Avogrado number and $\frac{dE}{dx}\big|_{\text{NIEL}}$ is the energy loss per unit length due to non-ionizing processes. This hypothesis arises from many observations where the performance degradation due to bulk damage scales with NIEL fluence, however at the moment there is no understanding of this phenomenon at the microscopic level. Since the displacement damage function is known for several particle types, as shown in Fig. 3.22, it is possible to scale the damage caused by one particle type to another. The usually used reference point (normalization factor) is the damage effect of 1 MeV neutron $D_n(E = 1\text{MeV}) = 95\text{MeV mb}$, therefore a scaling factor called hardness factor κ can be defined as:

$$\kappa = \frac{\int D(E)\phi(E)\text{d}E}{D_n(E = 1\text{MeV}) \int \phi(E)\text{d}E} \tag{3.22}$$

Then the equivalent fluence can be calculated as $\phi_{eq} = \kappa\phi$ (expressed in unit of n_{eq} /cm^2). The expected NIEL fluence at ATLAS experiment near the interaction point can reach levels of 1×10^{-16} n_{eq} /cm^2, as indicated in Fig. 2.8(c).
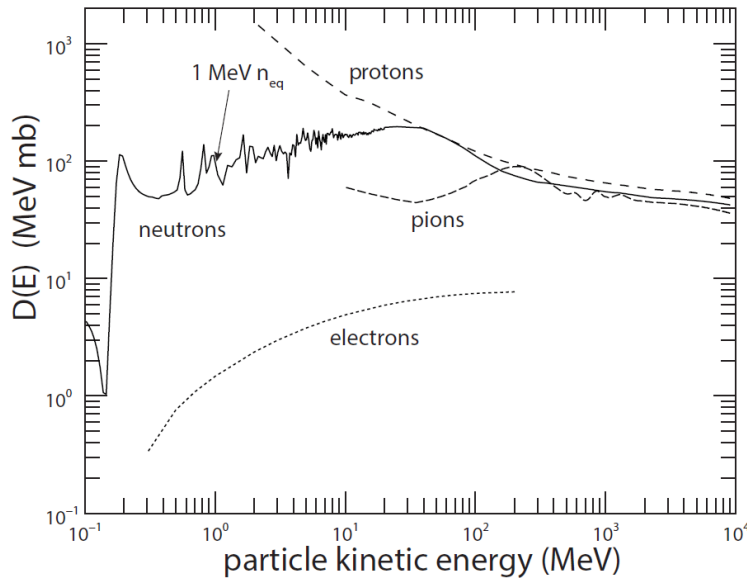

Figure 3.22: Damage function $D(E)$ for different particles. [24]

3.1.5.2 Surface damage

While the non-ionizing radiation is the main cause of damage to the silicon bulk, the surface of the silicon sensor as well as CMOS electronic circuits (especially the silicon dioxide and the Si-SiO$_2$ interface of MOSFETs) are mostly affected by the ionizing radiation. The ionization creates charge carriers in the oxides structures of the CMOS transistors, which do not fully recombine due to electric fields present in the structures. The holes' mobility in Si-SiO$_2$ is 10^6 times lower than the mobility of electrons. Therefore the holes stay much longer in the oxide and create a positive static charge regions. The negative effects can be sorted into two categories, depending on which oxide is affected:

- the charges trapped in the gate oxide of transistor influence its threshold voltage, which degrades the performance of the device and can lead to increased off-state current. In modern CMOS technologies (250 nm and below), however, the gate is very thin (few nanometres), which allows the holes to recombine thanks to tunnelling effects, thus reducing the positive space charge and improving radiation hardness

- the charges trapped in Shallow Trench Isolation (STI) or spacers between transistors can lead to activation of parasitic transistors and allow for leakage current to flow from source to drain. This happens even in modern CMOS processes, since the oxides in question are thick, as illustrated in Fig. 3.23(a). In this case the charge accumulation is influenced by the formation of dangling bonds at the oxide-Si interface which traps the negative charges. This can to some extent compensate for the positive space charge created in the oxide, however, the time scales of

those two process are different by several orders of magnitude, hence leading to complicated time- and technology-dependant interactions.

Temperature also plays a role in the damage mechanisms due to TID. A high temperature during irradiation causes an increase of the TID effects, and therefore should be avoided during the running of the chips in the detector. After a device sustains a TID damage some of it can be reduced by annealing.
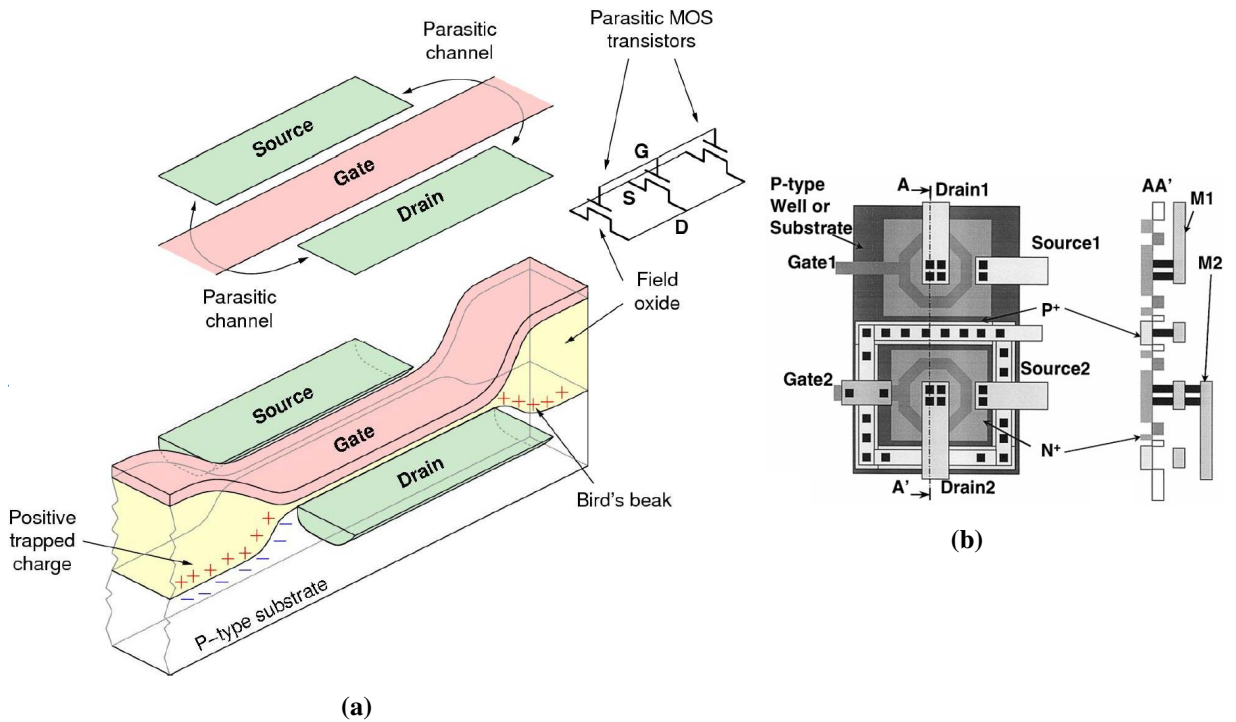

Figure 3.23: (a) Visualization of a positive charge trapped in the field oxide surrounding a transistor, which leads to creation of parasitic transistors[40]. (b) A technique of enclosed layout transistor geometry, which allows to avoid this issue Fig. 3.23(b)[41]. The p^+ guardring is used to stop leakage current paths between neighbouring transistors.

Radiation hardness of a transistor is influenced by more factors than just the thickness of oxides needed to construct it. Important factors are also its width and length. Fig. 3.24 presents the impact of the TID on the maximum current that a transistor can provide measured for several sizes of devices manufactured in 65 nm. While all nMOS transistors (Fig. 3.24(a)) behave quite similarly, the variation between pMOS devices (Fig. 3.24(a)) is very large. It can also be seen that the degradation of small pMOS transistors is so large, that after 1 Grad (TID expexted at the end of life of HL-LHC ATLAS detector) the transistor is in a permanent off-state. The performance degradation dependence on the size of the transistor has significant implications on the design of electronic circuits. Not only some sizes of transistors might be prohibited, but also it is possible that different parts of the circuit will degrade at different rates which has to be accounted for during the design and simulation phases. This requires the creation of models of irradiated transistors, which can be used by commercial circuit simulators. Such models are not provided by CMOS foundries since the majority of their customers are not interested in these effects, therefore they are created by HEP community based on measurements

on test structures [42].

Apart from using wide and long transistors, another way to improve radiation hardness of the design is by using non-standard layouts for transistors. The so-called Enclosed Layout Transistors (ELT), shown in Fig. 3.23(b), physically separates the source from the drain of a transistor. This radiation hardness enhancement has several drawbacks – restrictions on the possible width W, length L and $\frac{W}{L}$ of the transistor, increased and not equal parasitic capacitances at source and drain or larger occupied area. This makes it only feasible for implementation in analogue circuits. For digital, high density designs there is much less freedom in choosing transistor geometries, therefore the designers have to rely on accurate modelling of gates and careful parametrization of design corners [2].

(a) nMOS

(b) pMOS
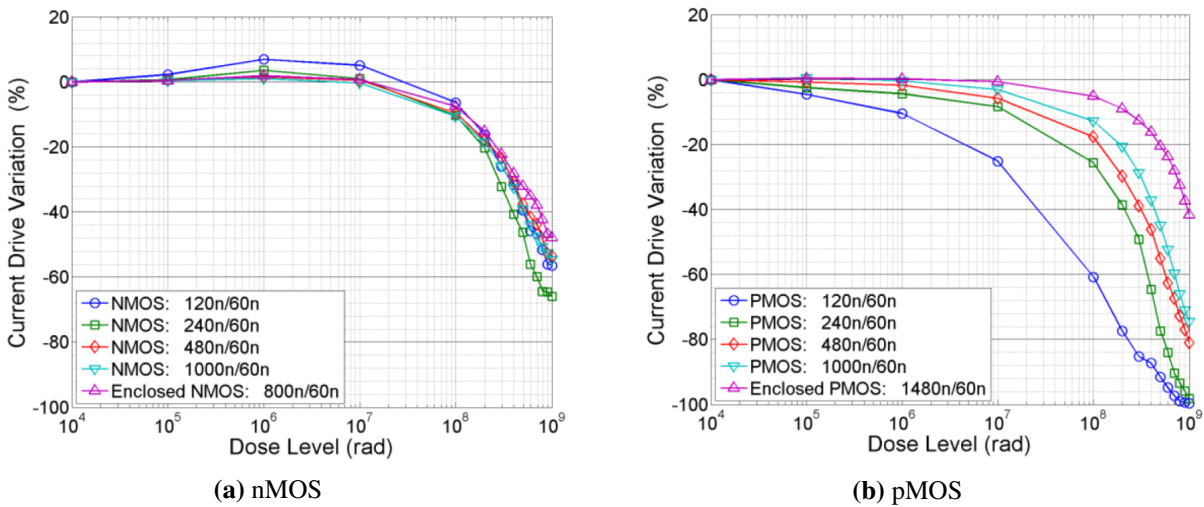
Figure 3.24: Measurement of the current drive (maximum current in on-state) as a function of TID for several device sizes manufactured in 65 nm process. [43]

An interesting feature of small technology nodes is that the surface damage depends not only on TID, but also on the rate at which the device is irradiated. The measurement results shown in Fig. 3.25 indicate that the degradation is more severe for devices irradiated slowly. The physical mechanism behind this phenomenon is not yet understood, however the practical implication of this discovery is the necessity to irradiate chip prototypes to much higher TID than the end-of-life TID of the detector (for practical reasons the prototypes are often irradiated using high dose rates).

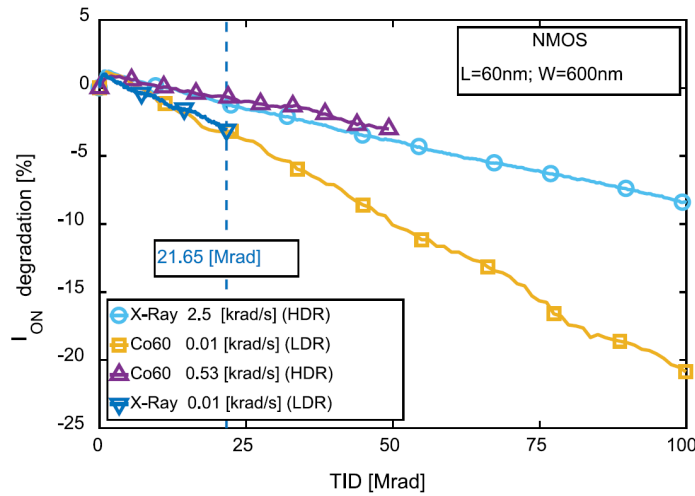[2] Definition of extreme cases of chip operation in term of power supply voltage drops and temperature variation

Figure 3.25: Comparison of the on-state current percentage degradation of nMOS transistor (65 nm process) for irradiation with X- and γ-rays at high (HDR) and low (LDR) dose rates. [44]

3.1.5.3 Single Event Effects

Single Event Effects (SEE) is a broad term, encompassing several types of possible changes in the microelectronics devices caused by radiation. The change is a result of the free charge carriers created by ionization and collected in a sensitive node of the circuit, usually reverse-biased p-n junctions for example in MOSFETs. The effect is undesired and detrimental to the operation of the circuit. The free charge carrier can be produced either directly by a passing ionizing particle or indirectly by a non-ionizing particle, which causes secondary interactions leading to charge generation (e.g. a neutron interacting with atoms producing a shower of secondary particles and a nuclear recoil). Depending on the amount of injected current and the timing with respect to other activities in the circuit a SEE can be classified as one of several types[3]:

- Non-destructive SEE

    - SET (Single Event Transient) - an unexpected change of the logic state of an output of a gate, which travels through combinatorial logic (a glitch). Depending on the timing relation to the clock it might either cause no issues (except increased switching activity) or lead to storing a wrong value in a memory element

    - SEU (Single Even Upset) - change of a value stored in a memory element

    - Latch-up - activation of a parasitic thyristor present between an nMOS and a pMOS transistor as shown in Fig. 3.26: one of the two bipolar transistors gets forward biased and it starts to force current through the base of the other transistor. This creates a positive feedback, leading to an increase of the consumed current. This condition usually requires the chip to be power cycled in order to restore normal operation.

---

[3] The presented list is not exhaustive, only the most common and relevant for this work types are mentioned

- Destructive SEE

  - Latch-up - if the current through the thyristor described in the previous point is large it can lead to permanent damage of the device due to electrical overstress (e.g. junction burn-out or oxide punch-through)

  - SEGR (Single Event Gate Rupture) - an event in which a single high energy particle (usually a heavy ion) strike leads to a breakdown and creation of a conducting path through the gate oxide of a MOSFET [45]
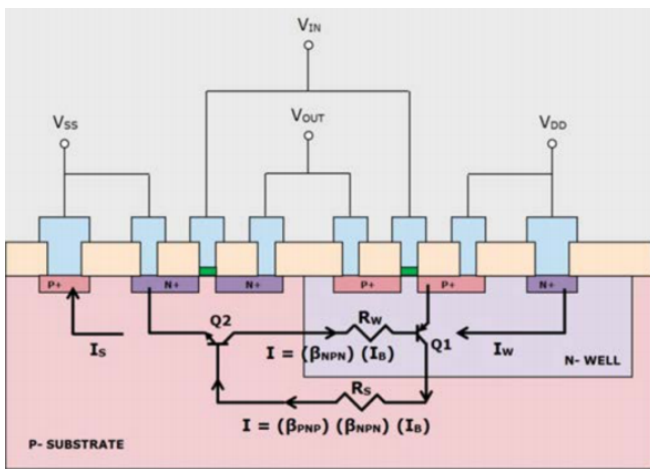
(a) The cross-section of a inverter in CMOS technology with indicated parasitic thyristor. [46]
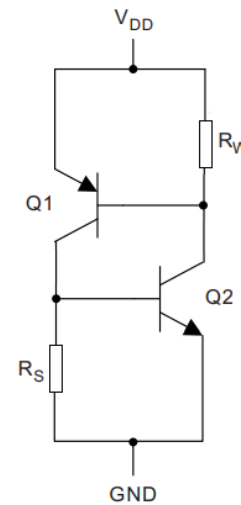
(b) Circuit equivalent of the parasitic thyristor.

Figure 3.26: Parasitic NPNP structure (thyristor) in the CMOS process which could cause latch-up.

SEE have always been a concern for space and satellite applications. Also manufacturers of standard consumer FPGAs are aware of this problem and produce SEE resistant device versions, since SEE can be induced by particle showers from comic rays interaction with Earth's atmosphere [47][48]. The HEP community did not pay much attention to this issue before the era of the LHC, however, considering the expected radiation levels it became apparent that the disruptions caused by SEE might not only corrupt parts of the data coming out of the detectors, but also would lead to problems with controlling the system. As a result the SEE hardness is an important design aspect of all chips that will be placed inside LHC detectors. The susceptibility of any given chip to SEE is very difficult to predict through simulations, therefore the only reliable way of determining SEE hardness is through measurement. Since it is not feasible to reproduce the exact radiation environment inside a detector before running the experiment, the HEP community has developed a method [49] which allows estimating the expected SEE error rate of a circuit in any radiation environment. In this technique first the evaluated circuit is placed into a testbeam and exposed to high fluences of heavy ions (one ion type at a time). The measured SEE error cross-sections σ as a function of the ion's LET (Linear Energy Transfer) is fitted with a Weibull curve σ_{fit} defined as:

$$\sigma_{\text{fit}} = \sigma_{\text{sat}} \left(1 - \exp\left(-\frac{\text{LET} - \text{LET}_{th}}{W} \right)^S \right) \tag{3.23}$$

where σ_{sat} is interpreted as a saturation cross-section, LET_{th} is the threshold LET (minimum LET required to cause SEE), W and S are shape parameters without physical meaning. An example of such a plot and fit is presented in Fig. 3.27. Next, the computational framework developed in [49] combines the Weibull fit parameters with radiation environment parameters of the real experiment obtained from simulation and provides estimated SEE error rate. This process is not perfect and the calculations rely on several assumptions (e.g. technology dependent sensitive volume of the circuit), however the results were proven to be accurate within an order of magnitude.
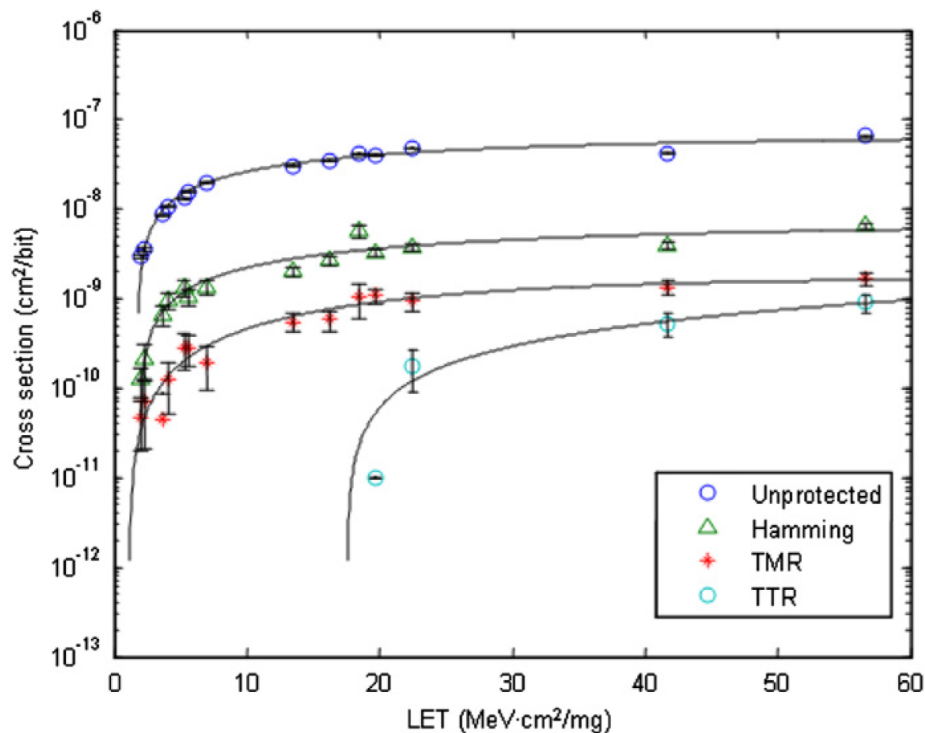
Figure 3.27: SEE error cross-section as a function of LET measured for four different architectures of a shift register (130nm CMOS process). Measurement points are fitted with Weibull curves. Explanation of the architectures is provided in the text. [50]

Circuit SEE hardness can be improved in a few ways. First, the architecture of the design can be changed. The four different curves in Fig. 3.27 represent four different architecture implementations of the same functionality (shift register) and it is clearly visible that the SEE hardness of the circuit changes. The general idea behind each of the evaluated architecture is rather universal and can be applied to any digital circuit and some analogue ones:

- encoding the data in a format providing redundant information, such that in case of a bit flip the full information can still be restored. In Fig. 3.27 the results are shown for Hamming encoding, but many different standards exist

- TMR (Triple Modular Redundancy) - an approach were most or all sensitive nodes of the circuit are triplicated and spread over some area (a good practice seems to be to keep at least $15\,\mu m$ distance between the triplicated elements when using $130\,nm - 65\,nm$ CMOS technologies) with addition of a voter gate, which compares the three logic states and produces a majority voted output (Fig. 3.28). As long as two of the three storage elements are not affected by SEE during a clock period the stored information will be kept correct.

- TTR (Triple Time Redundancy) - same as TMR approach, but with added spread on the clock input such that $\Delta t_0 < \Delta t_1 < \Delta t_2 \ll T_{CLK}$, where T_{CLK} is the clock period (Fig. 3.28). The added level of SEE protection comes from the information being written to each of the memory elements at slightly different time, thus allowing to avoid glitches coming from the preceding logic gates.

There are many more variations of SEE hardening approach by architecture changes e.g. adding feedback to memory elements or implementing multiple levels of majority voting. All of them come at a cost of increased area needed for implementation, higher power consumption and added complexity for design verification, therefore an appropriate approach has to be individually chosen for every design.

Another way of improving SEE hardness is increasing the current flowing through sensitive nets, such that a variation coming from SEE would cause less disturbance, or increasing the capacitive load of sensitive nets to make the SEE induced current create a smaller voltage swing. Both of these techniques are applicable mostly to slow or static analogue circuits, e.g. biasing current mirrors.
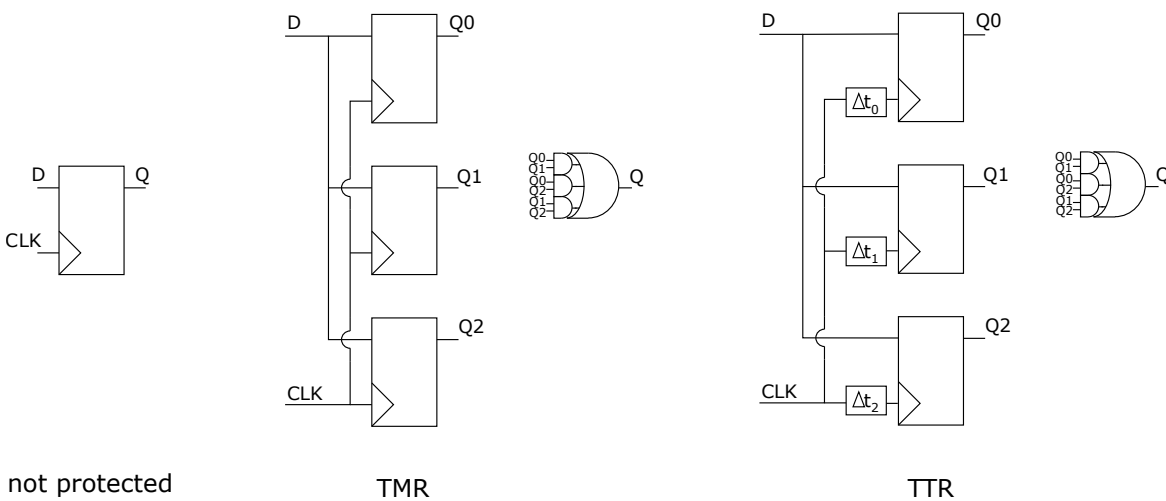

Figure 3.28: Simple examples of SEE hardening techniques: TMR (Triple Modular Redundancy) and TTR (Triple Time Redundancy). A not protected circuit is shown for reference.

3.2 Monolithic pixel detector concepts

As described in Section 3.1 the most commonly used approach for pixel detectors in HEP experiments at the moment is the hybrid technology. In this approach sensor and readout electronics chips are two

separate devices connected together by bump bonds, as shown in Fig. 3.29. The sensor is usually produced on a high purity silicon wafer using a sensor dedicated technology process and is later on thinned down to 100-300 μm. During operation the sensor is biased with sufficiently high voltage to ensure full depletion and therefore provide a large signal in response to the interaction with a traversing particle. The readout chip is manufactured using a standard, commercially available CMOS technology in order to implement required analogue and digital signal processing. An advantage of the hybrid approach is the possibility to develop the sensor and the readout chip in parallel, rather than independently of each other (a few basic specifications must be agreed upon, e.g. the pixel size or the detector capacitance), thus potentially speeding up the design of the pixel detector as a whole. The drawback of the hybrid approach is the high cost of the bump bonding process and the complexity of the assembly. An alternative way of constructing a pixel detector is a monolithic approach, were both the sensor and the readout electronics are combined into one physical entity.
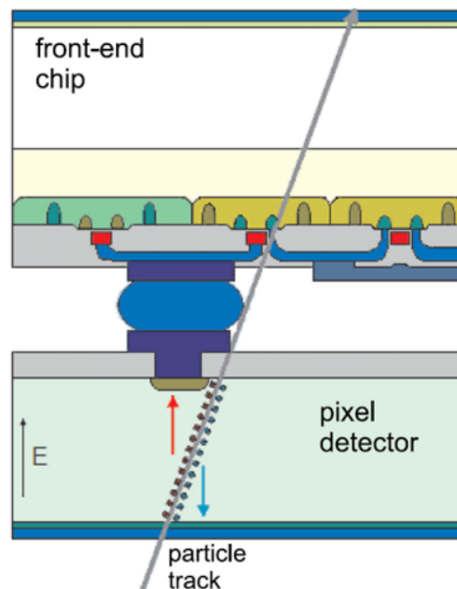

Figure 3.29: A cross section of a hybrid pixel detector with fully depleted sensor. [16]

3.2.1 Monolithic active pixel sensors

Monolithic sensors are presently widely used in commercial applications to detect visible light. So far however, they have not been widely adopted in HEP experiments, but they have been developed for particle detection since 1990s [51]. While the early attempts required complicated production steps e.g. double sided wafer processing or modifications of the manufacturing process, later on a Monolithic Active Pixel Sensor (MAPS) design was proposed using a standard commercial CMOS technology [52]. In that implementation a CMOS imaging process was used, where a thin (order of 10 μm) epitaxial silicon layer is grown on a standard silicon wafer. The CMOS electronics is implanted on top of the epitaxial layer, as shown in Fig. 3.30, and the circuitry is designed in a way which allows to use the epitaxial layer as the sensor.

While the first MAPS was an important step in the development of monolithic detectors, it suffered

from several problems. Due to low biasing voltage and low resistivity of the epitaxial layer the depletion zone is quite small, resulting in most of the charge being collected slowly though diffusion. This not only reduces the maximal operational readout rate, but also makes the detector less radiation hard. Additionally, due to the epitaxial layer being thin, the generated charge is much smaller compared to signals produced by thicker, fully depleted sensors used in hybrid designs. This poses a problem for the readout electronics, which must be designed for suitably low ENC in order to achieve satisfactory Signal-to-Noise and Signal-to-Threshold ratios. Another drawback of first MAPS was no possibility to use full CMOS logic, only nMOS devices were available [51] (a pMOS transistor requires an n-well implant, which would compete for charge collection with the actual sensor n-well implant). Later on this disadvantage was removed by using a deep p-well under the nMOS' n-well, offered by more advanced multi-well technologies. Despite the mentioned issues, MAPS detectors provide some advantages over hybrid detectors - lower production cost thanks to using only commercial CMOS process, lower processing cost (no bump bonding), easier module assembly and lower material budget. This makes MAPS an interesting alternative to hybrid detectors for HEP experiments with suitably low radiation levels and readout rates, leading to MAPS being successfully deployed in e.g. STAR detector at the Relativistic Heavy Ion Collider [53] and upcoming Inner Tracking System upgrade for the ALICE detector at LHC.
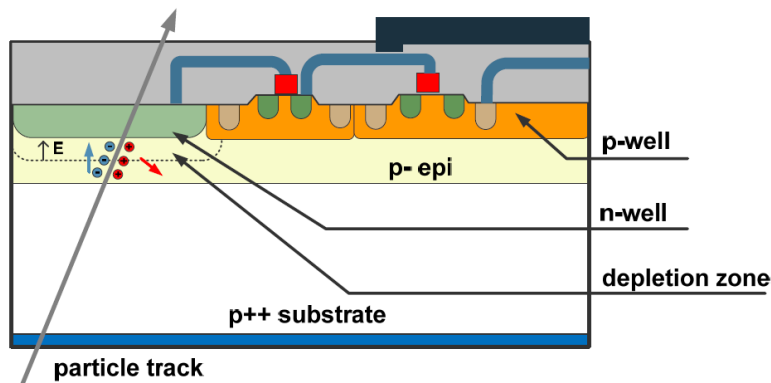

Figure 3.30: A cross section of a MAPS detector with the epitaxial layer used for charge collection. [16]

3.2.2 Depleted monolithic active pixel sensors

In recent years there has been a lot of effort within the HEP community to further develop monolithic detector designs and make them a viable candidate for building detectors for high radiation environments such as HL-LHC. An example of such an activity is a R&D collaboration aiming to replace the outer pixel layer of the HL-LHC ATLAS detector (ATLAS ITk L4) by monolithic sensors [10]. While ultimately that goal was abandoned due to several reasons (issues with the time schedule of building the ATLAS detector, lack of manpower), the produced designs show significant improvements over classical MAPS in such performance parameters as charge collection time, efficiency and radiation hardness. The so-called Depleted MAPS (DMAPS) designs achieved those improvements thanks to utilizing several technological advancements developed by the CMOS manufacturing industry in last years:

- multiple nested implant wells, allowing to implement both nMOS and pMOS transistors, and use deeper implant to isolate them from collection electrode therefore preventing disturbance of charge collection

- high resistivity silicon wafers (a few kΩcm instead of standard wafer resistivity in the range of a few Ωcm), which leads to a larger depletion zone (see Eq. (3.4))

- high voltage biasing, resulting in an increase of the depletion zone in accordance to Eq. (3.4). This is achieved by implementing an appropriate guardring structure around the chip, which modulates the electric field such that breakdown is avoided such that high enough voltage can be applied (breakdown above 250 V has been achieved in some designs)

When all those technology features are combined the resulting DMAPS detector can look like in Fig. 3.31 - CMOS electronics embedded in a collection well with the large volume of silicon bulk depleted. The device can be thinned down to a thickness which allows to fully depleted the entire volume (about 100-200µm is achievable), thus removing the slow diffusion component of charge collection. The large depletion volume leads to a large signal being produced by the traversing particle. Naturally, DMAPS designs also inherit all the benefits of the MAPS designs (easy production, low cost, low material budget).
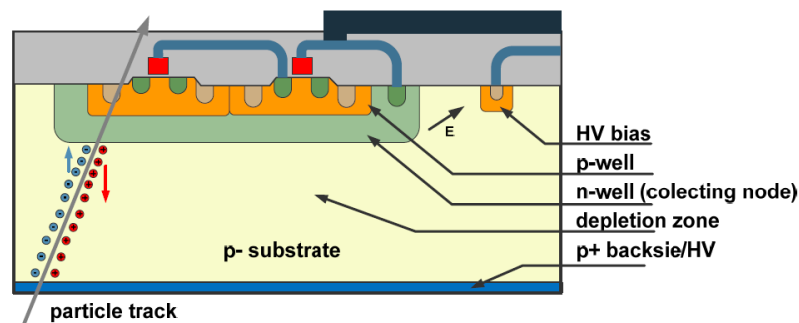

Figure 3.31: Cross section of a fully depleted DMAPS detector utilizing a high fill factor sensor design. [16]

3.2.3 DMAPS sensor fill factor

An important choice when designing a DMAPS sensor is defining the ratio of the sensor size to the whole pixel size, sometimes referred to as a fill factor. This choice is driven mostly by the required pixel size, radiation hardness and power consumption of the final DMAPS detector.

3.2.3.1 High fill factor design

A large collection electrode design (high fill factor) looks as shown in Fig. 3.31. The collection electrode is implemented as a large deep n-well, fully encasing all in-pixel readout electronics. This leads to a uniform electric field configuration under the pixel, allowing for an efficient charge collection

nearly independently of the position at which a particle passes through (similar to a standard planar sensor used in hybrid designs). Large collection electrode also means that the drift path of charges being collected is short, which reduces the trapping probability in radiation damaged devices, thus increasing the radiation hardness of the detector.

While the large size of the collection electrode has several benefits, it also brings some disadvantages. The large size of the charge collecting well leads to a large total capacitance at the input of the amplifier. This total capacitance has several contributors e.g. sensor-to-sensor or sensor-to-backside capacitances, which are present in all sensor designs, as well as large collection diode specific parts i.e. capacitance C_w between the collecting diode and the p-wells of the embedded electronics. Overall the detector capacitance C_d at the amplifier input can be in the order of several hundred fF. As discussed in Section 3.1.3 large C_d increases the ENC and degrades the timing performance. In order to improve those performance metrics the input transistor's transconductance must be increased, which is achievable by increasing the power consumption of the circuit. Additionally, large values of the C_w can lead to unwanted charge injections into collection well due to activity of the electronics as illustrated in Fig. 3.32. This can distort the real signal, increase the noise or even produce fake hits. Especially problematic are fast transient signals produced by the digital switching, therefore the in-pixel electronics have to be carefully optimized and much attention is needed for the layout (minimizing crosstalk, shielding sensitive nodes). A last potential problem with the large fill factor design is the increase of the pixel area due to manufacturing constrains for implanting all the well structures required for implementation of the large collection node. As a result, a pixel designed for a low fill factor architecture (described in the following section) can occupy a smaller area than a pixel with equivalent functionality using the large fill factor approach.
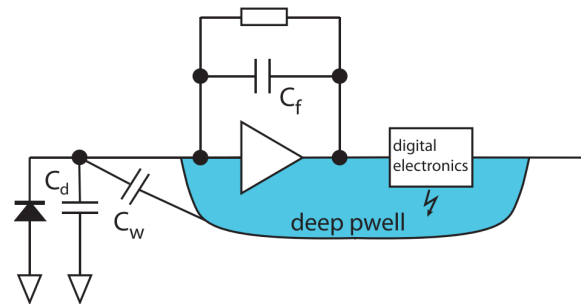

Figure 3.32: Illustration of a potential noise injection path from in-pixel electronics to the charge collecting node in a high fill factor DMAPS design. [54]

3.2.3.2 Low fill factor design

A small collection electrode design (low fill factor) is presented in Fig. 3.33. A small n-well is used to implement the charge collection electrode and it is placed outside a deep p-well housing all the CMOS electronics. This leads to several advantages when compared to large fill factor designs. Most importantly, the detector capacitance can be designed to be very small (order of few fF), which allows to achieve good readout performance at very low power consumption [55]. Secondly, since the sensing diode is no longer capacitively coupled to the readout electronics the risk of noise injection through the substrate is mitigated (careful routing of metal layers is still mandatory).

The disadvantage of this approach is the difficulty in achieving radiation hardness. Since the charge collection electrode is small the charge has to travel a longer (on average) distance when compared to the same size pixel implemented in high fill factor design. Additionally the electric field is not as uniform and the depletion zone is smaller than in the large collection diode approach. This forces the low fill factor designs to use only small pixel sizes. While small pixel size brings benefits in terms of better spatial resolution, it also imposes difficulties on electronics implementation and power distribution. Low fill factor designs for HEP applications have been in development for only few years and so far achieving radiation hardness levels suitable for outer layers of HL-LHC ATLAS has proven to be difficult. Improvements are achievable e.g. through modification of implanataion steps of a CMOS process [56] and prototypes utilizing that method show promising results [55] [57].
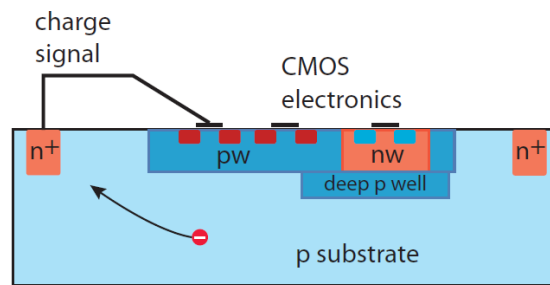
Figure 3.33: A cross section of a fully depleted DMAPS detector utilizing a low fill factor sensor design. [24]

3.3 DMAPS in LFoundry technology

Since 2013 University of Bonn and Centre de Physique des Particules de Marseille (CPPM) have been working together on the development of radiation hard, large fill factor DMAPS. In 2015 CEA Institute of Research into the Fundamental Laws of the Universe (IRFU) has joined the collaboration. The end goal was set for developing a full scale (1×2 cm^2) DMAPS chip with fast readout capability and high hit detection efficiency (above 95%) after sustaining radiation levels expected at the outer pixel layers of the HL-LHC ATLAS detector (10^{15} n$_{eq}$ /cm^2 NIEL, 80 Mrad TID). In order to achieve this goal several prototype chips have been made over the years, as presented in Fig. 3.34, with each next chip benefiting from the experience gained from preceding designs.

The technology chosen for this development line was the LFoundry 150 nm CMOS process (LF15A). It offers multiple nested wells, up to 6 aluminium metal layers and can be manufactured on high resistivity (above 2 kΩcm) p-type substrate. The cross section of this process is shown in Fig. 3.12(a). The manufacturer provided a detailed description of the technology and the processing steps, which enabled the collaboration to use Technology Computer Aided Design (TCAD) simulation tools in order to optimize the layout for HEP experiment needs, e.g. increasing breakdown voltage though guardring layout optimization [16]. This effort was part of a larger collaboration, named "CMOS Demonstrator", aiming to develop DMAPS suitable for HEP experiments (with focus on ATLAS detector) as well as qualifying several different CMOS processes for this purpose [10].

For transparency, it has to be pointed out that the work presented in this thesis regarding DMAPS prototypes mostly concerns the chip design aspects and was carried out in a close collaboration with
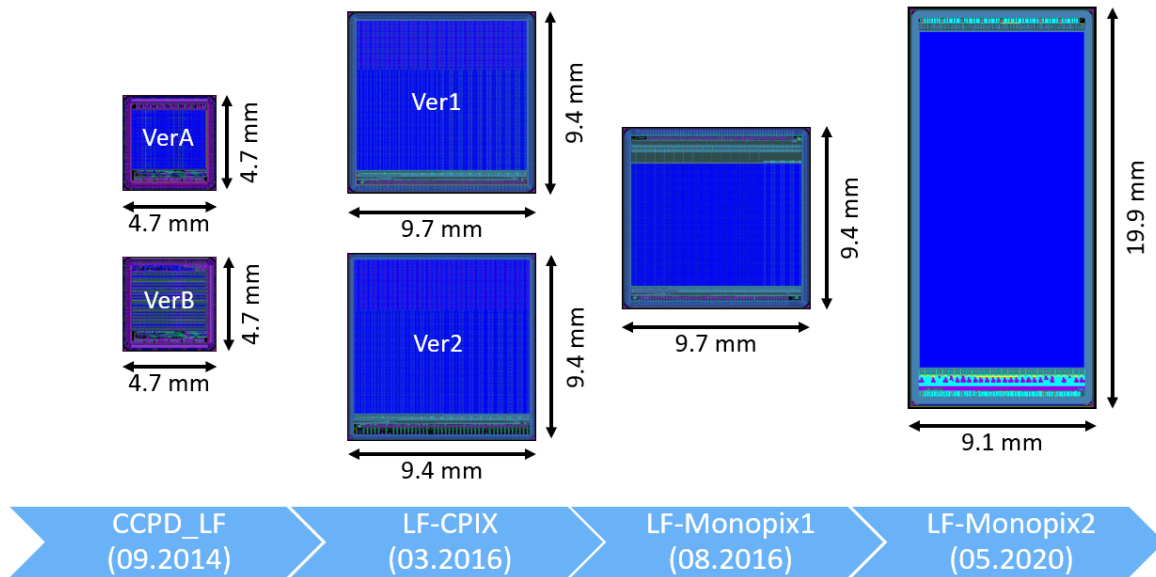
Figure 3.34: Layouts of all chips from the LFoundry DMAPS family with indication of the approximate chip size. Devices are arranged in chronological order with the submission date indicated at the bottom.

several other people from University of Bonn, CPPM and IRFU. Due to the relatively small size of the design team and the organization of the work it is fair to say that every team member had at some stage of the design contributed to every part of the chips. This means that while full chip descriptions are provided further on in this chapter (with a slight focus on tasks carried out by the thesis' author where possible), those devices were not made by a single person alone and individual contributions are not easy to point out. Carrying out the measurements of the prototypes was not a part of this thesis, but the results obtained by other people are cited in order to prove the functionality of the chips.

3.3.1 CCPD_LF

As the first device in the LFoundry DMAPS development line the focus of this design was to prove the feasibility of implementing electronics and a sensor in one physical device. A fast stand-alone digital readout was not foreseen for this prototype, the aim was to assess mostly the analogue stages (CSA and comparator) and either extract the data from chip with in-built slow readout or by connecting it to the FE-I4 readout chip [27]. In the latter configuration the end product is a hybrid detector, but with signal amplification and shaping stages built into the sensor. The connection between the two devices can be made by gluing them together, in which case the chips are AC coupled. Such detectors were already demonstrated to work [58] and are referred to as Capacitively Coupled Pixel Detector, CCPD in short, which led to naming this prototype CCPD_LF.

3.3.1.1 Sensor design

In this prototype two sensor architectures, version A and B, were implemented as shown in Fig. 3.35. Both of them use ultra deep n-well (DNW) as charge collection electrode with standard n-well and

deep n-well, called by the manufacturer NISO, used to connect to it. The electronics is implemented with standard n-well and p-well with a deep p-well (PSUB) added to separate it from DNW. Substrate is p-doped and the devices were manufactured on high resistive Czochralski wafers. An important difference between the two sensor versions is the size of the charge collecting well, which is approximately twice larger version A than in version B. This changes the detector capacitance C_d seen by the CSA, so it is expected that version B will have lower ENC and faster rise time. The drawback is a potential decrease in charge collection efficiency. Additionally, both sensors are biased in a different way, as illustrated in Fig. 3.35 in order to assure safe operating conditions of the readout circuitry. Due to the different biasing schemes, the two sensor types could not have been included in one matrix. Instead, two completely separate chips (5 mm × 5 mm) were made for this prototype submission. In both devices the pixel size is identical (33.3 μm × 125 μm), and both pixel matrices have the same number of pixels (2736 arranged into 24 rows with 114 columns). The same guardring structures surround the matrices (one n-well ring followed by ten p-well rings, the last of which sets the potential level of the substrate).

(a) CCPD_LF version A sensor

(b) CCPD_LF version B sensor.
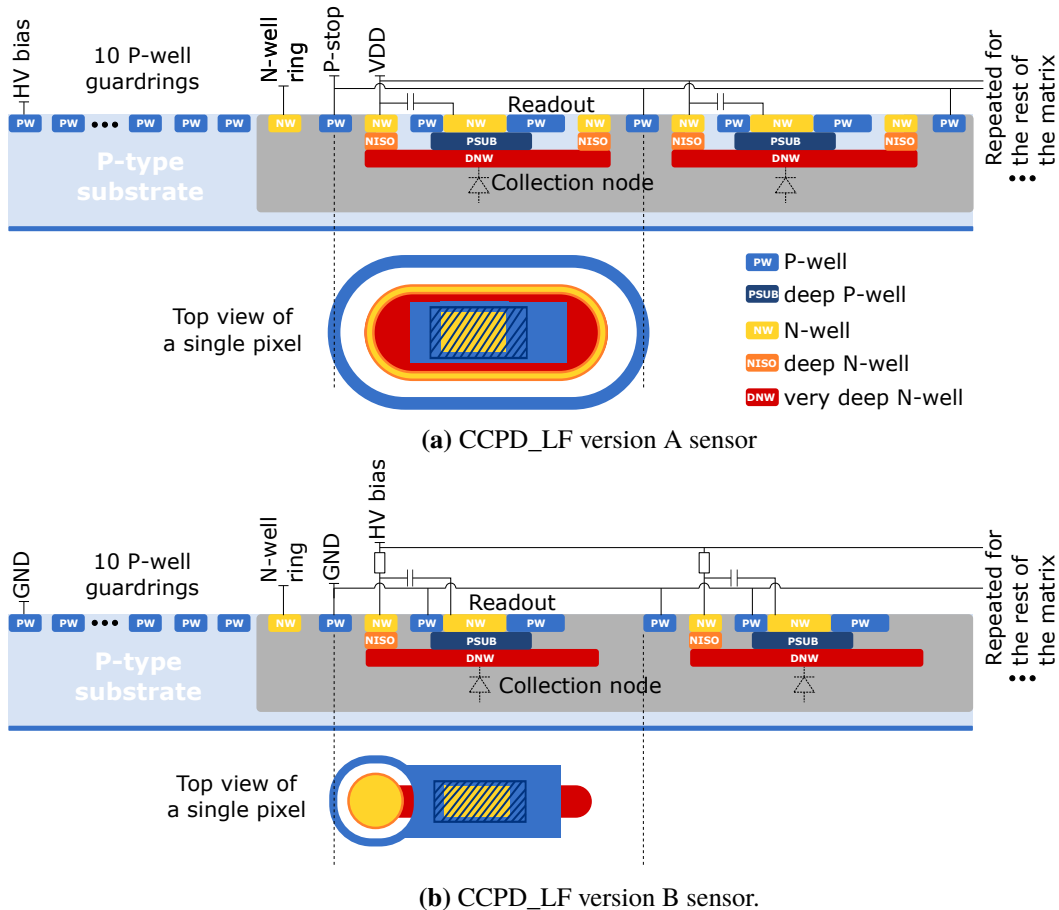
Figure 3.35: Simplified cross sections and top views of the sensor designs implemented in CCPD_LF chip. Depletion zone is indicated in grey (actual depth depends on applied bias). Based on [59]

While Fig. 3.35 shows a simplified view of the real layout for better clarity, it is interesting to point out that the guardrings and the charge collection well are indeed implemented with rounded corners. This is not a standard practice for CMOS design and it is difficult to achieve while complying to the manufacturer's design rules, but it was done in the hope of improving the electric field distribution and increasing the achievable breakdown voltage.

3.3.1.2 In-pixel electronics

Fig. 3.36(a) shows a simplified block diagram of the readout electronics implemented inside the pixels. For both versions A and B of the chip the circuitry is functionally the same, however the physical layouts are completely different in order to cope with different sensor designs. The collection electrode is AC coupled to the input stage of the CSA. A schematic of the implemented amplifier is shown in Fig. 3.36(b). It is a folded cascode design with an adjustable feedback current and 5 fF feedback capacitance. It was designed to consume $10\,\mu A$ of current. The main branch is powered by a separate power rail $VDDA_{PRE}$ in order to avoid any noise coupling from other parts of the circuit. ENC expected from simulations (typical corner) is 124 e, gain of $18\,\mu V/e^-$ and peaking time of 31 ns (assuming $C_d = 150$ fF and input charge of 4 ke). The CSA output is AC coupled to the input of the comparator to allow an external control of the baseline of the comparator's input (done by adjusting the BL voltage). A schematic of the implemented two-stage open loop discriminator is shown in Fig. 3.36(c). The comparator threshold is set globally by the V_{th} voltage and can be adjusted locally (independently for each pixel) using a 4-bit trim DAC. This trimming procedure allows reducing the threshold dispersion occurring due to e.g. mismatch of transistors or gradients in supply voltage distribution. The comparator output can be used in a few different ways, depending on the used mode of operation:

- Stand-alone readout mode: the hit information is stored in a D flip-flop and read out out via a global shift register running through the entire matrix (not shown in Fig. 3.36(a)). In this readout option, only the position of passing particles during the integration time (period when the HIT registers can store the hit signal from the discriminators) is recorded and the precise time information is not available (the shift register can be clocked at maximum frequency of ~1 MHz).

- Pixel monitoring mode: the analogue output of the CSA and the discriminator for any pixel can be monitored (one pixel at a time) using two analogue on-chip buffers. While this limits the readout to observing only a single pixel, it allows to extract precise information about the charge deposited by a particle (amplitude of the CSA output) and time when it happens

- CCPD mode: as mentioned above, the chip is designed to allow connecting it to the FE-I4 readout chip if faster and more sophisticated signal processing is desired, e.g. timestamping the hits with 25 ns resolution. In this case, the digital signal from the discriminator is sent to the output stage shown in Fig. 3.36(c), where its width is stretched (the falling edge of comparator output is also interpreted by the CSA of FEI-4, therefore it has to be delayed to avoid disturbing the real signal). After that, the second part of the output stage sends a signal of adjustable amplitude (this functionality is explained in the following section) through a coupling capacitor to a bond pad. The coupling capacitor is used here in order to allow connecting CCPD_LF with FE-I4 with metal bump bonds instead of glue, which provides a more uniform and better controlled interface at a cost of more expensive assembly process.

For testing purposes each pixel also has a circuit allowing injection of a known charge into the input of the CSA. This functionality is implemented in a very simple way - the V_{inj} line can be controlled externally, therefore when it is pulsed a charge is injected through C_{inj} into all pixels which have the switch EN_{inj} configured to conduct.
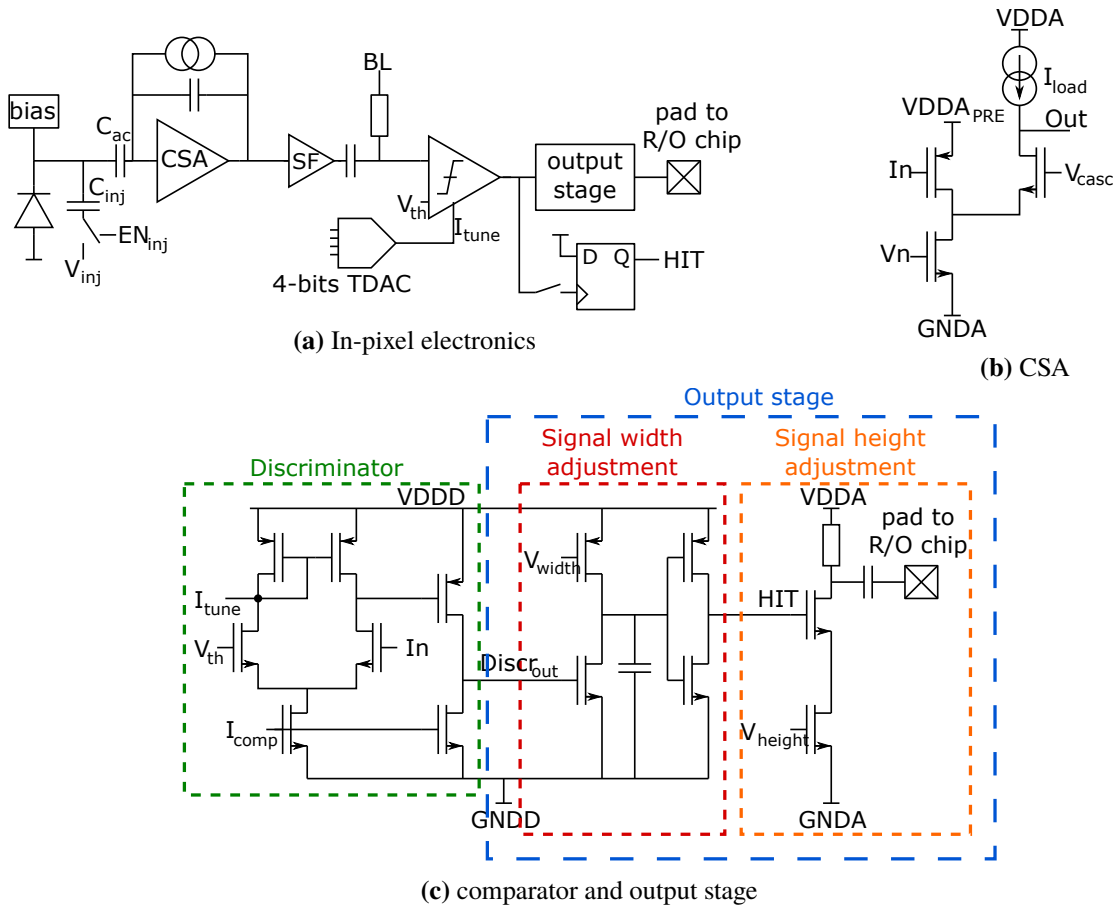

(a) In-pixel electronics

(b) CSA

(c) comparator and output stage

Figure 3.36: Schematics of electronic circuits implemented inside CCPD_LF pixels. [59]

3.3.1.3 Pixel arrangement in the matrix

The pixels of CCPD_LF are organized into groups of six. Inside such a group outputs of three pixels are connected to a single bond pad, through which in CCPD operating mode they are connected to a FE-I4 pixel, as shown in Fig. 3.37(a). This arrangement was used in order to test an idea of sub-encoding a large pixel of a readout chip to smaller pixels of the sensor. The differentiation between pixels is achieved by configuring them to output signals of different amplitude (controlled by the "height adjustment" circuit in Fig. 3.36(c), with V_{height} being programmable per pixel). The readout logic of FE-I4 has a 4-bit ToT functionality, therefore different amplitudes from the sensor are translated into different ToT codes. With the appropriate choice of the amplitude height it should be possible to even distinguish hit pixels in a hit cluster. This implementation also defines the pixel size

of the CCPD_LF - in order to fit six pixels into the area occupied by two FE-I4 pixels, the pixel size of $33.3\,\mu m \times 125\,\mu m$ has to be used.

The pixel matrix occupies most of the area of the chip, for both versions A and B it is consists of 114 columns and 24 rows. The rest of the space is occupied by global configuration, bias DACs and IO pads placed below the matrix. Although the sensor structure for all pixels in the given matrix version is the same, for the readout electronics a few flavours have been implemented. All flavours are based on the circuit described in Section 3.3.1.2, with only minor variations between them. Chip version A includes three kinds of pixels, differing by the length and design (standard linear geometry or ELT) of the transistor in CSA's feedback loop. The motivation for this was measuring the difference in performance degradation due to TID depending on the transistor type. In version B there are twelve pixel flavours, the differences between them are the length and design of the transistor in the CSA's feedback loop, the method of high voltage delivery (through diode or resistor), and amount of deep p-well in the pixel. The distribution of pixel flavours in the matrices is shown in Fig. 3.38.
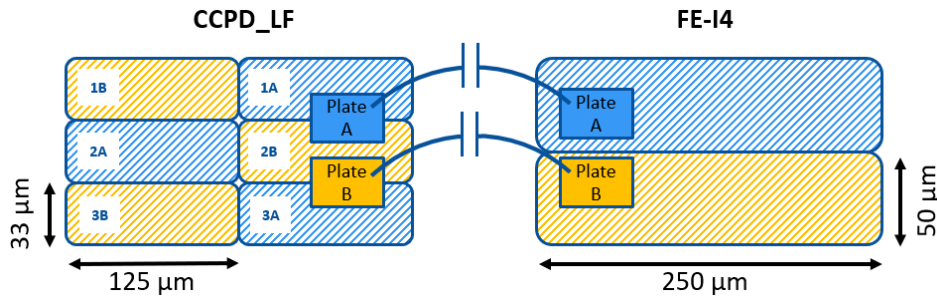
(a) Arrangement of connections between a group of six pixels of CCPD_LF and two pixels of FE-I4.
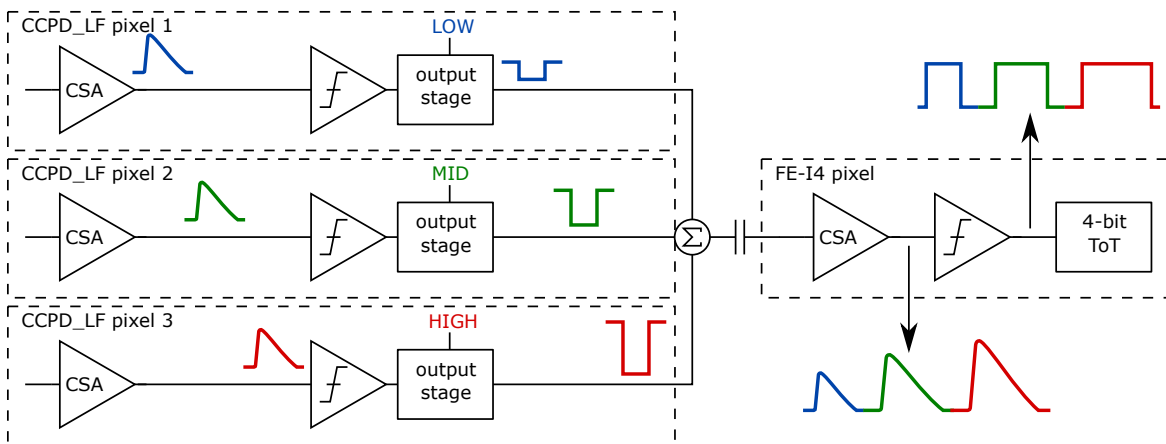
(b) Illustration of the idea of sub-encoding CCPD_LF pixels using programmable pulse height. Based on [60]

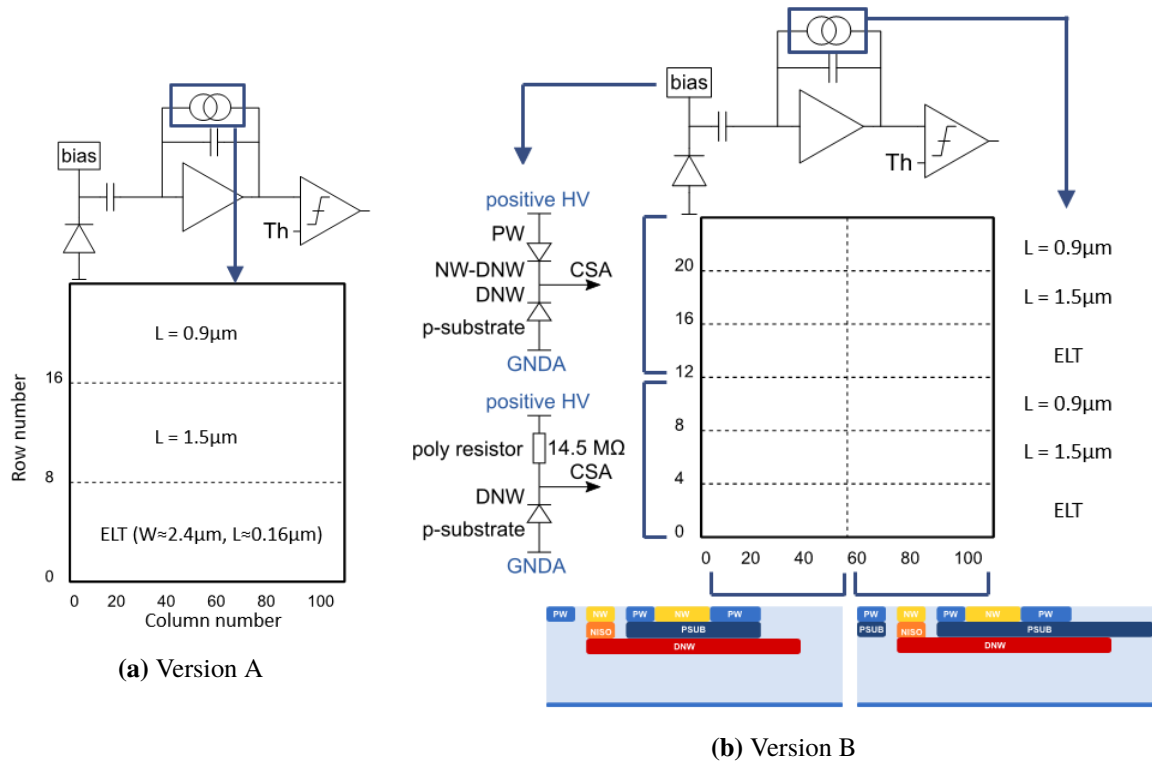**Figure 3.37:** Diagrams of connectivity between pixels of CCPD_LF and FE-I4.

(a) Version A

(b) Version B

Figure 3.38: Arrangement of pixel flavours inside matrices of CCPD_LF [59].

3.3.1.4 Measured pixel electronics performance

First, a single pixel response to a radioactive source was measured using the "pixel monitoring mode" i.e. observing the outputs of the CSA and the comparator via the analogue buffers. The source used was ^{55}Fe, which emits two X-ray photons with energies of $5.9\,\mathrm{ke^-}$ (K_α) and $6.5\,\mathrm{ke^-}$ (K_β, smaller intensity than K_α). Such photons are absorbed after a short distance i.e. large depletion zone is not required, therefore they always deposit a well known charge in the sensor (approx. $1.6\,\mathrm{ke^-}$ and $1.78\,\mathrm{ke^-}$, respectively). This provides a reliable way to calibrate the detector. Spectra obtained with CCPD_LF are presented in Fig. 3.39. Fitting a Gaussian to the signal peak allows finding the gain of the CSA (amplitude in volts per injected electron) and the ENC (from the standard deviation after taking into account statistical fluctuations of the signal i.e. Fano noise [24]). The values calculated from the measurement are: for version A the gain of $6.2\,\mathrm{\mu V/e^-}$ and ENC of $149\,\mathrm{e^-}$, while for version B the gain of $7.9\,\mathrm{\mu V/e^-}$ and ENC of $120\,\mathrm{e^-}$. Those values are comparable to simulation results (discrepancy with the numbers provided in Section 3.3.1.2 is caused by observing the signals through the analogue buffers). As expected, version B has a lower ENC thanks to a smaller detector capacitance C_d. The K_β peak of ^{55}Fe is not distinguishable in the measured spectra due to too high noise levels. The tail to the left of the K_α peak is caused by charge sharing between pixels.

The precise value of the injection capacitor C_{inj} (see Fig. 3.36(a)) can also be measured with an X-ray source by comparing the CSA amplitude for different injection voltage steps V_{inj} with an amplitude corresponding to a signal generated by photons from the source. While the value of C_{inj} is in principle known from the design, due to production non-idealities and difficult to simulate parasitic
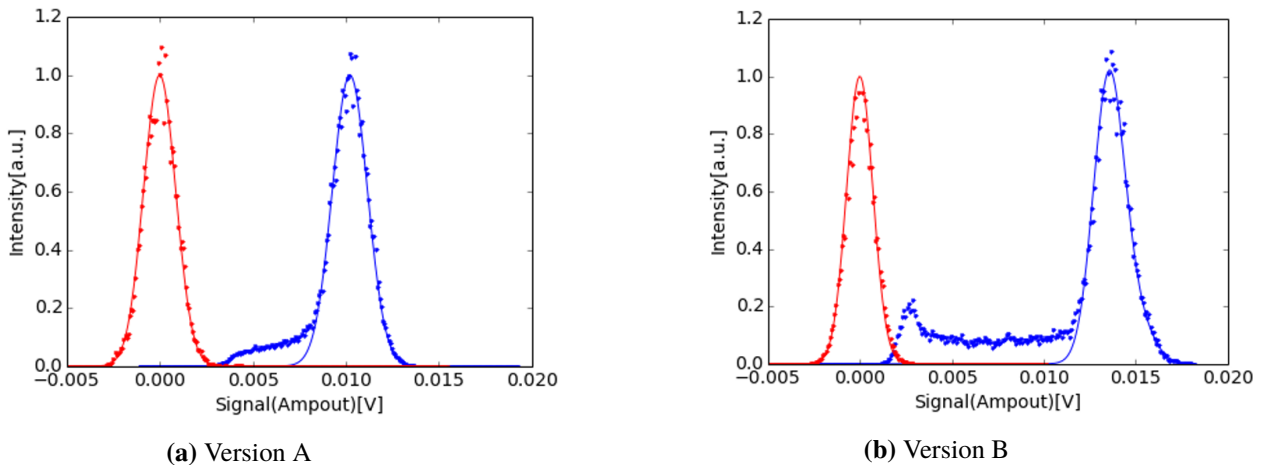
(a) Version A

(b) Version B

Figure 3.39: Single pixel charge spectra measured using the "pixel monitoring mode". Blue shows signals from ^{55}Fe, red shows the baseline. Measurement points (dots) are fitted with a Gaussian distribution (solid lines). [59].

effects, the real value has to be measured. Once this is calibrated (preferably using several X-ray sources) the injection circuit can be used to characterize the pixel electronics by injecting an arbitrary amount of charge at any time without the need for a radioactive source. This enables the usage of a threshold scan [23] in order to measure gain and noise of every pixel in the matrix. The gain values obtained this way match simulation values quite well (10% difference, which could be explained by e.g. not a precise value of C_d used in the simulation) and are homogenous across the matrices. The noise maps, however, look very different for two versions of CCPD_LF. The noise map of version A, shown in Fig. 3.40(a) is quite uniform (device-to-device differences are always present in CMOS processes), which is expected since the differentiation between the three pixel flavours is the type of transistor in the feedback loop. The noise map of version B presented in Fig. 3.40(b) shows a significant difference between the two regions of the matrix. Comparing with Fig. 3.38(b) indicates that the difference is caused by the different bias structure of the pixel. This could be explained by a large parasitic capacitance introduced by the polysilicon bias resistor. Such a capacitance was not visible in the post layout simulations, however, the resistor implemented in the pixel had a custom, hand-made layout (required to fit it into the available area), which most likely was not modelled sufficiently precise by the extraction tools.

One of the main problems found with the in-pixel electronics is a significant threshold dispersion of the comparator. Fig. 3.41(a) presents a histogram of the measured threshold dispersion together with a Gaussian fit. The standard deviation of the fit is 11.7 mV, which makes it too large to be corrected by the local 4-bit trim DAC (LSB$_{\text{TDAC}}$=0.78 mV). This does not prevent chip operation, but makes the characterization difficult. This effect is caused by the small size of the comparator's input transistors (making them very susceptible to process variation) and was later on reproduced in simulation Fig. 3.41(b).
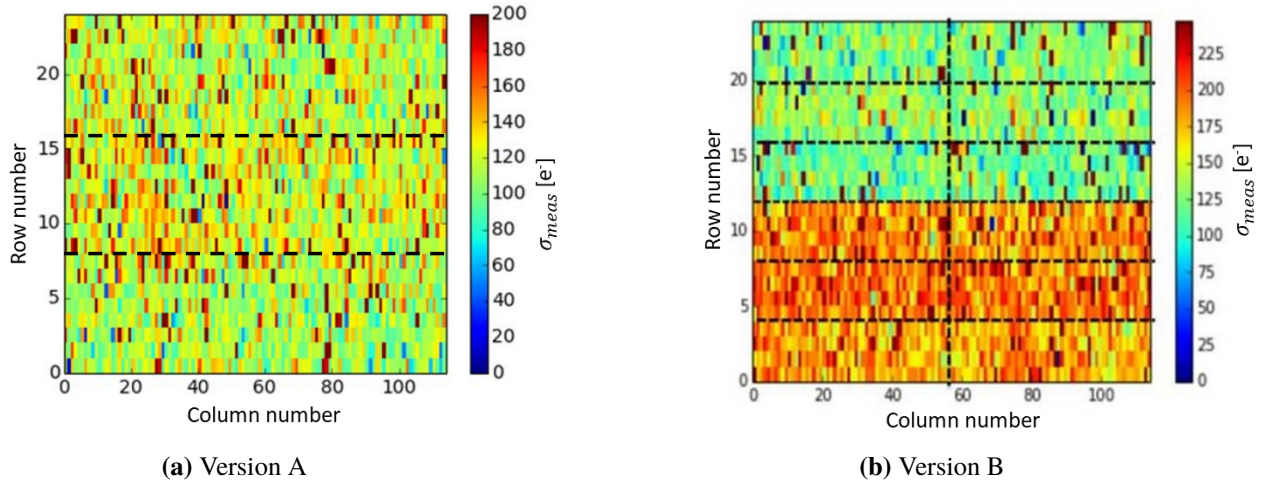
(a) Version A

(b) Version B

Figure 3.40: ENC maps for CCPD_LF measured employing the threshold scan method [23]. Black dashed lines indicate regions of different pixel flavours. [16].

(a) Measurement of version A

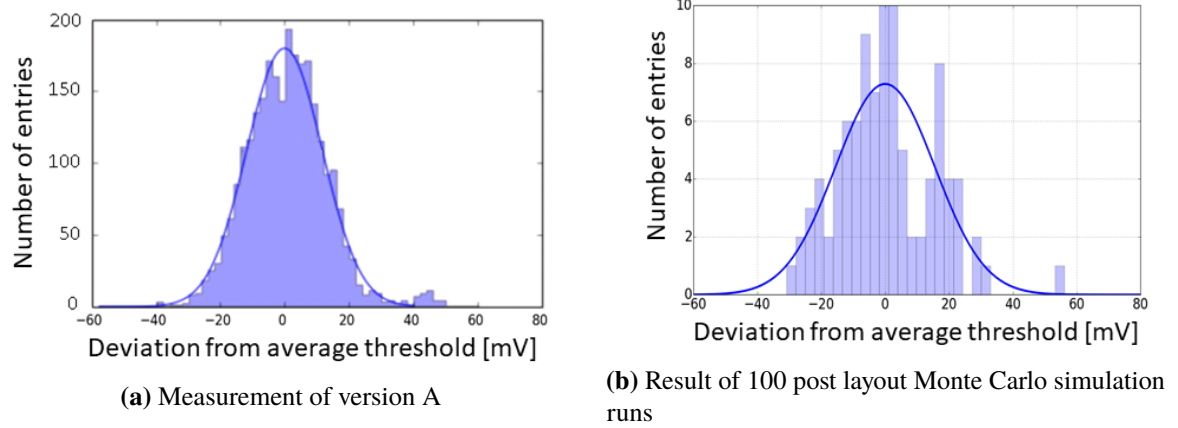**(b)** Result of 100 post layout Monte Carlo simulation runs

Figure 3.41: Histograms of comparator threshold dispersion. Bars represent measurement/simulation results, while the solid blue lines are Gaussian fits [59].

3.3.1.5 Breakdown voltage and depletion depth

As shown in Eq. (3.3) high bias voltage allows for a large depletion zone, therefore the chip structure (especially the guardrings) has to be optimized to allow as high breakdown voltage as possible. Since the used LF15A technology is aimed mostly at commercial imagers, a high voltage capable guardring structure is not provided by LFoundry and had to be developed within our collaboration based on available publications and TCAD simulations [16]. The guardring structure is the same for versions A and B, though the biasing schemes are different in order to accommodate different pixel sensor designs (see Fig. 3.35).

The measured I-V curves are shown in Fig. 3.42. Version A managed to achieve breakdown voltage of approximately 115 V. Using a photo emission microscope the first guardring (the one closest to the pixel matrix) was identified as the origin of a breakdown [23]. In version B the breakdown voltage is unfortunately limited by an AC coupling capacitor connected between the CSA input and the charge collection node (biased with high voltage). Because of a such connection, before the real junction breakdown can occur the dielectric of the mentioned capacitor will break and damage the transistor. For this reason the I-V curve in Fig. 3.42(b) is measured only up to 26 V (already showing much higher leakage current than version A) and the real achievable breakdown voltage is not observed. This shortcoming, along with a few smaller practical issues, ultimately decided about abandoning the sensor implementation from version B in favour of a version A style for the follow-up chips in LFoundry DMAPS family in order to ensure radiation hardness of the future devices.
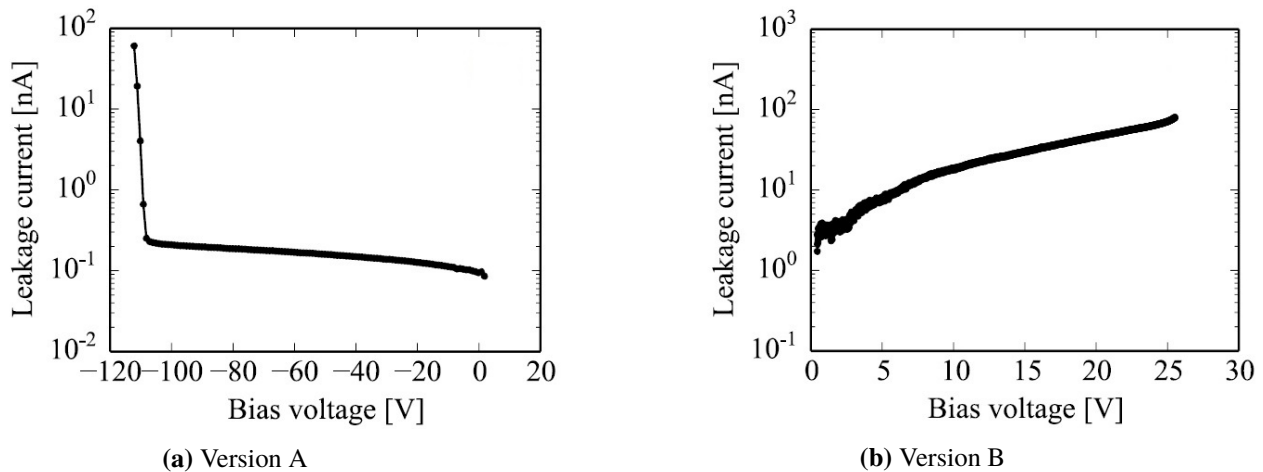

(a) Version A   **(b)** Version B

Figure 3.42: I-V curves of CCPD_LF chips. Please note different scale ranges [59].

The depletion depth was measured using a 2.5 GeV electron beam of an Electron Stretcher Accelerator (ELSA) at the University of Bonn. Spectra of ^{55}Fe, ^{109}Cd and ^{241}Am were used for calibration (energy depositions of 1.6 ke$^-$, 6.1 ke$^-$ and 16.5 ke$^-$, respectively). The single pixel spectra of the beam were measured for several bias voltages for both versions of CCPD_LF. The results are shown in Fig. 3.43. Each beam spectrum was fitted with a Langau function (convolution of Gaussian and Landau functions), the most probable value was translated to electrons and the depletion depth was estimated using Eq. (3.1). The maximal measured depletion depth for version A was 166 μm (at 110 V) and 85 μm (at 20 V) for version B. While the standard thickness of the silicon wafers is 725 μm, some of the CCPD_LF wafers were thinned down to 100 μm and remained fully functional.

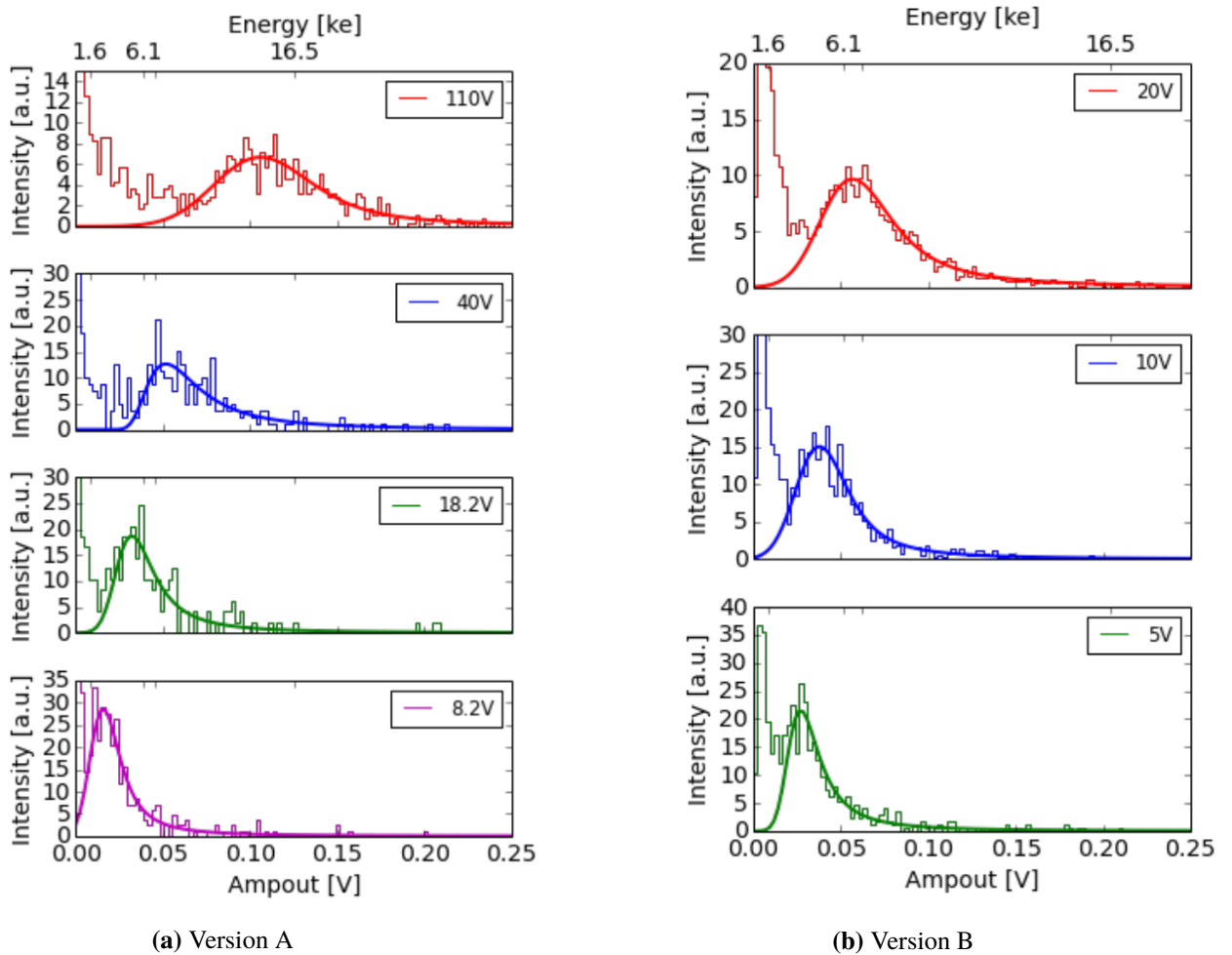This proves that already with this prototype a thinned down detector can be fully depleted.

Figure 3.43: Charge spectra of a single pixel of the CCPD_LF sensor exposed to an electron beam of 3.5 GeV. The bias voltages are indicated in the upper right corner of each plot. The upper axis indicates calibrated energies in electron charge. Solid curves are fits of the Langau function. [23].

Fig. 3.44 shows measured depletion depths as a function of the sensor reverse bias. The solid lines show theoretical depletion depths of a simple pn diode for several wafer resistivities (based on Eq. (3.4)). Measurements from both versions A and B follow quite well the behaviour of a 3 kΩcm case. This agrees with the specification from the wafer supplier stating a resistivity above 2 kΩcm. The observable differences between versions A and B can be caused by resistivity difference between wafers (the variability of resistivity is non-negligible from wafer to wafer, based on all performed measurements) or different configuration of the electric field inside the sensors.
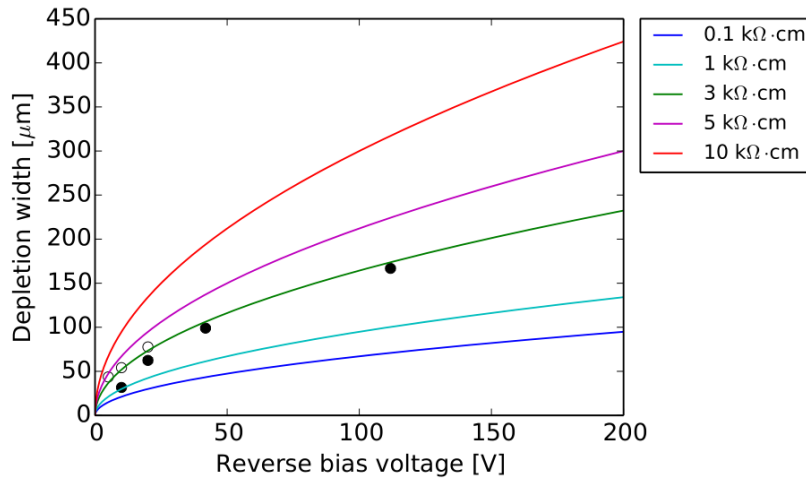
Figure 3.44: Depletion depths of the CCPD_LF estimated from the MPV of a Langau fit. Filled and open circles represent measurement data from versions A and B, respectively. The lines show the calculated depletion widths of planar silicon diodes with various resistivities. [23].

3.3.1.6 Subpixel encoding measurement

In order to test the subpixel encoding feature of CCPD_LF the device was bump bonded to the FE-I4 chip. The height of the output signal of each of the three subpixels in CCPD_LF was set such, that FE-I4 pixel should be capable of distinguishing between sensor pixels and assign them different ToT codes (values of 2, 4 and 8 were chosen, so their combinations should be distinguishable as well). The FE-I4 was tuned to provide as good separation between ToT codes as possible. Next, charge was injected into the sensor using the injection circuit such that every possible combination of the three subpixels occurred the same number of times. The results are presented in Fig. 3.45. It turns out that while the concept is valid and for some "good pixels" the encoding works (Fig. 3.45(a)), there are also pixels which produce wrong results visible as recording the same ToT codes for different subpixel combinations (overlapping bars in Fig. 3.45(b)). This problem is caused by the heights of the signals of three subpixels in the group being set globally, i.e. without the possibility of per-pixel amplitude trimming. Alternatively, a different readout chip than FE-I4 could be used with higher ToT resolution or more per-pixel tuning capabilities.
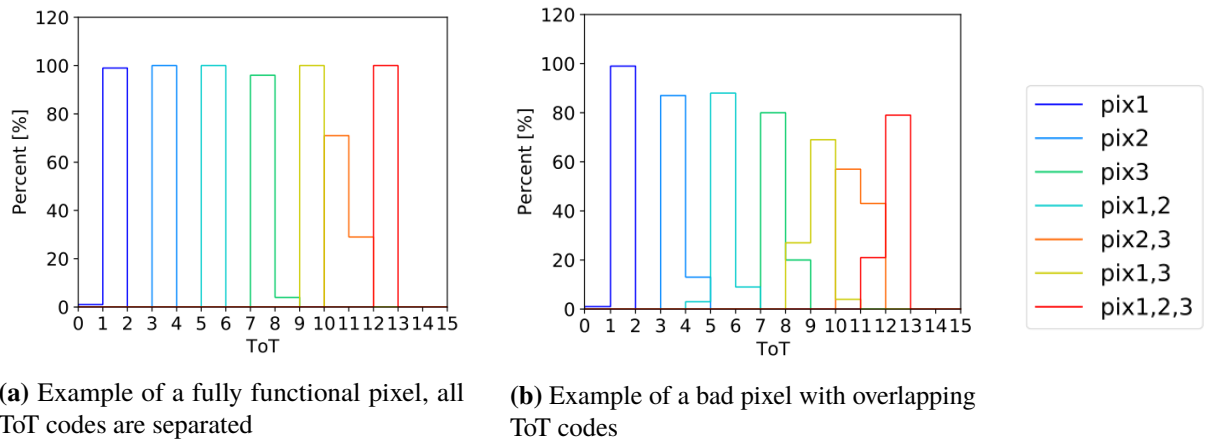
(a) Example of a fully functional pixel, all ToT codes are separated

(b) Example of a bad pixel with overlapping ToT codes

Figure 3.45: ToT values recorded by FE-I4 in response to test pulse injections to CCPD_LF subpixels. All possible subpixel injections patterns were used equally frequently.[60].

3.3.1.7 Time walk measurement

The time walk was measured using an electron testbeam at ELSA. The chip was operated in the "pixel monitoring" mode with the monitoring outputs of CSA and comparator from a single pixel connected to an oscilloscope. The triggering of the oscilloscope was done using a signal from a scintillator, which was placed behind the CCPD_LF in the beam. For this measurement the time walk was defined as the time dispersion of the comparator's output rising edge compared to the scintillator signal. In order for a detector to be suitable for ATLAS needs, the time walk cannot exceed 25 ns (one bunch crossing period) while operating at a threshold providing efficiency above 99% and noise occupancy (amount of fake hits caused by noise) below 10^{-6} hits /mm^2/s. The measurement results obtained for two different threshold settings are shown in Fig. 3.46. The fraction of "in-time" hits (time walk below 25 ns) for the higher threshold (2 600 e$^-$) was 79% and for the low threshold (190 e$^-$) it was 91%. Even the low threshold setting did not allow achieving high enough efficiency and the noise occupancy was 240 hits /mm^2/s, severely exceeding the ATLAS specification.

Version A of CCPD_LF was not measured in the same way, but due to a higher detector capacitance C_d the result would be worse. This shows that the time walk in this prototype is a significant issue and has to be improved for the next chips.
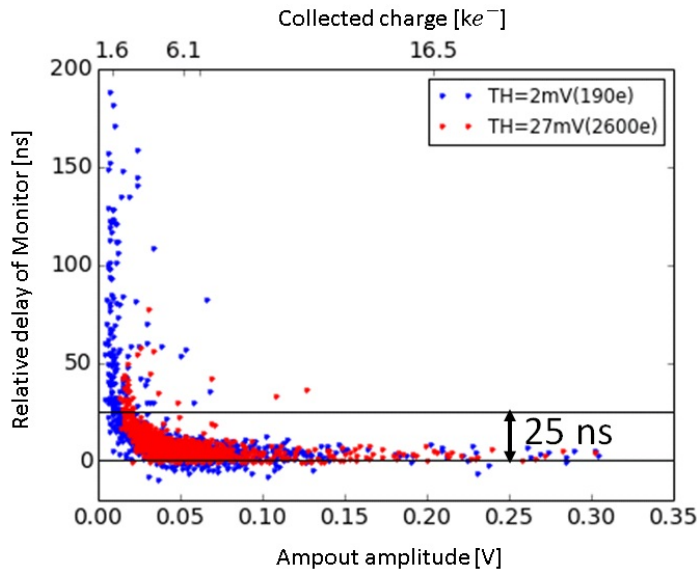
Figure 3.46: Time walk measurement for version B of CCPD_LF. The horizontal line marking 25 ns indicates the time limit for hits to be considered "in-time". [59]

3.3.1.8 TID degradation

In order to evaluate the electronics degradation due to the ionizing radiation the prototype version A was irradiated with X-rays up to 50 Mrad. The irradiation was carried out using an X-ray tube provided by the Institute of Experimental Nuclear Physics Irradiation Center at the Karlsruhe Institute for Technology (KIT) [61]. During the irradiation the chip was not cooled, the average temperature during the measurement was 27 °C. The dose rate was kept constant at 1.48 krad /h for the whole measurement period with short breaks for performing chip characterization at several TID steps.

As described in Section 3.3.1.3, the matrix of CCPD_LF version A was divided into three regions differing by the type of transistor used in the feedback of the CSA. The degradation of gain and noise due to TID for two pixels from each flavour is presented in Fig. 3.47, with the type naming convention: "L" - linear transitor W=0.35 μm L=1.5 μm, "S" - linear transitor W=0.35 μm L=0.9 μm, "ELT" - ELT transitor W≈2.4 μm L≈0.16 μm. Few observations can be made from those results:

- Despite a noticeable variation between pixels of the same flavour (expected based on pre-irradation measurements of dispersion) the differentiation between flavours is visible in gain: the devices with larger L and ELT transistor suffer less degradation than the short linear transistors. This suggest that the performance loss is induced by the transistor leakage current affecting the bias current (nominal current value is 400 pA, therefore even small alteration can be noticeable).

- The increase of ENC with TID follows approximately the same trend for all pixel types (the measurement uncertainties, not indicated on the Fig. 3.47(b) are in the order of 20%).

- While overall performance degradation is significant (40% gain loss and 75% ENC increase in worst case), it is not catastrophic and can be partially compensated for with changing the bias currents.

- The sudden drop in gain around 1 Mrad and partial recovery after approximately 10 Mrad was reported for other technopolis too [62] [63], but the physical origin of this phenomena is not clear.
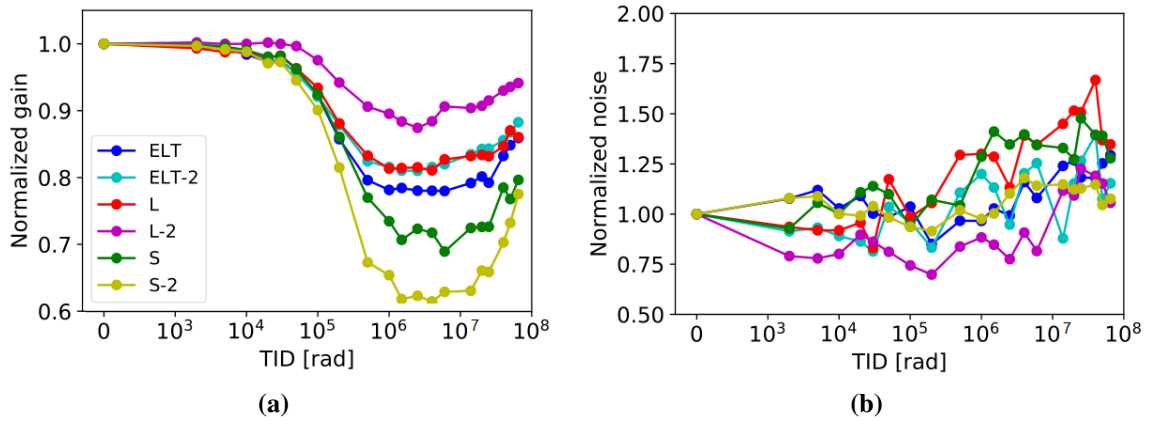

(a)

(b)

Figure 3.47: Degradation of gain and noise of the CSA due to TID in CCPD_LF version A. Both gain and noise values are normalized to their respective pre-irradiation values. Two pixels from each of the three flavours are shown, with the difference between flavours being the type of transistor in the CSA feedback (naming scheme explained in text). Chip configuration, including bias currents, was kept constant during irradiation [23].

3.3.1.9 Noise coupling between sensor and electronics

One of the big challenges of combining the sensor and electronics in a monolithic design is preventing injection of charge into the sensor due to CMOS circuit activity, as mentioned in Section 3.2.3.1. Such coupling can unfortunately be observed in the CCPD_LF version A when toggling a global digital line in order to disable or enable the "stand-alone readout" mode, as shown in Fig. 3.48. When this line switches a large charge (several ke⁻ with large chip-to-chip variation) is injected into the sensor, causing a large CSA response and the comparator indicates a hit. Since by design the "stand-alone readout" has to first be disabled in order to record hits, then enabled in order to read them out, the described coupling mechanism makes operating the chip difficult. A work-around for this problem was to temporarily increase the comparator threshold shortly before switching modes, however this limited the maximum operational readout rate and made testing difficult.

The coupling issue was much less severe in the version B of CCPD_LF, which has the same circuitry but completely different physical layout. This means that the coupling path was produced by the parasitic elements in the layout, though it was never fully reproducible in post-layout simulations.
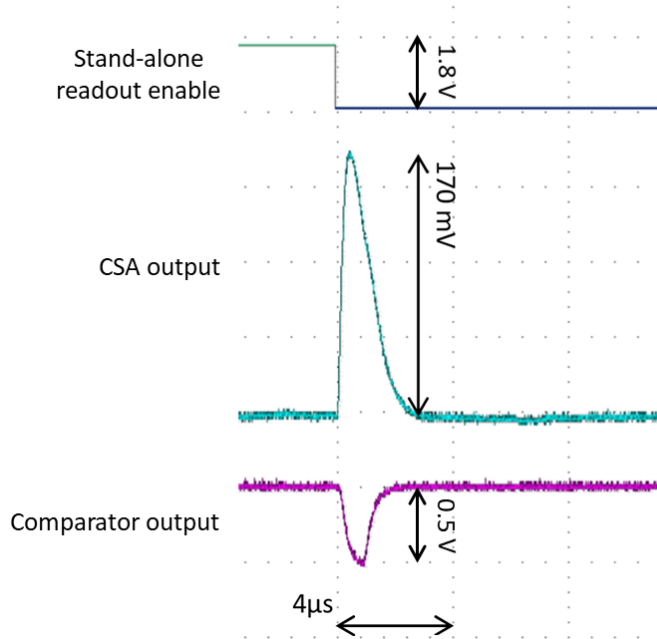
Figure 3.48: A screenshot from an oscilloscope showing a large CSA signal and comparator firing as a response to a global digital line (stand-alone readout enable) toggling. Based on [16].

3.3.2 LF-CPIX

Thanks to the encouraging results obtained with the CCPD_LF prototype the collaboration decided to proceed with large fill factor DMAPS designs. Instead of immaterially designing a monolithic chip with fast readout, first a stepping-stone design was made, to which the features required for fast readout would be added later on. This intermediate design was named LF-CPIX. The goal of this design was also to fix bugs found in CCPD_LF and test several different pixel variants.

The design and measurement of LF-CPIX was not part of this thesis due to a time conflict with the work needed on the Clock Data Recovery circuit described in Chapter 4. However, due to the importance of LF-CPIX in the overall LFoundry DMAPS family a short description of the design and measurement results is provided below. More details can be found in [23] [64] [65] [66].

- The size of the chip was increased to approx. 1 cm × 1 cm in order to test the feasibility of good power and biases distribution on a relatively large chip with the available metal stack.

- As shown in Fig. 3.34 two versions of LF-CPIX were produced, with the only difference between them being the guardring structure around the chip. The guardring structure was an evolution of the one used in CCPD_LF, the main change was increased distance between the innermost p-well guardring and n-well ring around the matrix. The breakdown voltage measurements showed improvement over the previous chip: version 1 of LF-CPIX had a breakdown voltage of 130 V and version 2 achieved 215 V.

- Since LF-CPIX was a stepping-stone prototype on the way to the fast monolithic detector, its pixel size had to be chosen to allow addition of fast readout logic in the next design. This

requirement lead to the pixel size increase to $50\,\mu m \times 250\,\mu m$. The space allocated to the readout circuitry of future design was filled in LF-CPIX by the power decoupling capacitors and the bump bonding pad.

- LF-CPIX has the same three modes of readout as CCPD_LF: a slow stand-alone one using a global shift register, single pixel monitoring mode and readout through bump bonding to FE-I4 (one-to-one pixel, without subpixel encoding). However, not all pixel types incorporated in the matrix can be used with all three modes.

- The matrix of LF-CPIX is composed out of 158 rows and 36 columns divided among several types of pixels, as illustrated in Fig. 3.49

    - Passive pixels include only the sensor and a charge injection circuit, no active electronics. These pixels were added in order to compare the LFoundry DMAPS sensor design to other sensor types when bump bonded to the FE-I4

    - "AnaDig" pixels have the same in-pixel electronics as CCPD_LF pixels (see Fig. 3.36) with some transistor size optimization. In addition to the CSA with pMOS-input transistor from CCPD_LF also a version with an nMOS-input device is implemented. The comparator architecture is identical as in CCPD_LF, however, the size of the input pair transistors was increased in order to reduce the threshold dispersion.

    - "Ana" pixels are similar to "AnaDig" pixels, but the comparator is removed in most of them and the CSA output is send either directly or after shaping to the FE-I4. This part of the matrix also hosts the third type of CSA, one with two input devices in parallel (one nMOS and one pMOS). Such design offers potentially a higher g_m than the other two flavours at the same power consumption. All three CSA are biased with $14\,\mu A$ in the main branch. A more detailed comparison of the front-ends is discussed in Section 3.3.3.

- LF-CPIX was irradiated with X-rays up to $50\,\mathrm{Mrad}$ following a similar procedure as for CCPD_LF. The degradation of gain and noise is presented in Fig. 3.50. When comparing to Fig. 3.47 the results for the linear transistor W=$0.35\,\mu m$ L=$1.5\,\mu m$ in the feedback should be considered, since all CSA in LF-CPIX use this configuration. The radiation hardness of the LF-CPIX is much improved, however, due to many changes of the circuit (higher bias currents, improved pixel layout, adjustments of CSA transistor sizes) it is difficult to point out the exact cause for this.

- When tested in the electron beam the LF-CPIX managed to achieve a hit detection efficiency of 99.4%. This was achieved with a fully depleted chip configured to operate with the threshold of $1\,700\,e^-$ (un-irradiated, without cuts on hit timing). The limiting factor on the threshold setting was the digital activity coupling, same as the one described in Section 3.3.1.9.
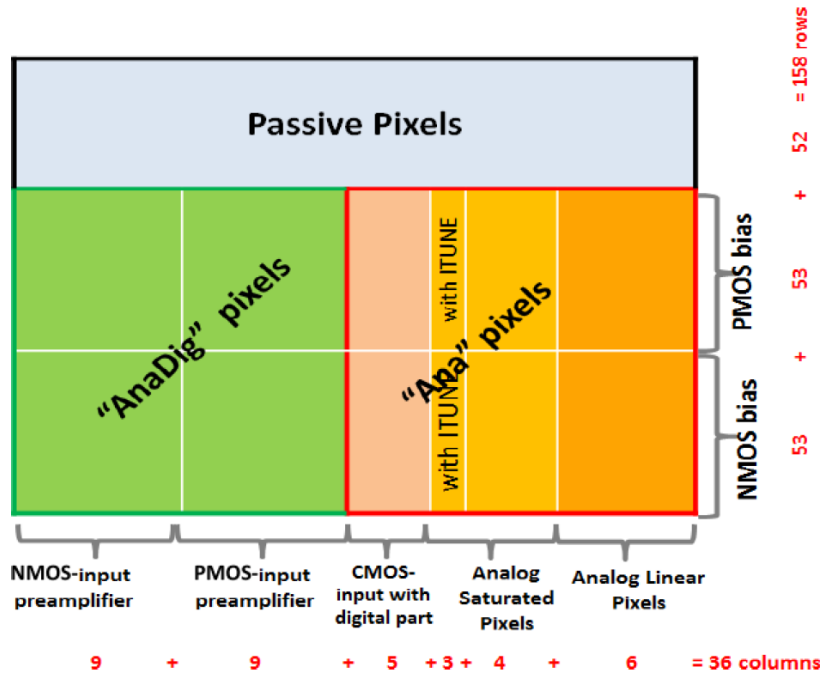
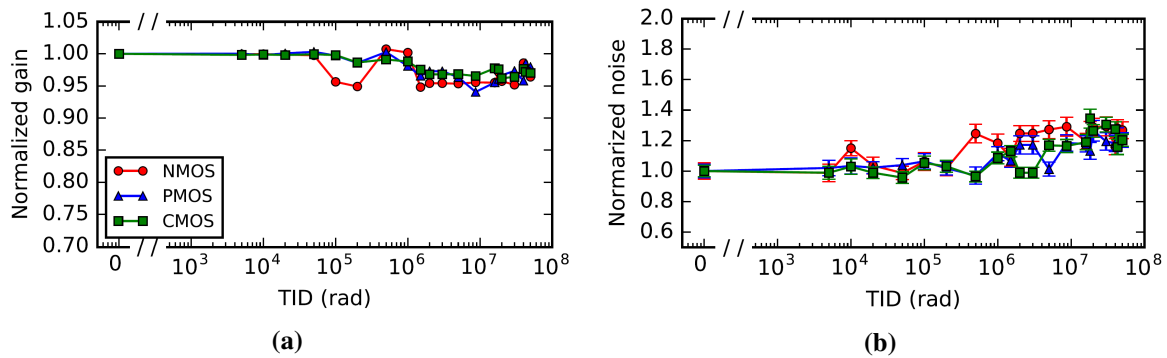**Figure 3.49:** Organization of the pixel types inside the matrix of LF-CPIX [64].

Figure 3.50: Degradation of gain and noise of the CSA due to TID in LF-CPIX. Both gain and noise values are normalized to their respective pre-irradiation values. Chip configuration, including bias currents, was kept constant during irradiation [67].

3.3.3 LF-Monopix1

The LF-Monopix1 is a large scale (approx. $1\,cm \times 1\,cm$) DMAPS prototype chip inheriting many features from LF-CPIX (sensor design, analogue electronics) with the addition of a fast integrated readout, allowing it to cope with hit rates of $100\,MHz/cm^2$ (depending on hit topology). The goal of this design was to prove the feasibility of such an approach. While achieving ITk ATLAS outer layer performance and radiation hardness needs were desired, the chip was still conceived as a test vehicle for several pixel variants.

At the time of chip submission for production (August 2016) it was the first device, to author's best knowledge, to incorporate both such complex logic and a radiation hard sensor.

3.3.3.1 Sensor design and pixel layout

LF-Monopix follows the large fill factor design approach of version A of CCPD_LF. The cross section of a pixel with indicated chip guardring is shown in Fig. 3.51(a). The very deep n-well (DNW) is used as the collection electrode with all the electronics placed on top of it. The electronics is isolated from the silicon bulk on the side thanks to n-well (NW) and deep n-well (NISO) implants around the DNW, with the transistor's n-well isolated from DNW thanks to deep p-well (PSUB). The p-well ring between pixels is a p-stop implant used to isolate neighbouring pixels and stop radiation-induced surface currents.

The top view of a pixel is shown in Fig. 3.51(b). As mentioned in Section 3.3.2 the LF-CPIX and LF-Monopix1 were designed from the begining to have the same pixel size of $50\,\mu m \times 250\,\mu m$. As can be seen from the layout, the sensor (DNW) occupies about 55% of the total pixel area, since the rest of the space was required in order to implement a p-stop ring around pixel which would comply with the manufacturing rules and not limit the overall chip breakdown voltage. Using NW, NISO and PSUB implants the "tub" in which the electronic circuits are placed is divided into two sections, separating the analogue and digital domains. This was done in order to minimize the interference from the digital circuit activity on the sensitive analogue parts of the readout chain. Additionally, instead of connecting the bulk of the digital part to the digital ground directly, it was tied to ground with a dedicated, separate metal line. This makes the potential of the digital bulk more stable, thus minimizing the chances of charge injection to the sensor through C_{pw} (see Fig. 3.51(a)).

As described in Section 3.2.3.1 the junction capacitance between DNW and PW/PSUB implants increases the detector capacitance C_d at the input of the CSA. While all the circuitry was hand-designed and the layouts were made in a way to minimize this effect, the increase is still substantial - based on TCAD simulations and test structure measurements [16] the C_d of LF-Monopix1 pixel is approximately $400\,fF$, while an equivalent completely passive sensor would have C_d of about $100\,fF$. This makes a significant impact on the design of the CSA and the power consumption if fast response time and low ENC are desired.

Same as previous prototypes in the LFoundry DMAPS family, the LF-Monopix1 was manufactured on high resistivity (above $2\,k\Omega cm$) p-substrate wafer. The guardring structure was copied from LF-CPIX version 2 with small modifications to the guardring spacing in order to further enhance the breakdown voltage.
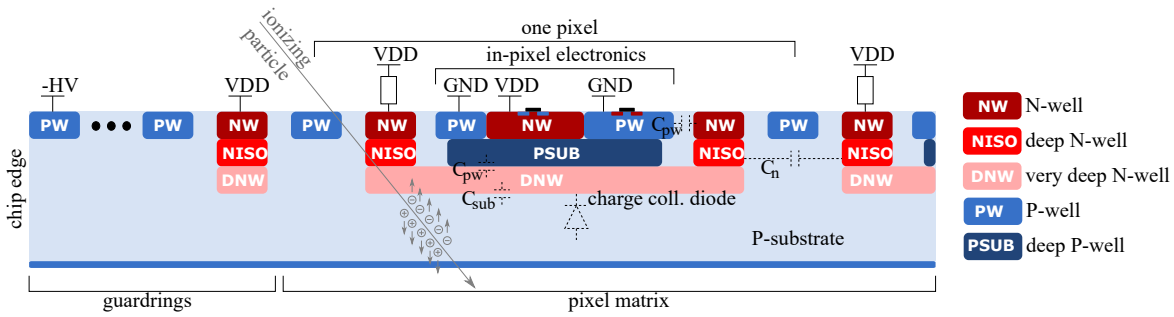
(a) Simplified cross-section of the sensor implementation. Parasitic capacitance contributions to the sensing node are indicated with respective symbols. The drawing is not to scale and only aims to indicate the relative positions of the implantation wells (in reality the depth of wells is few μm, while the bulk is several hundreds of μm thick). Based on [68]
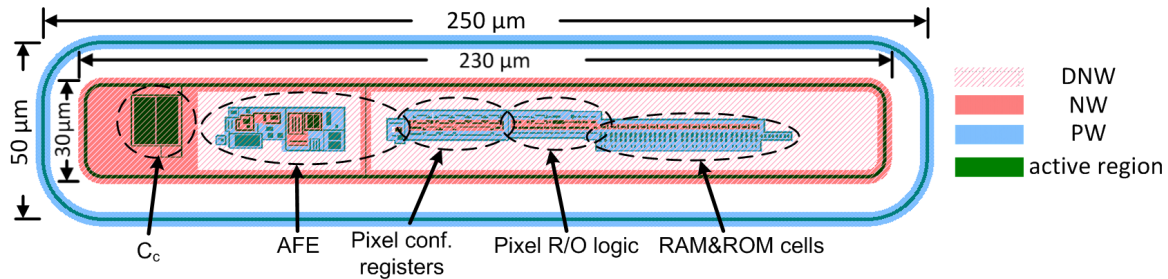
(b) Example of a layout of a pixel (most of the layers removed for clarity) with indicated placement of electronics circuits. The NW and DNW layers are connected with the NISO layer (not showed). The active region is used for implementing transistors or heavily doped contacts to the silicon. The layout is done in a way allowing to construct the matrix by abutting the pixels. [67]

Figure 3.51: Cross-section diagram and pixel top view implemented in LF-Monopix1.

3.3.3.2 In-pixel electronics

The schematic of in-pixel electronics implemented in LF-Monopix1 is shown in Fig. 3.52. It follows the general concepts described in Section 3.1.3. The signal created in the sensor is received by the CSA through the AC-coupling capacitor C_{ac}. The CSA works in a closed loop configuration with a capacitor C_f and a current source in the feedback branch. The output of the CSA is connected to the input of the comparator through an AC-coupling capacitor C_c. The inclusion of this capacitor allows to set the baseline input voltage of the comparator independently of the CSA DC-voltage output, making it easier to keep the comparator's input device in the desired operating region. The threshold voltage V_{th}, against which the CSA output is compared, is set globally for all pixels. Since a mismatch of devices is unavoidable in CMOS process, each comparator can be independently trimmed using an in-pixel 4-bit current DAC (the pixels are configured using a shift register running through entire matrix). If the signal generated by the CSA crosses the set threshold the comparator sets the HIT output to digital high state (VDDD = 1.8 V) and keeps it high until the CSA output goes back below threshold. The time of comparator firing and returning to zero is recorded by the readout (R/O) logic block with the precision defined by BCID change rate. Afterwards, the R/O logic indicates through

TOKEN lines to the rest of the chip that a hit has occurred and readout process should start. The readout procedure is controlled with the CONTROL lines and the hit timestamps together with the pixel address are readout via the DATA bus. The following paragraphs describe each of the building blocks in more detail.
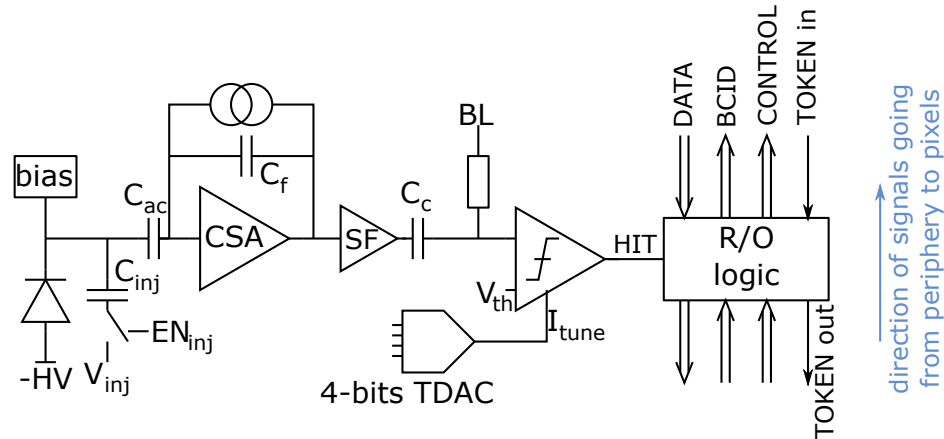

Figure 3.52: Schematic of the in-pixel logic implemented in LF-Monopix. The wide arrows connected to the R/O logic block indicates multi-bit buses.

CSA   Two types of CSA have been implemented in the LF-Monopix1. Both of them are based on a folded cascode architecture, but one of them uses an nMOS transistor as the input device (Fig. 3.53(a)), while the other one uses complementary pMOS and nMOS input devices (Fig. 3.53(b)). The benefit of the CMOS implementation is an increase of the transconductance for a given bias current ($g_{m\,\text{total}} = g_{m\,\text{nMOS}} + g_{m\,\text{pMOS}}$), which should provide better performance. The comparison of transient response to $4\,\text{ke}^-$ charge of both preamplifier types is shown in Fig. 3.54. For this comparison both preamplifiers were biased with the same current of $14\,\mu\text{A}$ and the CMOS CSA achieves noticeably faster peaking time of $25\,\text{ns}$ compared to $40\,\text{ns}$ for the nMOS CSA (when considering $400\,\text{fF}$ detector capacitance). The gain and noise performance of the preamplifiers are compared in Fig. 3.55, but here bias currents are set to default values for LF-Monopix1 ($I_{\text{bias CMOS}} = 15\,\mu\text{A}$, $I_{\text{bias nMOS}} = 17.5\,\mu\text{A}$). In this comparison again the CMOS CSA performs better.

The disadvantage of the CMOS preamplifier is the need to control the bias current with a dedicated regulator [64], indicated by the VDDPRE power domain in Fig. 3.53(b). This not only requires designing an additional, delicate block placed in the chip periphery, but also requires routing an additional power domain rail in the matrix in such a way that minimizing any potential noise coupling to it is ensured.
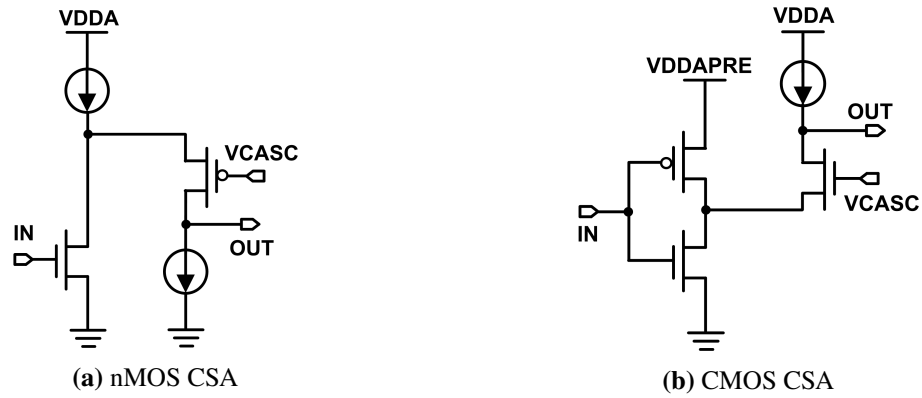
(a) nMOS CSA

(b) CMOS CSA

Figure 3.53: Schematics of two preamplifier circuits used in LF-Monopix1. [67].
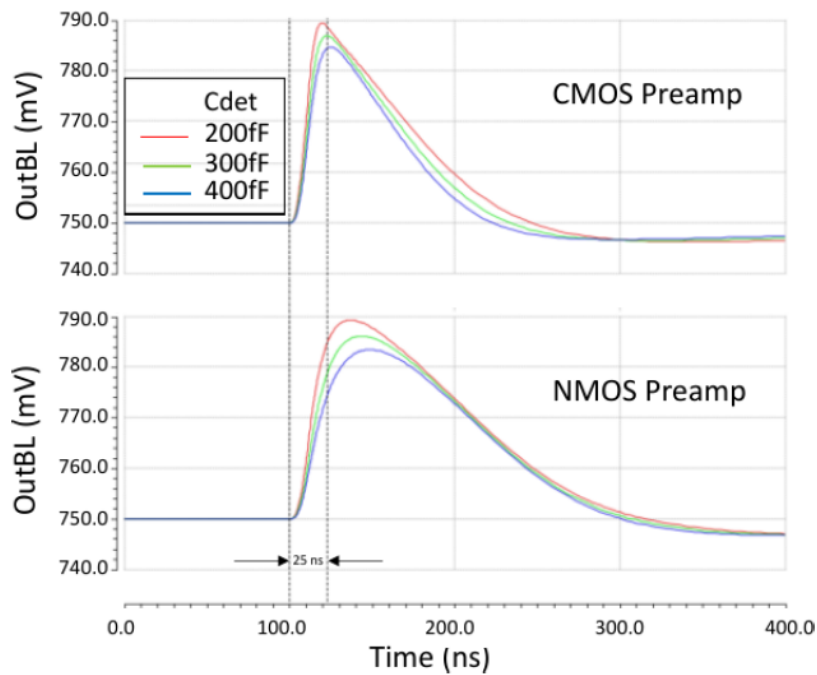

Figure 3.54: Simulated response of both preamplifier types to 4 ke⁻ charge injection for some typical detector capacitance values. Both CSA are biased with $14\,\mu A$. [64]
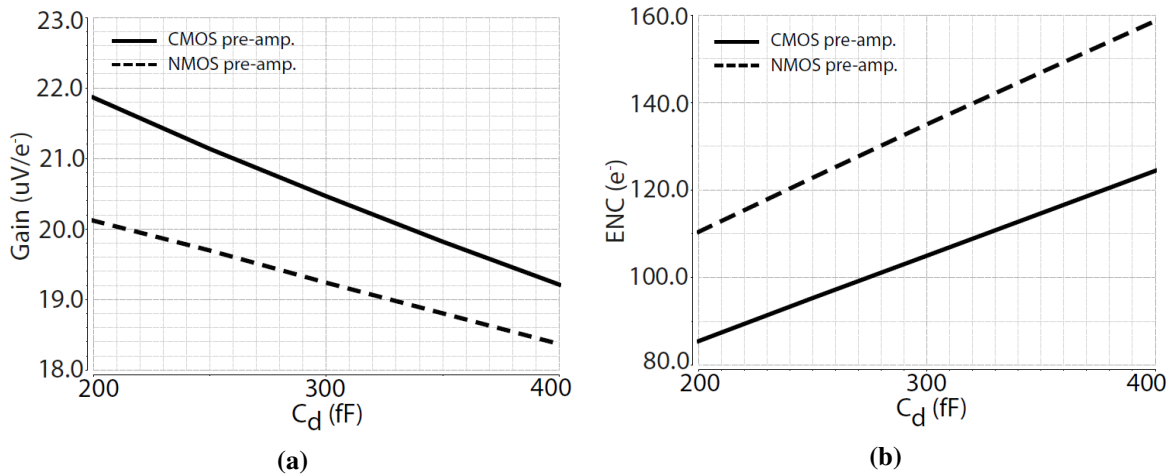
Figure 3.55: Comparison of the gain and the noise performance of the preamplifiers used in LF-Monopix1. [69].

Comparator LF-Monopix1 incorporates two different designs of the comparator. The two-stage open loop design shown in Fig. 3.56(a) is the same architecture used in previous LFoundry chips with increased size of the input transistors (mismatch reduction) and the bias current set nominally to 4.5 μA. The second design, shown in Fig. 3.56(b), is based on a self-biased differential amplifier with a CMOS output stage [70]. Similarly to the CMOS CSA, this discriminator also utilizes both nMOS and pMOS transistors as the input stage to take advantage of the increased transconductance for higher gain and quicker response time. The nominal bias current for this comparator is 3.5 μA. In simulation the timewalk achieved by the comparators was 34 ns for version 1 and 23 ns for version 2 (threshold set in both cases to $1\,500\,e^-$).
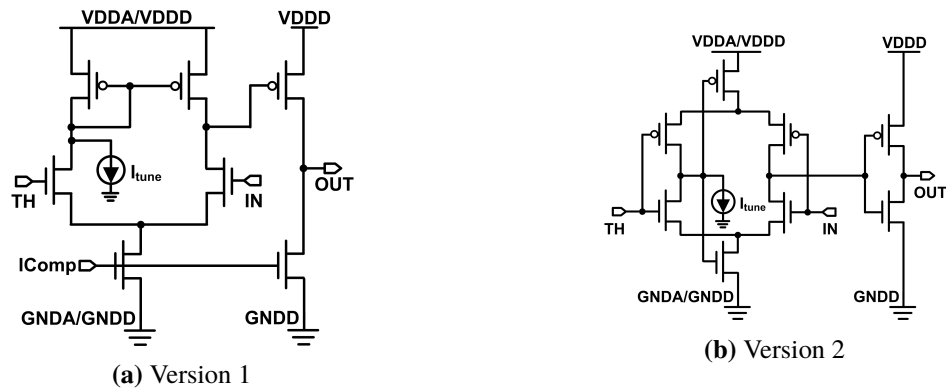
Figure 3.56: Schematics of comparators implemented in LF-Monopix1. [67].

The tuning current I_{tune} indicated in Fig. 3.56 is generated by the 4-bit trimming DAC shown in Fig. 3.52 and can be adjusted individually per pixel.

Because the comparator is the interface between the analogue (CSA) and the digital power domains it is not obvious from which domain its first stage should be powered. While the first comparator

stage is on its own a delicate analogue circuit, which should be capable of distinguishing the input crossing the threshold voltage precisely, at the same time it creates a fast transient large-swing signal, which could potentially couple back to the CSA and cause glitches. Such effects are difficult to simulate, therefore in LF-Monopix1 some pixels use comparators with the first stage powered from VDDA, while other pixels are powered from VDDD in order to compare the two configurations in measurements.

In-pixel readout logic The digital readout logic is based on the column drain architecture, which is used in the readout chips currently running inside the ATLAS pixel detector [30]. The schematic of the in-pixel logic and the timing diagram of one readout cycle are shown in Fig. 3.57.
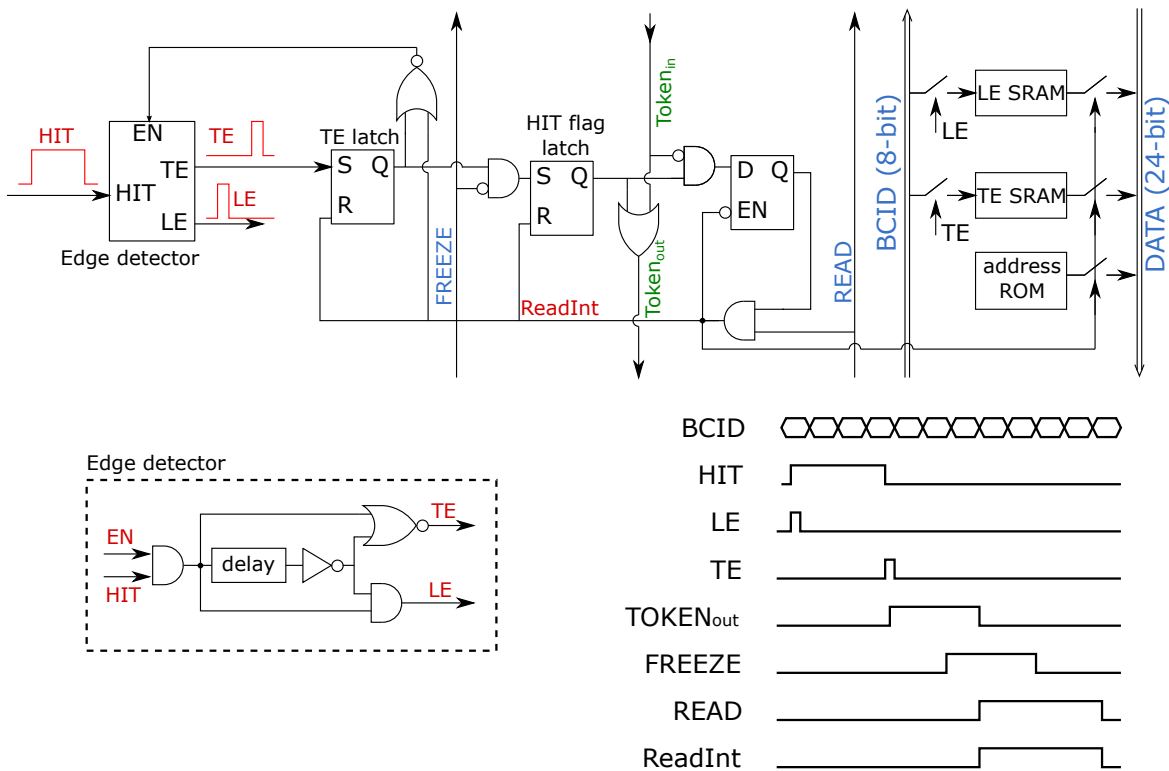
Figure 3.57: Schematic of the in-pixel readout block together with the timing diagram for the most important signals.

The readout logic is activated with the rising edge of the HIT signal (comparator output). The edge detector detects this event and creates a 1 ns long pulse LE (leading edge), which causes the LE RAM cell to save the current BCID value. The BCID (Bunch Crossing ID) is the timing information bus running at 40 MHz. It is 8-bit wide, therefore it rolls-over every $256 \times 25\text{ns} = 6.4\,\mu\text{s}$, which was deemed long enough to be sure that hits can be recorded and safely be read out. The falling edge of HIT is also detected by an edge detector, which makes a short pulse TE (trailing edge) causing TE RAM cells to store the BCID value. The LE timestamp therefore provides the Time of Arrival (ToA) of the hit, while the difference between TE and LE timestamps is the Time over Threshold (ToT) information.

The falling edge of the TE signals starts the procedure of reading out the stored data. The TE sets the "TE latch", which then sets the "HIT flag latch" (FREEZE is low at this time) high, causing "TOKEN out" signal to go high too. The TOKEN propagates through the column to the End of Column (EoC) logic block to indicate the hit happening in the given column. From the EoC the information about the hit happening is sent to the readout controller, which responds by first sending the FREEZE signal to the column. When FREEZE is set, the "HIT flag latch" is disconnected from "TE latch", so that any new hits do not disturb the data readout (they can still be recorded in other pixels by the "TE latch" and readout later). Afterwards the readout controller sends the READ signal to the column, which activates the ReadInt signal in the pixel with highest priority in the column (pixel with highest row number, which does not have "TOKEN input" active). Once the ReadInt is activated, it allows the data stored in the pixel (8-bit pixel row address stored in ROM memory, 8-bit LE timestamp, 8-bit TE timestamp) to be written to the column-common DATA bus. The READ signal is kept high during the DATA retrieval process. Once READ is deactivated, the in-pixel ReadInt of hit pixel goes low too, which clears the "HIT flag latch". This allows the TOKEN signal of given pixel to go down. The release of the TOKEN will propagate down the column to either be stopped by another pixel indicating hit, or completely release the readout cycle from the column. If more hits are present in the column the readout controller would send READ signals until all hits are drained from the column. If hits are indicated in multiple columns the priority is arbitrated in a similar manner, with the left-most columns (arbitrary choice) having higher priority.

Since the main goal of the LF-Monopix1 chip was to prove the feasibility of integrating fast digital processing in DMAPS device, a lot of attention and effort was put in this part. Some of the design considerations and used techniques were:

- All the logic gates used in the readout block were designed and layout was arranged by hand either from scratch or based on standard digital library cells. This was done in order to minimize the number and the size of transistors in effort to reduce the contribution of C_{pw} parasitic capacitance (see Fig. 3.51(a)).

- The BCID and DATA buses are designed to be differential in order to minimize the crosstalk and all lines expected to have fast switching transients were shielded from sensitive layout parts with grounded metal plains.

- The BCID in this chip is Gray coded, which reduces switching activity, therefore potentially saving power and reducing noise created by digital switching.

- While the TOKEN is shown to propagate through the OR gate in the pixel in Fig. 3.57, the real design uses alternating NOR and NAND gates. This provides the same overall functionality, while allowing to remove one inverter per pixel (CMOS OR gates are realized as a NOR gate plus a NOT gate), thus reducing TOKEN propagation delay along the column and reducing the area.

- While the readout logic is inactive in pixels without hits, the TOKEN propagation down the column causes switching activity in all the pixels it passes through. This could potentially cause charge injection to the sensor and result in fake hits. In an attempt to avoid this the logic gates of the TOKEN propagation chain were implemented using Current Steering CMOS (CS-CMOS) gates [71]. Schematic of a CS-CMOS NOR gate is shown in Fig. 3.58(b). It is similar to a traditional CMOS NOR gate (shown in Fig. 3.58(a)), but with an addition of a few elements to

reduce the current spikes during logic state switching. Transistor M_1 limits the total current which can be drawn by the gate (the current value is set globally). In steady state this bias current flows through either M_2 or M_3 (depending on output state Q) and charges the capacitor C_{CS}. When Q is switching value from 0 to 1 the M_2 transistor will turn off and the M_3 will turn on (the opposite will happen during Q switching from 1 to 0). During the switching time the current drawn by the gate is kept steady by capacitor C_{CS} discharging, bus keeping the total current drawn by the NOR gate constant. According to the simulation results shown in Fig. 3.59 the CS-CMOS scheme significantly reduced the parasitic charge injection to the sensor, thus avoiding generation of fake hits. The penalty is large increase in the gate area, because the C_{CS} has to be large to stabilize the current (in LF-Monopix1 the largest capacitor which could fit inside the pixel was 200 fF handmade on metal 2 layer, taking up the whole space right of R/O logic visible in Fig. 3.51(b)). Additionally the current consumption per pixel is increased, due to a constantly drawn bias current. For LF-Monopix1 the bias current is set to 3 μA, which allows to achieve a TOKEN propagation time from top to bottom of the column of 35 ns according to the results of post-layout simulations.
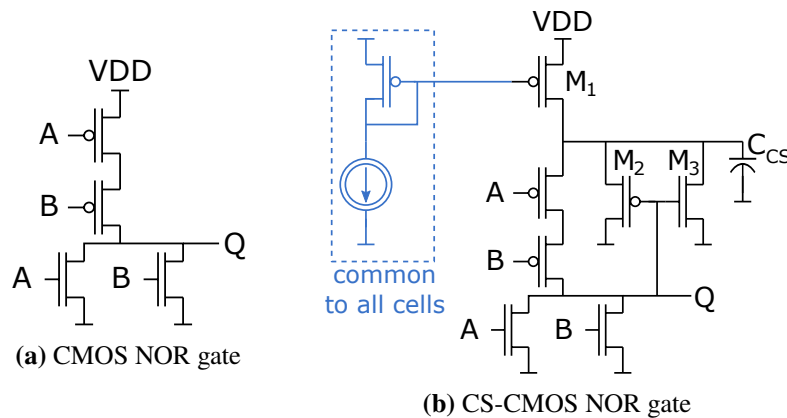
Figure 3.58: Schematics of NOR gates designed using standard CMOS approach and Current Steering CMOS (CS-CMOS).

- The SRAM (Static Random Access Memory) cell used to store the BCID timestamp shown in Fig. 3.52 conceptually is realized as two inverters connected back to back with switches connecting them to lines used for writing the data in and reading the data out, as shown in Fig. 3.60(a). The BCID and DATA lines are shared by all pixels in the column, therefore the SRAM cells have to be designed in a way that allows a reliable operation while interfacing with long metal lines (the extraction of the final layout showed that a capacitance of the 1 cm long line is about 2 pF). The BCID lines are driven by large buffers in the chip periphery, so the input switches of the SRAM cell can be small and reliable writing to SRAM is not problematic to achieve. However, when the pixel is hit and SRAM cells need to be readout (write to the DATA bus) the design challenge is larger. The readout has to be quick, so that little deadtime is generated and the effort of writing the correct value to the DATA bus cannot cause noise injection into the pixel. For LF-Monopix1 several different architectures of SRAM cells were studied in order to find the most suitable one:
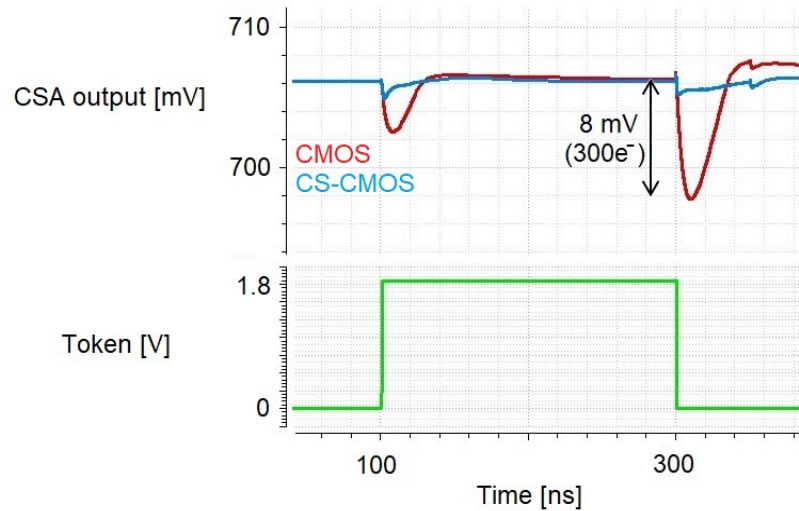
Figure 3.59: Results of transient simulation of the TOKEN signal passing through a pixel. The top waveforms show the output of the CSA as a reaction to the edges of the TOKEN signal. Simulation included layout parasitics for both CMOS and CS-CMOS cases.

– The simplest implementation uses only CMOS inverters, as shown in Fig. 3.60(b). This approach results in the smallest area occupied by a SRAM cell out of all considered designs, however during inverters switching a sharp current spike is generated in power rails, potentially leading to noise injection into sensor. Additionally, during writing the stored value to the DATA bus, the large capacitance of the bus line has to be driven by the inverter, which also leads to large current flow through the SRAM cell, can disturb the sensor and is slow.

– To ease the writing to the DATA bus a buffer can be inserted between the memory element (inverter) and the bus line. One possible implementation is a source follower, as shown in Fig. 3.60(c). Transistors M_1 and M_2 are placed in each SRAM cell, while the load resistors are placed only once, in the chip periphery. The area increase compared to pure CMOS implementation is therefore not large. Since M_1 and M_2 are nMOS transistors, in the pixel the current goes only through VDD terminal and though GND only in periphery, which is beneficial because the disturbance of GND in the pixel couples more easily to the sensor (see C_{pw} in Fig. 3.51(a)). The disadvantage of this approach is the necessity to bias the source follower, which increases the power consumption of the column.

– The last studied SRAM architecture is based on the Current Steering concept (the same as used for TOKEN). The schematic of a single cell is presented in Fig. 3.60(c). As described previously, the M_1 transistor controls the total current which the cell can draw, while transistors M_2 - M_5 and the capacitor C_{CS} assure that there are no sudden spikes in the drawn current. There are, however, several drawbacks of this approach. Firstly, the size of the cell is significantly larger than the two other implementations and since there are 16 SRAM cells in each pixels the total area penalty is large. Secondly, each SRAM cell consumes a static current, which causes a much larger impact than biasing each line of the bus as in the source follower based implementation. Lastly, the inverter of the memory

element still has to drive the DATA line during pixel readout.

All three approaches were studied in simulations. The SRAM circuits were optimized together with the parameters of the bus (used metal layers, their thickness and shielding scheme). An example of post layout simulation comparing all three designs is shown in Fig. 3.61. In this simulation column of pixels was simulated and the presented waveform shows the output of the CSA during SRAM activity. The pure CMOS implementation causes the largest injection to the sensor. The current steering architecture does not help much in this case, since the DATA line is driven by the inverter and therefore only a small improvement is achieved thanks to limiting the current that the cell can draw (1 µA was used in this case). The source follower implementation outperforms the other two very clearly and because the drawbacks of SF-CMOS are acceptable, all SRAM cells in LF-Monopix1 are implemented this way (the bias current was set to 20 µA per line).

(a) SRAM concept

(b) CMOS SRAM

(c) SF-CMOS SRAM
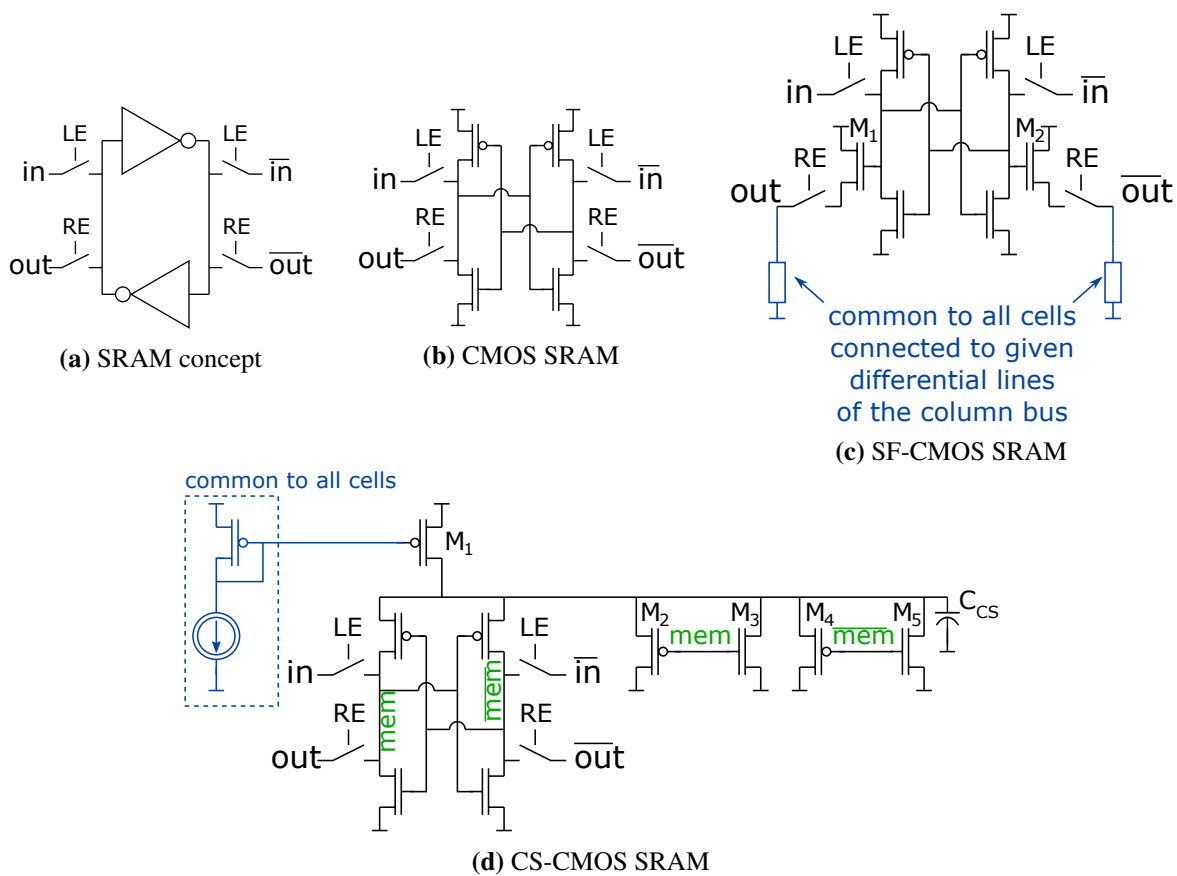
(d) CS-CMOS SRAM

Figure 3.60: Schematics of different implementations of the SRAM cell: pure CMOS, Source Follower (SF) CMOS and Current Steering (CS) CMOS. The input ports are connected to the BCID bus, while the output ports are connected to the DATA bus.

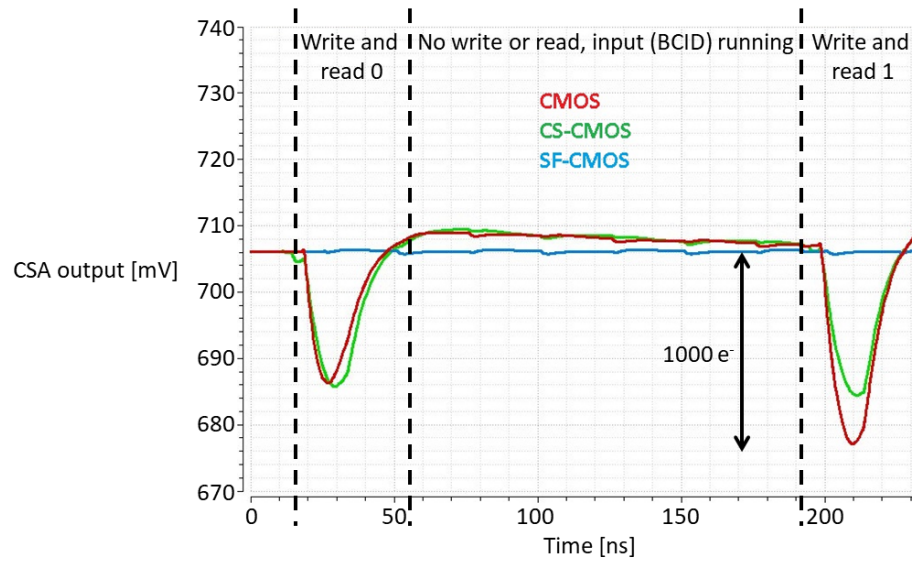• The R/O block placement impacts the design of the block itself, as well as the entire column. One

Figure 3.61: Results of transient simulation of a single pixel in the column being hit. The waveforms show the output of the CSA as a reaction to the activity in the SRAM block, simulated with post layout parasitics. The value indicated in electron units is an estimation based on a comparison to simulations of the charge injection into CSA.

approach is to place the R/O block inside the pixel, in which case the output of the comparator connects to it immediately and the BCID & DATA buses run along the whole column, as shown in Fig. 3.62(a). This was the approach used for all descriptions provided so far. Alternatively, the R/O blocks can be placed at the end of the column and are connected one-to-one with the pixels, as presented in Fig. 3.62(b). In this approach an additional buffer is needed after the comparator in order to drive the long metal line (worst case approx. 1 cm) to the periphery. Such architecture was used in other designs [72]. The benefits of removing R/O block from pixel are reduced detector capacitance C_d and no digital logic switching activity (lower in-pixel noise injection potential). The drawbacks are a need for a strong buffer to drive the line (source follower was used in this design) which increases power consumption, increase in the size of the insensitive to particles area of the chip and difficult to control and equalize propagation time of HIT signal from pixel to periphery. Additionally, the layout of the one-to-one connections must be done very carefully in order to minimize the cross talk between the lines. This can be a limiting factor on the length of the column realizable with this technique, since the routing resources are limited by the number of available metals and the width of the column.

3.3.3.3 Pixel matrix

The matrix of the LF-Monopix1 is composed out of 129 rows and 36 columns. Nine different pixel flavours were implemented, each of them occupying 4 full columns. The distribution of the pixel variants in the matrix is shown in Table 3.1. All the feature variants were described in the previous paragraphs. Configurations of pixels were chosen to allow studying the impact of a change of single feature at a time.
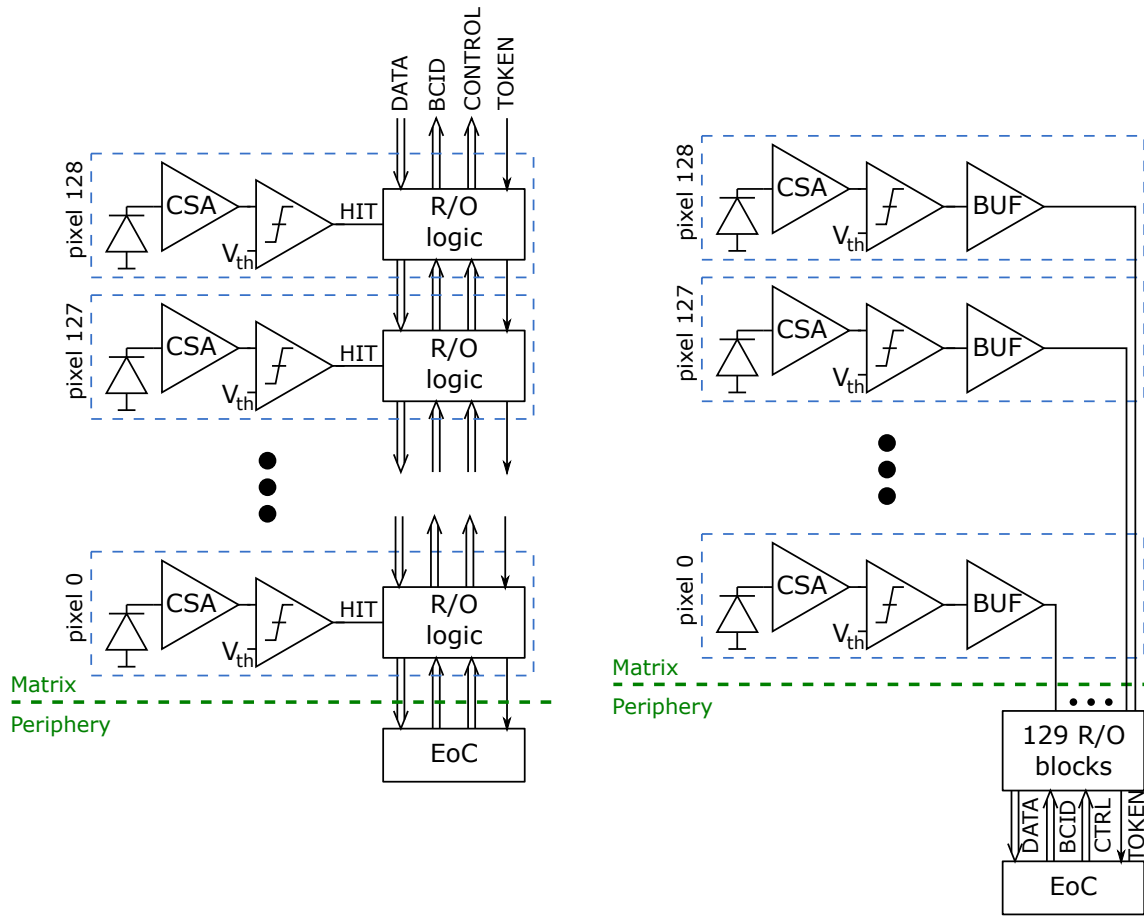
(a) Column with R/O block inside pixel

(b) Column with R/O block outside pixel

Figure 3.62: Different placements of readout blocks and its impact on the column design.

Table 3.1: Distribution of pixel flavours in the matrix of LF-Monopix1.

Column	35-32	31-28	27-24	23-20	19-16	15-12	11-8	7-4	3-0
Pixel R/O logic	inside pixel							periphery	
Preamplifier	nMOS		CMOS						
Discriminator	V1	V2	V1	V2	V1	V2	V1	V2	
Discri. power[a]	A+D				D		A+D	D	
Token logic	Current steering logic						CMOS logic		
Col. line driver[b]	N								P

[a] D: Digital domain   A: Analog domain
[b] P: pMOS source follower   N: nMOS source follower.

77

3.3.3.4 Chip architecture

The block diagram of the LF-Monopix1 chip is presented in Fig. 3.63. The lower part of the chip contains all the global bias DACs and current mirrors for bias distribution. The chip is configured using a shift register, with a portion of it dedicated for the chip periphery configuration and the rest inside the matrix for the configuration of all the pixels. The output of the CSA and comparator of any pixel can be observed through analogue buffers, one pixel at a time (same as for previous LFoundry DMAPS devices).

   The upper portion of the chip contains all the circuitry required for the column drain readout. At the end of each column there is an EoC (End of Column) block containing TOKEN priority arbitration logic, Gray counter with buffers for BCID distribution and sense amplifiers to receive data from DATA bus. The sense amplifier allows the data to be correctly read out, even if the differential lines for a given bit are not fully set at a logical value (few tens of millivolts difference between the two lines of a differential pair is enough for reliable operation). The 24-bits of data read from each hit (8-bit LE, 8-bit TE and 8-bit row number) is appended with 6-bits of column number information. Next the data is serialized and send off the chip at the rate of 160 Mbps with an LVDS driver. The data is received by an FPGA, where the readout controller is implemented. This simplifies the digital design of the chip and allows flexibility in adjusting the controller functionality during testing.
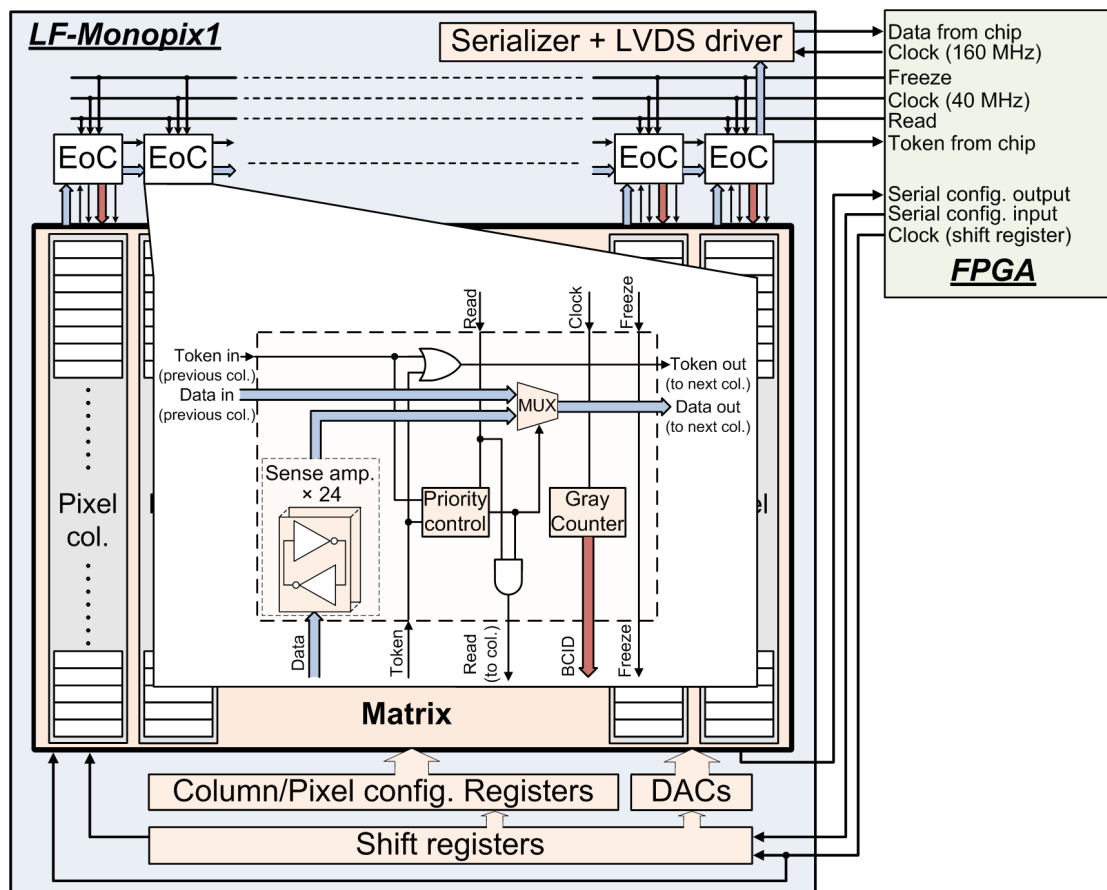
Figure 3.63: Block diagram of the LF-Monopix chip. I/O pads are omitted. [67]

3.3.3.5 Measured pixel electronics performance

The first measurements focused on verifying that the integrated column drain readout works as expected. It was quickly proven that when injecting a single pixel the serialized data from chip can be interpreted by the FPGA and addresses of pixels are encoded correctly. The chip also reacted in the expected way to the changes in global configuration.
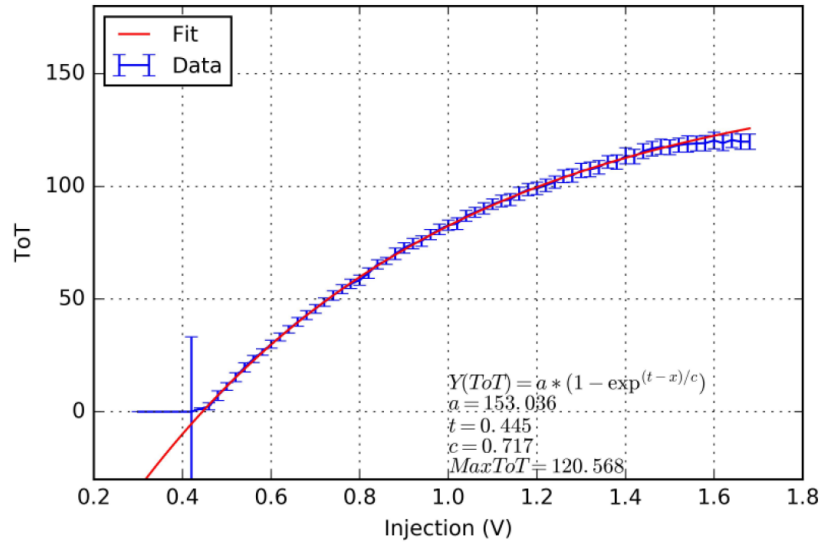
In order to measure the real value of the injection capacitance, the in-pixel injection circuit is used to calibrate the ToT value as a function of injection voltage. For this measurement a low feedback current setting is used in order to have a wide pulse output from CSA and therefore better calibration resolution. An example of such calibration curve together with a fit is presented in Fig. 3.64(a). Afterwards, the chip is exposed to a ^{241}Am X-ray source (59.5 keV) and X-ray fluorescence from ^{65}Tb (K_α=44.5 keV, K_β=50.4 keV). The spectra obtained from those sources together with peak fits are shown in Fig. 3.64(b). The information from plots in Fig. 3.64 allows finding correspondence between amplitude of voltage injection and the charge, which in turn allows calculating the value of injection capacitance in the given pixel $C_{inj} = \frac{Q}{V_{inj}}$. The same procedure was repeated for many pixels and while the shape of the injection voltage vs. ToT curves look different between pixels, the injection capacitance value is very consistent with average $C_{inj} = 2.7$ fF. This value is important, as it is used for ENC measurements and threshold determination. The injection capacitor is hand made (the manufacturer does not provide so small capacitor layout) and the capacitance value expected from post-layout simulation was 2 fF. This relatively large discrepancy between simulation and measurement is not fully understood, but it could be caused by an imprecise modelling of hand made metal-metal capacitor in the design software.

All pixel flavours have been characterized in terms of ENC and threshold dispersion, the results for two pixel variants can be seen in Fig. 3.65. The CSA in both variants is the CMOS architecture, but the comparators are different. The mean ENC value of 200 e⁻ is in good agreement with simulation and as can be seen in Fig. 3.65(a) the type of comparator does not influence the noise of CSA, which is expected.
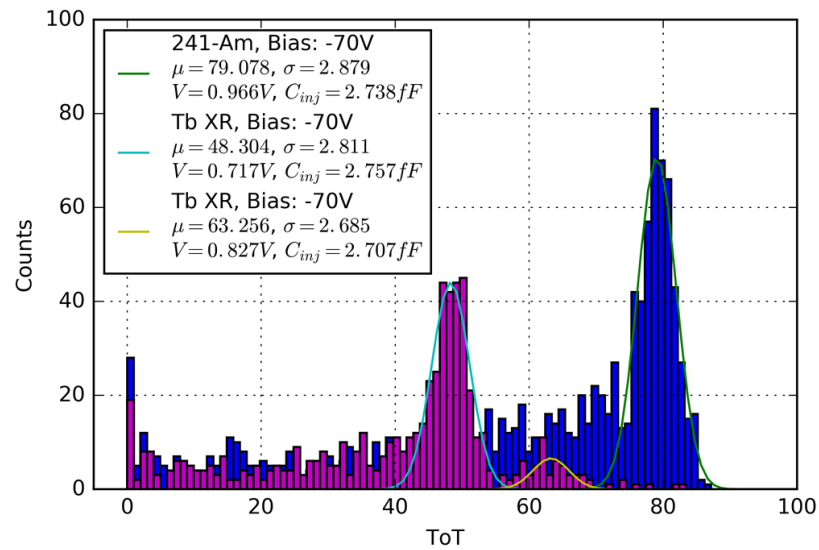
Fig. 3.65(b) shows the threshold distributions for the analysed pixels before and after tuning with the 4-bit in-pixel current DAC. The distribution for version 1 of the comparator is narrower than for version 2, both before and after tuning. The difference is caused by version 2 using smaller input transistors than the version 1, however both circuits can be tuned and achieve similar dispersion of approx. 100 e⁻. The tuning target thresholds visible in Fig. 3.65(b) are not the minimal achievable ones, the LF-Monopix1 pixels can be tuned down to threshold of 1 400 e⁻ with 100 e⁻ dispersion for un-irradiated samples and 1 700 e⁻ with 130 e⁻ dispersion for samples irradiated with NIEL fluence of 10^{15} n$_{eq}$/cm^2 [73].

The pixel performance is homogenous across any given flavour, which is illustrated for the CSA gain in Fig. 3.66. The ENC and threshold dispersion maps also do not show any noticeable patterns. The measured gain values of 10-12 µV/e⁻ are in agreement with simulation. The alternating pattern visible in the gain map is related to placement of discriminator flavours, because the version 2 of discriminator has smaller input transistor size, resulting in lower capacitive load on the CSA.

The results of TID irradiation of LF-Monopix1 are not published yet, but because the analogue circuitry is nearly identical to that of LF-CPIX, only little degradation is expected. Preliminary results seem to confirm this and the circuitry responsible for the column drain readout also appears to be working correctly after 80 Mrad TID.
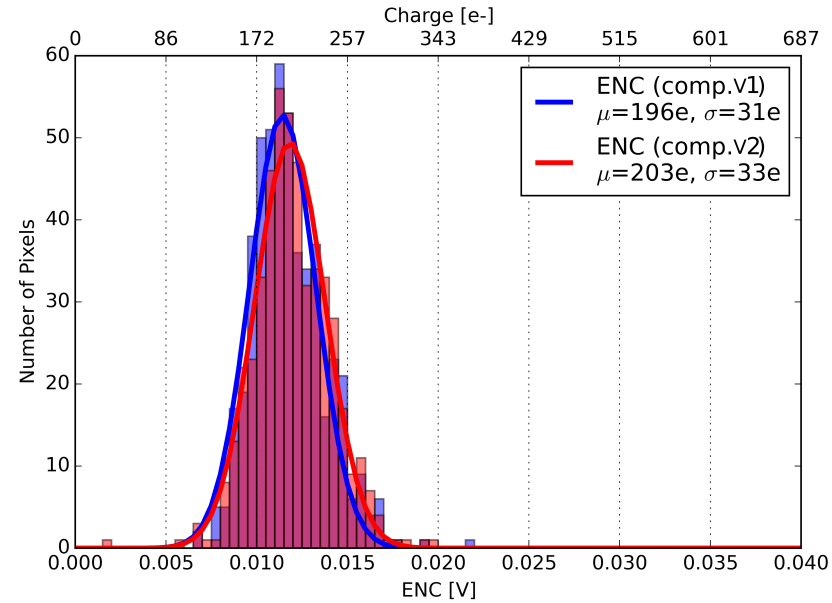
$$Y(ToT) = a * (1 - \exp^{(t-x)/c})$$
$a = 153.036$
$t = 0.445$
$c = 0.717$
$MaxToT = 120.568$

(a) Correspondence between ToT to injection voltage.

241-Am, Bias: -70V
$\mu = 79.078$, $\sigma = 2.879$
$V = 0.966V$, $C_{inj} = 2.738fF$
Tb XR, Bias: -70V
$\mu = 48.304$, $\sigma = 2.811$
$V = 0.717V$, $C_{inj} = 2.757fF$
Tb XR, Bias: -70V
$\mu = 63.256$, $\sigma = 2.685$
$V = 0.827V$, $C_{inj} = 2.707fF$

(b) Xray sources ToT spectra

Figure 3.64: ToT to charge calibration measurements for a single pixel. [68]

(a) ENC distribution
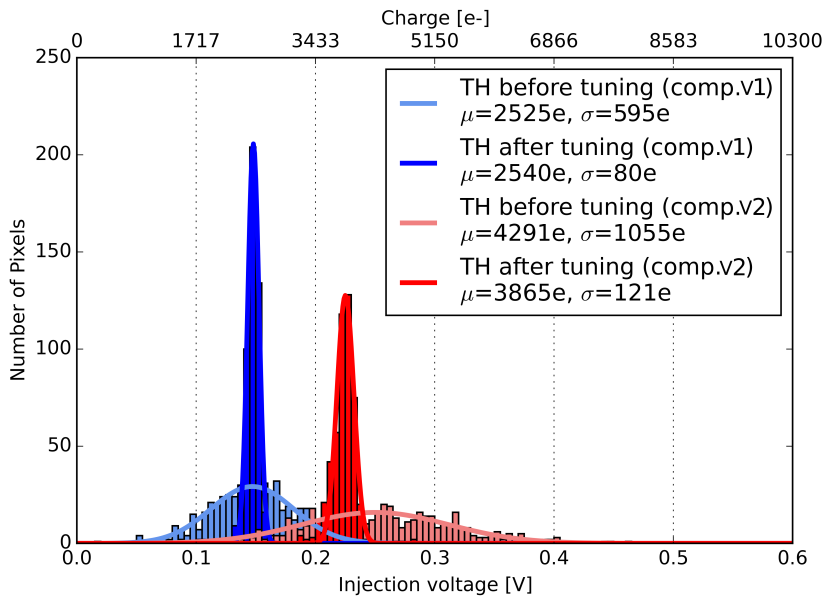
(b) Threshold dispersion before and after tuning

Figure 3.65: ENC and threshold dispersion of two variants of pixels: CMOS CSA with v1 comparator (blue) and CMOS CSA with v2 comparator (red). There are 516 pixels of each variant. [68]
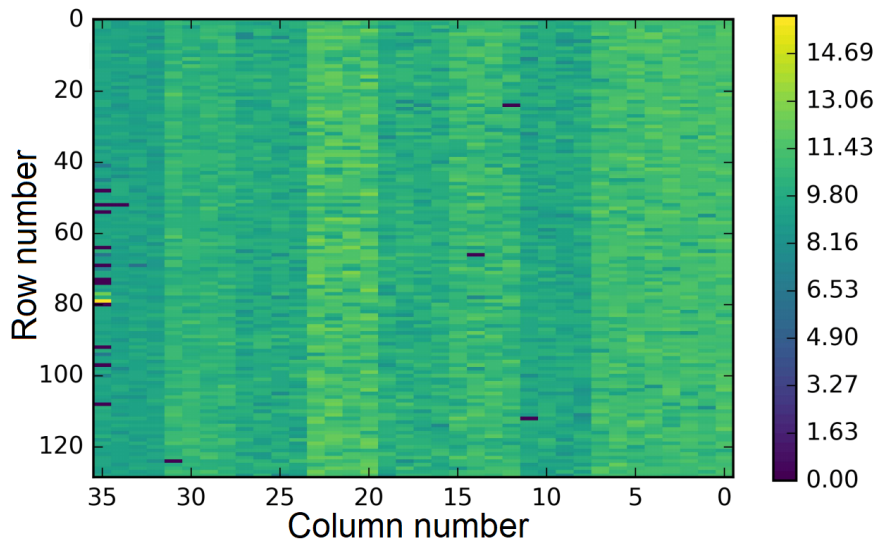
Figure 3.66: Gain map of LF-Monopix1. [73] (presentation showed in the conference)

3.3.3.6 Breakdown voltage and depletion depth

The LF-Monopix1 chip has a similar guardring structure to the previous LFoundry DMAPS chips (one n-well ring around the chip core plus eight p-well rings surrounding it, with the negative high voltage applied to the outermost p-well ring), with slight adjustment to guardring spacing.

The IV curves measured on un-irradiated devices are shown in Fig. 3.67(a). The 725 μm thick sample is a standard device, as delivered by the manufacturer. To produce the 200 μm and 100 μm samples the wafers were thinned down after delivery. The difference between achieved breakdown voltages can be explained by the wafer-to-wafer differences (e.g. resistivity) or influence of thinning down process (e.g. slight changes in implantation geometry due to diffusion during annealing steps). Nevertheless, a high voltage of above 250 V was achieved, which is an improvement over previous prototypes.

The IV measurement results for samples irradiated with neutrons and protons to various fluences as presented in Fig. 3.67(b). During the measurement the devices were kept at −25 °C in order to protect the chip from thermal runaway. During the measurement voltages only up to 200 V were applied to avoid damaging the devices in case of sudden current increase. It is clear, however, that even at 10^{15} n_{eq}/cm^2 a breakdown voltage above 200 V is achievable, which is more than sufficient to deplete the thinned down devices.

The depletion depth is measured in similar way to measurement done for CCPD_LF (Section 3.3.1.5). The ToT spectra of 2.5 GeV electron beam (ELSA) are taken at different chip bias voltages. Afterwards, a Langau curve was fitted to each spectra and the most probable values are translated to the amount of collected electrons based on the previously described calibration with X-ray sources. The amount of collected electrons as a function of applied bias is presented in Fig. 3.68. Higher bias voltages were not applied, because the sample used during the testbeam measurement exhibited unusually low breakdown voltage of 130 V. Based on Eq. (3.1) the depletion depth at 100 V bias is approx. 250 μm. The fit to Eq. (3.4) shows that the measured device has a silicon resistivity of 7.3 kΩcm, which is higher than measured on previous LFoundry DMAPS structures, but still within specification of the

(a) Un-irradiated samples
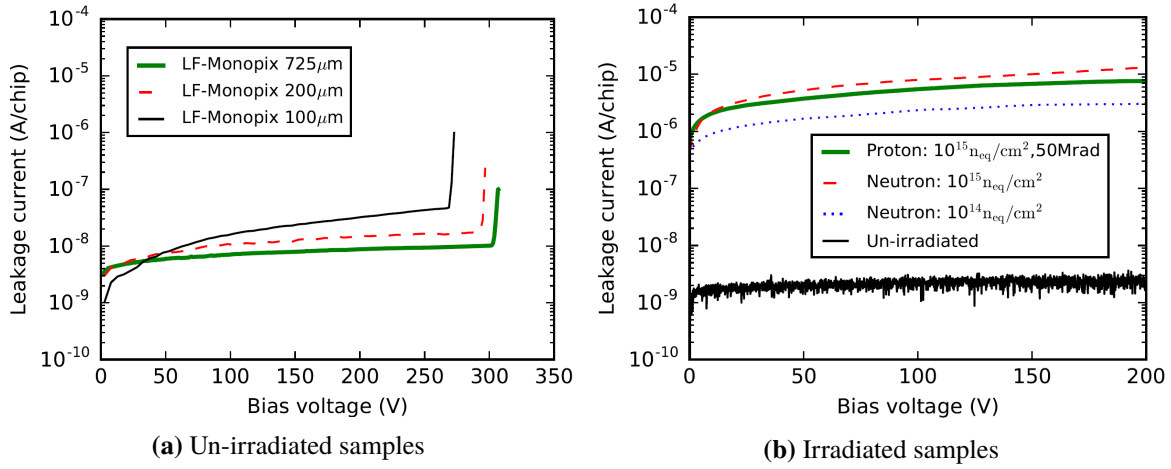
(b) Irradiated samples

Figure 3.67: Measured IV curves of LF-Monopix1. Based on [67]
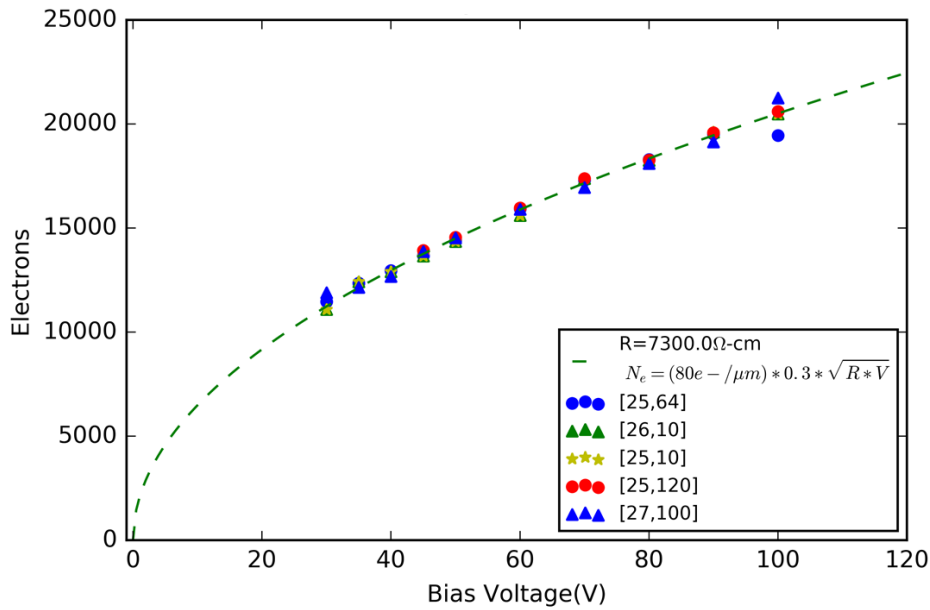
wafer manufacturer.

Figure 3.68: MPV of a Langau fit to charge spectra obtained from 2.5 GeV electron beam as a function of applied bias voltage. [73]

3.3.3.7 Fast readout performance and problems

As mentioned previously the column drain readout works as expected. An example of a readout sequence captured with an oscilloscope is shown in Fig. 3.69. The digital activity of TOKEN, FREEZE, READ and BCID CLK signals does not produce a significant noise on the CSA output. A

difference to the timing diagram shown in Fig. 3.57 is that the FREEZE signal is kept high for some time after the READ signal goes to zero. While this keeps the column inactive for longer, it was observed to be a safer solution in cases when new hits appear while the initial one is still being read out. Since the readout controller is implemented in FPGA, this change was easy to implement and did not require any modifications to LF-Monopix1.
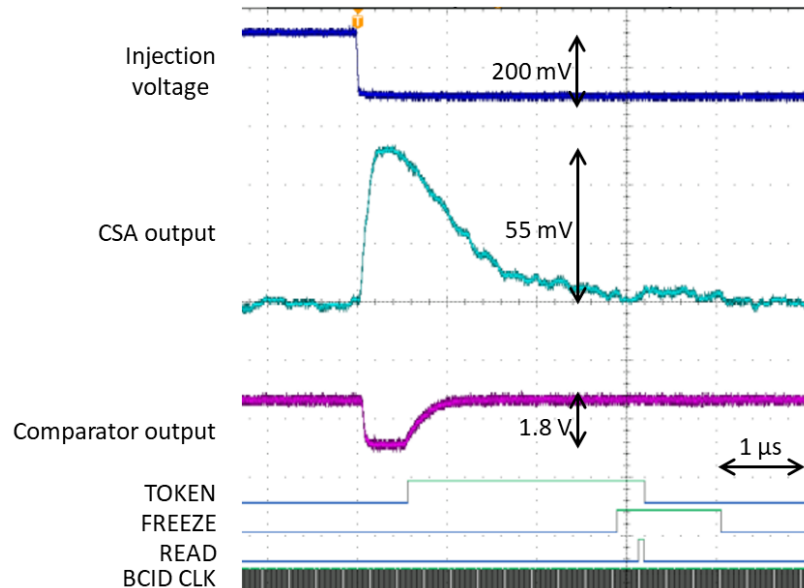
Figure 3.69: Readout process of a single injection to a pixel captured with an oscilloscope. CSA and comparator outputs are observed through on-chip analogue buffers. TOKEN, FREEZE, READ and BCID CLK signals were observed using digital probes (high value corresponds to 1.8 V). [23]

The performance difference of the Current Steering CMOS (CS-CMOS) TOKEN and pure CMOS TOKEN propagation chains has been investigated, but the final results are not yet available. So far it seems that the pure CMOS implementation does not inject any more noise into the pixels than the CS-CMOS version. This contradicts the simulation results presented in Fig. 3.59. Two potential explanations were found for this:

- The C_{CS} capacitance (see Fig. 3.58(b)) was custom made by hand in order to fully utilize the space available in the pixel. The capacitor is made on Metal2 with a grounded shield of Metal1 underneath it, but the sides of that capacitor are close to Metal1 ring directly connected to the charge collection DNW. It is plausible that during parasitic extraction the coupling to the sensitive net was not modelled correctly, which would hide crosstalk issues.

- The connection of the digital substrate to external ground in the pixel is not very robust, therefore it might be a relatively high impedance net locally in the pixel. In such case a variation in the current during switching of the CS-CMOS gate might cause larger disturbance to the sensor than switching of a simple CMOS gate (fewer transistors involved, faster transient with not too high current draw since the output capacitance for a gate in TOKEN chain is small).

More measurements are needed to fully understand this behaviour.

Fig. 3.70 shows ToT response to the test pulse injection for un-irradiated and NIEL irradiated (10^{15} n_{eq}/cm^2) LF-Monopix1. The difference between the curves is negligible, which shows that the full readout is capable of working correctly after sustaining NIEL damage expected for the ATLAS ITk pixel detector outer layers.
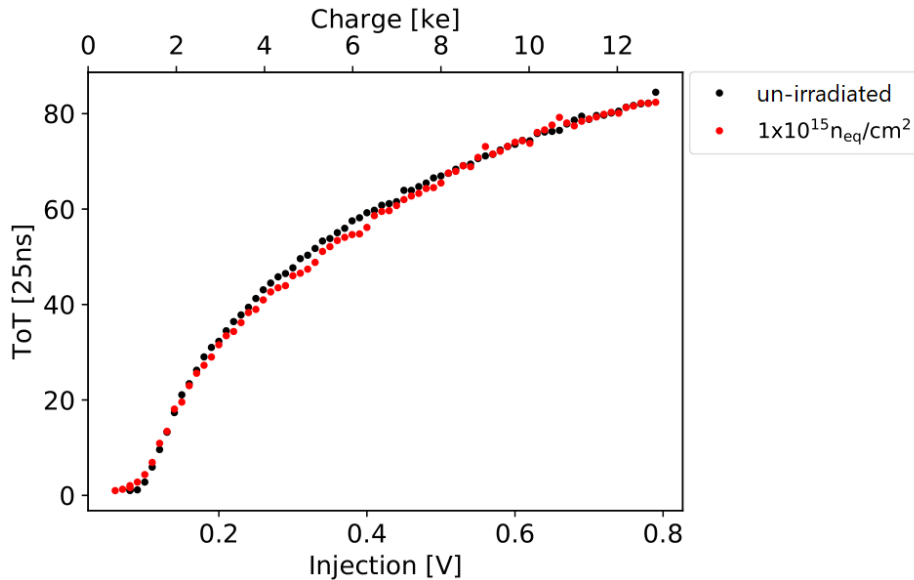

Figure 3.70: ToT response to the test pulse injection for un-irradiated and NIEL irradiated (10^{15} n_{eq}/cm^2) LF-Monopix1. [23]

The pixels with the column drain logic placed in the chip periphery does not show better performance than the pixels with readout logic placed inside pixel. Taking into account the difficulty of making robust one-to-one connections from pixels to periphery and how it can limit the scalability of the chip size, this solution will not be used for future chips.

3.3.3.8 Time walk

Fig. 3.71 shows the time walk measurement for discriminator design version 2. The measurement is done using the in-pixel charge injection circuit with the threshold set to $1\,550\,e^-$ and the response time of large signal (1.2 V injection, equivalent to approx. $20.2\,ke^-$ charge deposition) is used as the zero point for time walk calculation. All hits larger than $1\,890\,e^-$ have time walk lower than 25 ns (maximum tolerable in ATLAS experiment due to 25 ns bunch crossing period) and they are referred to as "in-time" hits. The difference between in-time threshold and the set minimum threshold is called "overdrive". When measured for all pixels in the matrix the average overdrive for version 1 discriminator is $664\,e^-$ and for version 2 it is $453\,e^-$. This agrees with the simulation expectation of version 2 comparator being faster. The overdrive was measured also on proton-irradiated chip sample (10^{15} n_{eq}/cm^2 and approx. 50 Mrad of TID due to background) and the observed degradation was only 10.3%. This shows an improvement over the time walk measured for CCPD_LF.
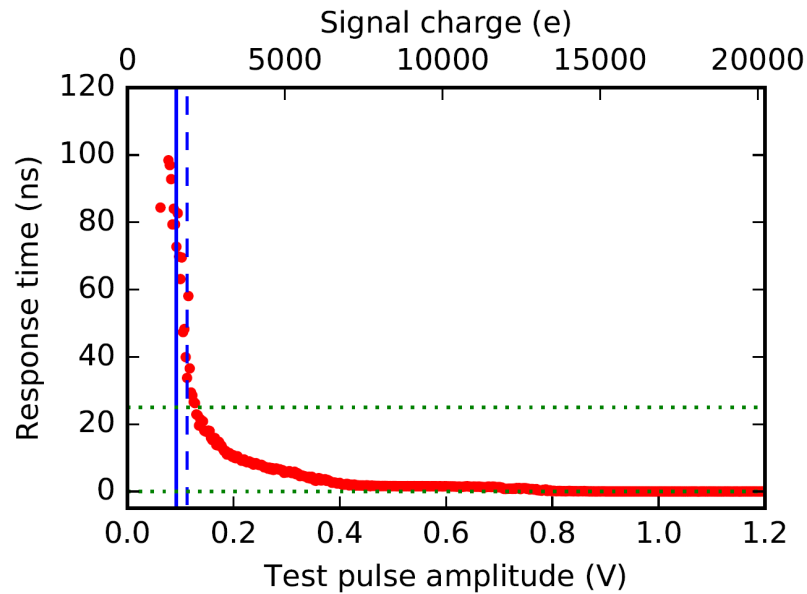
Figure 3.71: The time walk measurement of a single pixel of LF-Monopix1 with nMOS CSA and version 2 of comparator. The green dotted lines show the range of responses within 25 ns. The threshold was set at $1\,550\,e^-$ (solid blue line) and the in-time threshold was $1\,890\,e^-$ (dashed blue line). [67]

3.3.3.9 Hit detection efficiency

The hit detection efficiency of LF-Monopix1 was evaluated using a 2.5 GeV electrons beam at ELSA accelerator at the University of Bonn. The measurement was done using both un-irradiated and neutron irradiated ($10^{15}\,n_{eq}/cm^2$) samples with the comparator thresholds set to $1\,800\,e^-$ and $1\,600\,e^-$, respectively. This allows to keep the noise occupancy at the level of 40 Hz/pixel, order of magnitude lower than the requirement of ATLAS ITk. During the measurement the chips were biased at 200 V and kept in stable temperature of $-40\,^\circ$C. The measured hit detection efficiency maps are presented in Fig. 3.72. The telescope used for the measurement has a resolution of 20 μm and each hit in LF-Monopix1 was assigned to a reconstructed track if it was found within a distance of 300 μm for column- and row-wise direction of the pixel matrix. The average hit detection (after masking noisy pixels and without applying timing cuts) is 99.7% for un-irradiated chip and 98.9% for the irradiated one.
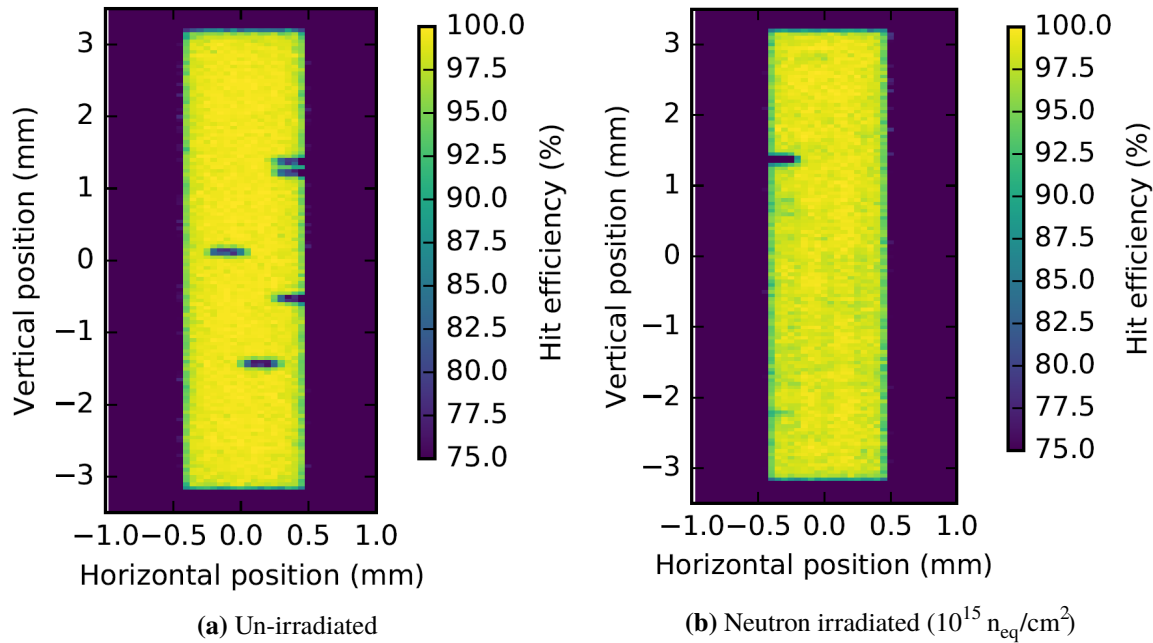
(a) Un-irradiated

(b) Neutron irradiated ($10^{15}\,n_{eq}/cm^2$)

Figure 3.72: Measured hit detection efficiency maps of LF-Monopix1. The dark blue spots correspond to masked pixels. [67]

3.4 Summary and outlook of the LF DMAPS project

Monolithic Active Pixel Sensors (MAPS) devices combine particle sensor and readout electronics into one chip. Such devices have already been used in HEP experiments, but their reliance on charge collection through diffusion makes them unsuitable for high rate and high radiation applications. For this reason the current state-of-the-art particle detectors use hybrid designs, where a depleted silicon sensor is connected to a dedicated readout chip. While hybrid detectors offer excellent performance, they are complicated and costly to build. A potential alternative approach would be Depleted MAPS (DMAPS). The depletion can be achieved using a commercial CMOS manufacturing processes and exploiting their technology add-ons such as multiple nested implantation wells or high resistivity wafers. By depleting the silicon bulk the charge created by an impinging particle can be collected through diffusion, which is much faster than drift (charge collection within few nanoseconds instead of many microseconds), leading to higher hit rate capability and better radiation hardness when compared to MAPS. The benefits of using a commercial CMOS technology are low cost and fast production of the devices.

This part of the thesis focused on exploring the feasibility of high fill factor DMAPS designed in 150 nm LFoundry CMOS technology. This effort was a part of the "CMOS Demonstrator" collaboration, aiming to develop DMAPS (using several different CMOS processes) capable of meeting the requirements of the ATLAS ITk Pixel Detector outer layers: 25 ns timing resolution, particle rate up to 100 MHz/cm^2, radiation hardness up to 50 Mrad TID and NIEL fluence of $10^{15}\,n_{eq}/cm^2$. In order to achieve this several prototype chips were made.

The first device designed in this prototypes line is CCPD_LF. It focused on testing two different approaches to the sensor design and integrating analogue readout chain (CSA and comparator) into

the pixel. The chip size was 0.5×0.5 cm^2, which allowed fitting a matrix of 24 rows and 114 columns using 125×33.3 µm^2 pixels. With the implemented guardring structure a breakdown voltage of 115 V was measured, enough to deplete 166 µm of silicon bulk. The in-pixel CSA behaved as expected from simulation, achieving gain of about 7 µV/e$^-$ and ENC of approx. 130 e$^-$. The comparator threshold dispersion was larger than initially expected (standard deviation of 11.7 mV), which was caused by large influence of mismatch on the input transistors and was improved for the next designs. The chip was designed to be readout either through slow, binary stand-alone mode or to be bump bonded to FE-I4 readout chip for fast readout. Both of the readout methods were verified and worked as expected. After exposure to X-ray radiation and accumulating 50 Mrad the chip's degradation was apparent (40% gain loss and 75% ENC increase in worst case), but not catastrophic and the device could still be operated.

The second device in the LFoundry family is LF-CPIX (the design and measurement of LF-CPIX was not a part of this thesis). Following the encouraging results of CCPD_LF, this device focused on further improving the breakdown voltage (215 V was measured) and the analogue in-pixel readout electronics. The chip was designed to be readout the same way as CCPD_LF (slow, binary stand-alone mode or to be bump bonded to FE-I4 chip), however the pixel size was increased to 250×50 µm^2 and designed from the beginning to be able to accommodate additional in-pixel logic for fast stand-alone readout, which would be implemented in the next prototype. The chip size was increased to 1×1 cm^2 and the matrix was composed out of 158 rows and 36 columns divided among several types of pixels. In addition to the CSA used in the previous chip, two new architectures were introduced and overall the analogue readout chain was optimized. After TID irradiation to 50 Mrad the degradation was smaller than that of CCPD_LF (5% gain loss and 40% ENC increase in worst case). The LF-CPIX hit detection efficiency was measured in an electron testbeam and at 1 700 ke$^-$ threshold 99.4% efficiency was achieved (without cuts on timing).

The last device belonging the the LFoundry DMAPS line described in this thesis is LF-Monopix1. The main focus of this chip was adding fast stand-alone readout into LF-CPIX-based design. The implemented readout architecture was "column drain", which is used in the FE-I3 readout chip currently working in the ATLAS detector. It allows timestamping the hit events with 25 ns precision and according to simulation studies should be able to cope with hit rates up to 100 MHz/cm^2. The design was proven to work well at full speed in measurements and the in-pixel digital processing activity does not cause interference with the sensor. Two CSA designs (nMOS and CMOS input device based) were copied from LF-CPIX and further optimized, achieving approx. 11 µV/e$^-$ gain and 200 e$^-$ ENC (for detector capacitance C_d=400 fF, much more than any of the previous chips). While the chip size is kept at 1×1 cm^2, the matrix is built out of 129 rows and 36 columns, which are divided into 9 pixel flavours. The measurement shows that the time walk was improved compared to CCPD_LF. The chip hit detection efficiency was measured in several test campaigns, resulting in 99.7% efficiency for un-irradiated devices and 98.9% for sample irradiated to 10^{15} n$_{eq}$/cm^2 (sensor bias was kept at 200 V, well below measured breakdown voltage of 250 V for safety reasons).

As can be seen in the timeline in Fig. 3.34 a successor to the LF-Monopix1 has been designed and manufactured. This chip is named LF-Monopix2, which was not a part of this thesis. It focuses on shrinking the pixel size to 150×50 µm^2, proving full functionality with twice longer columns (chip size is 1×2 cm^2), providing a more homogenous matrix (fewer pixel flavours for easier operation in testbeams) and improving the timing performance. Testing of the chip has started recently and first measurements show that the device is functional, but many more measurements are needed to fully verify all chip aspects.

Work presented in this thesis has proven the feasibility of the concept of DMAPS capable of working in high hit rate and high radiation environment. While the ATLAS experiment ultimately decided that DMAPS devices will not be used for ATLAS ITk Pixel Detector construction, the idea of DMAPS chips raised a lot of interest within the HEP community. Many developments are still on-going, exploring aspects such as low fill factor approach, improved timing capabilities or using smaller technology nodes to integrate more sophisticated readout architectures [55][57].

Clock and data recovery circuit for RD53

4.1 Motivation

The LHC High Luminosity upgrade will result in a significant change of environment in which particle detectors are going to operate, especially for devices very close to the interaction point like the pixel detector electronics, as described in Section 2.2.1. The performance requirement for the pixel readout chip resulting from those changes are very similar for ATLAS and CMS, therefore the groups decided to work together on the chip design. This collaboration, named RD53, started in 2013 [74] and over the years included members from over 24 institutes world-wide for chip design work and from many more institutes for chips testing. While the original mandate of the group was research and development of pixel readout chip suitable for HL-LHC environment, it was extended in 2018 [75] to designing the production versions of the chips for both ATLAS and CMS.

A short list of requirements for the pixel readout chip includes: chip size of approx. $4\,\text{cm}^2$ (ATLAS and CMS versions differ in size due to detector geometry differences), pixel size of $50 \times 50\,\mu\text{m}^2$, hit rate of $3\,\text{GHz/cm}^2$ (for the innermost detector layers), trigger rate of $1\,\text{MHz}$, trigger latency of $12.5\,\mu\text{s}$, power consumption below $1\,\text{W/cm}^2$ and radiation tolerance of at least $500\,\text{Mrad}$. This is a very challenging specification set, resulting in the RD53 chips being one of the most complex ASICs designed within the HEP community. In order to achieve that specification the RD53 chips need to be produced in a modern, high density CMOS technology and after extensive market survey TSMC 65nm process was chosen.

So far the RD53 collaboration has delivered two large chips. First one, named RD53A [28], is a half size ($1 \times 2\,\text{cm}^2$) design aiming at providing a proof-of-concept chip with many test features (three types of front ends, two types of readout architecture, large set of configuration options allowing to test several design aspects, etc.). The second one is a pre-production design for ATLAS named ITkPixV1 [31]. Here the size is increased to $2 \times 2.1\,\text{cm}^2$, the matrix was made homogenous by using only one front-end design and one readout architecture, larger effort was put on protection against Single Event Effects (SEE), bugs found during RD53A testing were fixed and many test features were removed. At the moment of writing this thesis a third large chip, the CMS specific design named CROCv1 has been submitted and very recently first samples were delivered by the manufacturer. While the CROCv1 is different than ITkPixV1 (it uses different type of front end, chip size is $2.16 \times 1.86\,\text{cm}^2$), both of them were designed using the same methodology, the same framework and in large part the same IP blocks. Additionally, the circuits designed as a part of this thesis were used in exactly the same way, in both

ITkixV1 and CROCv1, therefore in parts of this chapter those devices will be collectively referred to as the RD53B design. A full description of RD53 chips and measurement results obtained with them is beyond the scope of this thesis, instead in this chapter the focus will be solely given to the circuits designed as a part of this work, mentioning other RD53 chip aspects only when necessary.

A very simplified block diagram of a RD53 chip is presented in Fig. 4.1. While internally the differences between RD53A and RD53B designs are very significant and nearly all blocks were changed or improved in RD53B, at the high level of abstraction shown here the designs are the same. Most of the chip area is occupied by the pixel matrix, with the chip periphery and I/O pads taking approx. 1.8 mm vertically and spanning the full width of the chip. The RD53 chips are designed to require as few external inputs as possible and therefore should be fully functional by just connecting power and input command signals. In order to facilitate this requirement the periphery contains many IP blocks such as DACs, ADCs, power regulators, bias current generators, temperature and radiation sensors, CDR, complex digital processing, etc. Thanks to applying modern design methodology and advanced commercial chip design tools, each of those blocks can be designed and tested quite independently in parallel to each other and only later on everything is integrated into a full chip. Many of those IPs were prototyped as small chips before integrating into RD53 in order to verify the functionality in silicon and check radiation hardness. One of those blocks is the Clock Data Recovery (CDR) circuit, which is the subject of this chapter.
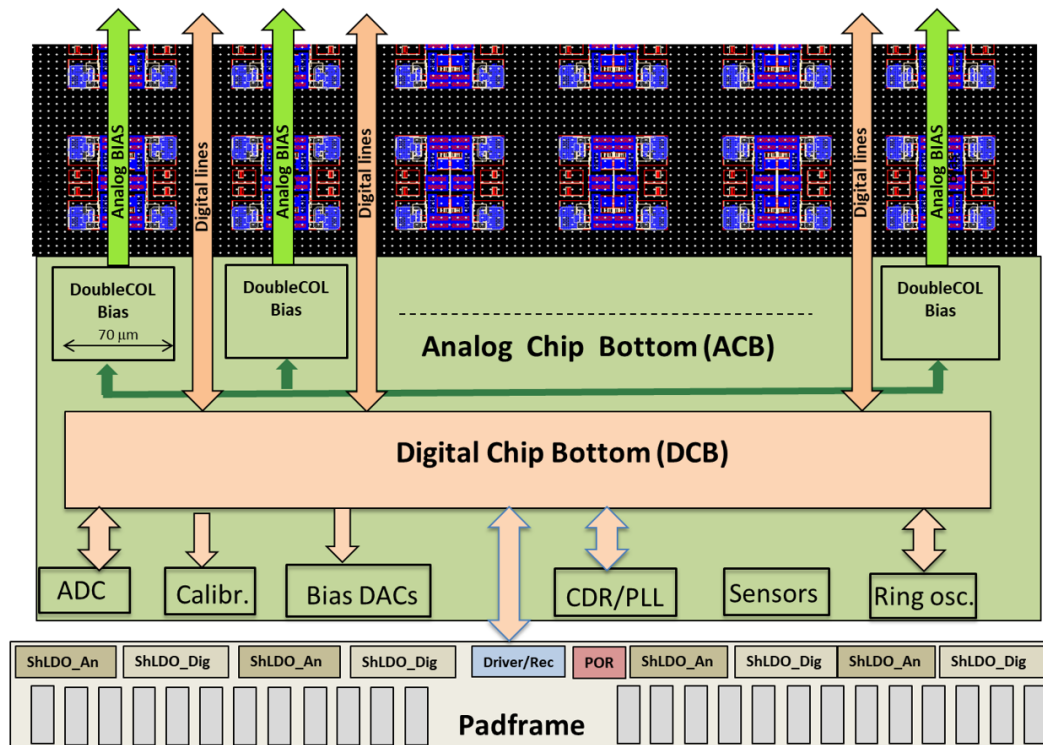
Figure 4.1: Block diagram of the structure of a RD53 chip. [31]

Due to very high hit rate expected in the inner most detector layer (approx. $3\,\mathrm{GHz/cm^2}$) the readout chips will produce much more data than their predecessors. This has a very strong impact on many aspects of the chip design e.g. hit data processing, pixel data readout architecture or data bandwidth

needed to send all obtained data off chip. In order to facilitate all the needed functionality the chip requires several clocks with different frequency ranging from 40 MHz up to 1.28 GHz. Since it is technically difficult to supply good quality clock signals to the chip through long, low mass cables (such as the ones required inside the ATLAS experiment due to material budget constrains), it is common practice to supply a chip with only one, low speed clock and synthesize all other needed ones inside the chip. In the previous generations of the pixel readout chips this was achieved by implementing a Phase Locked Loop (PLL) circuit, which would create faster clocks from the one provided, as depicted in Fig. 4.2(a). The RD53 chips takes it a step further: no dedicated clock signal is sent to the chip, instead the clock is extracted from the input data signal using a Clock Data Recovery (CDR) circuit, as visualized in Fig. 4.2(b). The advantages of this approach are:

- Lowering the material budget of the detector. While removing one input cable going to a chip might not seem like a big saving, it amounts to a large saving overall, since the ATLAS pixel detector will be composed out of more than 34000 readout chips [10]. The reduction in electrical cabling material budget expected for ATLAS ITk (shown in Fig. 2.7) is achieved partially by switching from PLL to CDR based chip interface (the second contributor to the reduction is a new scheme of powering the chips).

- Simplifying the routing inside the detector. In case of using PLL the input command data (CMD DATA) and command clock (CMD CLK) are sent separately and have to arrive at the chip input without accumulating a large phase difference along the way. This means that the lines have to be carefully designed and matched in length. When using a CDR, only the command (CMD) is send to the chip and the clock is extracted inside the chip, which simplifies the detector design.
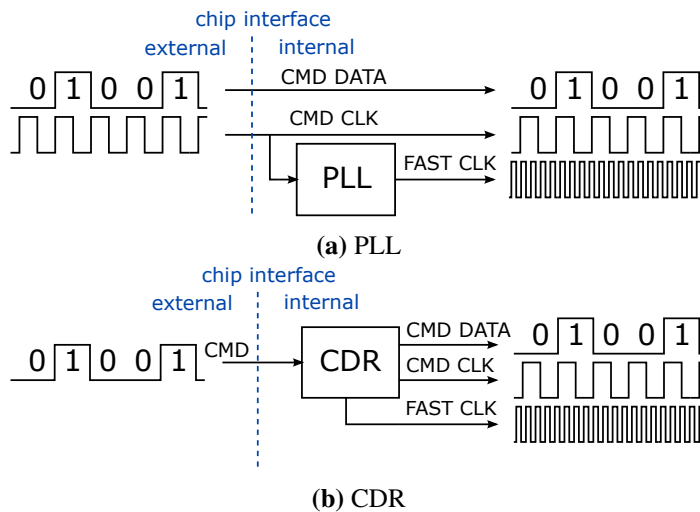

(a) PLL

(b) CDR

Figure 4.2: Comparison of usage of a PLL and a CDR in an ASIC. The main difference is the necessity of supplying both the *CMD DATA* and *CMD CLK* to the chip when using a PLL, while in case of CDR providing just the *CMD DATA* is enough, as the clock is generated internally by CDR.

While the technical details of the CDR circuit implementation changed significantly from RD53A to RD53B, as it will be explained further on in this chapter, its overall role and specification remained mostly the same because they are defined by the needs of the data link of the whole pixel detector.

A simplified data link configuration of the pixel detector that is expected to be implemented in the ATLAS experiment is presented in Fig. 4.3. The commands for controlling the FE chips are sent via optical fiber from the DAQ in control room to the VTRx+ chip [76]. There the conversion from optical to electrical signals happens and the data is propagated at 2.56 Gbps to the lpGBT chip [77], where the data is decoded. Next the appropriate data for the addressed RD53 chip is sent via an electrical link at 160 Mbps. This input command is received inside the RD53 chip by the LVDS receiver and propagated to the CDR. The 160 MHz clock is extracted from the input command and after resynchronization with the extracted clock both of them are propagated further into the RD53 chip. At the same time the CDR synthesizes a 1.28 GHz clock and sends it to the data serializer. Aurora encoded [78] data streams are fed into a 20:1 serializer, which combines them into a single 1.28 Gbps data stream that is sent off-chip by the Current Mode Logic (CML) driver [79]. In the RD53 chip there are 4 independent serializer + CML driver sets, therefore a full data output bandwidth of the RD53 chip is 5.12 Gbps. Those streams are then received by the lpGBT chip, which combines data from several chips into a 10.24 Gbps stream. This data is sent to VTRx+, where conversion from electrical to optical signal is performed and the data collected from particle collision is finally sent back to DAQ in experiment control room.
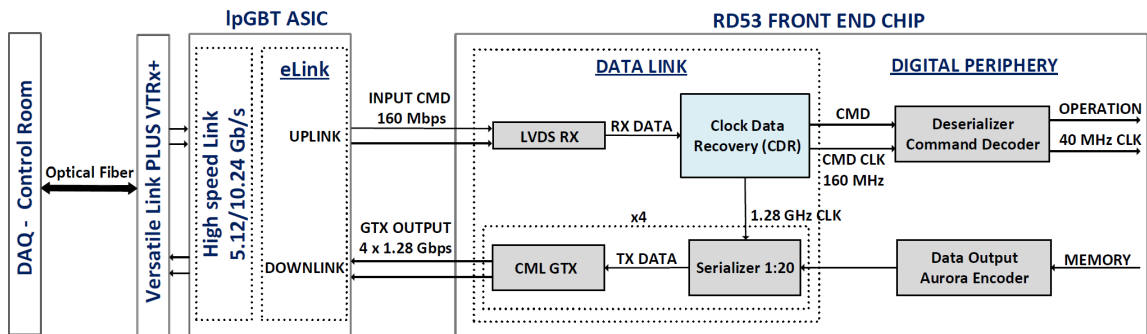
Figure 4.3: Data link configuration of the HL-LHC Phase-II upgrade ATLAS pixel detector.

In reality this scheme is much more complicated: each VTRx+ communicates with several lpGBT chips, and each of the lpGBT chips communicates with several RD53 chips. The numbers of RD53 chips inside the pixel detector of HL-LHC ATLAS is estimated to be around 34000 and the amount of data produced by each of them depends on their position in respect to the interaction point. According to the detector simulations the pixel readout chips close to the interaction point will have to utilize the full 5.12 Gbps output bandwidth, while chips in outer layers will produce smaller amounts of data e.g. one 1.28 Gbps link could be shared by four RD53 chips (that would be achieved by grouping the RD53 chips into quads and having three RD53 chips send their data directly to the fourth one, where the data will be combined and send out). Overall, defining the number of links and signal processing chips needed is a very complicated tasks, far out of the scope of this thesis. At the moment of writing this text the exact numbers are not yet defined, however, the values relevant for the CDR design (RD53 chip input and output data rates) are not expected to change.

This chapter presents the author's work on design and testing of the CDR circuits used in the prototypes and the RD53 chips. During the course of this thesis two prototype chips have been designed named CDR53A and CDR53B. The CDR circuits that each of them contain were later on integrated into RD53A and RD53B, respectively. The timeline of this development is presented in Fig. 4.4.

Figure 4.4: Photos of CDR prototypes and RD53 chips. The devices are arranged in the chronological order with the submission date indicated at the bottom. Chip sizes are very different: 2 (4) µm^2 for CDR53A (CDR53B) and 2 (4) cm^2 for RD53A (RD53B).

For transparency it should be pointed out that the work presented in this chapter was done with help and guidance from colleagues from University of Bonn.

4.2 Basic concepts

4.2.1 Eye diagrams

An eye diagram is a tool used to visualize the quality of signal e.g. its timing jitter, voltage noise of logic levels or slew rate of transitions. The diagram is constructed by recording the transient waveform of a signal, dividing it into equal pieces (usually two bit periods long) and overlaying them on top of each other as presented in Fig. 4.5. The result will resemble a shape of an eye. Usually the eye diagrams will be distorted by the noise present in the measured signal, leading to a reduction of the horizontal opening E_h (mostly caused by timing jitter) and vertical opening E_v (mostly due to voltage noise). This makes eye diagrams a very efficient tool for quick evaluation of the signal integrity by simply checking how "closed" the eye is. The opening of the eye is obviously influenced by all noise sources, both deterministic and non-deterministic, therefore it is important when measuring a circuit performance to keep in mind the potential influence of other system components on the signal quality. For example a very low mass cables, such as those planned to be used in HL-LHC ATLAS, can significantly dampen the signal and limit the transmission channel bandwidth, resulting in degradation of the system performance. Measuring eye diagrams in several important points of the system (close to transmitter and receivers) can provide a lot of information about the sources of signal degradation in the system and is therefore crucial for detector design.

4.2.2 Timing jitter

One of the most important performance metrics for a CDR circuit is the timing jitter. A simple, general definition of jitter is provided in [80]: "Jitter is defined as the short-term variations of a digital signal's significant instants from their ideal positions in time". While this definition allows for an easy understanding of the essence of this phenomenon, a few things need to be specified more before jitter can be used as metric of circuit (or data link) performance:
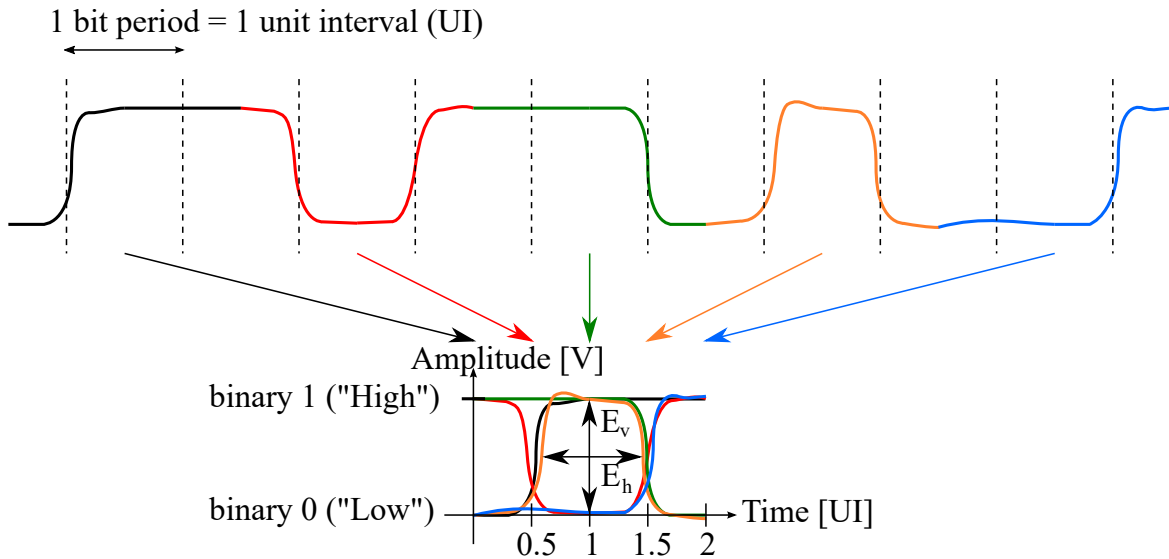
Figure 4.5: Visualization of an eye diagram construction concept.

- "short-term": any variation occurring at very low frequency is considered as wander, while any variation more frequent is classified as jitter. The frequency threshold value between the two is arbitrary, but a commonly used value is 10 Hz [81]. Wander is considered as inconsequential in most serial link, because it is usually eliminated by CDRs and PLLs present in the system.

- "significant instant": since jitter is most commonly characterized for digital signals and in vast majority of cases this means binary signals (two voltage levels), the most significant instances in the waveform are the transitions between logic levels. Any edge-sensitive circuit will detect such an event when the signal in question crosses a pre-defined threshold. The absolute value of the threshold depends on the technology used and the circuit architecture, but in first order approximation it will be close to average between logic low and high voltage. Any noise on the signal when it is crossing the threshold level will increase the jitter seen by the receiving circuit.

- "ideal position": defining ideal positions for a clock signal is conceptually easy - they correspond to the positions of the significant instances of an ideal (jitter free) clock with constant frequency and phase (same as the one analysed). In case of a data stream, where there are no transitions when the same bit value is transmitted one after another, it is practically more difficult to define ideal positions, but the concept is the same - jitter-free clock signal should be defined with appropriate frequency and phase.

Keeping those definitions in mind it is possible to measure jitter in several different ways (notations refer to Fig. 4.6):

- period jitter $J_p(i) = t_i^a - t_{i-1}^a$: measurement of the values of periods of the analysed waveform. This is the simplest measurement to perform and it does not require any knowledge about the ideal reference signal. While it allows to asses the frequency stability of observed signals it does not reveal any longer-term effects.

- cycle-to-cycle jitter $J_{cc}(i) = J_p(i) - J_p(i-1)$: shows differences between consecutive periods. While it does not provide much additional information compared to J_p, when plotted it sometimes allows to more easily observe the instantaneous behaviour of the signal.

- k-cycle jitter $J_k(i) = t_i^a - t_{i-k}^a$: extension of J_{cc} to k-periods of the signal, effectively averaging the instantaneous variations it allows to observe a longer-term behaviour

- Time Interval Error $\text{TIE}(i) = t_i^r - t_i^a$: measure of how far a given edge is from its ideal position. This is the only metric of the ones mentioned here that requires the knowledge about the ideal signal, which makes this value much more complicated to calculate.

This is not a comprehensive list, more can be found e.g. in [82]. For the rest of this chapter whenever jitter is mentioned it should be understood as TIE, unless otherwise specified.
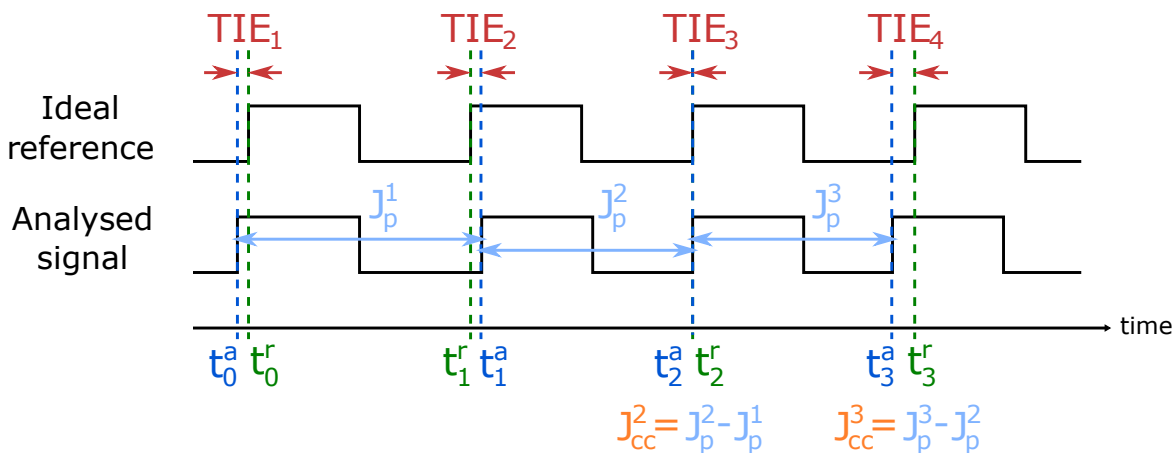
Figure 4.6: Example of comparison between and ideal clock signal and a jittery one with indicated several types of jitter.

Jitter is usually caused by several sources. Histogramming the observed jitter values can help find the cause of timing noise, since specific jitter types produce distinct histogram shapes. Basic classification of the jitter types with corresponding histogram shapes is presented in Fig. 4.7, each category can be in short explained as:

- Random Jitter (RJ) - timing noise which cannot be predicted, it does not have any pattern. Primarily caused by thermal noise and can be well described by fitting a Gaussian distribution. It is unbounded i.e. can theoretically attain any value and the peak-to-peak values will grow with increasing the observation time. This type of jitter is present in every physical system, an example of an eye diagram showing purely random jitter is shown in Fig. 4.8(a)

- Deterministic Jitter (DJ) - repeatable and predictable component of the jitter. They are bounded in values it can attain and can be subdivided into several categories:

    - Duty Cycle Distortion (DCD) - caused either by difference between slew rate of the rising and falling edges, or by shifted decision threshold. The end result is a distorted eye diagram with crossing away from usual threshold level (as shown in Fig. 4.8(b)), visible in jitter histogram as two peaks

- Data Dependent Jitter (DDJ) - sometimes also referred to as Inter-Symbol Interference (ISI). It manifests itself as a correlation between edge transition time and the pattern of bits prior to it. Can be caused by the frequency response of a cable or device (different transmission properties depending on the frequency of the transitions). The resulting eye diagram will show few distinct lines during the bit transition (example shown in Fig. 4.8(c)), leading to few spikes in the jitter histogram.

- Bounded Uncorrelated Jitter (BUJ) - disturbance not correlated with the rate of tranistions in the analyzed signal. Usually caused by crosstalk either from other signal lines or the power rails. It can be subdivided into two types:

  * Periodic Jitter (PJ) - caused by a periodic noise source e.g. a clock signal. If the noise is sinusoidal it produces a very characteristic histogram shape, as shown in Fig. 4.8(d)

  * Non Periodic Jitter (NPJ) - bounded jitter which does not exhibit any timing correlation
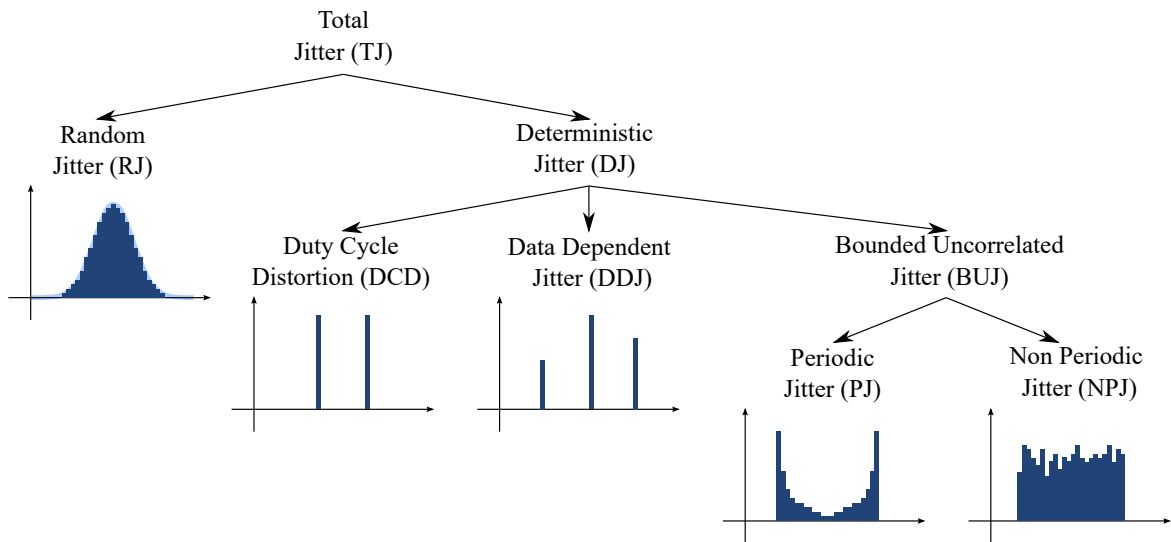
Figure 4.7: Decomposition of jitter into most common types together with typical shape of histogram associated with each jitter type.

In a real system it is rather rare to observe signals with purely random jitter or deterministic jitter. In most cases several jitter sources will be present, leading to the jitter histogram being a convolution of the basic shapes described above, as shown in Fig. 4.9. Due to this effect the two peaks in the jitter histogram of Fig. 4.8(b) look like Gaussian distributions.

Jitter quantification

In order to compare jitter performance of data streams or clock signals several different metrics can be used. One of the more commonly used is Total Jitter at Bit Error Rate (TJ@BER). It describes with what confidence level the jitter will be kept within a given value. For example a TJ = 100 ps @ BER = 10^{-9} means that one out of 10^9 edges of the signal will exceed 100 ps jitter. The required BER value is system specific, however a standard used in the industry is BER = 10^{-12}

(a) Purely random jitter.

(b) Duty cycle distortion with random jitter component.

(c) Data dependant jitter.

(d) Periodic (sinusoidal) jitter.
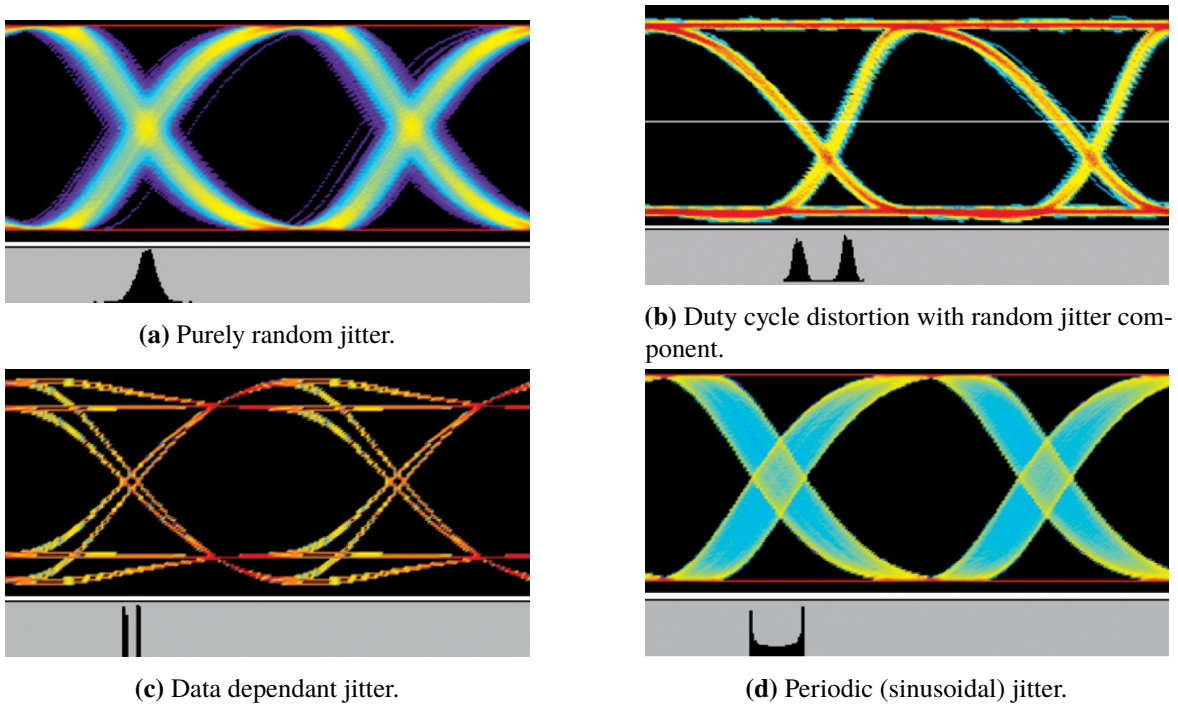
Figure 4.8: Examples of influence of different type of jitter on the eye diagrams. Histogram of jitter at first crossing point is shown in the grey field under each diagram. [80]
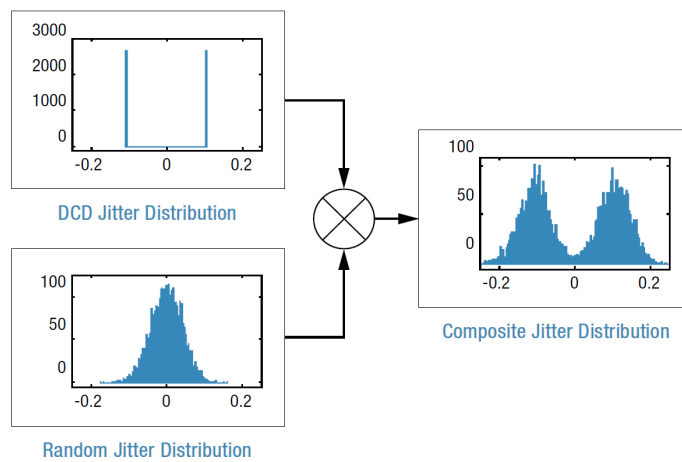

Figure 4.9: Illustration of an idea of convolution of two jitter types. [80]

and some applications go even to BER = 10^{-15}. Measuring jitter for such low BER can be very time consuming, therefore high speed oscilloscopes designed for measuring fast data streams are often equipped with software capable of estimating expected total jitter at any BER based on a relatively small waveform sample[1].

In this work however the TJ@BER metric was not used, because the BER required in the detector system has not been defined so far. For this reason the jitter for the rest of this chapter will be given as a peak-to-peak jitter and, if possible, as the root mean square (rms) of the Gaussian fit of the jitter histogram. All measurements presented further on were done by gathering enough statistics for the result to be stable and reproducible. However, since all real signal streams exhibit random jitter, which follows the Gaussian distribution, the total jitter will grow with the number of observed waveforms (theoretically reaching infinity after infinite observation time). This means that this approach is not as robust and indicative as using TJ@BER. In hindsight choosing an arbitrary BER value and following the TJ@BER metric would have been a better approach, but this conclusion was made much too late for repeating of all the measurements to be feasible.

4.2.3 CDR working principle

The aim of this section is to provide a qualitative overview of the working principle of an analogue, PLL-based CDR circuit and introduce its main building blocks. The general ideas described here are applicable to all CDR circuits designed within this thesis, however the implementation details of the blocks vary significantly between CDR53A/RD53A and CDR53B/RD53B implementations, therefore they will be discussed separately in appropriate sections. This section considers only an analogue PLL-based CDR circuit. This is one of the most commonly used architectures, however many alternatives exit, some of which can even be implemented fully using the digital design flow. An overview of several architecture types together with their advantages, disadvantages and typical applications can be found in [83].

Fig. 4.10 presents a generic CDR schematic. The main input signal to a CDR is a data stream *CMD* (for RD53 chips this data carriers the commands to the chip). The phase of the bit transition edges in *CMD* steam is compared with the phase of the transition edges in the recovered clock *REC CLK* produced internally by the CDR. Since a clock signal has twice the number of edges compared to data stream, usually only rising clock edges are considered. In steady state this phase relation is constant, usually set for the edges to be either aligned to each other or for the clock edges to be in the middle of data stream bit values, depending on the requirements of the system. If the phase relation is not as desired the CDR needs to correct it.

The comparison between the phases of *CMD* and *REC CLK* is performed by the Phase Detector (PD) circuit. After every comparison the outputs of the PD get activated according to the observed situation:

- if *REC CLK* is too slow, the *UP* signal is set high

- if *REC CLK* is too fast, the *DN* signal is set high

---

[1] This is done by decomposing jitter measured in the sample into different jitter categories (described in previous paragraphs), predicting the jitter for each category at required BER separately and finally combining all the elements together. The software and algorithms used for those calculations differ between oscilloscope manufacturers and therefore the end result can be different depending on the device used for the measurement.

- if *REC CLK* is in the correct position both outputs remain low (in practice this is a very rare outcome, since due to jitter and circuit detection uncertainties there nearly always will be a phase difference)

The duration for which the *UP* or *DN* signals are activated depends on the architecture of the Phase Detector. It can for example be proportional to the phase difference between *CMD* and *REC CLK* (this is the case for CDR53A/RD53A) or it can have fixed length (CDR53B/RD53B implementation). The *UP* and *DN* are received by the Charge Pump (CP) circuit, which converts those digital signals into current pulses (output through *CPout* port). The current from CP is integrated by a Low Pass Filter (LPF) resulting in creation of the *Vctrl* voltage. This voltage is the input to the Voltage Controlled Oscillator (VCO) block, which outputs a clock signal (*VCO CLK*) with frequency proportional to the *Vctrl*. The high frequency VCO output is sent to the Frequency Divider (DIV) block, where a slower clock *REC CLK* is produced and the division factor *N* is such that the newly created clock has the appropriate frequency ($f_{VCO\,CLK} = N \cdot f_{REC\,CLK}$). Thus, the adjusted *REC CLK* is created, completing the loop.

At first glance the scheme of using CP and LPF might seem unnecessarily complicated, since they only translate two voltage signals (*UP*, *DN*) into one voltage (*Vctrl*), but they are actually crucial for good operation of the loop. The *UP* and *DN* signals are digital, therefore on their own they would only allow to set VCO into few output frequencies. In order to slowly vary the VCO output frequency across a wide range of frequencies the *Vctrl* needs to be created in such a way that every comparison result from PD varies it only by a small amount. While it means that the VCO output will follow the phase changes of the *CMD* with a delay, this is usually not a problem since in steady state operation no sudden large phase jumps of *CMD* are expected. The wide tuning rage with high granularity of adjustment for the VCO is needed for practical reasons, because there are many factor which will influence the voltage-to-frequency relation of the VCO in the real world. On one hand the range has to be wide enough to accommodate any variations caused by CMOS process manufacturing non-idealities, while on the other hand the control voltage has to be constantly adjusted to correct for jitter and environmental changes (temperature, variations in power supply voltage, radiation damage, etc.).

One important detail not mentioned so far, is the lack of frequency comparison between *CMD* and *REC CLK* in the PD block. This means that before the steady state is achieved, the desired operating point of the loop (V_{ctrl} voltage and resulting frequency of *VCO CLK* and *REC CLK*) has to be achieved by some other means. The approach to this startup issue was very different between CDR53A/RD53A and CDR53B/RD53B implementations, therefore it will be described in their appropriate sections.

Every element of the circuit contributes to the total jitter performance. However, since a CDR is a closed loop feedback system, contributions from different places in the loop have different effects on the output jitter. This is a very complex topic, more details can be found in [82], [84] or [85].

Additionally, since a CDR is a system with feedback loop, it has to be kept stable i.e. the phase margin of the loop gain has to be kept above 45° (typically, for safety above 60°, in order to reduce jitter peaking) [82].
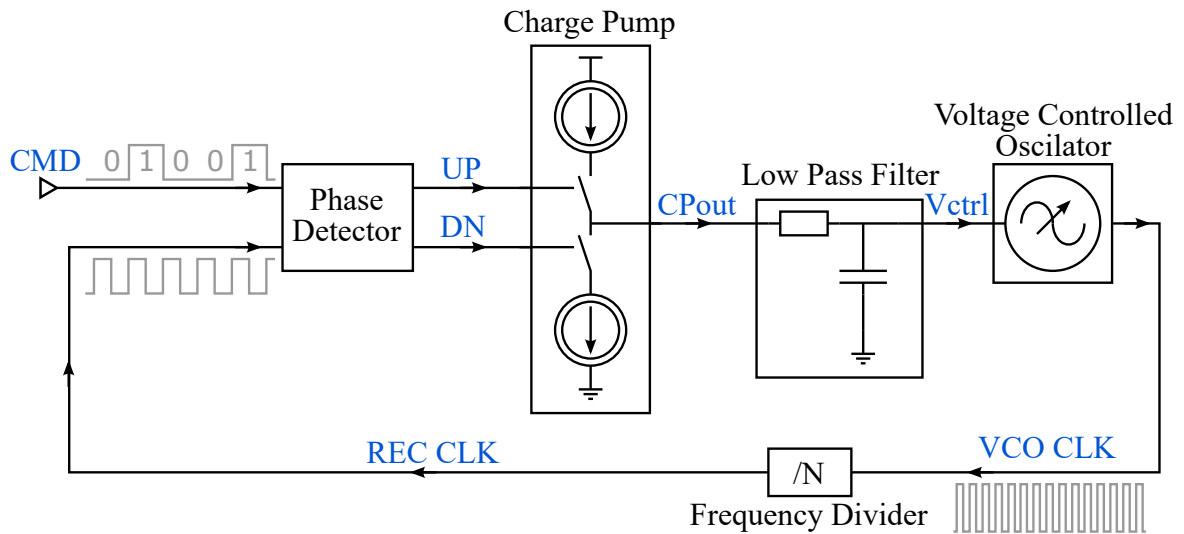
Figure 4.10: Generic block schematic of an analogue CDR.

4.3 CDR53A prototype

At the point in time when CD53A was designed, the performance requirements for the I/O link of RD53 chip were unknown. The only certain specification point was rate of input *CMD* (160 Mbps), while the jitter requirements were completely unknown and the data rate was supposed to be either 1.28 Gbps or 1.6 Gbps. Therefore, the CDR53A was designed on "best effort basis", without aiming for a specific performance metrics.

4.3.1 CDR architecture and building blocks design

This section will go over implementation details of each of the CDR's building blocks (as described in Section 4.2.3) followed by the simulation results of the whole circuit and description of the CDR53A prototype chip.

4.3.1.1 Phase detector

In order to understand the design of Phase Detector used in CDR53A, it is beneficial to first consider a Phase-Frequency Detector (PFD) shown in Fig. 4.11. The action of this circuit is quite simple: a rising edge of *CMD* (*REC CLK*) triggers a D-Flip Flop causing *UP* (*DN*) to be set high. When both *UP* and *DN* are high the *RST* goes high, clearing the D-Flip Flops and setting *UP* and *DN* to low. Example waveforms of circuit operation are shown in Fig. 4.12, where Fig. 4.12(a) presents case of *REC CLK* lagging behind *CMD*, while Fig. 4.12(b) shows the case of *REC CLK* being too early in respect to the *CMD*. From the waveforms it is also clearly visible that the width of *UP* and *DN* signals is proportional to the phase difference between *CMD* and *REC CLK*.

This circuit is very popular in PLL designs because of its simplicity and good performance. However, it is not suitable for CDR operation since the input (*CMD*) to a CDR is not a clock but a data stream, which prevents the circuit from working correctly due to lack of edges when the same bit values are transmitted consecutively. In order to overcome this limitation the PFD was modified by adding a
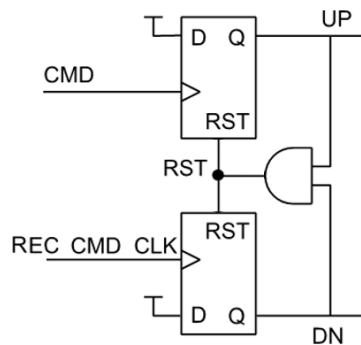
Figure 4.11: Schematic of a Phase-Frequency Detector commonly used in PLLs.

clock gating circuitry in the path of the *REC CLK* as well as delay in the path of *CMD*, as shown in Fig. 4.13. An example of simulation result showcasing the operation of this circuit is presented in Fig. 4.14. The *CMD* is used to create the *GCLK EN* signal, which enables gating of the *REC CLK* signal. Thanks to this the PFD only receives *REC CLK* when there was a rising edge in the *CMD*. If the gated clock *GCLK* was being compared directly with *CMD* the CDR circuit would not function properly because *DN* could never be asserted (the *GCLK* would never have any edges preceding *CMD* in phase). Therefore, it was necessary to add a delay element in the path of the *CMD* and use the delayed version *CMD del* as the input to PFD.

The idea of clock gating PFD is reasonable, however, during the measurements of CDR53A and RD53A shortcomings of the implementation where discovered e.g. issues with locking to correct frequency after startup. This topic is discussed in more detail in Section 4.4.2 and ultimately lead to abandoning this approach in CDR53B/RD53B in favour of a more industry standard PD architecture.
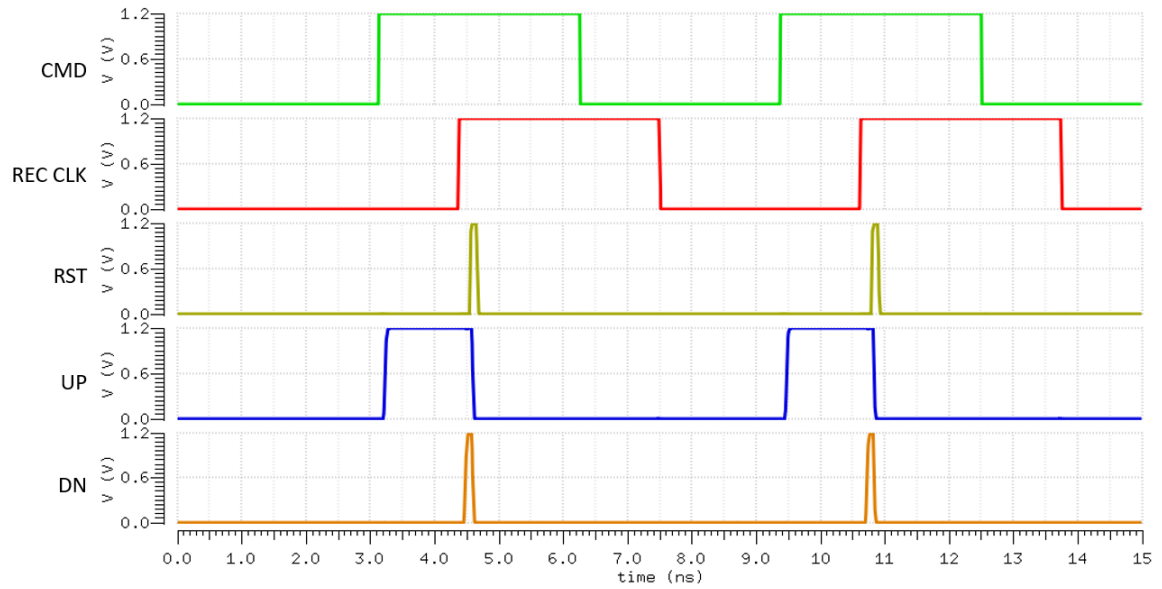
(a) *REC CLK* lagging behind *CMD*.
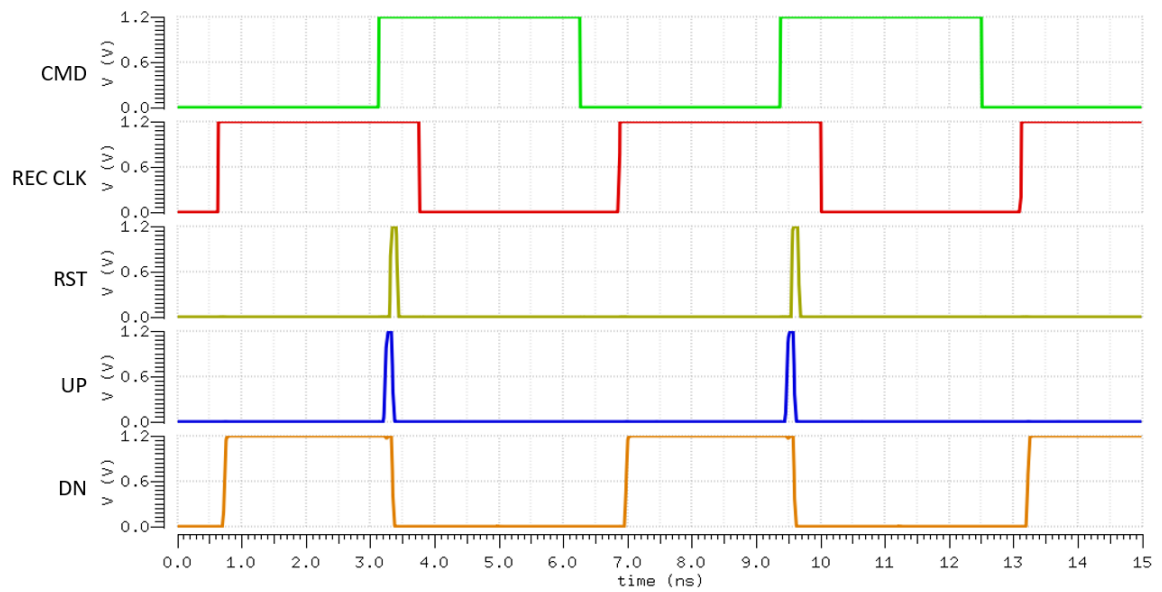
(b) *CMD* lagging behind *REC CLK*.

Figure 4.12: Results of a simulation of the PFD circuit for two different phase relations between *CMD* and *REC CLK*.
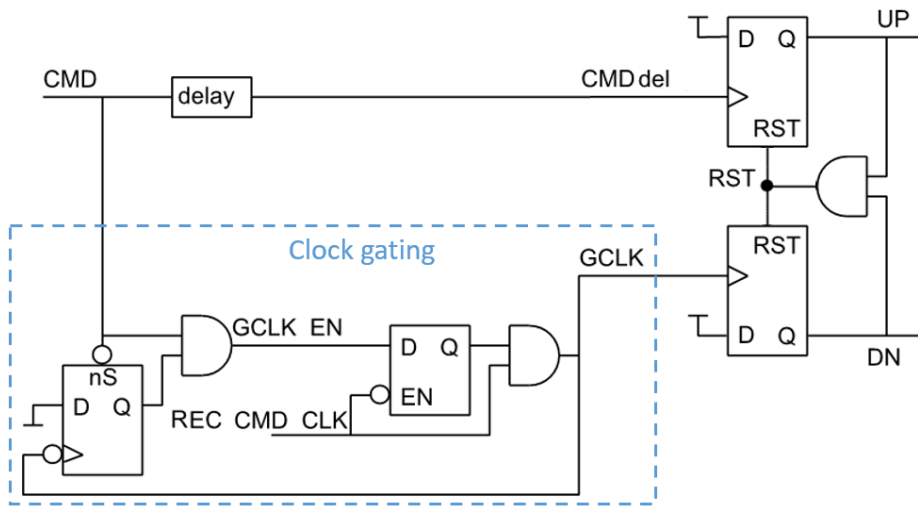
Figure 4.13: Schematic of a clock gating Phase Detector implemented in CDR53A.
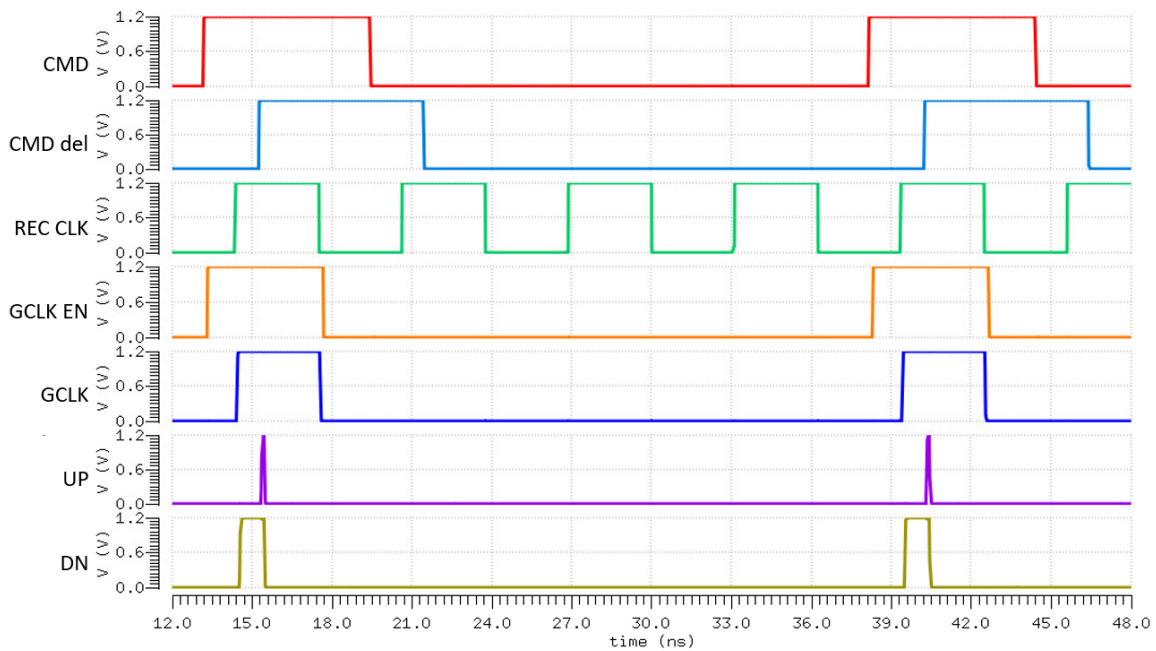
Figure 4.14: Result of a simulation of the clock gating Phase Detector.

4.3.1.2 Charge pump

The schematic of implemented charge pump is shown in Fig. 4.15. The current sources CS_1, CS_2 charge and discharge the Low Pass Filter based on the status of *UP* and *DN* signals. Since it is important that the amount of current injected into LPF is well defined, the current sources are kept on all the time (turning them on and off takes time and causes output current to vary), but the path of the current flow is steered differently depending on the state of *UP* and *DN* signals:

- *UP*=0, *DN*=0 - transistors M_2 and M_4 are turned off (no current is sent to LPF), transistors M_1 and M_3 are turned on which lets the current to flow from VDD to the dummy load branch $(R_{1,2}, C_{1,2})$ or to GND

- *UP*=1, *DN*=0 - transistors M_1 and M_4 are turned off, transistor M_2 is conducting and allowing CS_1 to charge LPF, transistor M_3 is conducting to let CS_2 current flow from the dummy branch

- *UP*=0, *DN*=1 - transistors M_2 and M_3 are turned off, transistor M_4 is conducting and allowing CS_2 to discharge LPF, transistor M_1 is conducting to let CS_1 current flow into the dummy branch

- *UP*=1, *DN*=1 - transistors M_1 and M_3 are turned off, transistors M_2 and M_4 are turned on which lets the current to flow from VDD to the LPF or to GND. This is not a desired situation, since the current flow to LPF is not controlled, however the time duration of *UP*=1 and *DN*=1 is very short for the used Phase Detector architecture, therefore this is not a big problem

Such current flow manipulation technique is called drain switching and is commonly used in charge pumps.

The dummy load branch $(R_{1,2}, C_{1,2})$ is implemented only to help with the current flow. Its component values were chosen such that the impedance of this branch matches the impedance of the implemented Low Pass Filter.

The timing of turning on/off transistors $M_1 - M_4$ is very important for keeping all current pulses equal. Since inverters are needed only for transistors M_2 and M_3, the delay they introduce is compensated for by adding transmission gates (configured to be always on) for transistors M_1 and M_4.
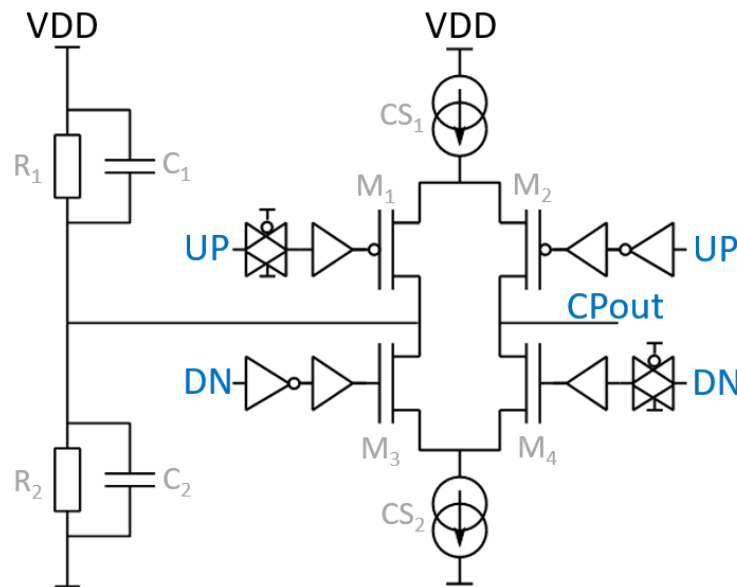
Figure 4.15: Schematic of the charge pump implemented in CDR53A.

4.3.1.3 Low Pass Filter

As mentioned in Section 4.2.3, the loop bandwidth of the CDR has to be chosen such that it assures stable operation as well as appropriate jitter filtering and performance. For this design a simple passive third order RC filter was chosen (shown in Fig. 4.16). While for stability of the loop even simpler second order filter (Fig. 4.16 without R_f and C_f) would have been enough, the additional filtering was supposed to lead to better performance. The filter elements were chosen based on guidelines from [84], which suggest to set the filter bandwidth to 10% of input rate.
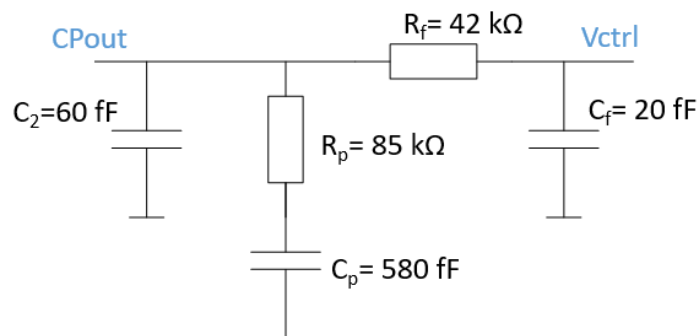
Figure 4.16: Schematic of the third order Low Pass Filter implemented in CDR53A.

4.3.1.4 Voltage controlled oscillator

The voltage controlled oscillator implemented in CDR53A is a ring oscillator design, as shown in Fig. 4.17. The three differential stages form the core of the ring oscillator. Each of those stages

is essentially an inverter and since there are three of them connected in a loop, the structure will produce a self sustaining switching signal at the *RING OSC* nodes. The frequency of that switching is controlled by the control voltage *Vctrl* and bias voltages *VBP, VBN*. Since the ring oscillator is made out of differential stages, a differential-to-single ended converter and a digital buffer are needed to produce the single-ended *VCO CLK* used in the rest of the CDR.
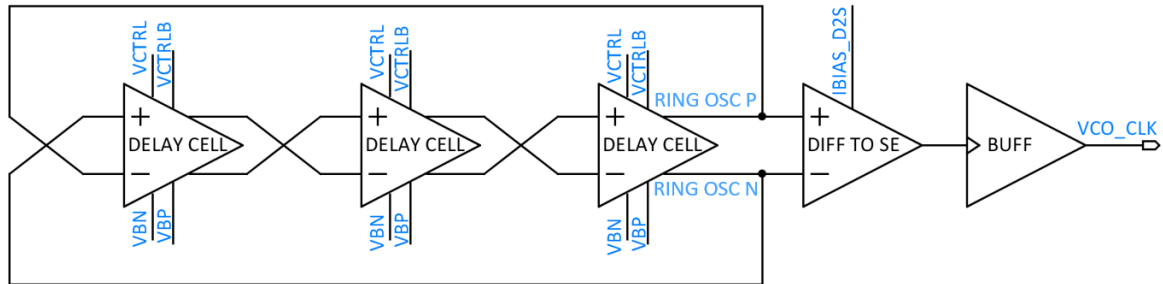
Figure 4.17: Schematic of voltage controlled oscillator implemented in CDR53A.

The schematic of the delay cell used in the ring oscillator is presented in Fig. 4.18. It is a differential pair (transistors $M_1 - M_2$) with pMOS load devices ($M_3 - M_6$). Devices $M_7 - M_8$ provide the tail current to the input pair. Transistors $M_9 - M_{12}$ form a cross-coupled load stage, which allows for rail-to-rail output voltage switching and better performance. The architecture of this delay stage was based on [86] and [87]. The delay with which the output of this circuit responds to input change can be adjusted either by changing static biasing voltages *VBB, VBP* or by a change in the control voltages *VCTRL, VCTRLB*. By changing this delay the frequency of VCO output is changed. The control voltage does not come from the Low Pass Filter directly, rather there is a gain adjustment circuit (shown in Fig. 4.19) between VCO and LPF. The role of this circuit is twofold:

- conversion of the LPF output voltage (*VCTRL IN*) into two complementary voltages (*VCTRL, VCTRLB*)

- adding the possibility of configuring the VCO to have different gain factors K_{VCO} (amount of frequency change in response to the control voltage change). This option was implemented in case of performance issues due to degradation after TID damage

The simulated characteristic curve of the VCO (output frequency as a function of the control voltage) is presented in Fig. 4.20. In the figure all possible gain settings are represented. The default gain setting is 1, leading to $K_{\text{VCO}} = 2\frac{GHz}{V}$. The range of achievable frequencies is quite wide in order to account for manufacturing process variations as well as for the influence of low power supply voltages and temperature.
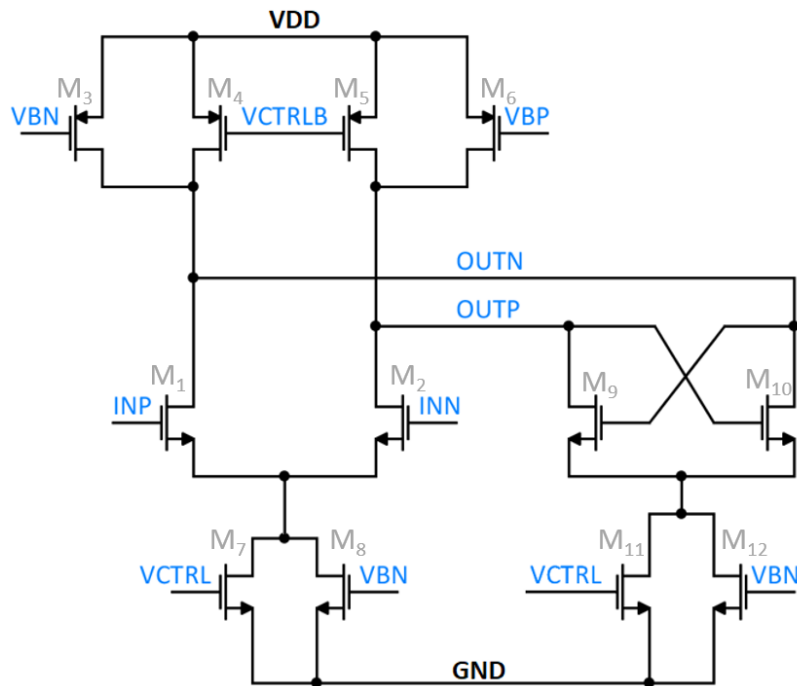
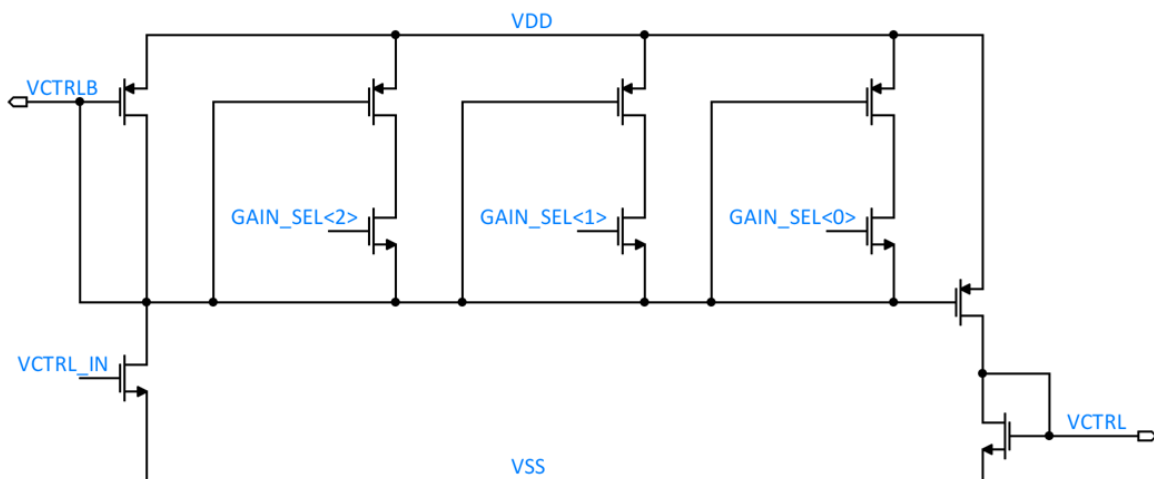**Figure 4.18:** Schematic of a single delay cell used in the CDR53A's ring oscillator.
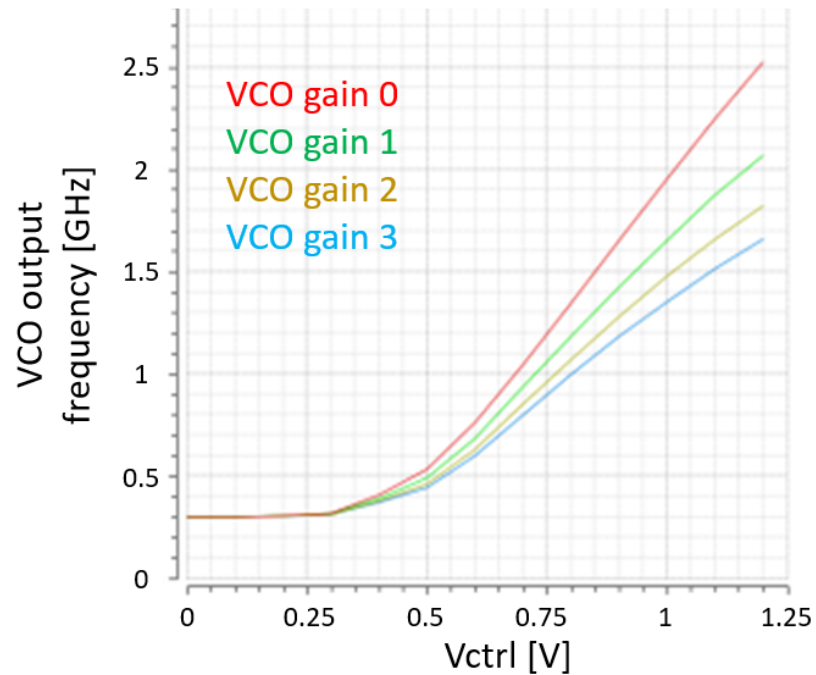

Figure 4.19: VCO gain selection circuit.

Figure 4.20: Characteristic curves of the CDR53A VCO for all available gain settings.

4.3.1.5 Clock divider

At this point in time the required frequency of the *VCO CLK* was supposed to be configurable between 1.28 GHz and 1.6 GHz. Since the *CMD* rate is 160 Mbps this means that two different clock division ratios are needed: 8 and 10. For this reason CDR53A included two clock dividers. Only one of them is active at the time. The schematics of implemented circuits are presented in Fig. 4.21. Those are classical, textbook design of synchronous clock dividers. The synchronous approach was chosen, since it produces a less jittery divided clock than an asynchronous approach.

(a) Synchronous clock divider by 8

(b) Synchronous clock divider by 10
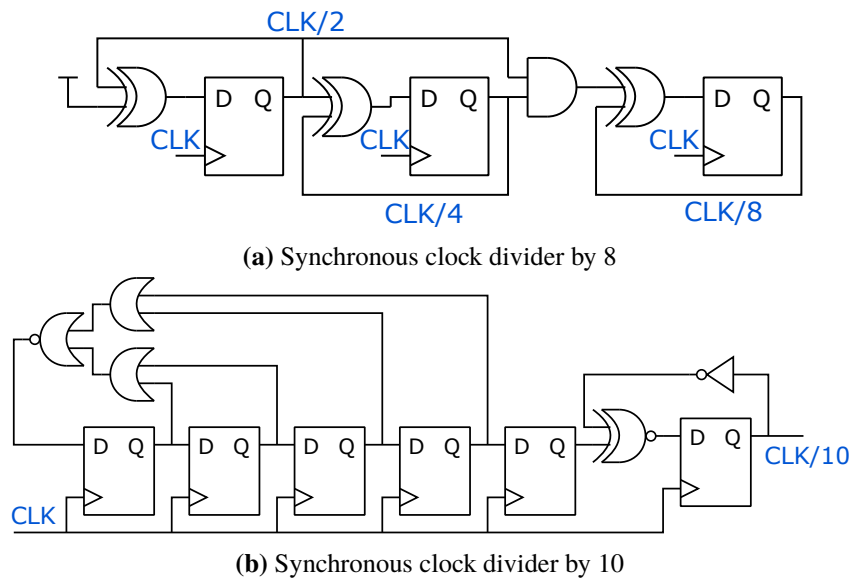
Figure 4.21: Schematics of clock dividers used in CDR53A.

4.3.1.6 CDR simulation results

After each block was designed and verified on its own, they were combined together into the final circuit. The designed turned out to be quite compact, as shown in Fig. 4.22. The placement of the major blocks in respect to each other was optimized for best connectivity between blocks and good power network distribution.
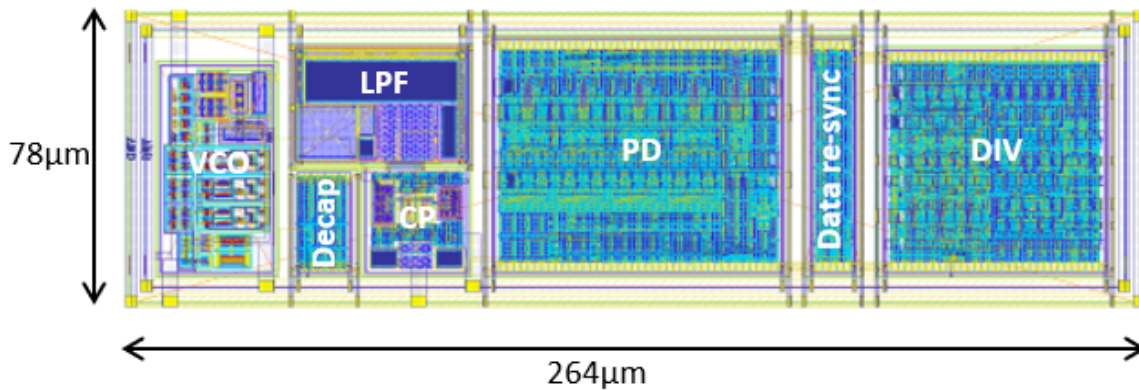
Figure 4.22: Layout of CDR53A CDR circuit with indicated major blocks.

The CDR circuit was simulated and verified, including post-layout parasitic effects. At this point in time the only relatively simple simulation techniques were used i.e. transient simulations with optional inclusion of thermal noise. An example result from such simulation is presented in Fig. 4.23, where the VCO output frequency is plotted as a function of time. During the first 500 ns the input *CMD* is a training pattern (010101..., effectively a 80 MHz clock). Using a training pattern is a common

practice, since it makes it easier for the CDR to lock to the correct frequency at startup thanks to maximal amount of edges present in the *CMD*. In this example it takes approx. 150 ns for the circuit to lock to the correct frequency (1.28 GHz). After the training phase, the *CMD* is changed to a PRBS7 pseudorandom pattern, which has the same run length as commands used to operate RD53 chips. The starting VCO output frequency visible it Fig. 4.23 is 1.08 GHz, while comparison with Fig. 4.20 would suggest that after startup (when *Vctrl* = 0 V) it should be approx. 300 MHz. This discrepancy comes from the fact that in order to reduce the simulation time an initial condition was set on *Vctrl*, bringing it close to correct value at startup. The jitter measured in this simulation with the PRBS7 input was 84 ps peak-to-peak (12 ps rms).
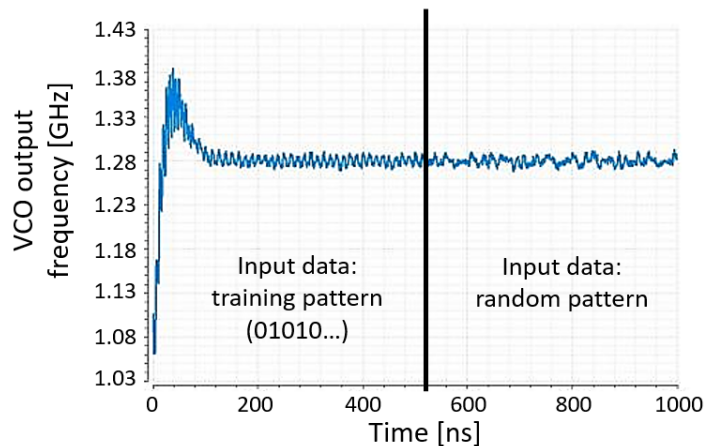
Figure 4.23: Result of transient simulation of the CDR53A circuit.

4.3.1.7 CDR53A test chip

In order to verify the simulated performance of the CDR circuit a small test chip was designed. This chip, called CDR53A, contains not only the CDR but also other circuits needed for the I/O interface of the RD53 chip i.e. data serializer (SER) and current mode logic cable driver (CML) [88]. The schematic overview of the CDR53A chip is presented in Fig. 4.24. All the blocks are connected together in a way allowing the characterization of them as a full chain or separately. The input *CMD* to the CDR is received with standard CMOS input pads. All the configuration bits and bias currents are delivered from the outside through I/O pads. Since all the main blocks are rather small, the test chip size is determined by the number of required I/O pads, as can be clearly seem in the image of the layout of the chip shown in Fig. 4.25. In order to utilize the free space, the area was filled with power decoupling capacitors.
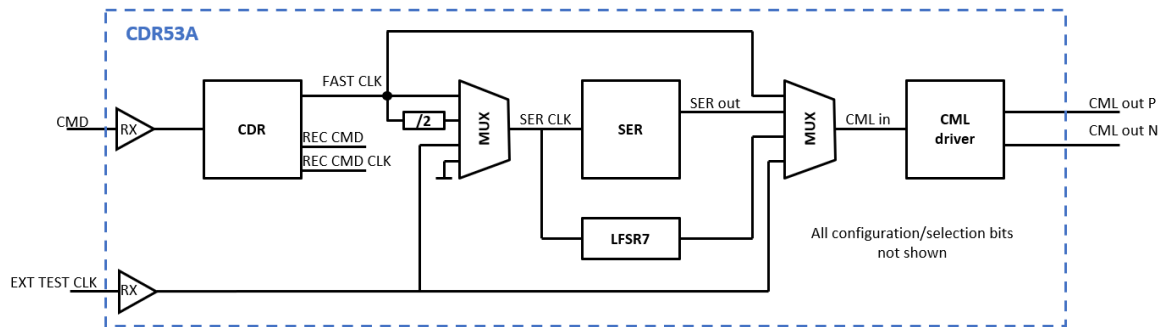
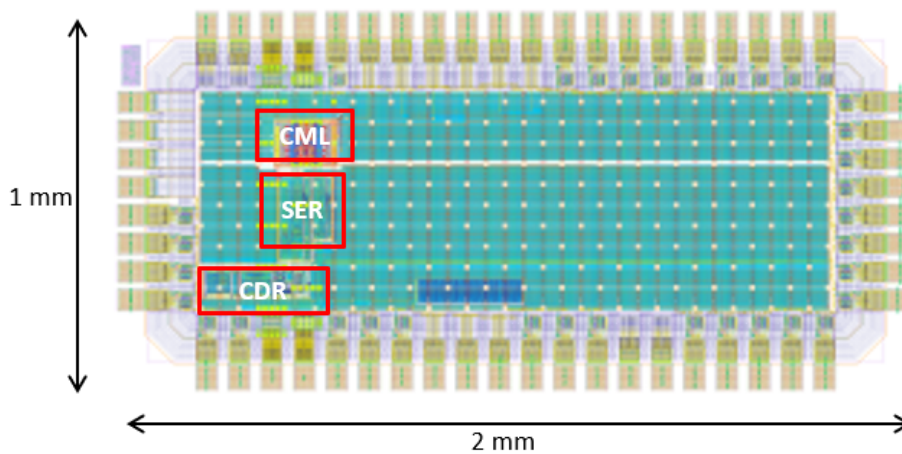**Figure 4.24:** Simplified schematic of circuitry included in CDR53A test chip.

Figure 4.25: Layout of CDR53A test chip with highlighted position of the main building blocks and indicated layout dimensions.

4.3.2 Measurement results

4.3.2.1 Experimental setup

The setup used to control the CDR53A chip during characterization consists of three main components (shown in Fig. 4.26):

- MIO3 board [89], which contains an FPGA and handles the communication with a PC through USB and Ethernet

- General Purpose Analogue Card (GPAC), which contains components such as power supply channels, current sources and level shifters (required for translating control signals from FPGA to voltage levels safe for the CDR53A chip)

- CDR53B carrier PCB

The software and firmware used in this setup is based on Basil [90] - a modular testing and data acquisition framework written in Python and Verilog. The *CMD* input to the CDR is provided either from MIO3/GPAC or from a high precision signal generator Agilent 81134A. The signal generator was

used especially for jitter measurement, since it can provide very clean signals with down to 2 ps rms jitter. The jitter measurements are carried out mainly using a Tektronix MSO 70804 high bandwidth oscilloscope and its DPOJET link quality measurement tool.
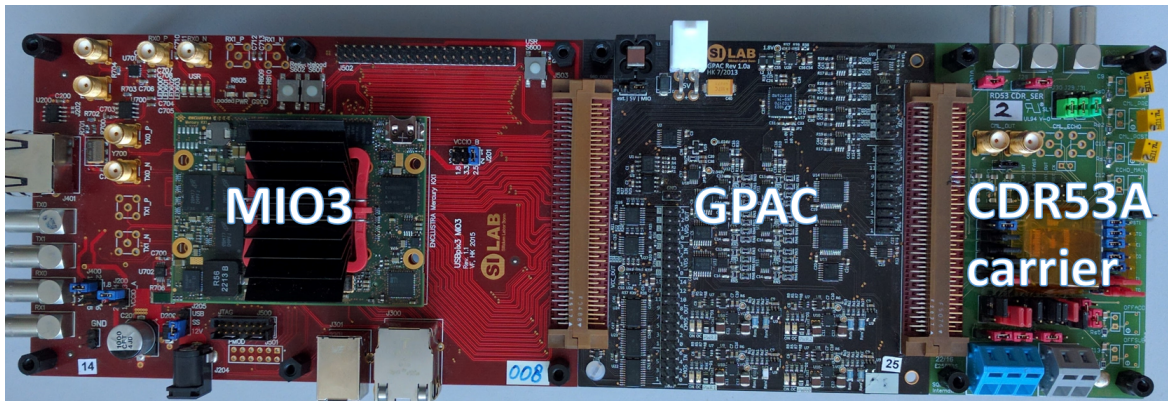

Figure 4.26: Photo of the measurement setup used for characterizing CDR53B.

4.3.2.2 Electrical performance

The testing showed that the CDR53A chip is functional, consumes expected the amount of power (5 mW) and reacts correctly to configuration changes. Operation at both 1.28 Gbps and 1.6 Gbps output rates is possible. However, the jitter performance did not match the simulation.

The first example of the measured signal quality is shown in Fig. 4.27. In this measurement the CDR53A was supplied with a high quality (2 ps rms jitter) PRBS7 pattern at the *CMD* input. The 1.28 GHz clock generated by the CDR was sent to the Linear Feedback Shift Register (LFSR), which generated a PRBS7 pattern. The PRBS7 data was sent off chip by the CML driver and received directly by the oscilloscope. The measured jitter is 19.8 ps rms (164 ps peak-to-peak), which is nearly twice larger than the simulated jitter figures. The jitter histogram is can be well fitted with a Gaussian, suggesting that the jitter performance is dominated by random jitter.

The second example of the measured signal quality is shown in Fig. 4.28. In this case the *CMD* is again a PRBS7 (high quality) stream, but the CDR-generated 1.28 GHz clock is used by data serializer (the same SER circuit is used by RD53 chip), which produces PRBS7 pattern. Same as in previous case, the CML driver send the data off-chip. The resulting eye-diagram has much higher total jitter and a clearly visible lines in the bit transitions, causing two-peak structure of the TIE histogram. This distortion is a result of the serializer circuit utilizing both rising and falling edges of the 1.28 Gbps clock (double data rate SER architecture), which turned out to not have ideal (50%) duty cycle.

After long investigation several reasons for the jitter performance discrepancy between simulation and measurements were found:

- The simulation and jitter calculation methods used were not robust and extensive enough. Additionally, the simulation time for the whole CDR simulations was not long enough to capture a suitable amount of data.

- The bandwidth of the loop was set too high. The bandwidth should be set as a compromise between cleaning the input jitter (better with lower bandwidth) and following the input phase
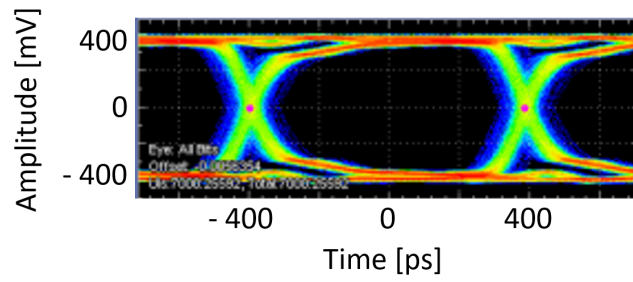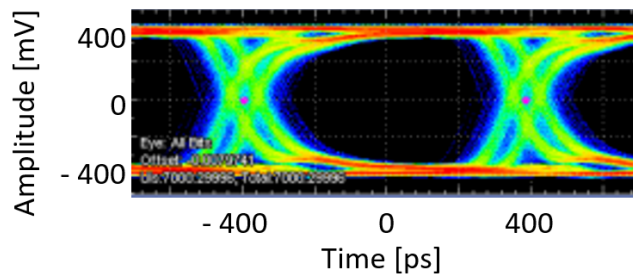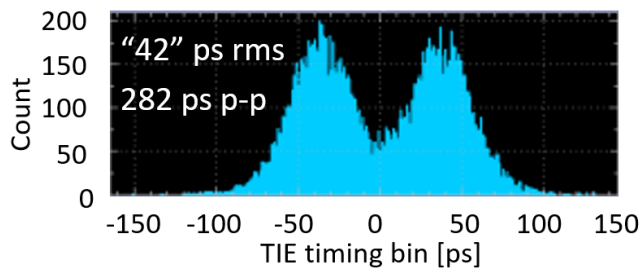
(a) Eye diagram.
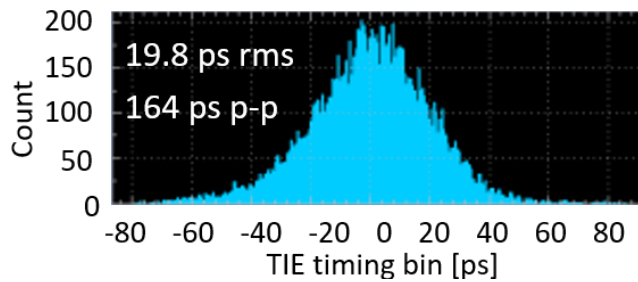
(b) TIE histogram with an indication of the measured jitter values.

Figure 4.27: Signal quality measurement results of 1.28 Gbps PRSB7 data stream generated with LFSR7 using CDR generated clock. The input to the CDR is a PRSB7 pattern with 2 ps rms jitter.

(a) Eye diagram.

(b) TIE histogram with an indication of the measured jitter values. Due to two-peak structure of the histogram (caused by DCD in SER), the rms value is not realistic.

Figure 4.28: Signal quality measurement results of 1.28 Gbps PRSB7 data stream generated with SER using CDR generated clock. The input to the CDR is a PRSB7 pattern with 2 ps rms jitter.

changes (better with higher bandwidth). Additionally, it also affects the contribution of each of the CDR loop's building blocks to the total output jitter.

- The single-ended CMOS receivers used for receiving the *CMD* input turned out to not be very suitable for operation at 160 Mbps and ended up injecting additional jitter into the CDR

4.3.2.3 TID hardness

The CDR53A prototype was irradiated at Karlsruher Institut für Technologie using X-ray source (tungsten target X-ray tube biased at 60 kV and 30 mA, thin aluminium filter was used to condition the radiation spectrum [61]). The irradiation took 180 h, which allowed to reach 490 Mrad TID. First 2 Mrad were achieved using radiation dose rate of 300 krad/h, for the rest the dose rate of 3.9 Mrad/h was used. During the irradiation the chip sample was not cooled (temperature measured with an NTC close to chip was in the range of 24 °C - 28 °C). The measurement setup used during the irradiation was similar to the one used for electrical characterization, but in order to avoid damaging MIO3 and GPAC cards, the CDR53A carrier PCB was redesigned and connected to the other boards with a 3 m IDC cable. Since the radiation damage to a transistor depends, among many other factors, on its biasing point, the chip was operated all the time during irradiation.

Due to problems with the irradiation setup (X-ray tube's high voltage cable inducing noise into power cables of the CDR53A chip, resulting in significant jitter increase) as well logistic issues, during the irradiation the jitter was characterized mainly using the Xilinx IBERT IP [91][2] implemented in the FPGA of MIO3. The measurement results are presented in Fig. 4.29. The overall link quality degrades with TID, however, the circuit functioned properly with the default configuration up to approx. 210 Mrad. At that point the bias current of the differential-to-single ended converter of the VCO had to be increased from 40 μA to 60 μA in order to restore link stability. While this change allowed the circuit to function for some more time, at higher TID additional configuration changes were needed: increase of *VDD* to 1.3 V at 300 Mrad, increase of differential-to-single ended converter bias current to 80 μA at 380 Mrad and finally increase of *VDD* to 1.4 V at 490 Mrad. Those results suggest that the TID was causing strong VCO degradation, however due to non-ideal conditions during irradiation (high temperature compared to operation in ATLAS, issues with measurement setup) it is difficult to reach a definitive conclusion. Replicating those results in simulation was not possible at that time, due to lack of models of irradiated transistors.

---

[2] This macro is provided by Xilinx. IBERT stands for Integrated Bit Error Ratio Test. It is capable of quantifying the quality of data a stream by sampling the waveform at several thresholds and estimating the BER at each of sampled points. It does not allow to measure jitter in picoseconds, rather it assigns the measured stream a score based on the estimated BER.
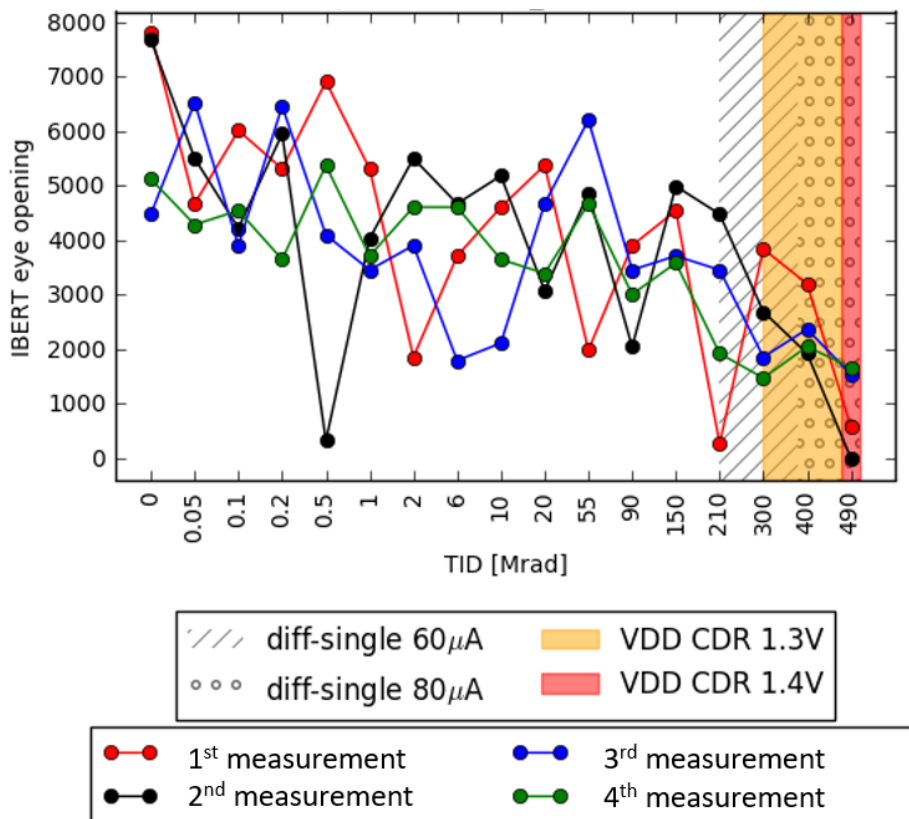
Figure 4.29: Degradation of 1.28 Gbps output link quality of CDR53A due to TID damage as measured by IBERT IP. At each of the indicated TID steps the eye quality measurement was taken four times, as indicated by different coloured curves. The 160 Mbps input PRBS7 *CMD* was supplied by MIO3/GPAC. Failure points indicated by coloured regions. Meaning of IBERT score is described in text.

4.4 RD53A CDR

This section presents only topics related to the CDR of RD53A, both in terms of design implementation and the measurement results. Details of other aspects of the chip can be found in numerous publications e.g. [28] [92] [93], to name a few.

4.4.1 CDR implementation

The CDR circuit integrated into RD53A is nearly identical to the one from CDR53A. While it was clear that the overall performance of the CDR would have to be improved for the proper operation inside the detectors, the performance of CDR53A circuit was deemed enough to allow operation and characterization of the RD53A chip. The only modifications introduced to RD53A's CDR were:

- Removal of clock divider by 10. It was unnecessary, since before RD53A submission it was decided that all RD53 chip moving forward will only utilize 1.28 Gbps output data rate.

- Added the possibility to set the VCO control voltage externally. The reason to add it was twofold. Firstly it was a test feature, helpful with measurement of the VCO. Secondly, it was a safety feature to be used in case of chip startup issues (more details about this aspect are given in Section 4.4.2.2).

- Increased decoupling on biases. This is a general good practice, which was not fully implemented in CDR53A.

- Not directly connected to CDR design, but nevertheless important for the output link quality - the duty cycle distortion issue in the serializer (described in Section 4.3.2.2) was fixed by adding a re-timing D Flip-Flop at the SER output.

4.4.2 Measurement results

4.4.2.1 Measurement setup

Two DAQ systems for operating RD53 chips were developed within the RD53 collaboration:

- YARR [94]: PCIe-based readout system with software written in C. The DAQ used for ATLAS ITk will most likely be based on this system.

- BDAQ53 [95]: Ethernet-based verification and readout system with software written in Python. The focus of this DAQ is ease of use, providing fast debugging features and covering many measurement scenarios (single chip electrical testing, operating chips in testbeam environment, multi-chip module testing, wafer probing, etc.)

The chip carrier PCB for the RD53A chip is called a Single Chip Card (SCC) and is common to both DAQ system. The SCC contains only passive components needed for the chip's operation and configuration, as well as cable ports for providing power and I/O signals. A photo of BDAQ53 system connected to a RD53A SCC is shown in Fig. 4.30. All measurements described in this chapter were obtained with BDAQ53 system, supplemented with Agilent 81134A signal generator and Tektronix MSO 70804 high bandwidth oscilloscope.
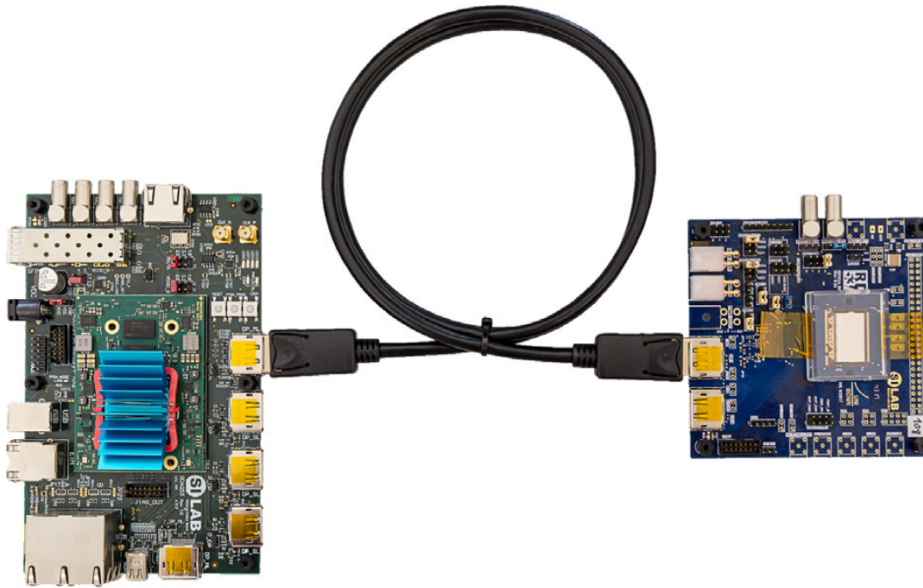
Figure 4.30: BDAQ53 base board (left) with Mercury+ KX2 daughter board, connected to an RD53A Single Chip Card (right) via a DisplayPort cable. [95].

4.4.2.2 Startup reliability

Reliable startup (correct behaviour after powering the circuit up) is vital for proper operation of the pixel detector. In case of the CDR, the expected behaviour is to lock to correct frequency and provide stable clocks to the rest of the chip. If something goes wrong e.g. CDR locks to a harmonic of the base frequency or does not lock at all, the chip needs to be power cycled. For a single chip operation and characterization a problem with startup is inconvenient, but can be usually solved quickly and easily. However, for the actual operation of the pixel detector a problem with startup is a much bigger issue, since due to the chosen powering scheme for pixel readout ASICs (serial powering [96]) it is only possible to power on/off an entire stave of modules, not a single chip. As a result, the startup reliability of the CDR is just as important as good jitter performance.

The startup reliability measurement procedure for the CDR is quite simple, as shown in Fig. 4.31. The environmental settings of particular concern are the *VDD* voltage, chip temperature, TID level, as well as chip settings. The number of power cycle repetitions *N* is arbitrary, but should be high enough to give statistically significant and reproducible results.

The initial measurements showed a rather poor reliability, where the success rate was 38% for $VDD > 1.15V$ and 0% for lower *VDD* voltages (at room temperature with default chip settings and un-irradiated). This issue was caused by the Phase Detector. Through simulations it was discovered that depending on the initial conditions of the D Flip-Flops after startup, as well as the initial phase difference between *CMD* and *REC CLK*, it is possible for the Phase Detector to enter a "no-return state". In this state the PD continuously sets *UP* signal to high, which in turn causes VCO control voltage to raise to *VDD* voltage level. As a result, the *VCO CLK* settles at maximum frequency it can achieve (approx. 1.9 GHz) and is not correlated in phase to input *CMD* at all.

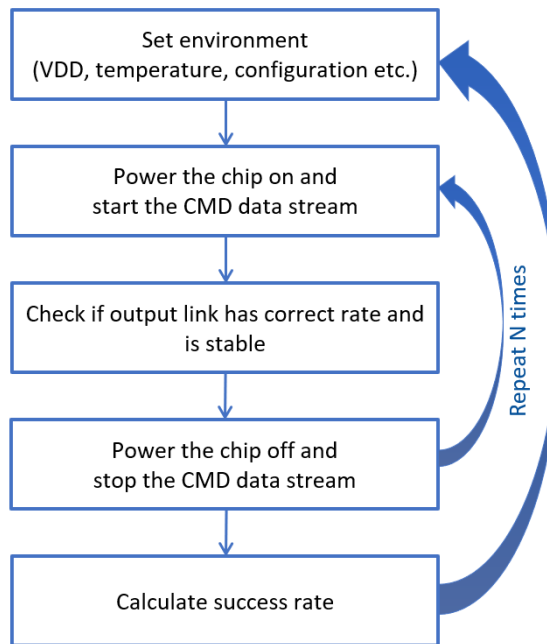This problem was fixed for the RD53A chip by utilizing the output of the Power on Reset (PoR)

Figure 4.31: Flow diagram of CDR startup reliability measurement of the RD53A.

circuit [97] to change the startup behaviour of the CDR. The action of PoR circuit is shown in Fig. 4.32 - it provides a signal, which for some time after chip power up stays in a known state (low in this case) and afterwards changes to the opposite state (high in this design). [3] While it was not originally envisioned to use the PoR signal to reset the CDR by default, provisions were made to make it possible i.e. a pMOS transistor was added to the CDR, which when enabled allows overwriting the VCO control voltage with an external one. When this overwrite mechanism is enabled, the Phase Detector is disabled, so it cannot disturb the overwriting. The SCC has the capability to connect the PoR output to the gate of the reset pMOS, as shown in Fig. 4.33. The external VCO control voltage was set to approx. $800\,\mathrm{mV}$, which is close to the voltage resulting in *VCO CLK* frequency of $1.28\,\mathrm{GHz}$. As a result, after *VDD* is ramped up at power up, the CDR is reset and set close to correct operating point. After PoR signal goes high, the CDR reset is disabled and it works normally. This modification of SCC configuration lead to the startup success rate of 100% for $VDD > 1.15V$, therefore solving the issue for the RD53A.

---

[3] Inclusion of PoR circuit is quite common in ASIC designs, since a chip after powering up is in an unknown state and has to be rest to the expected default condition, which can be achieved by utilizing the PoR output.
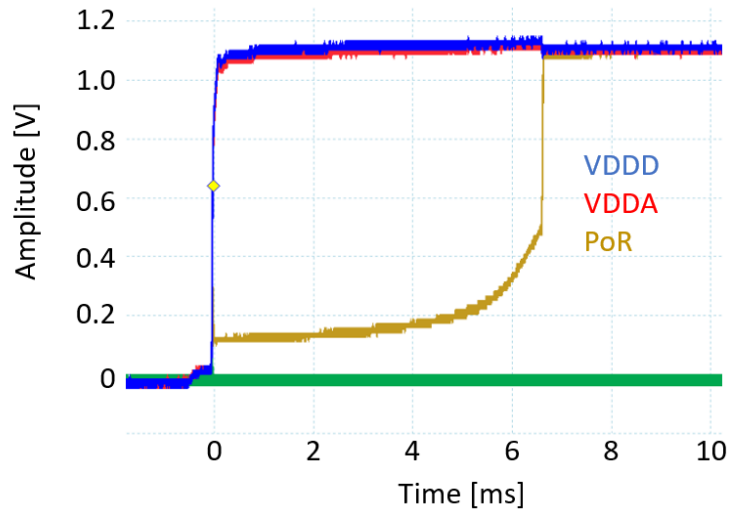
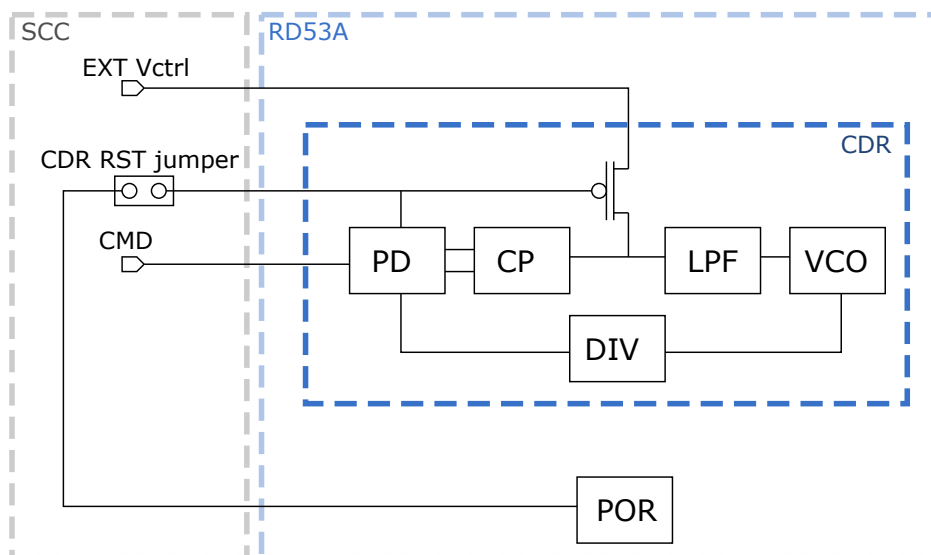**Figure 4.32:** Measurement of PoR output during powering the chip up [98].

Figure 4.33: Simplified schematic of connection between CDR and PoR circuits.

4.4.2.3 Jitter performance

Initial measurements of the output link performance of RD53A showed results similar to those of CDR53A, which was expected. An example of an eye diagram measurement is shown in Fig. 4.34(a), where the measured jitter is 13.5 ps rms. The observed signal is a 640 MHz clock (created from *VCO CLK* of 1.28 GHz), which is the highest frequency signal coming directly from the CDR which is accessible from the CML outputs [4]. When compared with Fig. 4.27(a), the lower jitter numbers can be explained by using a better receiver for the *CMD* (CMOS in case of CDR53A, LVDS for RD53A) as well as more stable biases (all bias voltages and currents in RD53A are generated internally). This measurement, however, was taken with the RD53A in low activity state i.e. with the pixel matrix turned off. In quickly became apparent that turning the matrix on significantly degrades the output link jitter performance, as shown in Fig. 4.34(b) where the measured jitter was 25.9 ps rms. While a slight increase would be understandable (large number of digital switching in the chip is bound to disturb power supply rails or couple noise through metal lines or substrate), such a large difference was surprising.

A long investigation of this issue was carried out through measurements of the chip with different configuration settings and careful re-examination of the CDR implementation. The problem seemed to be caused by the lack of power domain crossing buffers for the bits controlling the gain setting of the VCO (*GAIN_SEL<0-2>* in Fig. 4.19). The long metal lines carrying those bits are coming directly from the digital domain and pick up noise along the way, which then couples to the delicate analogue circuit controlling the VCO. While this hypothesis was supported by the measurement results, it is not possible to prove it with a standard RD53A chip because it is not possible to remove the coupling noise from the gain selection lines. Since it was important to have the definite proof of the cause of this issue, so it can be avoided in the next chips, it was decided to find a way to modify a few RD53A chips in a way that would allow validating the noise coupling hypothesis. Such a modification could theoretically be done by changing the chip design and manufacturing the new version, but this would be prohibitively expensive and time consuming. Instead a Focus Ion Beam (FIB) [99] technique was used in order to physically alter a few samples of RD53A. FIB method allows cutting individual metal traces in the chip as well as making new ones e.g. to attach the cut lies to other metal traces. In this case the modification consisted of cutting the VCO gain selection bit lines and hardwireing them to CDR's power supply lines. Since none of the institutes within RD53 collaboration had their own FIB machine, the modification was done by an external company (MASER Engineering [100]). After the modification the chips were tested - Fig. 4.34(c) shows an eye diagram of a FIB modified chip with pixel matrix turned off, where jitter was measured at 13.3 ps rms. For the same chip with matrix turned on, the jitter was 15.2 ps rms (Fig. 4.34(d)), which confirms the hypothesis of large amount of digital noise coupling to VCO through the gain selection bit lines.

Since the DAQ systems used to operate the RD53A could cope even with the large amount of jitter present in the I/O link during maximum matrix activity, it was decided that for this chip the CDR design will not be modified i.e. no alteration of production masks for further chip was done.

---

[4] The 1.28 Gbps Aurora output is of course the default CML output and its quality is in the end the most important for chip operation. However, in order to understand the jitter performance of the CDR it is easiest to examine its output (*VCO CLK*) directly, eliminating any impact of the Aurora logic circuit and serializer on jitter jitter numbers.
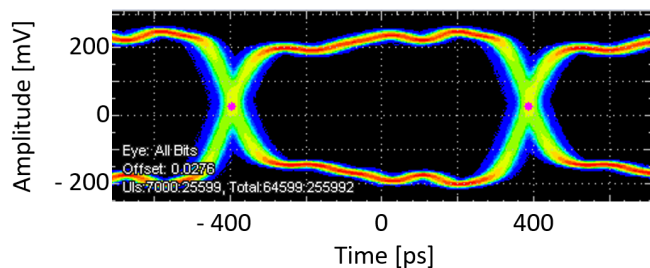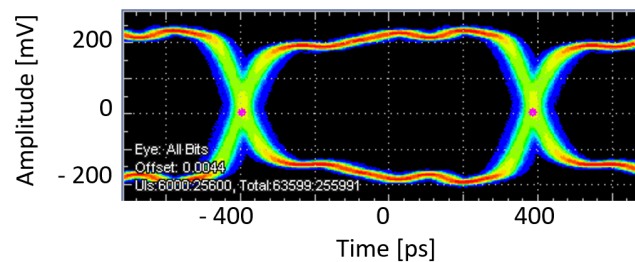
(a) Standard chip (before FIB), pixel matrix inactive. Measured jitter: 13.5 ps rms.
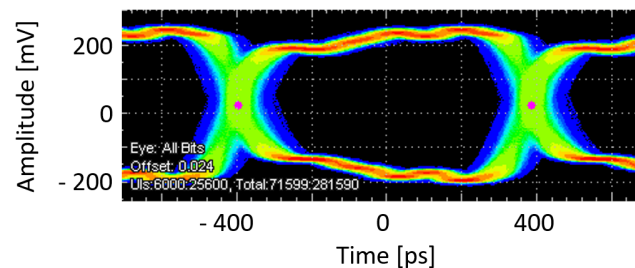
(b) Standard chip (before FIB), entire pixel matrix active. Measured jitter: 25.9 ps rms.
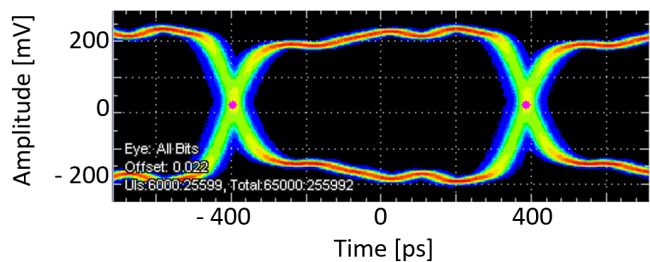
(c) Chip modified with FIB process, pixel matrix inactive. Measured jitter: 13.3 ps rms.

(d) Chip modified with FIB process, entire pixel matrix active. Measured jitter: 15.2 ps rms.

Figure 4.34: Comparison of eye diagrams taken with different configurations of RD53A chip. Each measurement is a results of measurement 640 MHz clock signal generated by the CDR. The input to the CDR is a training pattern (010101..., equivalent to 80 MHz clock) with 2 ps rms jitter.

4.4.2.4 Outcome of RD53A testing

The two previous sections highlighted the biggest issues discovered with the RD53A CDR. While significant, they did not prevent chip characterization. The RD53A was extensively measured in terms of electrical performance, operation in testbeams while bonded to sensors and during TID irradiation campaigns. Overall, hundreds of chips were successfully used in various test scenarios. It was, however, clear that the CDR required significant improvements in order to be suitable for operation in the ATLAS ITk system. For this reason measurement of TID hardness and SEE susceptibility of the RD53A I/O link were abandoned in favour of focusing on the new design - CDR53B.

4.5 CDR53B prototype

4.5.1 Definition of RD53 link quality requirements

Shortly after first samples of the RD53A were tested, the lpGBT ASIC [77] group had finalised the specifications of their design. Since lpGBT will provide the input *CMD* to RD53 and receive it outputs (see Fig. 4.3), the jitter specification of lpGBT effectively defines the requirements of the RD53 I/O link design. The specification is:

- the 160 Mbps *CMD* at the output of lpGBT should have 5 ps rms or less random jitter.

- the 1.28 Gbps *GTX OUTPUT* at the input of lpGBT should have jitter lower than 10 ps rms (60 ps pk–pk) or 40 ps rms (200 ps pk–pk), depending on the mode of data stream sampling of lpGBT (automatic or manual). The decision of choosing between those two modes is not taken yet, as it requires system level measurement.

While this specification gives an indication for the requirements of the link quality, one important element is still unknown - the influence on data transmission of the long, low mass cables used in the ATLAS and CMS experiments for connecting the two ASICs. Low mass cables will introduce signal damping and lead to Inter Symbol Interference (ISI), in turn increasing the jitter observable in the link. As the result, RD53 chips can expect total jitter higher than 5 ps rms at input and have to have lower jitter than 10 ps rms / 40 ps rms at the output in order for the whole system to work.

4.5.2 CDR architecture change and building blocks redesign

Based on the results from CDR53A and RD53A it was clear that in order to meet link quality needs of the lpGBT (and in turn ATLAS and CMS experiments), the CDR circuit had to be significantly improved. In order to fix the discovered shortcomings of the CDR, it was decided to completely change the architecture of several blocks (PD, LPF, DIV), re-optimized the others (CP, VCO) and add a dedicated startup circuitry. Additionally, the simulation methodology of the CDR closed loop had to be improved. All those changes are described in this section.

4.5.2.1 Phase detector

The Phase Detector architecture is completely different than the one used in CDR53A (Section 4.3.1.1). It uses the so-called "bang-bang PD" or "Alexander PD", which was introduced in 1975 [101] and has since been commonly used in CDR designs. Its main feature is that it provides only early/late

information about the phase relation between *CMD* and *REC CLK*, instead of linear measure as the circuit from CDR53A. While this reduces the amount of information and has a significant impact on the overall CDR architecture, e.g. LPF needs to be much larger to achieve stability, thanks to its simple design the circuit is very reliable and does not have "no-return" state issue as the one described in Section 4.4.2.2. The bang-bang PD implementation used in this design is based on [102] and the schematic is shown in Fig. 4.35(a). It is an improvement over the original design, as it adds combinatorial logic for generating *UP, DN* signals which has equalized propagation delay and node loading.

The circuit working principle is taking three consecutive samples of the *CMD* input data using the rising and falling edges of the *REC CLK*. Based on the comparison of taken samples $S1 - S3$ it is possible to identify if a transition was present in the *CMD* and if the *REC CLK* is too early or too late (see Fig. 4.35(b))[5]:

- If the *REC CLK* is too early in relation to the *CMD*, the sample values will be $S1 = S2 \neq S3$. This will result in setting the *DN* signal high and setting the *UP* low in order to reduce the *REC CLK* frequency, thus delaying the phase of its next edge.

- If the *REC CLK* is too late in relation to the *CMD*, the sample values will be $S1 = S3 \neq S2$. This will result in setting the *UP* signal high and setting the *DN* low in order to increase the *REC CLK* frequency, thus advancing the phase of its next edge.

- If the *REC CLK* is in perfect position in relation to the *CMD* a metastable state occurs and the outcome will be random. This state in practice is extremely rare and has no overall impact on the CDR behaviour.

- If no *CMD* transitions are present, than all samples will have the same value, resulting in both *UP, DN* staying low.

While the bang-bang PD can provide reliable information on the phase difference of the two signals, it cannot work properly in case the frequencies of those signals are wrong. This means that after the circuit is powered and the VCO frequency is wrong, some other circuit needs to first lock the loop to correct the frequency and only afterwards the bang-bang PD can function. The solution to this problem is described in Section 4.5.2.6.

Additionally, a rotational Frequency Detector (FD) was added to the CDR. This circuit can detect deviations for the desired frequency of random data, as long as the deviation does not exceed ±25% of the nominal frequency (thus, it cannot solve the startup issue). The circuit architecture is based on [103]. In normal operation this circuit should never be active, however, in case of SEE-induced upsets it might help keeping the CDR operational.

---

[5] It is important to keep in mind that the samples are taken in sequence, so they are not available simultaneously to calculate the *UP* and *DN*. To solve this problem it is necessary to delay by one clock cycle the sample $S1$ and by half clock cycles the $S2$, which are the roles of the DFF1 and the DFF2 D Flip-Flops. The DFF3 and the DFF4 are the sampling D Flip-Flops.

(a) Circuit schematic.
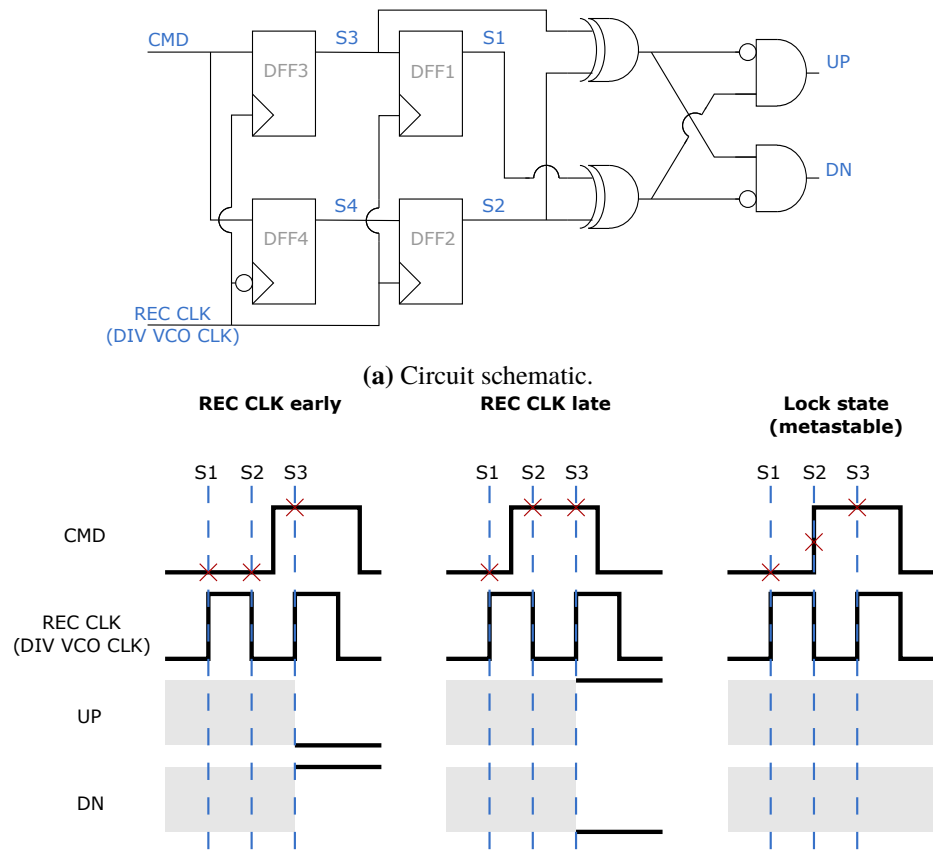
(b) Visualization of circuit operation concept. The grey regions in the waveforms of *UP, DN* symbolize unknown state.

Figure 4.35: Bang-bang Phase Detector used in CDR53B.

4.5.2.2 Charge pump

The charge pump architecture (drain switching) remained mostly unchanged compared to Section 4.3.1.2. The only alteration to the design was removing the dummy branch in favour of adding a unity gain buffer (*OP* in Fig. 4.36). This is a common technique, which helps to keep voltages of both CP's branches equal (similarly to the dummy branch in CDR53A), but also reduce the dynamic glitches during switching the transistors $M_1 - M_4$ on/off. The buffer is realized as a class-A differential pair buffer with power consumption of 0.7 mW. The output current of charge pump (when active) remained unchanged at $I_{CP} = 1\,\mu\text{A}$.
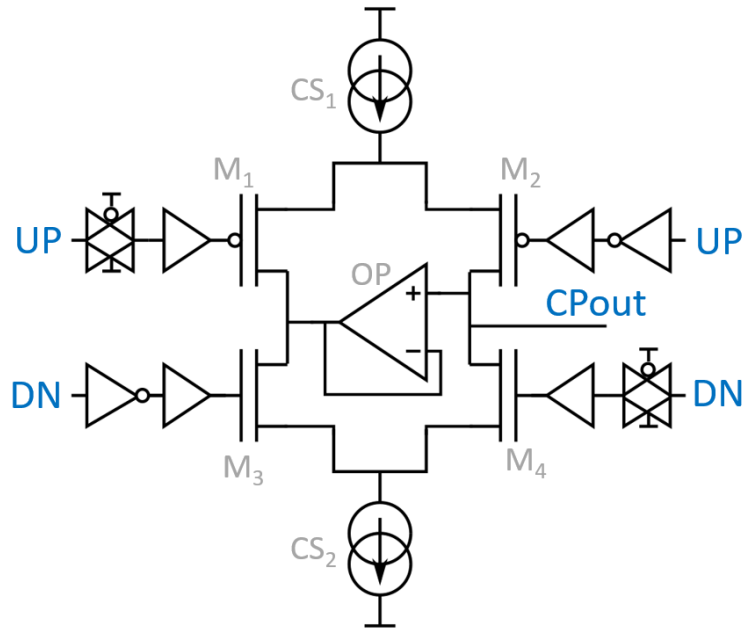
Figure 4.36: Schematic of the Charge Pump implemented in CDR53B.

4.5.2.3 Low Pass Filter

The LPF was significantly altered in comparison to the one described in Section 4.3.1.3. The new filter is second order, instead of third, as the filtering capability of the last stage was not worth the extra thermal noise introduced by it. The CDR53B's LPF is also much larger, with the main capacitor of 300 pF. The filter resistor R_F was made configurable in the range of $50\,\Omega - 1\,k\Omega$ with $50\,\Omega$ step, as visible in Fig. 4.37. The default resistance value is $400\,\Omega$, the configurability was added in order to adjust the bandwidth in case problems are observed during measurements. The secondary capacitor is approximately 75 fF and it is not implemented explicitly, rather the parasitic capacitance of the metal wires is utilized for it.

Such filter configuration allows achieving a low bandwidth of approximately 3 MHz, which is needed for a good jitter performance. The drawback of the chosen values is very large area (250 µm × 800 µm, approximately 80% of entire CDR size) needed for implementation of the LPF. While the same bandwidth could have been achieved by increasing the resistor size while reducing the capacitor, which would result is lower area needed, such trade-off is not desirable from the point of jitter performance. Since the PD uses the bang-bang architecture, during every *UP,DN* cycle the VCO control voltage will change by $Vctrl_{step} = I_{CP} \cdot R_F$, resulting in a frequency step of $\Delta f_{VCO} = Vctrl_{step} \cdot K_{VCO}$. Since the frequency in steady state has to be stable, the value of R_F has to be kept low. This is a simplified overview, more details can be found in [102].
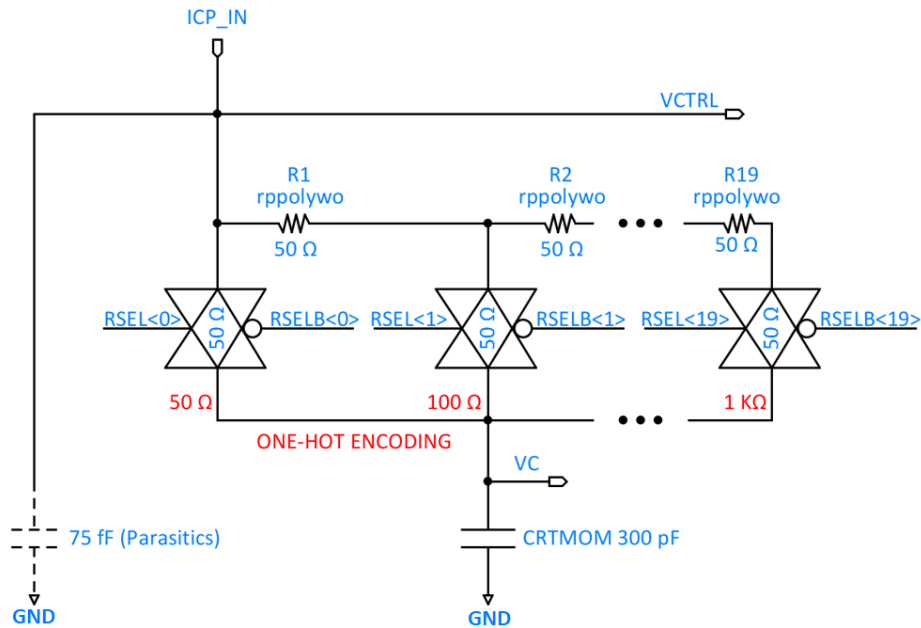
Figure 4.37: Schematic of the configurable Low Pass Filter implemented in CDR53B.

4.5.2.4 Voltage controlled oscillator

The architecture of the VCO remained unchanged from CDR53A i.e. it is based on a ring oscillator built from differential buffers, as shown in Section 4.3.1.4. The transistor sizing was adjusted in order to change the VCO's characteristic curve (output frequency dependence on *Vctrl*), such that the default VCO gain is lowered to 1.5 GHz/V (helps reducing jitter) and the achievable output frequencies are not harmonics of the nominal output frequency (1.28 GHz), which helps with avoiding locking to false frequency. The comparison between the simulated characteristic curves of CDR53A and CDR53B is shown in Fig. 4.38. The same figure also present the simulated impact of manufacturing process and power supply variations (thin lines), as well as prediction of degradation after 500 Mrad TID based on models of irradiated transistors [42].
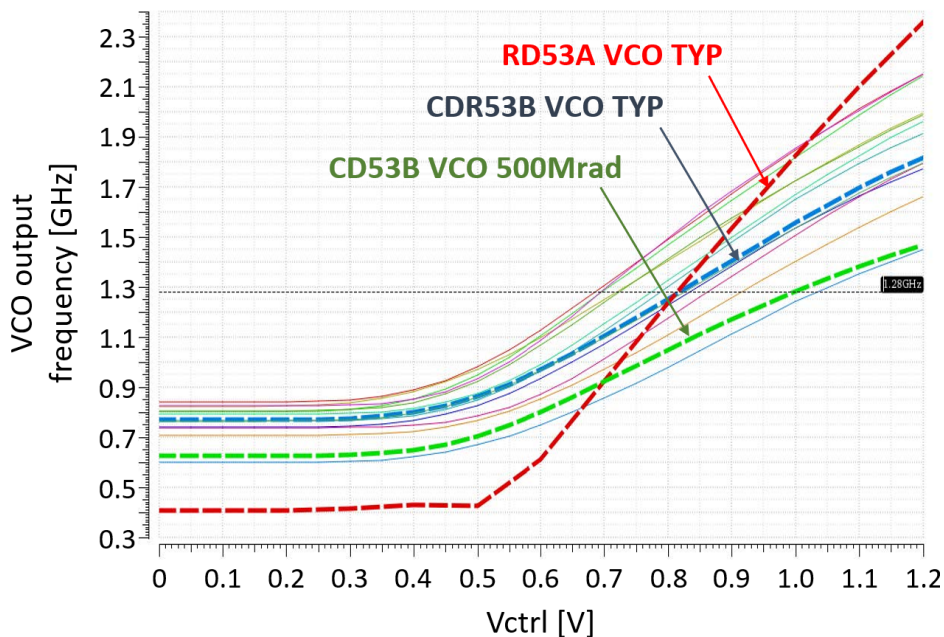
Figure 4.38: Characteristic curve of the CDR53B's VCO. The default (typical manufacturing corner, default power supply voltage of 1.2 V) curve is indicated with thick, dashed blue line. The thin lines represent impact of manufacturing corners and power supply variation of ±10%. The thick, dashed green line show simulation result using models of irradiated transistors. The RD53A VCO curve shown for reference.

4.5.2.5 Clock divider

The clock divider architecture is similar to the one implemented in CDR53A/RD53A - a synchronous divider by 8. However, the design was altered in order to be more resistant to SEE-induced upsets. This is achieved by triplicating all the D Flip-Flops as well the combinatorial logic at their inputs, as shown in Fig. 4.39 (in order to simplifying the drawing only an example of triplication is shown, in reality all divider stages are triplicated).

In addition to standard transient simulations of the divider, done also for previous designs, the circuit was also simulated using models of the transistors irradiated to 500 Mrad TID. Those simulations revealed that the D Flip-Flop degrade significantly enough, that when clocked at 1.28 GHz they do not function correctly. While it is known that the available models of irradiated devices are overly pessimistic and in reality the circuit might work fine, such a big performance loss was worrying. In order to fix this issues D Flip-Flops from two other standard libraries were simulated: 9 track low threshold voltage (LVT) and 12 track LVT [6]. The results of the comparison as shown in Fig. 4.40 (dashed lines indicated output of irradiated devices). It is clearly visible that using LVT cells on its own is not enough, and only larger 12 track library cells remain functional after irradiation. Based on this simulation it was decided that the whole clock divider as well as any other standard cell toggling at high speed will be implemented as the 12 track LVT cell.

---

[6] The "track" nomenclature refers to the size of physical layout of a standard cell. Larger track number means larger cell due to larger (wider) transistors used. The default cell size used in RD53 digital design is 9 tracks.
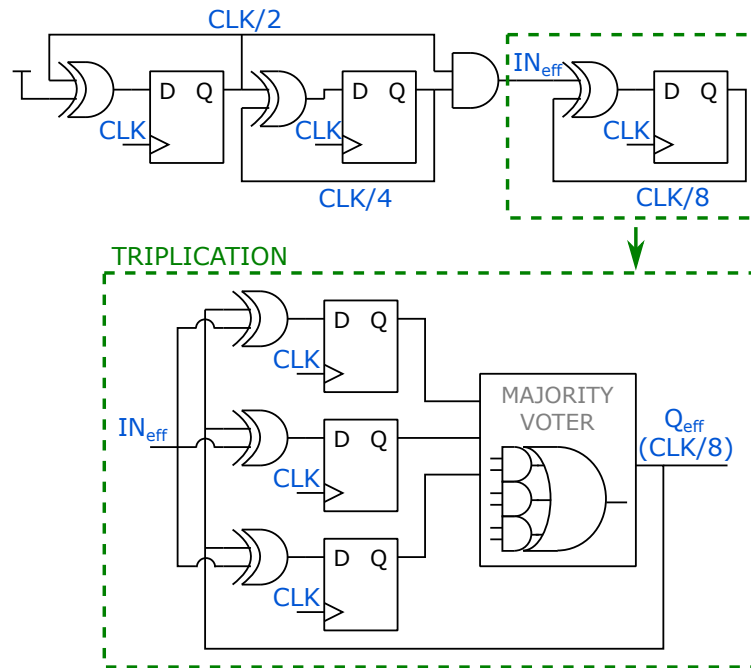
Figure 4.39: Schematic of clock divider by 8 implemented in the CDR53B with indication of used triplication approach. The triplication of only one D Flip-Flop and combinatorial logic at its input is shown for simplicity, in the actual circuit implementation all parts are triplicated.
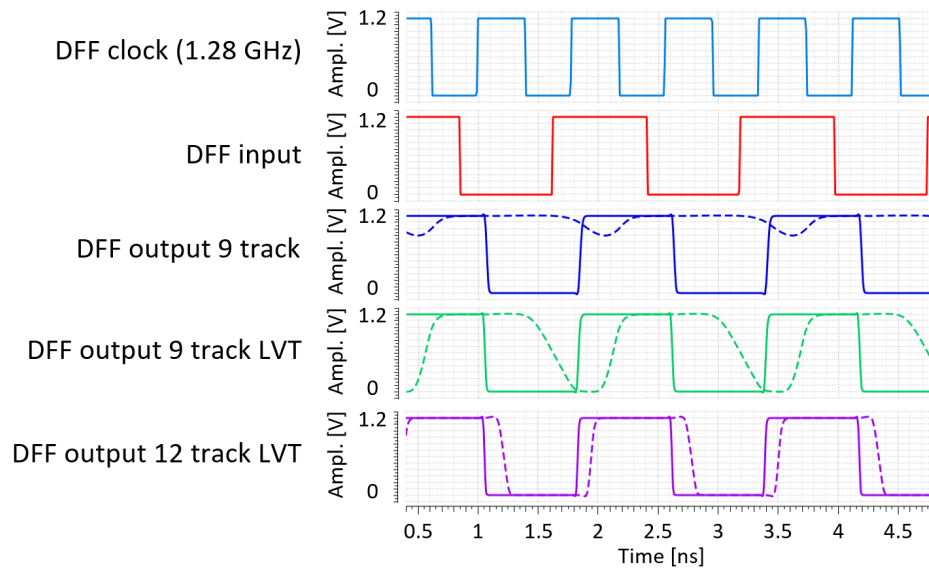

Figure 4.40: Comparison of simulated performance of single D Flip-Flops from three different standard cell libraries (9 track, 9 track LVT, 12 track LVT). In this simulation each D Flip-Flop is clocked at 1.28 GHz and their outputs are loaded with 20 fF capacitance. Dashed lines indicate performance after 500 Mrad TID.

4.5.2.6 Working modes: startup and normal operation

As explained in Section 4.4.2.2 a reliable startup of the CDR circuit is crucial for proper operation of the RD53 chip. One of the main goals of the redesign of the CDR circuit was improving the startup reliability. In order to achieve this a dedicated startup circuitry was added to the CDR and two operation modes were introduced:

- Startup mode shown in Fig. 4.41. This mode is used during power-up and for some time afterwards in order to ensure locking to proper frequency. This is achieved by using a Phase Frequency Detector (PFD, as shown in Fig. 4.11 and described in Section 4.3.1.1), instead of the bang-bang PD, effectively turning the CDR into a PLL with infinite frequency capture range (given enough time the loop will always lock to correct frequency). When this mode is active the expected *CMD* input is a training pattern (01010..., equivalent to 80 MHz clock). In order to lock to the desired frequencies, additional divider stage is activated, such that the feedback clock to PFD is the the *VCO CLK* divided by 16.

- Normal operation mode shown in Fig. 4.42. After the frequency lock is achieved the circuit switches the PFD off and hands over the control to the loop to the PD and FD circuits. At this point there is no expectation of the contents of the *CMD*, except that the run-length (number of consecutive bits of the same value) should not exceed 10.

The switch from the startup mode to normal operation mode is controlled by the startup counter. This block counts the number of edges of *VCO CLK* divided by 40 and once it reaches a pre-defined value CNT_{max}, the working mode is changed to normal operation. A safe CNT_{max} was found through simulation to be 30000, resulting in the startup mode being active for approximately 800 μs after power-up.
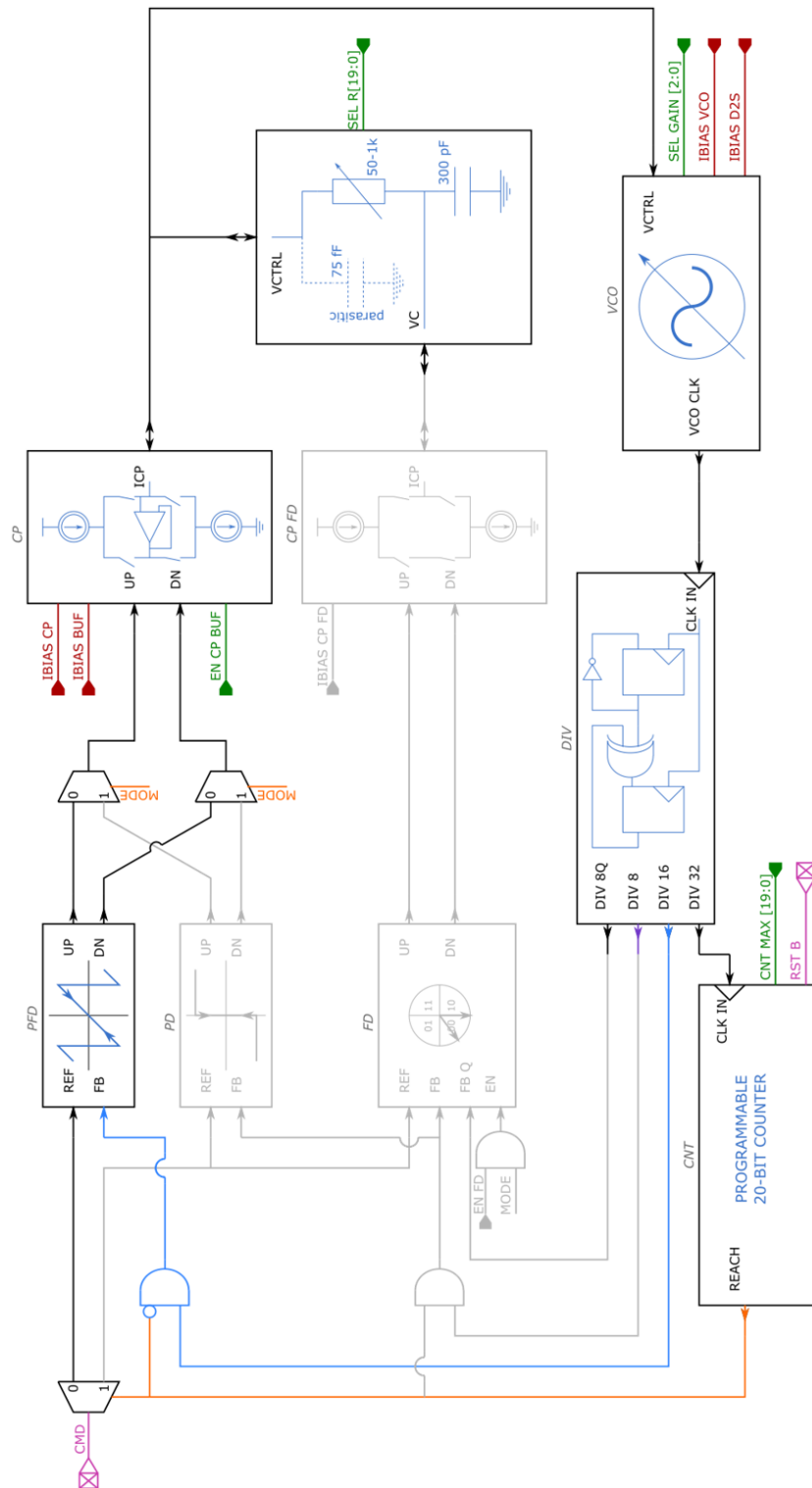
Figure 4.41: Startup mode, expected *CMD* input is the training pattern (010101...). They greyed out parts of the circuit are inactive.
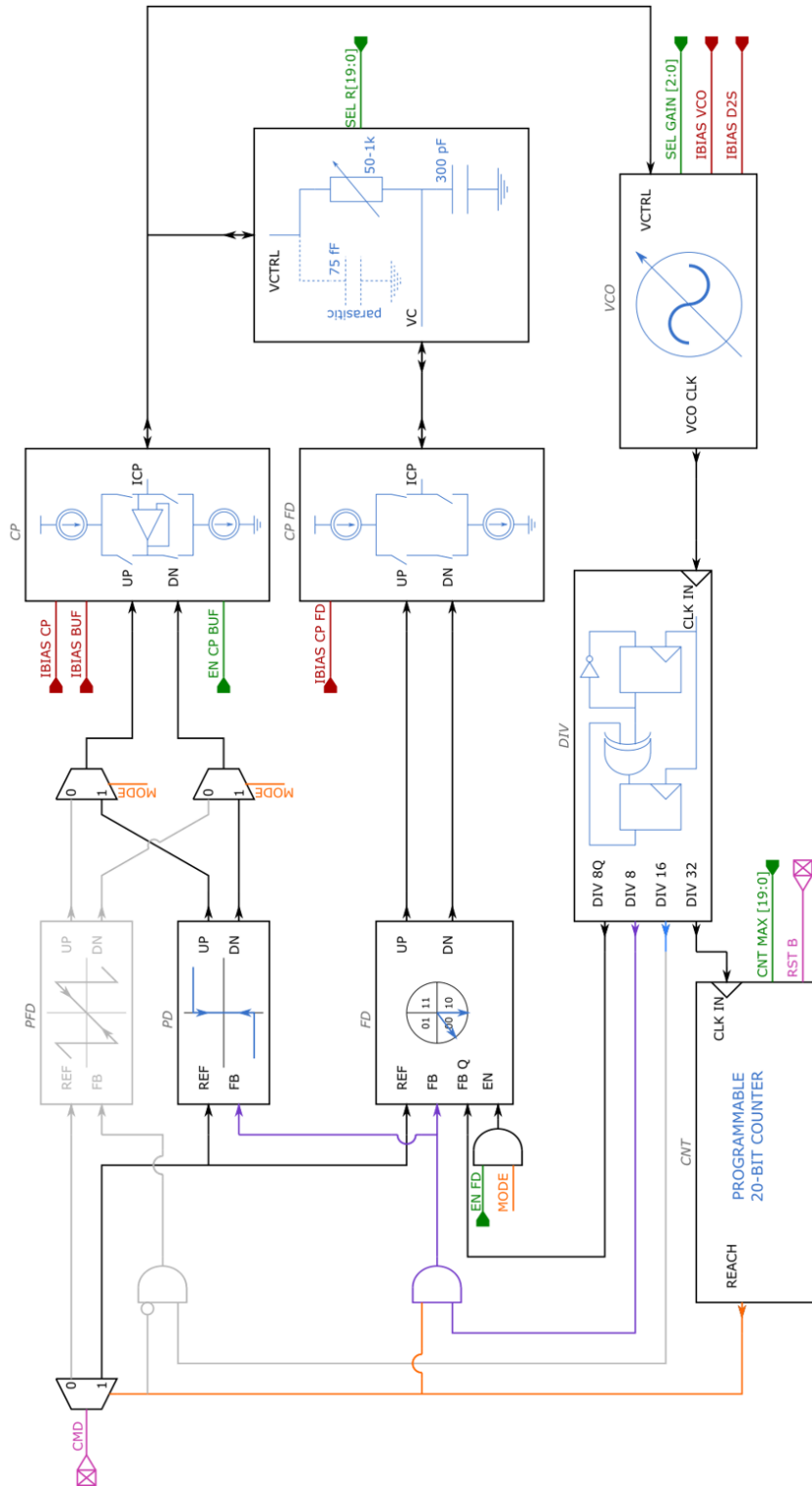
Figure 4.42: Normal operation mode, no specific pattern is expected for *CMD* input. They greyed out parts of the circuit are inactive.

4.5.2.7 CDR simulation results

In an effort to better predict the CDR performance the simulation methodology was significantly improved compared to CDR53A. Each block was simulated and optimized on its own, same as before. However, now also VerilogA and Verilog models of each block were created based on the simulation results. Those models were then used to create models of the whole, closed loop CDR. Using models has a significant impact on simulation time, for example simulating 10 μs of CDR operation takes about 2 hours when using a real schematic or 10 days when post-layout parasitics and transient (thermal) noise are include, while the same simulation using VerilogA models takes 1.5 minutes or only few seconds when Verilog models are used. The drawback is a loss of precision – while models can give a good prediction of performance, the speed-up is achieved by simplifying the behaviour of each block and neglecting some non-idealities.

Both Verilog and VerilogA models were created, because they had different use cases:

- Verilog models were used for long simulations e.g. full startup after power-up or several milliseconds of CDR run time. The simulations were done using purely digital circuit simulation tools, which is unusual for simulating an analogue circuits such as CDR, but allows for shortest simulation times [104].

- VerilogA models were used mostly to simulate the steady state of CDR operation and simulations up to one millisecond long. This type of simulations were used for parametric optimization and jitter estimations.

An example of a TIE histogram obtained from Verilog CDR model simulation is shown in Fig. 4.43. In this simulation the *CMD* input was 160 Mbps PRBS5 data stream with 5 ps rms jitter, which lead to *VCO CLK* TIE of 7.08 ps rms. Post-layout simulation of the whole circuit, including transient noise effects, for the same type of input resulted in *VCO CLK* TIE of 6.5 ps rms.
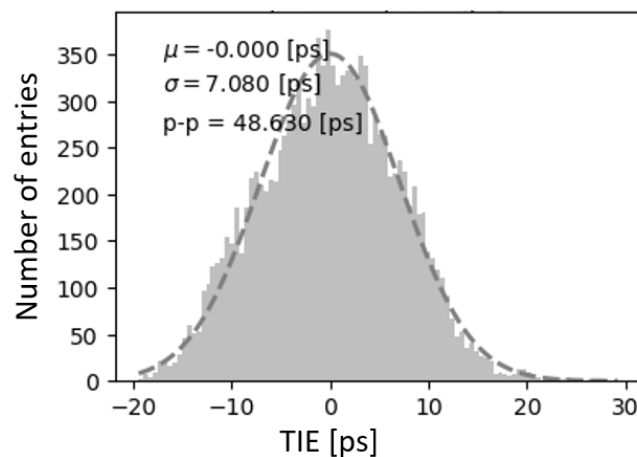

Figure 4.43: TIE histogram obtained from simulation of the Verilog model of the CDR. The *CMD* input was 160 Mbps PRBS5 data stream with 5 ps rms jitter. Dashed line show Gaussian fit with the fit parameters written on the plot.

4.5.2.8 CDR53B test chip

Since it would be too big of a risk to introduce significant circuit changes and implement the new design directly into the next big RD53 chip, a prototype ASIC was needed for the CDR53B. In addition to the CDR itself, the chip contains several other circuits from the RD53 design: bandgap reference, current DACs for biases, LVDS receivers, CML output drivers and data serializer. This way the CDR is in an environment very similar to the one in the RD53B chip, therefore the performance should be similar for both chips. Additionally, the CDR53A circuit was included as well, in order to compare its performance with the new design in the same conditions. The diagram of the prototype is presented in Fig. 4.44. All the blocks connections are designed in a way to allow characterization of several of them stand-alone (CDR, LVDS receiver, serializer and CML driver). The configuration for all the blocks is provided through the SPI bus, which controls a bank of triplicated registers. The layout of the ASIC is shown in Fig. 4.45. It can be clearly seen that the CDR53B design is much larger than the CDR53A, mainly due to the need for 300 pF LPF capacitor.
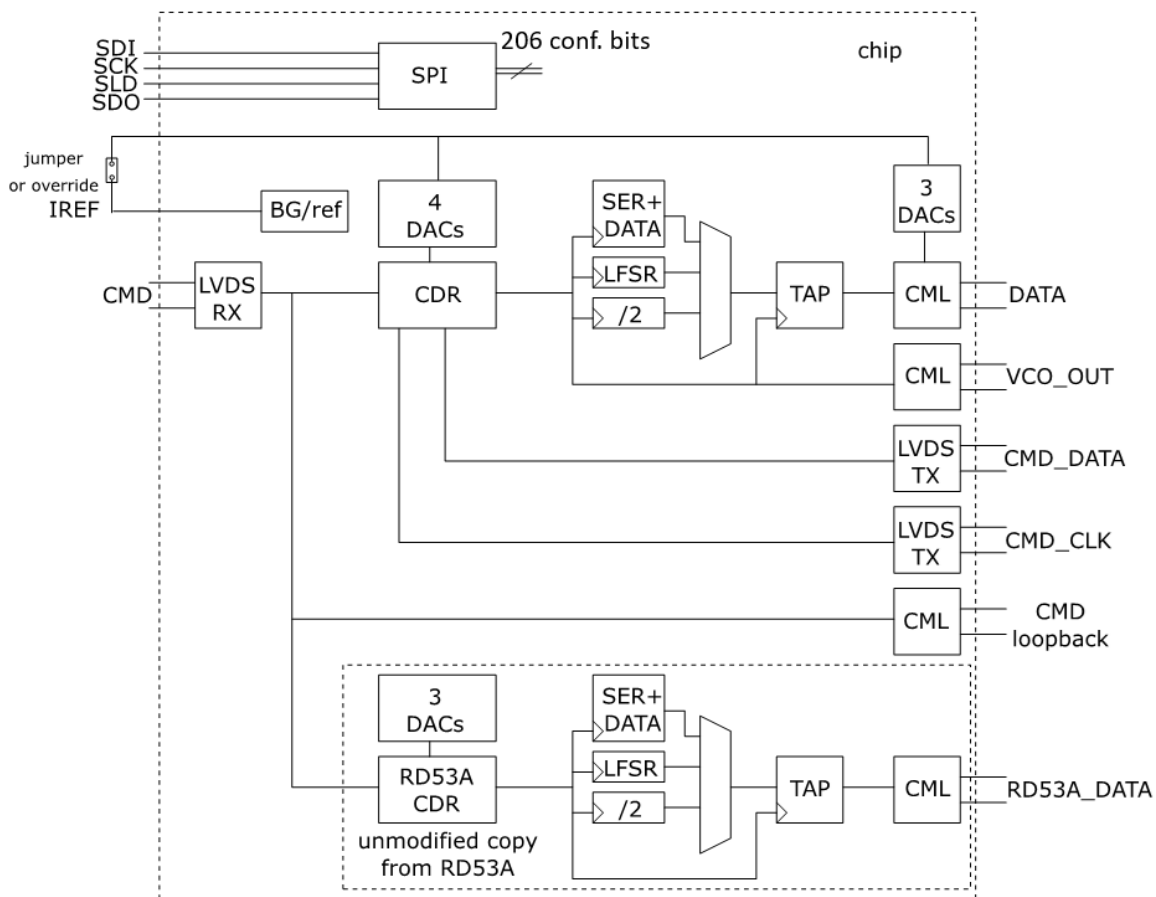

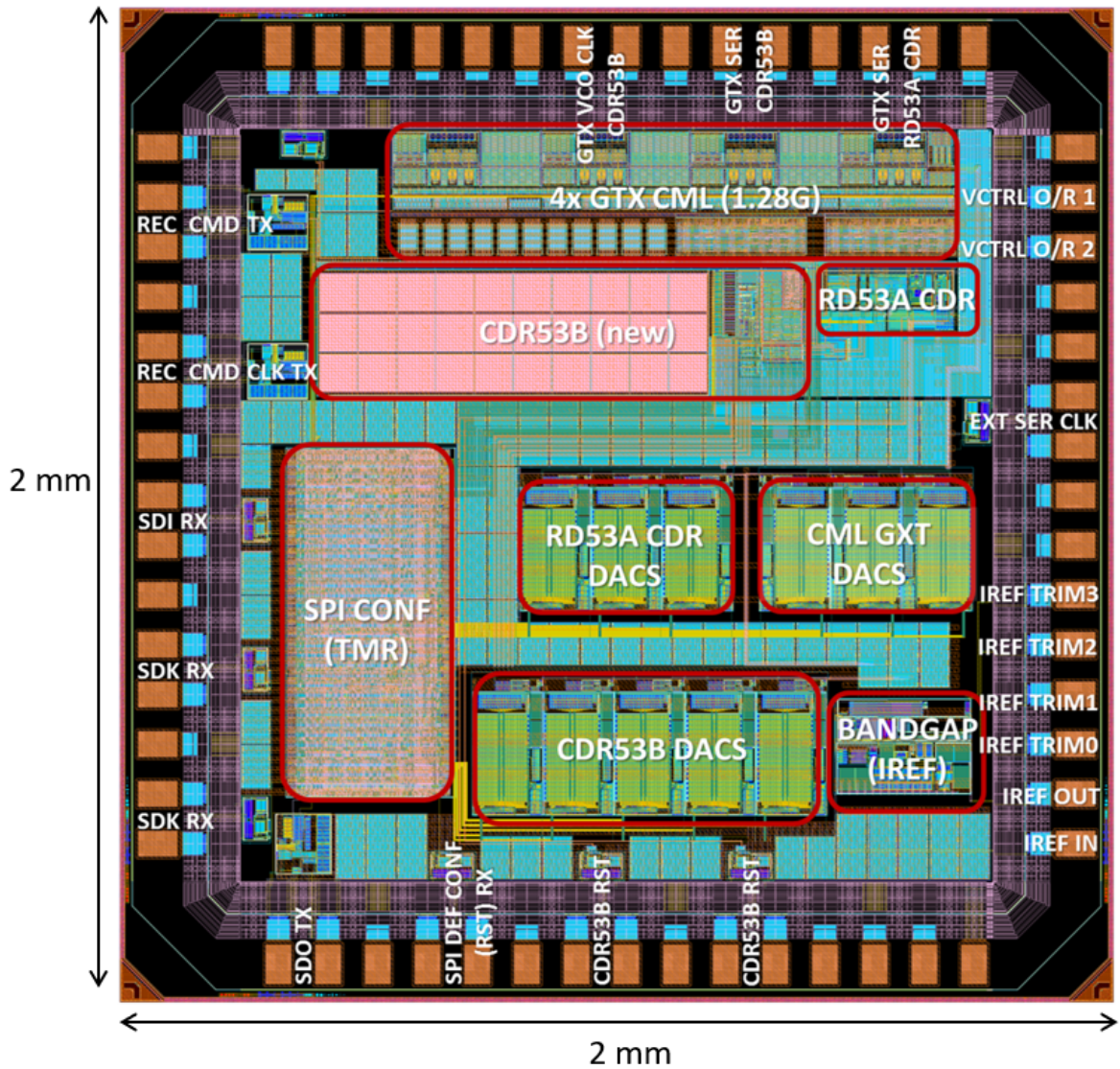**Figure 4.44:** Simplified schematic of the CDR53B test chip.

Figure 4.45: Layout of the CDR53B test chip with major blocks highlighted. The space between the highlighted blocks is filled with decoupling capacitors for power rails.

4.5.3 Measurement results

4.5.3.1 Experimental setup

The setup used for characterization of the CDR53B chip is presented in Fig. 4.46. The setup consists of two main components:

- BDAQ53 board with an FPGA module. This board provides the communication between the control PC and the measurement setup, and generates the signals needed for controlling the CDR53B ASIC. It is the same hardware that is used for characterization of the RD53 chips, as mentioned in Section 4.4.2.1.

- The carrier PCB for the prototype chip.

The setup was designed to be flexible enough to be used in all measurement scenarios – electrical testing, X-ray irradiation for verifying TID hardness and SEE testing facilities. The cable adapters visible in the photo and connections between the boards are changed, depending on the needs of measurement scenario.

The firmware for the FPGA was based on the BDAQ53 firmware, however it was significantly changed in order to suit the needs of the CDR53B ASIC. The software for controlling the setup and analysing the measurement results was written in Python. The *CMD* input to the CDR is provided either from the BDAQ53 board or from a high precision signal generator Agilent 81134A. The signal generator was used especially for jitter measurement, since it can provide very clean signals with down to 2 ps rms jitter. The jitter measurements are carried out mainly using Tektronix MSO 70804 high bandwidth oscilloscope and its DPOJET link quality measurement tool.
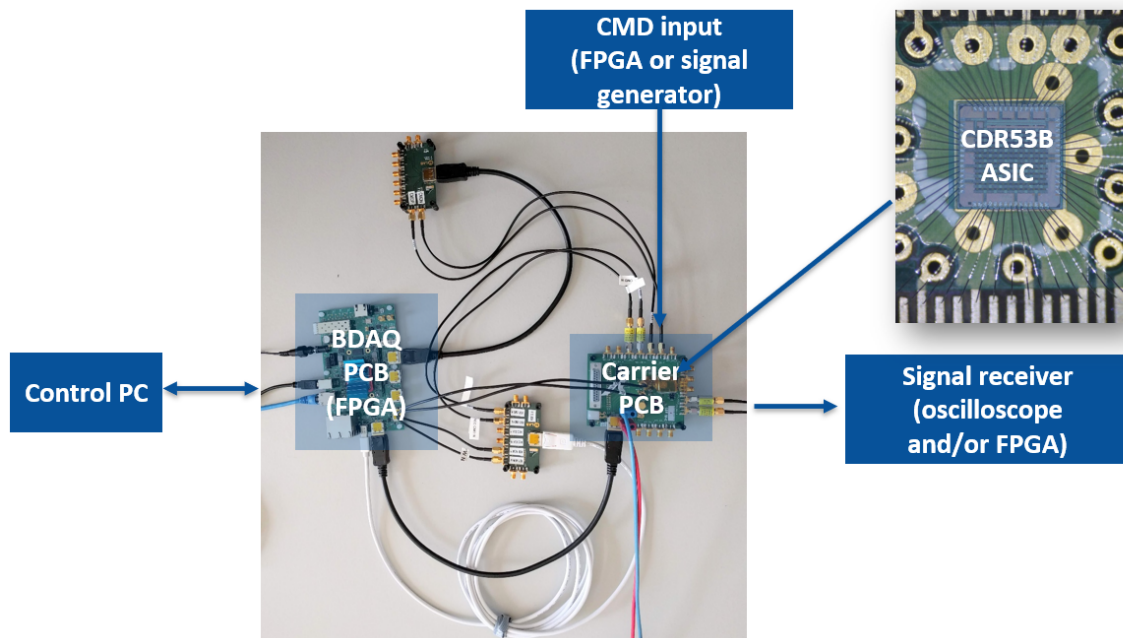

Figure 4.46: Measurement setup used for characterization of the CDR53B prototype chip.

4.5.3.2 Electrical performance

The chip is fully functional and responds to all changes in the configuration as expected. The power consumption of the new CDR itself is 6.5 mW, which matches the simulation prediction. The VCO characteristic curve measurement agrees well with the simulation results, as shown in Fig. 4.47. The slight offset of the measured curve compared to simulated one can be explained by difference in temperature (measurement was done at 21 °C, while simulation was done for 27 °C) or a small deviation of the manufacturing process from the simulated typical corner.
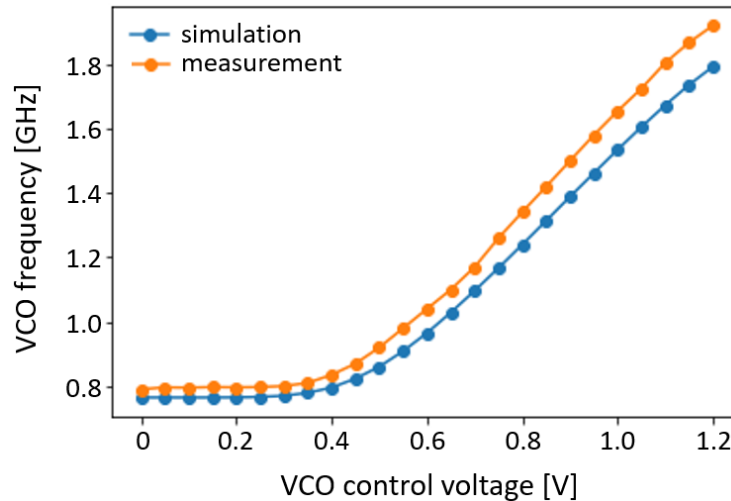

Figure 4.47: Comparison between measured and simulated characteristic curves of VCO.

Startup reliability   Startup reliability was one of the downfalls of the CDR53A, therefore it was very important to improve it in CDR53B. The startup reliability measurement procedure used for this chip is presented in Fig. 4.48. It is very similar to the one used for RD53A (Fig. 4.31), the only difference was providing the training pattern (010101..., equivalent to 80 MHz clock) right after power-up, and changing it to random commands only after some time.

The measurement was done for several chip samples both in room temperature and at −20 °C. In all cases the startup reliability was 100% for $VDD > 0.9V$. This is a significant improvement when compared to CDR53A. The startup problems at $VDD < 0.9V$ are understandable, since with this supply voltage the VCO cannot reach the needed 1.28 GHz output frequency. This is not a problem for the RD53 chips, since the supply voltage should never drop below 1.1 V when chips are running in the ATLAS or the CMS detectors.
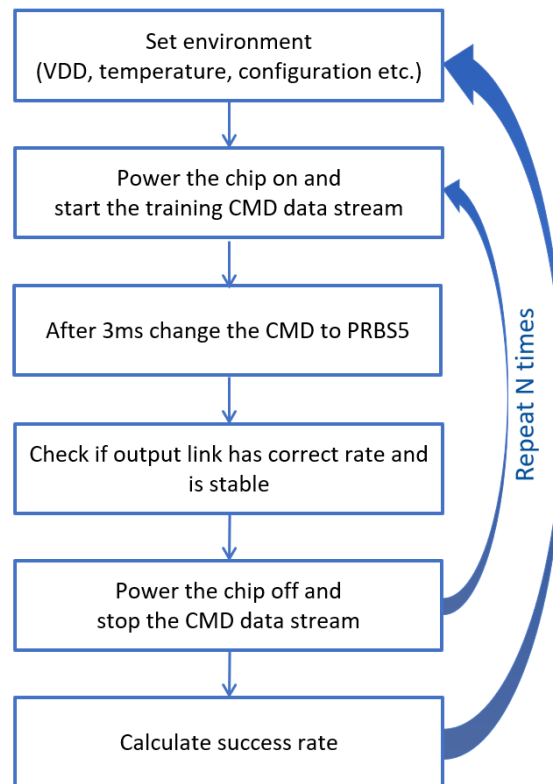
Figure 4.48: Flow diagram of CDR53B startup reliability measurement.

Jitter performance   First jitter measurements results did not match the expectation, as the measured jitter was nearly twice larger than the simulation predictions. The cause turned out to be a duty cycle distortion (DCD) of the *CMD* introduced by the LVDS receiver. It was confirmed both in measurement and simulation of this circuit. Since the Phase Detector used in the CDR53B uses timing information of both the rising and falling edges in the *CMD* stream, a DCD translates to a noticeable jitter increase. This phenomenon is shown Fig. 4.49. The DCD for this measurement was produced with the signal generator by changing the duration of high logic states in the data stream, while keeping the duration of low states ideal (6.25 ns). The measurement points indicate a clear minimum, which corresponds to the DCD from signal generator perfectly counteracting the DCD caused by the LVDS receiver.

Once this was understood, further measurements of jitter performance of CDR53B were done with compensation of the DCD. Two examples of quality of 1.28 Gbps PRBS15 data streams generated with CDR53B are shown in Fig. 4.50 and Fig. 4.51. Those examples differ by the jitter of the input PRBS5 *CMD*: 2 ps rms (representing the best case scenario with the available equipment) and 5 ps rms (machining random jitter expected from lpGBT), respectively. Both cases match very well with simulated performance [105] and are a big improvement over the CDR53A.
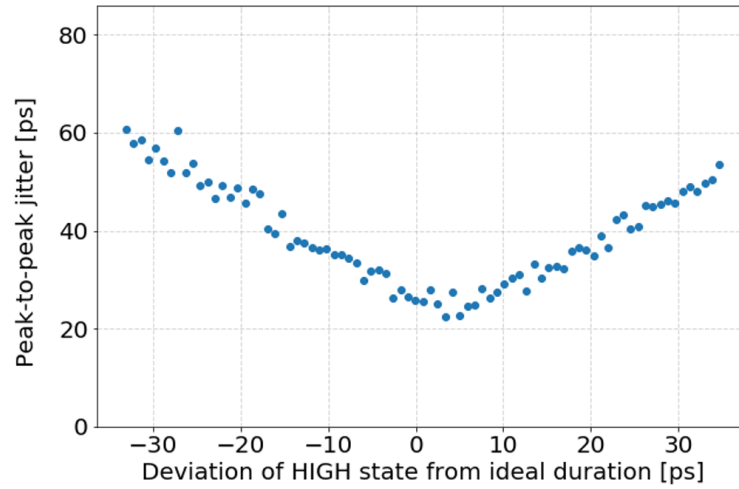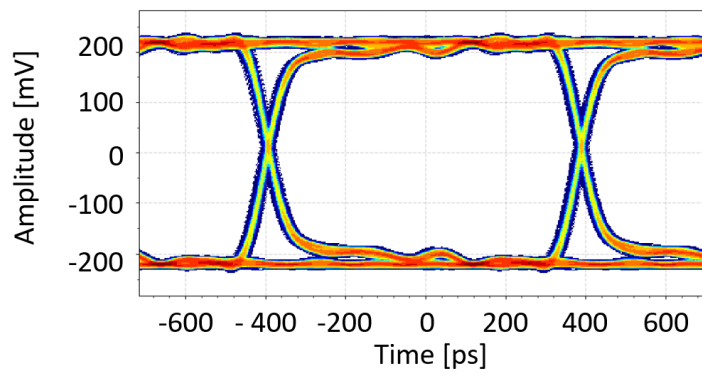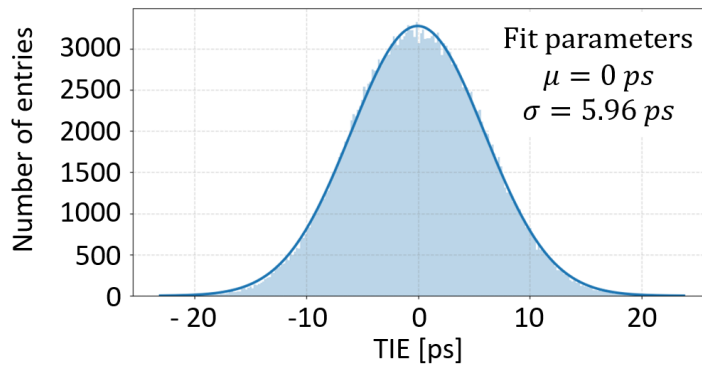
Figure 4.49: Measured influence of DCD in the *CMD* input on the jitter of 1.28 GHz *VCO CLK*.

(a) Eye diagram.

(b) TIE histogram (light blue) fitted with a Gaussian (dark blue line).

Figure 4.50: Jitter performance of 1.28 Gbps PRBS15 data stream generated by SER using CDR-generated clock. The *CMD* input is PRBS5 data steam with jitter of 2 ps rms.
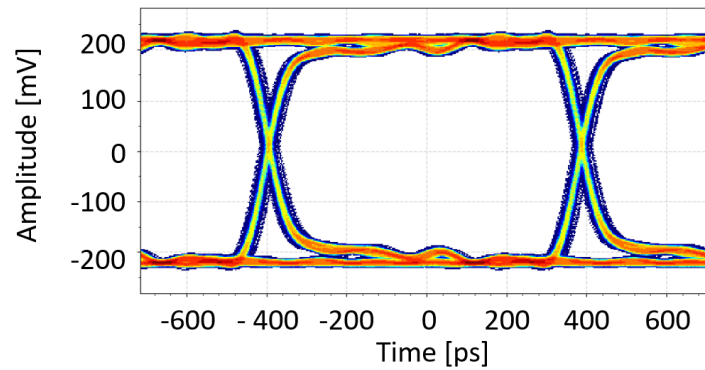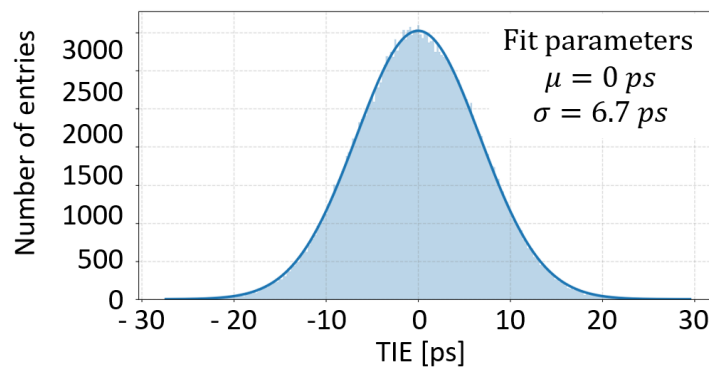
(a) Eye diagram.

(b) TIE histogram (light blue) fitted with a Gaussian (dark blue line).

Figure 4.51: Jitter performance of 1.28 Gbps PRBS15 data stream generated by SER using CDR-generated clock. The *CMD* input is PRBS5 data steam with jitter of 5 ps rms.

4.5.3.3 TID hardness

In order to evaluate the effects of ionizing radiation damage on the circuit, the CDR53B ASIC was irradiated using an X-ray source at the University of Bonn (tungsten target X-ray tube operated at 60 kV and 30 mA, a thin aluminium filter was used to condition the radiation spectrum [106]). During the irradiation the chip was kept at constant $-14\,°C$ in a dry N_2 atmosphere. The irradiation lasted for 7 days and the ASIC received 600 Mrad TID [7]. During the irradiation the chip was operated continuously using an automated setup performing several types of measurements in a loop, as shown in Fig. 4.52.

The chip remained fully functional after the irradiation. Fig. 4.53 presents a comparison of VCO's characteristic curves before and after X-ray irradiation. Since the VCO is the fastest switching circuit inside the CDR and it contains several small transistors (necessary to achieve good performance), it is potentially the most susceptible to TID damage part of the design. However, comparing the gain numbers before and after the irradiation, a degradation of only 6%–10% was measured, which does

---

[7] The official requirement for TID hardness of RD53 chip and its components is 500 Mrad. However, due to variation of the effective damage to the electronic circuit depending on many factors (temperature, radiation dose rate, as mentioned in Section 3.1.5.2), it is a common practice to apply a safety margin and irradiate to a higher TID.
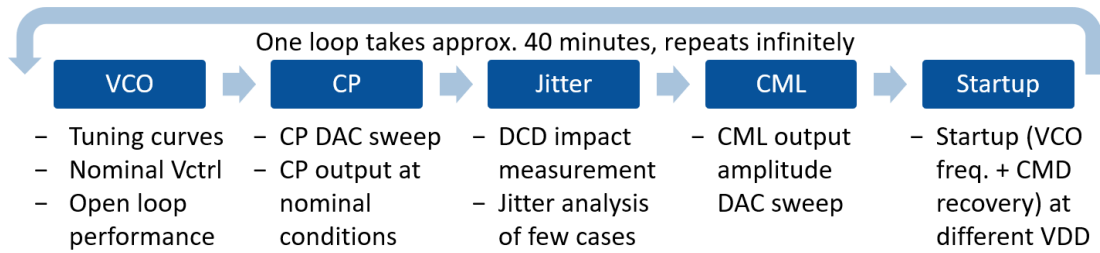
One loop takes approx. 40 minutes, repeats infinitely

VCO	CP	Jitter	CML	Startup
– Tuning curves – Nominal Vctrl – Open loop performance	– CP DAC sweep – CP output at nominal conditions	– DCD impact measurement – Jitter analysis of few cases	– CML output amplitude DAC sweep	– Startup (VCO freq. + CMD recovery) at different VDD

Figure 4.52: Diagram of the measurements done continuously during X-ray irradiation. Additionally, several parameters were monitored at each step e.g. power consumption, chip temperature, ambient temperature and humidity inside the cold box.

not have a significant impact on the overall jitter performance. One negative effect of the VCO's degradation is a higher power supply voltage VDD needed to achieve a 100% reliable startup: before the irradiation the chip would startup correctly with $VDD > 0.9V$, after 190 Mrad $VDD > 0.95V$ was needed and after 590 Mrad a $VDD > 1.0V$ was required. While this is a noticeable degradation, it is not a problem for operating in the ATLAS environment, as VDD is never expected to drop below 1.1 V.

In terms of jitter performance, a gradual degradation was observed throughout the whole irradiation campaign, ending with 13% higher peak-to-peak jitter on the 1.28 Gbps output as shown in Fig. 4.54. This is a very good result, since such jitter increase due to TID is negligible compared to the expected jitter introduced by low mass cables connecting RD53 and lpGBT in ATLAS.
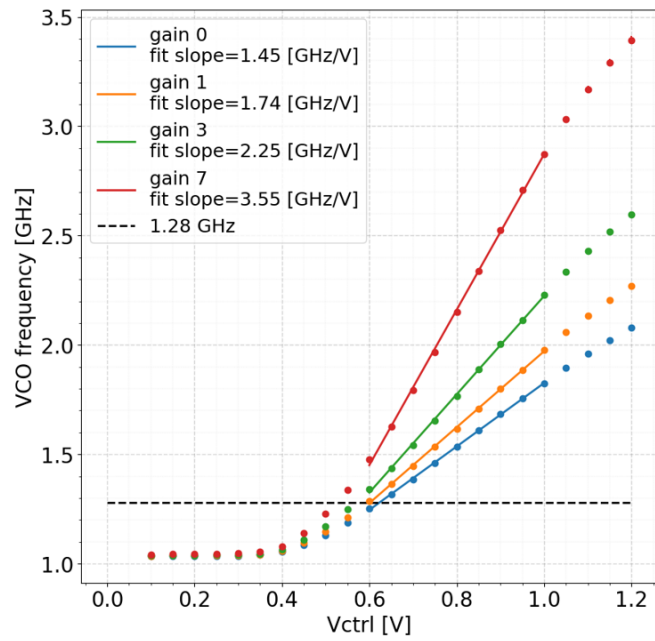
(a) Before irradiation
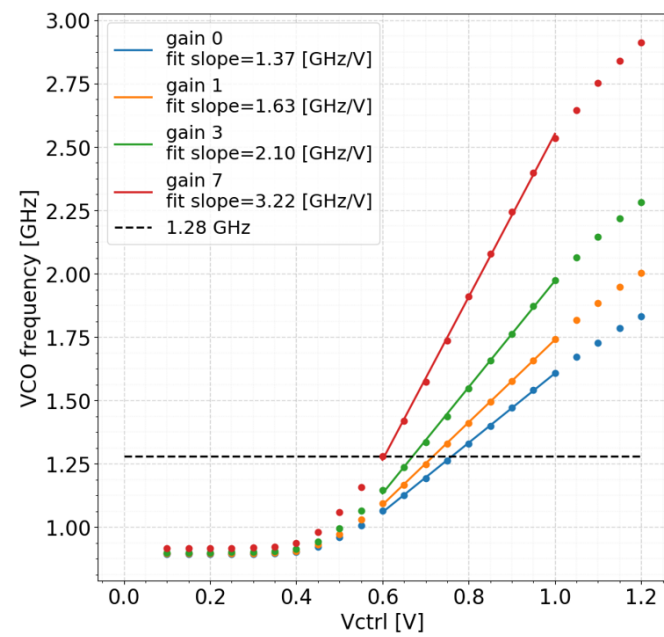
(b) After 600 Mrad TID

Figure 4.53: Comparison of VCO's characteristic (tuning) curves before and after X-ray irradiation to 600 Mrad TID.
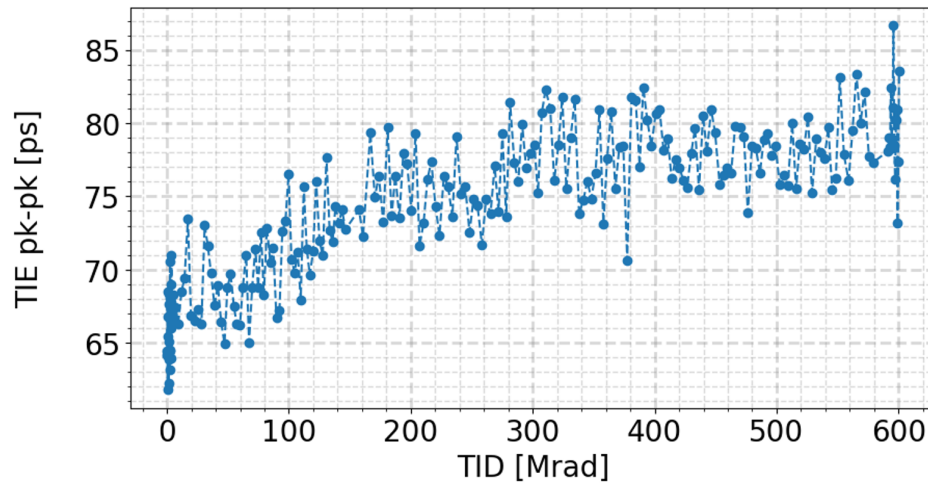
Figure 4.54: Changes of TIE peak-to-peak of 1.28 Gbps PRBS15 output due to accumulated TID. The *CMD* input for this measurement was PRBS5 data stream with 5 ps rms jitter.

4.5.3.4 SEE hardness

Resistance to Single Event Effects is crucial for a reliable operation of any electronic circuit in the radiation environment of ATLAS or CMS, as described in Section 3.1.5.3. During the design of CDR53B this issue was taken into account and several measures were taken in order to improve SEE immunity e.g. the clock divider was triplicated, all wells have a good contact to substrate to prevent latchup and MOSFET capacitors were avoided. Additionally, with the help of collogues from Seville University, the AFTU SEE simulation and verification tool [107] was used to examine and improve the main building blocks of the CDR53B. Despite all that, predicting SEE hardness of a design through simulation is not possible and measurements are required.

Heavy ions testbeam   In order to assess the susceptibility of a device to SEE it is a common practice to use ion beams. The device under test is exposed to several types of ions (one kind at a time) with different Linear Energy Transfers (LET). Based on the obtained error cross-section for different LET it is then possible to estimate the number of errors expected in the proton–proton collision environment inside LHC [49]. The CDR53B was tested in this way in the Heavy Ion Facility of Université catholique de Louvain [108]. Some of the parameters of the ions used for testing are described in Table 4.1. The thickness of the used chip sample was 300 μm, therefore all the ions were stopped completely and deposited their energy inside the device.

In order to capture upsets caused by the SEE, the measurement setup described in Section 4.5.3.1 was used with an addition of a Tektronix MSO54 oscilloscope. During the measurement the chip sample was placed inside a vacuum chamber, all the cables were passed through a special adapters on the flange of the chamber and connected to the rest of the rest of the setup. Throughout the measurement the oscilloscope was monitoring the VCO output clock (1.28 GHz) and a 1.28 GHz reference clock *REF CLK* generated by the FGPA. In the normal state (without SEE induced upsets) the two signals were aligned in the way shown in Fig. 4.55(a), where the rising edge of the *REF CLK* is in the middle of the high state of the *VCO CLK*. When an upset happens, the *VCO CLK* would

Ion	LET on device $\left[\frac{MeV}{mg/cm^2}\right]$	Range in Si $[\mu m]$
$^{13}C^{4+}$	1.3	269.3
$^{22}Ne^{7+}$	3.3	202.0
$^{27}Al^{8+}$	5.7	131.2
$^{36}Ar^{11+}$	9.9	114.0
$^{53}Cr^{16+}$	16.1	105.5
$^{58}Ni^{18+}$	20.4	100.5
$^{84}Kr^{25+}$	32.4	94.2
$^{103}Rh^{31+}$	46.1	87.3
$^{124}Xe^{35+}$	62.5	73.1

Table 4.1: Ions types available in Heavy Ion Facility used for SEE testing of CDR53B.

change phase and/or frequency, resulting in the change of position of its edges in reference to the *REF CLK*. If the upset is large enough to move any *VCO CLK* edge by more than 195 ps, a rising edge of the *REF CLK* would happen during a low state of the *VCO CLK*, which would trigger the oscilloscope (example shown in Fig. 4.55(b)). Since the FPGA was placed outside of the ion beam, the only source of triggers were upsets of the CDR53B. The oscilloscope was setup to trigger on every observed event and store the waveforms for later analysis. At the same time the FPGA was comparing the send *CMD* and the *REC CMD* received back from CDR53B in order to check if they are identical. This measurement method is not perfect since the small phase changes below 195 ps would not be captured. However, such small errors are not expected to cause problems for the RD53 chip operation in ATLAS or CMS, therefore the obtained results should give a good representation of the expected performance.

(a) Normal state

(b) Visualisation of SEE causing upset in *VCO CLK*, resulting in trigger condition of the oscilloscope being met
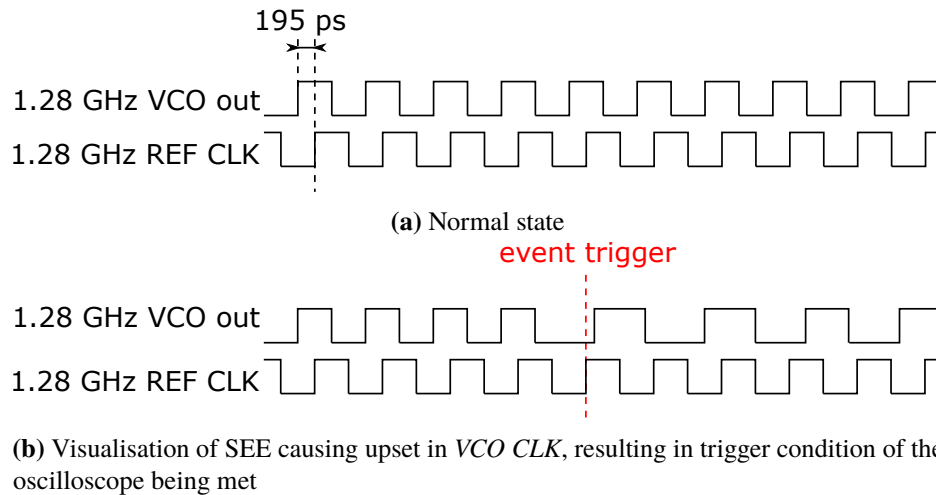
Figure 4.55: Visualisation of the oscilloscope triggering concept used during SEE testing of CDR53B.

All ion types mentioned in Table 4.1 were used during the test. Additionally, for a few ions the chip was tilted by 30°, allowing more charge to be deposited close to the surface and thus increasing the

effective LET. For each ion type the beam had a flux of $\approx 6.5 \cdot 10^3 \left[\frac{particle}{s \cdot cm^2} \right]$ and exposure time was 10 minutes. The beam had a diameter of 25 mm, allowing a full, homogenous coverage of the entire ASIC.

All captured events were analysed, both on individual basis and in a statistical way. An example of a captured event is shown in Fig. 4.56. The upset starts at time 0, causing both the frequency and phase of the VCO output to change. The SEE induced disturbance finishes within 1 μs, but the CDR loop then needs to correct itself and align to the input *CMD*. This recovery process last 7 μs in this case and during this time the *VCO CLK* frequency and phase go up and down, until they settle to the correct values. An example of a statistical analysis is shown in Fig. 4.57, where the histograms of the events durations are shown for all ions used. The behaviour is as expected - the heavier the ion, the more SEE upsets occur and their duration is longer. The average event duration is 5 μs.

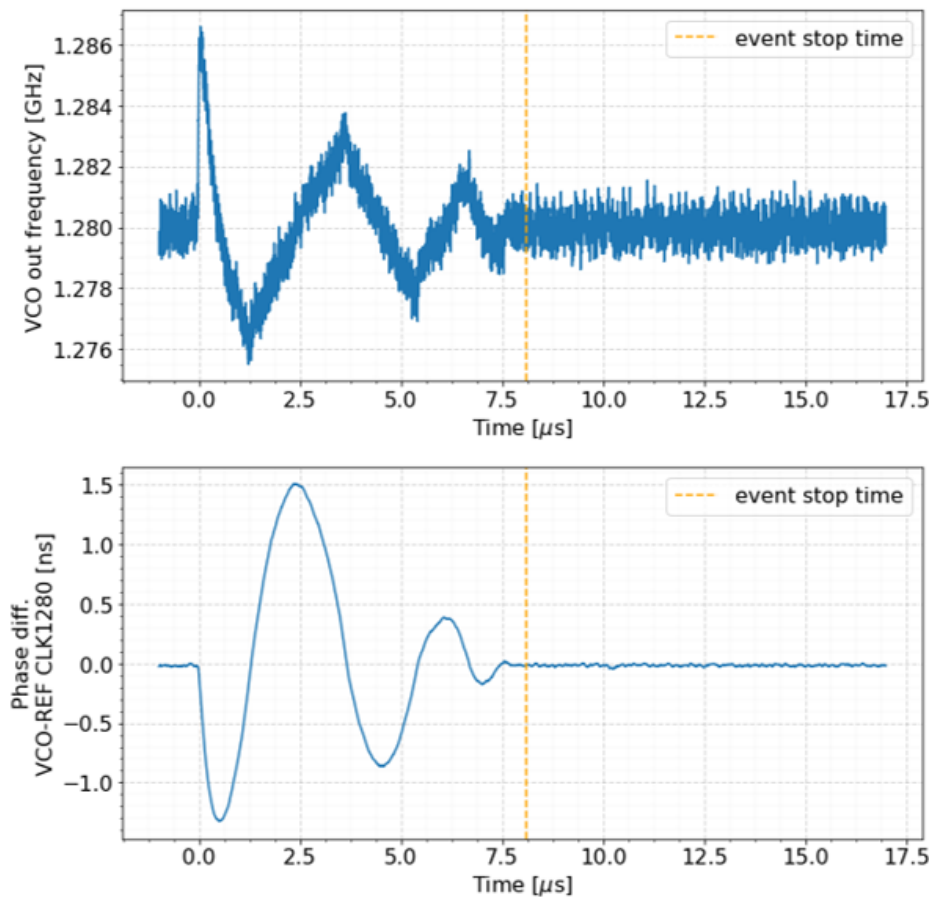**Figure 4.56:** An example of a SEE-induced upset cause by a $^{53}Cr^{16+}$ ion. The top plot shows the change of 1.28 GHz *VCO CLK* frequency in time. The bottom plot shows the change of the phase relation between *VCO CLK* and the FPGA-generated reference clock. The "event stop time" is defined as time after which the frequency returns to its nominal value for longer than 200 ns.
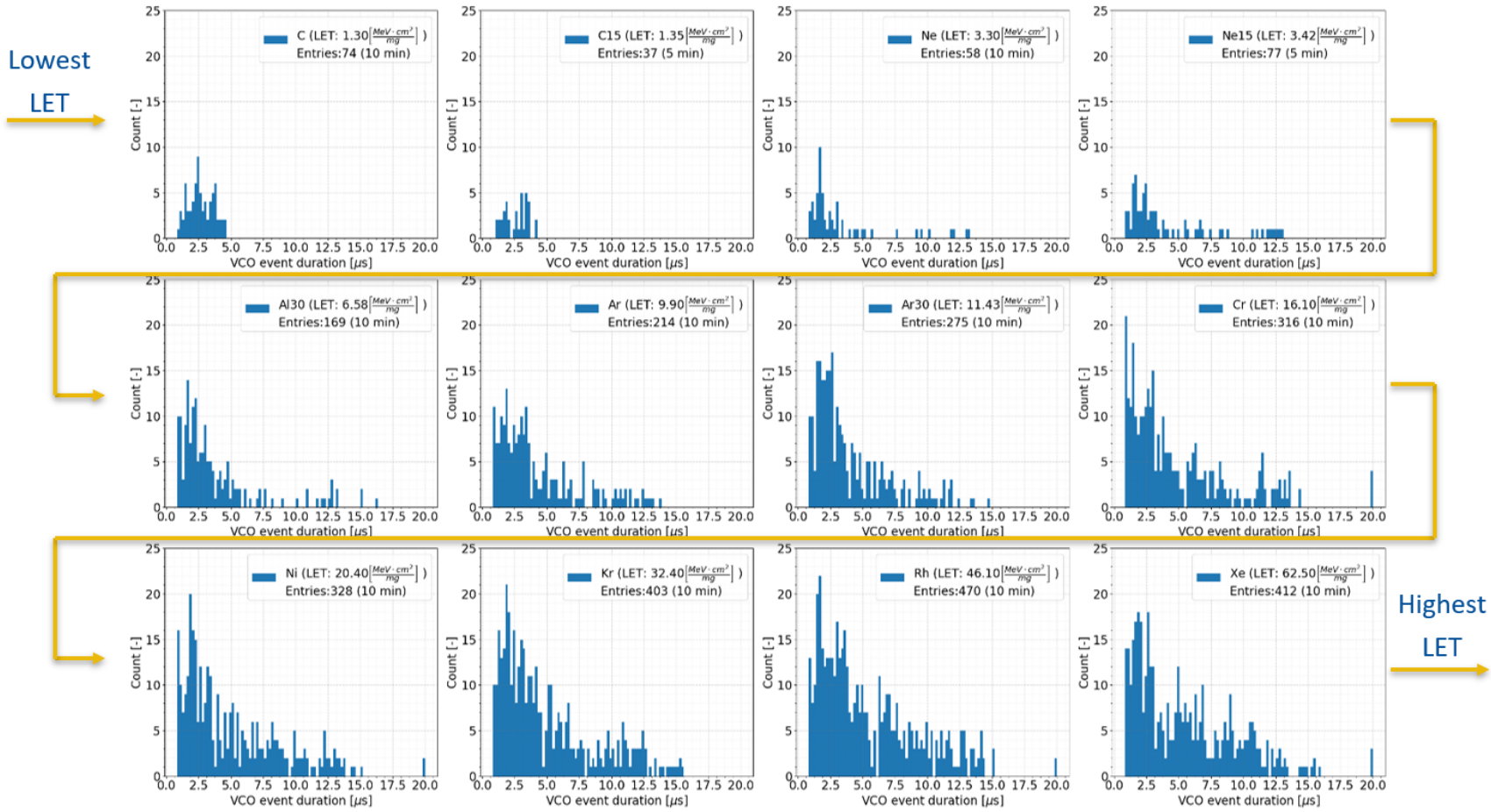
Figure 4.57: Histograms of the durations of SEE induced upsets for all ions used during the testbeam campaign.

The most important result from the SEE testing is shown in Fig. 4.58 – the measured SEE error cross-sections σ as a function of the ion's LET (Linear Energy Transfer) is fitted with a Weibull curve σ_{fit} (described by Eq. (3.23)). The four different data sets represent different categorizations of the observed upsets:

- Blue set includes all the events observed.

- Orange set excludes events caused by biasing circuitry, not the CDR directly.

- Green set includes only CDR-related events which lead to *VCO CLK* phase change larger than 390 ps. This threshold value was chosen, because such upsets are guaranteed to cause issues to the lpGBT (wrong sampling of data).

- Red set includes only the events which cause the *REC CMD* to differ from the *CMD* send.

For the operation of the detector system the green data set is most relevant. The red data set shows events which would lead to wrong interpretation of the *CMD* in RD53 chip. Since such error happens only when a large upset occurs inside the CDR, it is understandable that the cross-sections are lower.

The obtained Weibull curve fit parameters were given to Federico Faccio (CERN), who translated the results obtained with heavy ions into predicted upset rate in HL-LHC ATLAS layer 0 based on the expected radiation environment (described in Fig. 2.8(b)) using the method explained in [49]. The outcome of this analysis is that four errors per minute should be expected, two of them having phase jump larger than 390 ps. While those results should be treated as a rough estimate (the analysis method is based on many assumptions and simplifications), it gives an idea about the expected performance inside the detector. This seems reasonable, but at this time it is not possible to determine if this result is sufficient for reliable operation inside ATLAS. An upset in the CDR not only causes problem in the output data link, but also might disturb the digital logic. Additionally, system level test with RD53, lpGBT and DAQ are needed in order to determine the upset rates and recovery times of the entire data link.
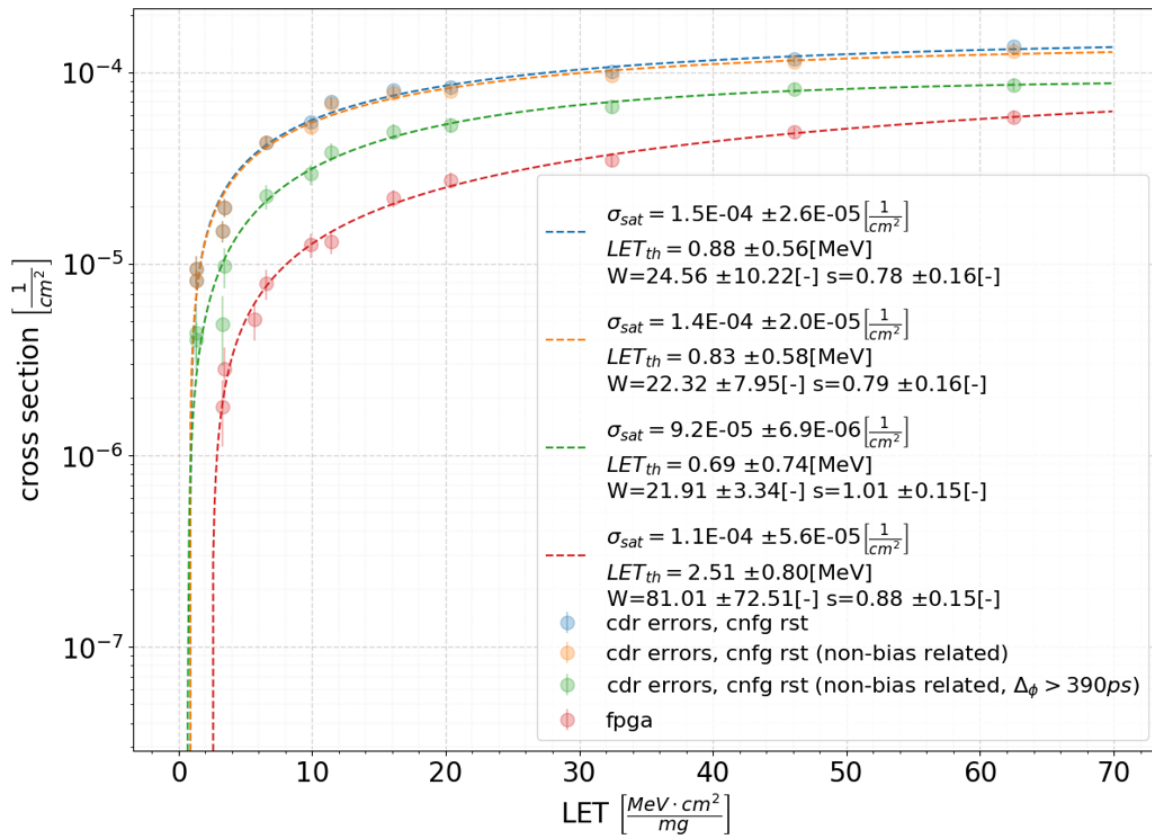
Figure 4.58: Cross-sections of the observed events as a function of ion's LET (circles) fitted with a Weibull curve (dashed lines, fit parameters in legend). The description of different data sets is provided in the text.

Two Photon Absorption (TPA) laser In order to improve the SEE resistance of the design the sensitive nodes had to be identified. For this purpose the PULSCAN system [109] in the RELY Lab of the Katholieke Universiteit Leuven [110] was used. It is a Two Photon Absorption (TPA) [111] laser setup combined with infrared camera and precise stepping motors. This machine allows injecting charge into a precisely defined region of a chip (1 μm laser spot size, 100 nm of stepping resolution), emulating the effects of charge deposition by an ionizing particle. The injection energy can be varied, the value typically used is between 1 and 2 nJ (1.4 nJ was used for CDR53B testing) with the frequency of injections being approximately 100 Hz. The chip is injected from the back, as shown in Fig. 4.59, since the metal layers on the top side of the ASIC would impede the laser. During the injection scan the ASIC DAQ has to be running and capable of reporting the number of upsets observed in the given step back to the PULSCAN system. Those upset numbers are used to create a sensitivity map and overlay it on top of an infrared photo of the chip taken in real time. The DAQ used for the CDR53B was the same as the one used for testing with heavy ions.

As an example the result of scanning the VCO is shown in Fig. 4.60(a), the layout corresponding to the photographed area is presented in Fig. 4.60(b) and the schematic of one stage of ring oscilator (part of the VCO) is shown in Fig. 4.60(c). The VCO is one of the most most sensitive parts of the CDR,
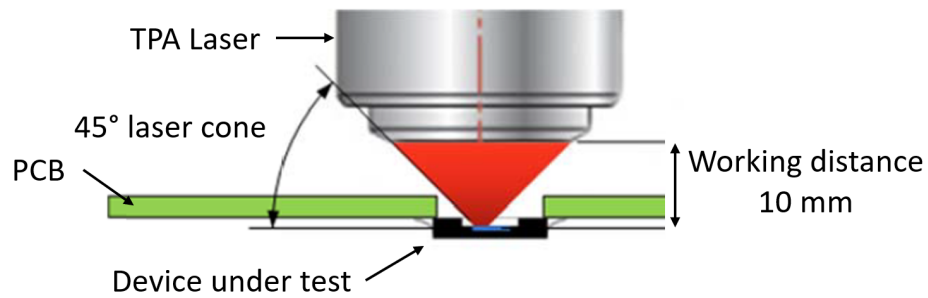
Figure 4.59: Drawing of an example TPA laser setup. It is important to note that the tested chip is injected from the back, therefore the PCB needs to have big enough hole to allow the laser to reach the regions of interest.

therefore several susceptible points were identified. The full list of found sensitive areas is presented in Table 4.2. An example of sensitive devices is highlighted by the red circles in Fig. 4.60, which demonstrates that the results obtained from PULSCAN system allow to identify sensitive transistors in the schematic of the design.

The sensitivity to laser pulses is not directly translatable to sensitivity to ions due to e.g. different charge distribution. However, the TPA laser injection method is very useful for improving the SEE hardness of a design.

Nr.	Block	Sensitivity
1	Differential-to-Single ended VCO output buffer	14
2	Configurable resistor in the LPF	9 (short events)
3	VCO Vctrl overwrite circuit	9 (short events)
4	VCO output digital buffer	9
5	VCO gain selection	8
6	Clock frequency divider	8 (long events)
7	Input pair of the VCO's ring oscillator stage	7
8	VCO load balancing dummy transistors	5
9	VCO's ring oscillator bias structure	4
10	VCO current mirror	4
11	Startup counter	1

Table 4.2: CDR design points sensitive to TPA laser SEE injections. The sensitivity value indicates maximal number of upsets observed during the measurement for a single step of the laser.
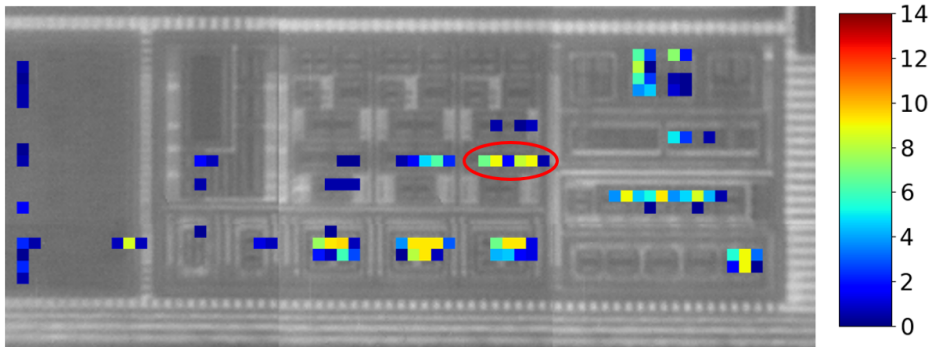
(a) IR photo with indicate sensitive areas. The colour coding represents number of observed errors.
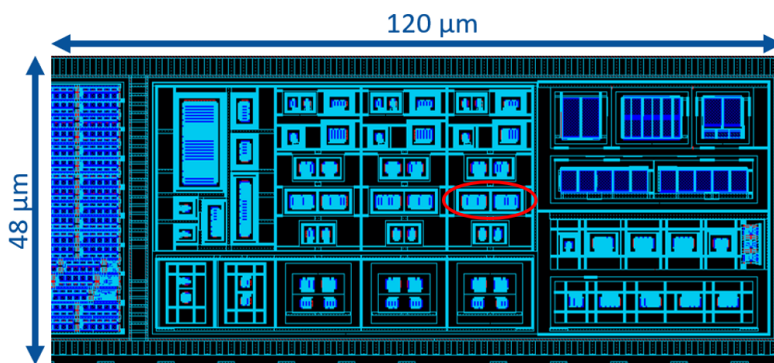
(b) Layout of the VCO showing polysilicon (red), diffusion (dark blue) and first metal (light blue) layers.
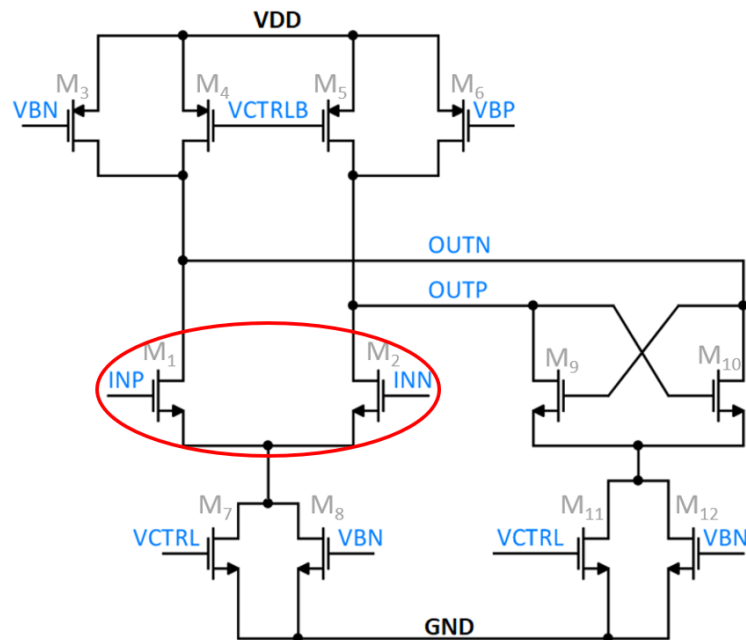
(c) Schematic of a single stage of VCO's ring oscillator.

Figure 4.60: Comparison of infrared photo taken with TPA laser injection system with the layout and schematic of the VCO. The red circle drawn in all figures indicate the same transistors (input pair of one stage of the ring oscillator).

151

4.6 RD53B CDR

This section presents only topics related to the CDR of RD53B, both in terms of design implementation and the measurement results. As mentioned in Section 4.1, the RD53B chip is made in two version: one for ATLAS (ITkPixV1) and one for CMS (CROCv1). The two designs have several differences (type of analogue front-end, chip size), but in many aspects they are the same, including the CDR design. Therefore, in this chapter the common name of RD53B will be used.

4.6.1 CDR implementation

Since the measurement results of the CDR53B prototype were very good, the CDR design from CDR53B prototype chip was integrated into RD53B chips with only minor modifications:

- The Phase Detector logic was slightly modified in order to allow possibility of using only the rising edge of the input *CMD* for phase aligning. This removes the sensitivity of the Phase Detector's to the duty cycle distortion in the *CMD* input described in Section 4.5.3.2.

- The clock divider triplication was improved based on the sensitive points found with the TPA laser. The new architecture is shown in Fig. 4.61. The two main changes are:

  - Spread the VCO generated clock into three separate branches immediately at VCO output and clock each D Flip-Flop in triplicated macro separately.

  - Addition of more majority voters, such that an upset in any single voter does not cause errors in the following stages

- Decoupling of several current mirrors was increased

- Based on the characterization of the CDR53B prototype the most suitable parameters of the circuit were determined and the option for their configuration was removed or restricted. This includes the LPF resistor (set to $400\,\Omega$), all bias currents (restricted to ranges which are considered safe) and the CNT_{max} value of the startup counter (set to 30000). Removal of those configuration options is expected to improve SEE hardness.

Overall, most of the sensitive nodes listed in Table 4.2 were improved. Only the input pair of the VCO's ring oscillator stage and the ring oscillator's bias structure were not modified, since such modification would involve a major redesign which was considered too risky.

Once the changes were implemented, the CDR was integrated into the RD53B, which meant adjusting the layout (adding power decoupling and building power grid compatible with RD53B), adding domain crossing buffers for all configuration bits and adding small auxiliary blocks needed for RD53B chip functionality (data deserializer for data merging functionality and precision phase shifters for adjusting phases of few internal signals [31]).

After the integration a dedicated Verilog model of the CDR macro was made, which was used by the designers of the digital circuitry of the RD53B for simulating the chip. Next, the CDR block was simulated in a mixed-signal environment with VCD stimuli generated from the whole chip simulation in order to confirm the correct connectivity between the CDR block and the rest of the chip. Finally, a voltage drop and electromigration (IR/EM) verification were performed on the final layout in order to identify potential issues with power distribution or inadequate metal wiring. The IR/EM simulation did
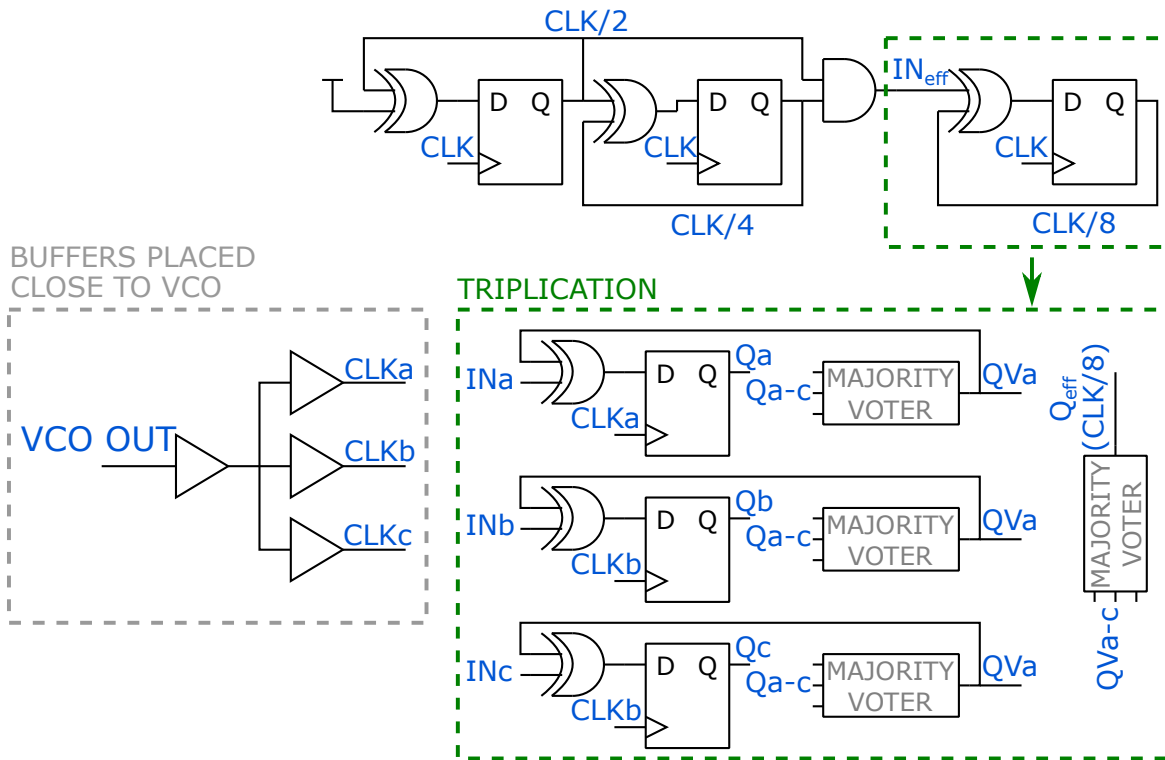
Figure 4.61: Schematic of clock divider by 8 implemented in RD53B with indication of used triplication approach. The triplication of only one D Flip-Flop and combinatorial logic at its input is shown for simplicity, in the actual circuit implementation all parts are triplicated.

not find any problems, the maximal observed voltage drop was 23 mV which is completely acceptable and the electromigration robustness analysis showed safety margin of 40% in the worst case.

4.6.2 Measurement results

The RD53B chip is characterized using the same two DAQ system used for RD53A, described in Section 4.4.2.1. The measurements described in this section were obtained with the BDAQ53 DAQ [95], however, they are not yet as extensive as the measurements done on CDR53B. Overall, while the RD53B was used by many people in many different testbenches, no results are published at the moment of writing this thesis.

The startup measurements were carried out following the same procedure as for CDR53B (Fig. 4.48). Several chip samples were measured in room temperature and at $-20\,°C$. The results are the same as for CDR53B i.e. the startup is 100% reliable for $VDD > 0.9V$. Several RD53B chips were irradiated to high TID levels (up to 1 Grad) in low temperatures and no startup issues were observed.

The 1.28 Gbps Aurora output link quality is shown in Fig. 4.62, where 11.2 ps rms jitter was measured. In this measurement the input *CMD* had jitter of 5 ps rms, so for comparison with CDR53B result Fig. 4.51 should be used. It is clear that the jitter measured for RD53B is larger. A small increase would be understandable, since a big chip with large amount of digital logic is likely to have larger noise on power supply rails than a small prototype, but not by this much. No precise answer was found yet, but some observations are:

- Due to a known bug in the ITkPixV1.0 the power consumption is much higher than expected. It is caused by a faulty latch design present in all pixels in the matrix, leading to a large static current. Workaround exists for reducing this power consumption, thought the chip cannot be fully operated in such state. However, the reduction of the static current also leads to a reduction on the output jitter by 1 ps rms, suggesting that large amount of noise is injected into the CDR.

- Stopping the clock distribution inside the pixel matrix also reduces the output jitter by approximately 1 ps rms, independently of the power consumption issue. This suggests a non-negligible amount of cross-talk from the digital circuitry to the CDR.

While more investigation is needed to understand to cause of increased jitter, the output eye quality is already a significant improvement over RD53A. Moreover, even the performance of RD53B currently is considered to be sufficient for the needs of the ATLAS and the CMS experiments.

The SEE hardness characterization is currently ongoing. First measurement campaigns suggest that the SEE susceptibility of the CDR itself is similar or lower to the CDR53B. However, since RD53B chip is a far more complicated design, there are more possible upset scenarios, it is more difficult to understand the SEE testing results and the DAQ plays a more significant role.
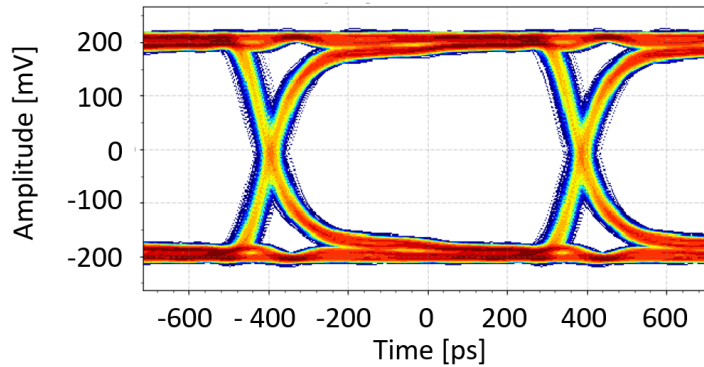
(a) Eye diagram.
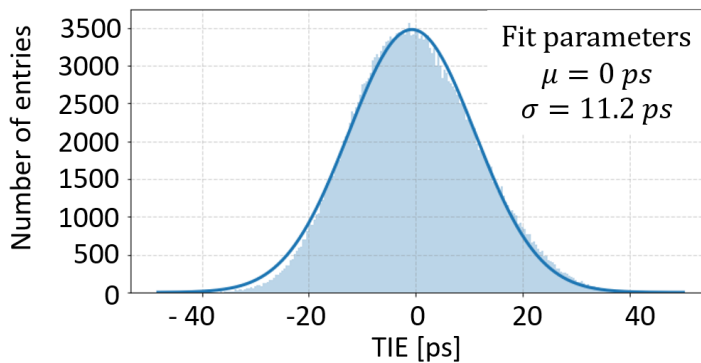
(b) TIE histogram (light blue) fitted with a Gaussian (dark blue line).

Figure 4.62: RD53B jitter performance of 1.28 Gbps Aurora data stream generated by SER using CDR-generated clock. The *CMD* input is PRBS5 data steam with jitter of 5 ps rms.

4.6.3 Summary and outlook of RD53 CDR project

This part of the thesis focuses on designing and testing a Clock Data Recovery circuit in 65 nm TSMC CMOS technology. This circuit is an important part of the I/O interface of the RD53 chips family. The CDR has to recover a 160 MHz clock from the input 160 Mbps *CMD* data stream, synchronize those two and pass them on to the chip's digital logic. At the same time a 1.28 GHz clock signal has to be synthesised and provided to the serializer block, so the data collected by the chip can be sent out.

All the CDR designs in this thesis use analogue, PLL-based architecture with ring oscillator based VCO. The first CDR design used a novel clock-gating Phase Detector design, which tried to combine the advantages of Phase-Frequency Detectors used in PLLs (infinite frequency pull-in range, good performance) with the ability to work with random data as an input. The clock divider was configurable, allowing the VCO to produce either 1.28 GHz or 1.6 GHz output in the locked state. In order to test this design in silicon an ASIC was made, called CDR53A, which combined the CDR, the data serializer and the CML cable driver. The measurement results only partially matched the simulation. The power consumption was as expected 5 mW, the chip was functional and responded to configuration changes as expected. However, the measured jitter of 19.8 ps rms (for input with 5 ps rms jitter) is nearly two times larger than the simulation predictions. The causes of this discrepancy turned out to be

inappropriate CDR loop parameters optimization and not thorough enough simulation methodology.

Despite its shortcomings, the CDR53A CDR circuit was integrated into the RD53A design (the first big chip produced by the RD53 collaboration). Only minor changes were made to the CDR, the most notable one being removal of clock divider configurability, as it was decided that only 1.28 Gbps output rate will be used in the RD53 chips. First measurement results showed a jitter performance of 13.5 ps rms (for input with 5 ps rms jitter) on the output link. This improvement over the prototype is attributed to using a better performing receiver for the input *CMD*, as well as improvements to the circuit biasing. However, further measurements revealed that the jitter degrades significantly when the activity inside the chip increases. With the entire pixel matrix enabled, the jitter increased to 25.9 ps rms. This phenomenon was caused by missing power domain crossing buffers on configuration bits controlling the VCO. Furthermore, issues with a correct startup of the chip were observed, which were traced back to a problem with CDR's Phase Detector architecture. While this issue had to be fixed for future designs, for the RD53A it was solvable with proper configuration of the carrier PCB. Despite the observed problems the performance of the CDR was good enough to allow extensive characterization of the RD53A chip - hundreds of chip were successfully tested in various scenarios including single chip testbenches, irradiations to 500 Mrad, hybrid assemblies operated in testbeams or multi-chip modules.

In order to fix the issues observed in CDR53A/RD53A design, it was decided to significantly modify the CDR. The architecture of several blocks (PD, LPF, DIV) was changed, others were re-optimized (CP, VCO) and a dedicated startup circuitry was added. The most impactful change was replacing the clock-gating Phase Detector with a commonly used bang-bang Phase Detector, as this forced changes to several other blocks e.g. a very big increase to the capacitor in the Low Pass Filter. Additionally, the simulation methodology was significantly improved based on the suggestions of the CDR design experts from CERN. The new CDR design was prototyped in an ASIC named CDR53B. The measurement results are significantly better compared to previous circuit – the CDR startup is 100% reliable for $VDD > 0.9V$, the jitter is 6.7 ps rms for input with 5 ps rms jitter and the 5.9 ps rms for input with 2 ps rms jitter, which matches the simulation results very well. The prototype was irradiated to 600 Mrad TID and was fully functional afterwards, with only a small jitter performance degradation of 13%. The chip was also tested for SEE hardness using heavy ion beams. The outcome is that approximately 4 upsets should be expected every minute in the HL-LHC ATLAS layer 0 environment. While this SEE hardness might already be sufficient for the experiment needs (not clear at the moment), in order to improve it the CDR53B was tested with TPA laser injection system, which allowed identifying several SEE sensitive nodes.

The CDR53B was integrated into the RD53B chips (pre-production candidates for ATLAS and CMS pixel detector) with only small changes (mostly based on the TPA laser injection measurements). After integration the correct connectivity and functionality was extensively verified in simulations. The measurements of the RD53B chips show results similar to those of CDR53B. The startup is 100% reliable for $VDD > 0.9V$, the jitter of the 1.28 Gbps output link is 11.2 ps rms for input with 5 ps rms jitter. The increase in the jitter is most likely caused by excessive noise on the power supply rails present due to a design mistake inside the pixel matrix, or by cross-talk from digital activity to the CDR. However, this performance is a significant improvement over the RD53A and is most likely already sufficient for the needs of ATLAS and CMS readout. Overall, several hundreds of RD53B chips were tested already in many different environments including irradiations up to 1 Grad TID. While the results in most cases are not published yet, no problems with the link quality or stability were reported, which is a very good result. The SEE hardness measurement are presently ongoing and

first result are encouraging and suggest performance similar or better than the CDR53B.

For the future RD53C chips, which will be the final production ASICs for the experiments, no changes are expected in the CDR design unless severe problems will be uncovered during system tests in SEE environment.

Conclusions

The High Luminosity upgrade of the Large Hadron Collider imposes challenging requirements for the future tracking detectors. In particular the devices closest to the interaction point, like the pixel detectors of ATLAS and CMS experiments, will experience very high radiation and will have to deal with high hit rates. At the same time the area covered by the pixel detectors is expected to be substantially larger than in the detectors used presently. In order to cope with those requirements development of new readout electronics and sensors are needed. All the work in this thesis was done in this context, specifically for the future ATLAS Inner Detector (ITk) and covers two different topics. First one is the development of Depleted Monolithic Active Pixel Sensors (DMAPS) in LFoundry 150 nm CMOS technology. The second one is the development and characterization of a Clock Data Recovery circuit of a pixel readout chip manufactured in TSMC 65 nm technology.

The aim of the DMAPS project was exploring the feasibility of combining the sensor and the readout electronics into one physical device capable of meeting the requirements of the ATLAS ITk Pixel Detector outer layers: 25 ns timing resolution, particle rate up to $100\,\text{MHz/cm}^2$, radiation hardness up to 50 Mrad TID and NIEL fluence of $10^{15}\,\text{n}_{eq}/\text{cm}^2$. In a DMAPS device the charge created by an impinging particle is collected through drift thanks to depletion of the silicon bulk. Afterwards, the charge is processed by the integrated readout electronics. Such a device can be produced using a commercial CMOS manufacturing processes by exploiting their technology add-ons, therefore low cost and fast production of large number of devices can be achieved. The work done in this thesis concentrated on high fill factor DMAPS using 150 nm LFoundry CMOS technology and lead to designing two prototype chips:

- CCPD_LF, the first chip in LF-DMAPS development line, focused on testing two different approaches to the sensor design and integrating analogue readout chain (CSA and comparator) into the pixel. The chip size was $0.5 \times 0.5\,\text{cm}^2$ with a matrix of 24 rows and 114 columns using $125 \times 33.3\,\text{μm}^2$ pixels. The in-pixel electronics behaved as expected from simulation. The chip was designed to be readout either through slow, binary stand-alone mode or to be bump bonded to FE-I4 readout chip for fast readout. Both of those methods were tested and worked well. The chip was irradiated to 50 Mrad TID and while the performance degradation was visible (40% gain loss and 75% ENC increase in worst case) the device could still be operated.

- LF-Monopix1 was the third chip in LF-DMAPS development line (the second one, LF-CPIX, was not a part of this thesis). The main focus of this chip was adding a fast stand-alone

readout into LF-CPIX-based design. The implemented readout architecture is a "column drain", which is used in the FE-I3 readout chip currently working in the ATLAS detector. It allows timestamping the hit events with 25 ns precision and according to simulation studies should be able to cope with hit rates up to 100 MHz/cm^2. The design was proven to work well at full speed in measurements and the in-pixel digital processing activity does not cause interference with the sensor. The chip size is 1×1 cm^2 and the matrix is build out of 129 rows and 36 columns, which are divided into 9 pixel flavours. The chip hit detection efficiency was measured in several test campaigns, resulting in 99.7% efficiency for un-irradiated devices and 98.9% for sample irradiated to 10^{15} n$_{eq}$/cm^2. To the authors' best knowledge, the LF-Monopix1 was the first device to successfully incorporate both complex, fast readout logic and radiation hard sensor.

A successor to the LF-Monopix1 has already been designed and manufactured. The LF-Monopix2 (not a part of this thesis) focuses on shrinking the pixel size to 150×50 μm^2, proving the full functionality with twice longer columns (chip size is 1×2 cm^2), providing a more homogenous matrix and improving the timing performance. Testing of the chip has started and preliminary results show that the device is functional, but more measurements are needed to fully verify all chip aspects. Overall, the LF-DMAPS project was successful in proving the feasibility of the concept of DMAPS capable of working in high hit rate and high radiation environment. While the ATLAS experiment ultimately decided that DMAPS devices will not be used for ATLAS ITk Pixel Detector construction, the idea of DMAPS chips raised a lot of interest within the HEP community and many developments in this field are still on-going.

The second part of this thesis focused on designing and testing a Clock Data Recovery circuit in 65 nm TSMC CMOS technology. This circuit is an important part of the I/O interface of the RD53 chips family. The CDR has to recover a 160 MHz clock from the input 160 Mbps *CMD* data stream, synchronize those two and pass them on to the chip's digital logic. At the same time a 1.28 GHz clock signal has to be synthesised and provided to the serializer block, so the data collected by the chip can be sent out. In total two very different CDRs were designed in this thesis (both using analogue, PLL-based architecture with ring oscillator based VCO) and each of them was implemented into a dedicated prototype chip and a big RD53 ASIC:

- The CDR53A was the prototype chip containing the first CDR design. This CDR used a novel clock-gating Phase Detector. The clock divider was configurable, allowing the VCO to produce either 1.28 GHz or 1.6 GHz output in the locked state (the 1.6 GHz option was abandoned in the later designs). The measurement results only partially matched the simulation. The power consumption was as expected 5 mW, the chip was functional. However, the measured jitter of 19.8 ps rms (for input with 5 ps rms jitter) is nearly two times larger than the simulation predictions. The causes of this discrepancy were identified as inappropriate CDR loop parameters optimization and issues in the simulation methodology.

- The RD53A large scale chip included the CDR53A CDR design without major changes. A jitter performance of 13.5 ps rms (for input with 5 ps rms jitter) on the output link was measured, however, when large amount of activity happens the jitter increases to 25.9 ps rms. This problem was caused by missing power domain crossing buffers on configuration bits controlling the VCO. Additionally, problems with a correct startup of the chip were observed, which turned out to be caused CDR's Phase Detector architecture. This issue was solved for RD53A by adjusting

the carrier PCB configuration. Despite the observed problems, the performance of the CDR was good enough to allow extensive characterization of hundreds of RD53A chips, including irradiations to 500 Mrad TID.

- The CDR53B was a second small prototype and it contained a new, re-designed CDR circuit. Based on the lesons learned from previous submissions the architecture of several blocks (PD, LPF, DIV) was changed, other circuits were re-optimized (CP, VCO) and a dedicated startup circuitry was added. Additionally, the simulation methodology was significantly improved. The measurement results show much improved performance – the CDR startup is 100% reliable for $VDD > 0.9V$, the jitter is 6.7 ps rms for input with 5 ps rms jitter, which matches the simulation results. The prototype was irradiated to 600 Mrad TID and was fully functional afterwards. The chip was also tested for SEE hardness using heavy ion beams and the outcome suggest that it might be robust enough for operation inside ATLAS (system level test are needed for better evaluation). The CDR53B was also tested with TPA laser injection system, which helped to further improve SEE hardness.

- The RD53B large scale chip included the CDR53B CDR. Only small changes were made to the design based. The measurements of the RD53B chips show results very similar to those of CDR53B. The biggest difference is higher output jitter (11.2 ps rms for input with 5 ps rms jitter), which can be caused by excessive noise on the power supply rails present due to a design mistake inside the pixel matrix. Several hundreds of RD53B chips were tested already in many different environments including irradiations up to 1 Grad TID and no CDR-related issues were observed. The SEE hardness measurement are presently ongoing and first result are encouraging.

For the future RD53C chips, which will be the final production ASICs for ATLAS and CMS, the CDR design will remain unchanged, unless the SEE testing reveals severe problems.

Bibliography

[1]   L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008).

[2]   *The ATLAS experiment homepage*, URL: `https://www.atlas.cern`.

[3]   *The CMS experiment homepage*, URL: `https://www.cms.cern`.

[4]   *The LHCb experiment homepage*, URL: `https://www.lhcb.web.cern.ch/lhcb/`.

[5]   *The ALICE experiment homepage*,
       URL: `https://www.alice-collaboration.web.cern.ch`.

[6]   *The TOTEM experiment homepage*, URL: `https://www.totem.web.cern.ch/Totem/`.

[7]   *The MoEDAL experiment homepage*, URL: `https://www.moedal.web.cern.ch/`.

[8]   *Technnical Proposal for the CERN LHCf Experiment : Measurement of Photons and Neutral
       Pions in the Very Forward Region of LHC*, CERN-LHCC-2005-032, CERN, 2005.

[9]   *The HL-LHC project homepage*, URL: `https://www.hilumilhc.web.cern.ch/`.

[10]   K. Einsweiler and L. Pontecorvo (editors), *Technical Design Report for
       the ATLAS Inner Tracker Pixel Detector*, CERN-LHCC-2017-021, CERN, 2017.

[11]   M. Tanabashi et al. (Particle Data Group), *Review of Particle Physics*,
       Physical Review D **98** (2018).

[12]   V. L. Highland, *Some practical remarks on multiple scattering*,
       Nuclear Instruments and Methods **129** (1975) p.497–499.

[13]   A. La Rosa, *ATLAS Pixel Detector: Operational Experience and Run-1 to Run-2 Transition*,
       Proceedings of Science (Vertex 2014) (2014).

[14]   A. La Rosa, *The ATLAS Insertable B-Layer: from construction to operation*,
       JINST **11** (2016) C12036.

[15]   D.-L. Pohl, *3D-Silicon and Passive CMOS Sensors for Pixel Detectors in High
       Radiation Environments*, BONN-IR-2019-01, PhD Thesis: University of Bonn, 2019,
       URL: `https://www.hep1.physik.uni-bonn.de/results/theses`.

[16]   T. Hemperek, *Exploration of advanced CMOS technologies for new pixel detector concepts in
       High Energy Physics*, BONN-IR-2018-04, PhD Thesis: University of Bonn, 2018,
       URL: `https://www.hep1.physik.uni-bonn.de/results/theses`.

[17]   B. Henke, E. Gullikson and J. Davis, *X-ray interactions: photoabsorption, scattering,
       transmission, and reflection at E=50-30000 eV, Z=1-92*,
       Atomic Data and Nuclear Data Tables **54** (1993) p. 181 –342.

[18]   *p-n junction*, URL: `https://en.wikipedia.org/wiki/P%E2%80%93n_junction`.

[19] L. Rossi, P. Fisher, T. Rohe and N. Wermes, *Pixel Detectors From Fundamentals to Applications*, Springer, 2006.

[20] H. Spieler, *Semiconductor Detector Systems*, Oxford University Press, 2005.

[21] W. Shockley, *Currents to Conductors Induced by a Moving Point Charge*, Journal of Applied Physics **9** (1938) p.636–636.

[22] S. Ramo, *Currents Induced by Electron Motion*, Proceedings of the IRE **27** (1939) p.584–585.

[23] T. Hirono, *Development of depleted monolithic active pixel sensors for high rate and high radiation experiments at HL-LHC*, BONN-IR-2019-03, PhD Thesis: University of Bonn, 2019, URL: https://www.hep1.physik.uni-bonn.de/results/theses.

[24] H. Kolanoski and N. Wermes, *Particle Detectors: Fundamentals and Application*, Oxford University Press, 2020.

[25] M. Daas, *Characterization and Irradiation Studies of Diode Test Structures in LFoundry CMOS Technology*, BONN-IR-2016-04, MSc Thesis: University of Bonn, 2016, URL: https://www.hep1.physik.uni-bonn.de/results/theses.

[26] E. R. Fossum, *CMOS active pixel image sensors*, Nucl. Instr. Meth. A **395** (1997) p.291–297.

[27] M. Garcia-Sciveres et al., *The FE-I4 Pixel Readout Integrated Circuit*, Nucl. Instr. Meth. A **636** (2011) p.S155–S159.

[28] RD53 Collaboration, M. Garcia-Sciveres et al., *The RD53A Integrated Circuit*, CERN-RD53-PUB-17-001, CERN, 2017.

[29] M. Karagounis, *Analog Integrated CMOS Circuits for the Readout and Powering of Highly Segmented Detectors in Particle Physics Applications*, PhD Thesis: University of Hagen, 2010.

[30] I. Perić et al., *The FEI3 readout chip for the ATLAS pixel detector*, Nucl. Instr. Meth. A **565** (2006) p.178–187.

[31] RD53 Collaboration, M. Garcia-Sciveres, F. Loddo, J. Christiansen et al., *The RD53B Manual*, CERN-RD53-PUB-19-002, CERN, 2019.

[32] Y. N. N. Doerin, *Handbook of Semiconductor Manufacturing Technology*, Taylor & Francis Ltd., 2017.

[33] *LFoundry*, URL: http://www.europractice-ic.com/technologies_LFoundry.php.

[34] *Copper Interconnects The Evolution of Microprocessors*, URL: https://www.ibm.com/ibm/history/ibm100/us/en/icons/copperchip/.

[35] *RD53: Development of pixel readout integrated circuits for extreme rate and radiation*, URL: https://ep-news.web.cern.ch/content/rd53-development-pixel-readout-integrated-circuits-extreme-rate-and-radiation.

[36] M. Havránek, *Development of pixel front-end electronics using advanced deep submicron CMOS technologies*, BONN-IR-2014-11, PhD Thesis: University of Bonn, 2014, URL: https://www.hep1.physik.uni-bonn.de/results/theses.

[37] M. Moll, *Radiation damage in silicon particle detectors: Microscopic defects and macroscopic properties*, PhD Thesis: University of Hamburg, 1999,
URL: `http://cds.cern.ch/record/425274`.

[38] M. Moll, *Displacement Damage in Silicon Detectors for High Energy Physics*,
IEEE TNS **65** (2018) p.1561–1582.

[39] M. Karagounis et al., *Development of the ATLAS FE-I4 pixel readout IC for b-layer Upgrade and Super-LHC*,
Proceedings of the Topical Workshop on Electronics for Particle Physics (2008).

[40] S. Löchner, *Radiation Tolerance of Electronics*, URL:
`https://indico.cern.ch/event/176795/contributions/291275/attachments/230088/321970/Radiation_Tolerance_of_Electronics-Sven_Loechner.pdf`.

[41] W. Snoeys et al., *Layout techniques to enhance the radiation tolerance of standard CMOS technologies demonstrated on a pixel detector readout chip*,
Nucl. Instr. Meth. A **429** (2000) p.349–360.

[42] A. Nikolaou et al., *Modeling of High Total Ionizing Dose (TID) Effects for Enclosed Layout Transistors in 65 nm Bulk CMOS*, International Semiconductor Conference (2018).

[43] M. Menouni et al., *1-Grad total dose evaluation of 65nm CMOS technology for the HL-LHC upgrades*, JINST **10** (2015).

[44] G. Borghello et al., *Dose-Rate Sensitivity of 65-nm MOSFETs Exposed to Ultrahigh Doses*,
IEEE TNS **65** (2018) p.1482–1487.

[45] J. L. Titus and C. F. Wheatley, *Experimental Studies of Single-Event Gate Rupture and Burnout in Vertical Power MOSFETs*, IEEE TNS **43** (1996) p.533–545.

[46] *Latch-Up*, Texas Instruments White Paper **SCAA124** (2015).

[47] *SEU Mitigation Techniques for SRAM based FPGAs*,
URL: `https://indico.cern.ch/event/357738/contributions/848841/attachments/1161500/1675463/Chapman_ID240.pdf`.

[48] *Introduction to Single-Event Upsets*, ALTERA White Paper **WP-01206-1.0** (2013).

[49] F. F. M. Huhtinen, *Computational method to estimate Single Event Upset rates in an accelerator environment*, Nucl. Instr. Meth. A **450** (2000) p.155–172.

[50] A.Ranieri et al., *Latest results of SEE measurements obtained by the STRURED demonstrator ASIC*, Nucl. Instr. Meth. A **626** (2011) p.82–89.

[51] C. J. Kenney et al., *A prototype monolithic pixel detector*,
Nucl. Instr. Meth. A **342** (1994) p.59–77.

[52] R. Turchetta et al., *A monolithic active pixel sensor for charged particle tracking and imaging using standard VLSI CMOS technology*, Nucl. Instr. Meth. A **458** (2001) p.677–689.

[53] J. Schambach et al., *A MAPS Based Micro-Vertex Detector for the STAR Experiment*,
Physics Procedia **66** (2015) p.514–519.

[54] M. Garcia-Sciveres and N. Wermes, *A review of advances in pixel detectors for experiments with high rate and radiation*, Rep. Prog. Phys. **81** (2018) 066101.

[55] K. Moustakas et al., *CMOS Monolithic Pixel Sensors based on the Column-Drain Architecture for the HL-LHC Upgrade*, Nucl. Instr. Meth. A **936** (2019) p.604–607.

[56] W. Snoeys et al., *A process modification for CMOS monolithic active pixel sensors for enhanced depletion, timing performance and radiation tolerance*, Nucl. Instr. Meth. A **871** (2017) p.90–96.

[57] H.Pernegger et al., *Radiation hard monolithic CMOS sensors with small electrodes for High Luminosity LHC*, Nucl. Instr. Meth. A **986** (2021) 164381.

[58] I. Peric, *Active pixel sensors in high-voltage CMOS technologies for ATLAS*, JINST **7** (2012) C08002.

[59] P. Rymaszewski et al., *Prototype Active Silicon Sensor in 150 nm HR-CMOS technology for ATLAS Inner Detector Upgrade*, JINST **11** (2016) C02045.

[60] T. Hirono et al., *Characterization of Fully Depleted CMOS Active Pixel Sensors on High Resistivity Substrates for Use in a High Radiation Environment*, IEEE NSS-MIC (2017) p.1–4.

[61] M.Guthoff et al., *Geant4 simulation of a filtered X-ray source for radiation damage studies*, Nucl. Instr. Meth. A **675** (2012) p.118–122.

[62] L.Gonella et al., *Total Ionizing Dose effects in 130-nm commercial CMOS technologies for HEP experiments*, Nucl. Instr. Meth. A **582** (2007) p.750–754.

[63] C. Bespin et al., *DMAPS Monopix developments in large and small electrode designs*, Nucl. Instr. Meth. A **978** (2020) 164460.

[64] Y. Degerli et al., *Pixel architectures in a HV-CMOS process for the ATLAS inner detector upgrade*, JINST **11** (2016) C12064.

[65] L. Vigani et al., *Study of prototypes of LFoundry active CMOS pixels sensors for the ATLAS detector*, JINST **13** (2018) C02021.

[66] Z. Chen et al., *Test results of irradiated CMOS pixel circuits in 150 nm CMOS technology for the ATLAS Inner Tracker Upgrade*, Proceedings of Science (TWEPP2018) **343** (2019).

[67] M. Barbero et al., *Radiation hard DMAPS pixel sensors in 150nm CMOS technology for operation at LHC*, JINST **15** (2020) P05013.

[68] P. Rymaszewski et al., *Development of Depleted Monolithic Pixel Sensors in 150 nm CMOS technology for the ATLAS Inner Tracker Upgrade*, Proceedings of Science (TWEPP2018) **313** (2018).

[69] T. Wang et al., *Development of a Depleted Monolithic CMOS Sensor in a 150 nm CMOS Technology for the ATLAS Inner Tracker Upgrade*, JINST **12** (2017) C01039.

[70] M. Bazes, *Two novel fully complementary self-biased CMOS differential amplifiers*, IEEE Journal of Solid-State Circuits **26** (1991) p. 165–168.

[71] A. Taparia, *CS-CMOS: A Low-Noise Logic Family for Mixed Signal SoCs*, IEEE Transactions on VLSI Systems **19** (2011) p. 2141 –2148.

[72] I. Perić et al., *High-voltage pixel detectors in commercial CMOS technologies for ATLAS, CLIC and Mu3e experiments*, Nucl. Instrum. and Meth. A **731** (2013) p. 131 –136.

[73] I. Caicedo et al., *The Monopix chips: depleted monolithic active pixel sensors with a column-drain read-out architecture for the ATLAS Inner Tracker upgrade*, JINST **14** (2019) C06006.

[74] J. Christiansen and M. Garcia-Sciveres, *RD Collaboration Proposal: Development of pixel readout integrated circuits for extreme rate and radiation*, CERN-LHCC-2013-008, CERN, 2013.

[75] J. Christiansen and F. Loddo, *Extension of RD53*, CERN-LHCC-2018-028, CERN, 2018.

[76] F. Vasey, *Versatile Link PLUS project*, URL: `https://espace.cern.ch/project-Versatile-Link-Plus/SitePages/Home.aspx`.

[77] P. Moreira, *The lpGBT project*, URL: `https://espace.cern.ch/GBT-Project/LpGBT/default.aspx`.

[78] *Xilinx Aurora 64B/66B Protocol Specification SP011*, URL: `https://www.xilinx.com/support/documentation/ip_documentation/aurora_64b66b_protocol_spec_sp011.pdf`.

[79] T. Wang et al., *A high speed transmitter circuit for the ATLAS/CMS HL-LHC pixel readout chip*, Proceedings of Science **343** (2019) p.098.

[80] Tektronix, *Understanding and Characterizing Timing Jitter*, URL: `https://www.tek.com/primer/understanding-and-characterizing-timing-jitter-primer`.

[81] *Definitions and Terminology for Synchronization Networks*, G.810, International Telecommunication Union, 1996.

[82] K. Kundert, *Predicting the Phase Noise and Jitter of PLL-Based Frequency Synthesizers*, 4th ed., The Designers's Guide Community, 2012.

[83] M. Hsieh and G. E. Sobelman, *Architectures for multi-gigabit wire-linked clock and data recovery*, IEEE Circuits and Systems Magazine **8** (2008) p. 45–57.

[84] B. Razavi, *Design of Monolithic Phase-Locked Loops and Clock Recovery Circuits - A Tutorial*, Wiley-IEEE Press, 1996.

[85] D. Banerjee), *PLL Performance, Simulation, and Design*, URL: `https://www.ti.com.cn/cn/lit/ml/snaa106c/snaa106c.pdf`.

[86] T. Kishishita et al., *Prototype of a gigabit data transmitter in 65nm CMOS for DEPFET pixel detectors at Belle-II*, Nucl. Instrum. and Meth. A **718** (2013) p. 168–172.

[87] T. V. Cao et al., *Low Phase-Noise and Wide Tuning-Range CMOS Differential VCO for Frequency ΔΣ Modulator*, IEEE Computer Society Annual Symposium on VLSI (2009).

[88] T. Kishishita and H. Krüger, *Prototype of simultaneous bidirectional data-transmitter in 65 nm CMOS*, JINST **16** (2021).

[89] V. Filimonov, *Development of a serial powering scheme and a versatile characterization system for the ATLAS pixel detector upgrade*, PhD Thesis: University of Bonn, 2017, URL: `https://bonndoc.ulb.uni-bonn.de/xmlui/handle/20.500.11811/7265`.

[90] Silizium Labor (University of Bonn), *Basil*, URL: `https://github.com/SiLab-Bonn/basil`.

[91] Xilinx, *IBERT for 7 Series GTX Transceivers*,
URL: `https://www.xilinx.com/products/intellectual-property/ibert_7series_gtx.html`.

[92] L. Gaioni et al., *Test results and prospects for RD53A, a large scale 65 nm CMOS chip for pixel readout at the HL-LHC*, JINST **936** (2019) p. 282–285.

[93] L. Gaioni, *Development of a Large Pixel Chip Demonstrator in RD53 for ATLAS and CMS Upgrades*, Proceedings of Science **313** (2018).

[94] T. Heim, *YARR: Yet Another Rapid Readout*, URL: `https://yarr.web.cern.ch/yarr/`.

[95] M. Daas et al., *BDAQ53, a versatile pixel detector readout and test system for the ATLAS and CMS HL-LHC upgrades*, Nucl. Instrum. and Meth. A **986** (2021) 164721.

[96] S. Orfanelli et al., *Serial Powering Optimization for CMS and ATLAS Pixel Detectors within RD53 Collaboration for HL-LHC: System Level Simulations and Testing*, Proceedings of Science **313** (2018).

[97] E. López-Morillo et al., *Design of a Radiation Hardened Power-ON-Reset*, IEEE Transactions on Nuclear Science **65** (2018).

[98] A. Dimitrievska, *Powering Measurements of RD53A*, RD53 collaboration internal meeting (2018).

[99] S. Reyntjens and R. Puers, *A review of focused ion beam applications in microsystem technology*, J. Micromech **11** (2001) 287.

[100] MASER Engineering, URL: `https://maserengineering.eu`.

[101] J. D. Alexander, *Clock Recovery From Random Binary Data*, Electronics Letters **11** (1975) p. 541–542.

[102] P. Moreira, *Lectures on Microelectronics*, URL: `http://paulo.moreira.free.fr/microelectronics/microelectronics.htm`.

[103] F. Tavernier et al., *The eCDR, a Radiation-Hard 40/80/160/320 Mbit/s CDR with internal VCO frequency calibration and 195 ps programmable phase resolution in 130 nm CMOS*, JINST **8** (2013) C12024.

[104] K. Moreira (CERN, *Private communication*, (2018).

[105] K. Moustakas et al., *A Clock and Data Recovery Circuit for the ALTAS/CMS HL-LHC Pixel Front End Chip in 65 nm CMOS Technology*, Proceedings of Science **370** (2020).

[106] Silizium Labor (University of Bonn), *SiLab X-ray irradiation machine*, URL: `https://github.com/SiLab-Bonn/X-ray_machine/wiki`.

[107] F. Márquez et al., *Automatic Single Event Effects Sensitivity Analysis of a 13-Bit Successive Approximation ADC*, IEEE Transactions on Nuclear Science **62** (2015) p. 1609 –1616.

[108] L. Standaert, N. Postiau, M. Loiselet, *UCL irradiation facilities status*, 17th European Conference on Radiation and Its Effects on Components and Systems (RADECS) (2017) p. 1–3.

[109] PULSCAN, *A Modular System for Laser Stimulation of Scaling Technologies*, URL: `https://www.pulscan.com/pages/pulsys.php`.

[110] Katholieke Universiteit Leuven), *RELY Lab*,
       URL: https://iiw.kuleuven.be/onderzoek/advise/advise-lab.

[111] D. McMorrow et al., *Subbandgap Laser-Induced Single Event Effects: Carrier Generation via Two-Photon Absorption*, IEEE Transactions on Nuclear Science (2002) p. 3002 –3008.

Acknowledgements