Karlsruher Institut für Technologie

Charakterisierung von monolithischen HV-CMOS-Pixelsensoren für Teilchenphysikexperimente

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN
(Dr. rer. nat.)

von der KIT-Fakultät für
Physik
des Karlsruher Instituts für Technologie (KIT)
genehmigte

DISSERTATION

von

M.Sc. Felix Michael Ehrler

aus Künzelsau

Tag der mündlichen Prüfung:
16.04.2021

Hauptreferent: Prof. Dr. rer. nat. Marc Weber
Korreferent: Prof. Dr. rer. nat. Ivan Perić

Karlsruhe Institute of Technology

Characterization of monolithic HV-CMOS pixel sensors for particle physics experiments

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN
(Dr. rer. nat.)

von der KIT-Fakultät für
Physik
des Karlsruher Instituts für Technologie (KIT)
genehmigte

DISSERTATION

von

M.Sc. Felix Michael Ehrler

aus Künzelsau

Tag der mündlichen Prüfung:
16.04.2021

Hauptreferent: Prof. Dr. rer. nat. Marc Weber
Korreferent: Prof. Dr. rer. nat. Ivan Perić

Für meine Eltern

Eidesstattliche Versicherung gemäß § 13 Absatz 2 Ziffer 3 der Promotionsordnung des Karlsruher Instituts für Technologie (KIT) für die KIT-Fakultät für Physik:

1. Bei der eingereichten Dissertation zu dem Thema Charakterisierung von monolithischen HV-CMOS-Pixelsensoren für Teilchenphysikexperimente handelt es sich um meine eigenständig erbrachte Leistung.
2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.
3. Die Arbeit oder Teile davon habe ich bislang nicht an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt.
4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich.
5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt.

Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

Karlsruhe, den 14. Juni 2021

. . . . . .

(**Felix Michael Ehrler**)

Contents

1. Introduction

High-energy physics experiments have formed our knowledge of small and smallest elementary particles. Enormous accelerators and detectors were built to search for new particles and interactions were investigated with unprecedented accuracy. The technology necessary to build these machines did not exist at the time of their initial design. It had to be invented and developed during the process to enable the science community to study this aspect of nature.

Silicon sensors are already dominating the sector of particle tracking. A widely used implementation is the combination of two layers of silicon, one for sensing and one for readout. Both have to be produced in highly optimized technologies and are connected by high-precision bump bonds. These hybrid sensors suit even the most challenging experimental conditions, but are very expensive and are therefore only a good choice for rather small or very demanding experiments.

High Voltage Complementary Metal-Oxide-Semiconductor (HV-CMOS) technology is a silicon sensor concept invented by Prof. Dr. Ivan Perić for high-energy physics in 2007 [1]. The sensing element and the readout electronics are both implemented on the same piece of silicon which is produced in a commercial standard process. This concept captivates due to its performance in efficiency, availability, time- and spatial resolution [2] at comparatively low costs. Especially the latter is a strong advantage over other approaches, considering the vast area tracking detectors of large-scale experiments have to cover. But also the reduced material and increased spatial resolution, compared to hybrid approaches, are strong arguments. Its radiation tolerance grants a considerable advantage over other hybrid-alternatives.

This thesis investigates cutting-edge HV-CMOS Application Specific Circuits (ASICs) in the context of high-energy particle physics, namely the High-Luminosity Upgrade of the Large Hadron Collider at CERN (Switzerland), the Mu3e experiment at Paul-Scherrer-Institute (Switzerland) and KArlsruhe TRItium Neutrino Experiment (Germany).

Besides utilization in fundamental research, HV-CMOS technology promises advantages for medical applications: The minimal material budget of monolithic HV-CMOS sensors allows for beam monitoring of therapeutic particle beams. To date, they have to accept the limitations of wire chambers in surveillance of medical irradiation sessions in favor of a detector that affects the beam as little as possible. HV-CMOS sensors offer a leap forward in spatial- and time resolution, while keeping the material budget low. Unlike wire chamber detectors, it has not been observed that HV-CMOS sensors were sensitive to magnetic fields (cf. Lorentz force), enabling live monitoring of patients by magnetic resonance imaging.

The development of HV-CMOS sensors has started with a number of experimental test chips, exploring the opportunities offered by this novel concept. As the design evolved, sensor architecture became more complex and also the size grew as much as to full reticles

size ($\approx 2 \times 2$ cm^2). Further growth is only possible by specialized post-processing steps, such as stitching or interconnection layer. From every sensor generation, information was gained on HV-CMOS technology and circuit design that was incorporated in the next generation. The latest generation is a pair of full reticle size prototypes, ATLASpix3 and MuPix10. Findings from this thesis had a large influence on their design.

During my PhD, I studied the latest HV-CMOS sensors within the Karlsruhe Institute of Technology ASIC and Detector Laboratory (KIT ADL). The characterization of novel HV-CMOS sensors, designed by Prof. Perić, my fellow PhD students and designers outside KIT, required extensive groundwork prior to the actual measurement. This groundwork has three aspects:

Framework Design From foundry, the bare chips are delivered and require an ASIC-specific characterization setup, comprising hardware, firmware and software.

Data Collection I performed measurements on the sensors in laboratory with radioactive sources, electrical test signals, so called electron injections or at particle sources, such as beam test facilities and X-ray tubes.

Data Analysis The collected data had to be analyzed, using various tools and custom programs, for filtering, fitting and visualization.

Additionally to this experiment design work, some experimental results and their interpretation will be presented, to answer the key questions that have to be posed to every particle detector: The questions for efficiency of particle detection, time resolution, spatial resolution, rate capability and radiation tolerance.

The content of my thesis is arranged in six parts. Those dealing with sensor characterization are arranged chronologically, by date of sensor production:

- Part I introduces particle detectors in general and emphasizes their importance to high-energy physics (Chapter 2), before delving into the concept and basics of HV-CMOS sensors, with special respect to monolithic active pixel sensors (Chapter 3).
- Part II presents my work on H35Demo ASIC and its context, the High-Luminosity Upgrade of the Large Hadron Collider and associated experiments at CERN, Switzerland (Chapter 4). The sensor itself is described in chapter 5, the application-specific setup in chapter 6. Chapter 7 holds the actual characterization, comprising various tests and experiments together with their analysis.
  The Belle II experiment has expressed special requirements on its tracking detector (Chapter 8). The suitability of HV-CMOS sensors for these modules, is investigated in chapter 9 on selectively thinned multi-reticle H35Demo modules.
  The characterization results of H35Demo are summarized in chapter 10.
- Part III deals with two sensors: ATLASpix1 and MuPix8. Their design is in many aspects similar or even identical and both are part of the same submission, produced on the same wafers with the same technology.
  As the name indicates aims the ATLASpix1 ASIC (Chapter 11) for application at LHC's ATLAS detector.
  Application in the Mu3e experiment (Chapter 12) at Paul-Scherrer-Institute, Switzerland, is the goal of the MuPix8 ASIC development (Chapter 13).
  For characterization of both sensors, I have developed a reusable modular characterization environment, the Multi-purpose Adapter Board (MAB) system. It is introduced in chapter 14.
  Characterization of both sensors and analysis of measurement results is conducted in chapter 15.
  At a later point, the need for a multi-sensor readout system converged into the

characterization ecosystem GEneric Configuration and COntrol, GECCO, presented in chapter 16. Its parallel capabilities have been exploited in characterization of H35Demo multi-sensor module and an ATLASpix1 beam telescope.
Chapter 17 investigates, if HV-CMOS sensors are applicable in medical ion beam monitoring and their possible advantage over alternative detectors. It includes measurements on a GECCO beam telescope.
The characterization results on ATLASpix1 and MuPix8 are summarized in chapter 18.

- Part IV presents my work on the KIT TRISTAN Integrated Circuit. Unlike the chips described in the previous parts, it is not a monolithic sensor, but a sensor assembly consisting of silicon drift diode and readout ASIC. This approach is necessary for maximal efficiency and energy resolution, required for the TRISTAN (TRitium Investigation on STerile (A) Neutrinos) upgrade of KATRIN (KArlsruhe TRItium Neutrino experiment). Both KATRIN and TRISTAN are introduced in chapter 19. This in assembly more complex and more costly approach, is feasible for the small number of channels projected for TRISTAN. It is described in chapter 20.
  Chapter 21 presents the characterization and results that are summarized in chapter 22.

- Part V concludes my work, by putting the findings on the individual sensors in relation and sketches an outlook for the HV-CMOS technology.

Part I

Particle detectors in high-energy physics experiments

2. Particle identification in HEP

Generally speaking, experiments in high-energy physics (HEP) at particle accelerators collide particles at highest energies and observe the elapsing reactions. The reactions are then analyzed e.g. to identify new particles or branching ratios of processes, to test the Standard Model of particle physics or to expand it. However, only long-living particles can be detected, as all others decay before they have the chance of reaching a detector. Consequently, what detectors can observe, are decay products of the original particle reaction. Reconstruction of the original process of interest is only possible by precise characterization of as many particles as possible: type, energy and origin.

For this purpose most detectors in HEP are constructed after the same principle: A magnetic field bends the trajectory of particles according to their charge. These curved (charged particles) or straight (neutral particles) flight paths are tracked by sub-detectors with high spatial resolution at minimal material budget to minimize deflection of the particles (Figure 2.1). The second sub-detector is located behind the tracker and measures the total energy of the emerging particles. These calorimeters typically stop all particles (with few exceptions) and record the deposited energy.

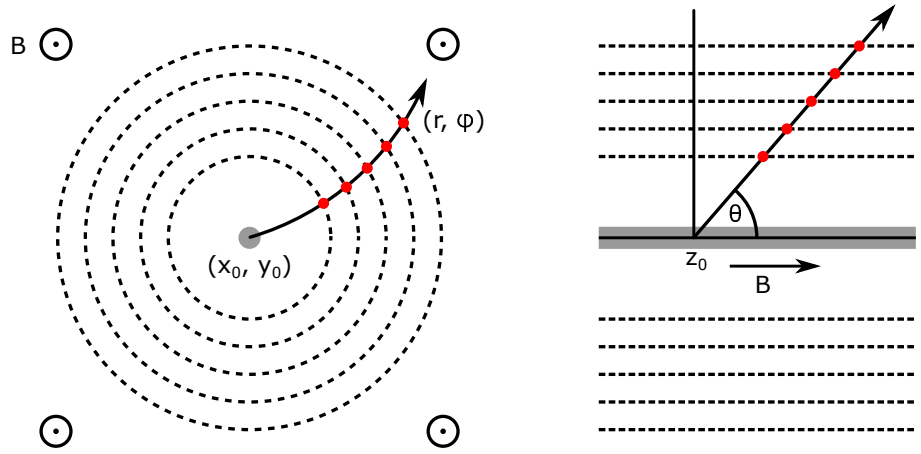


Figure 2.1: Schematic sketch of a tracking detector around the interaction point (x_0, y_0, z_0) to illustrate the principle. The magnetic field parallel to the beam pipe (gray) bends the particle trajectory, which leaves signals (red) in the barrel-shaped detector layers (dashed lines). Their coordinates r, φ and θ are used to determine the particles charge and momentum. (After [3])

These principally simple measurements are sufficient to identify particles [3]: Charged particles in a magnetic field travel on a helical path, the projection perpendicular to the magnetic field is a circle with radius

$$R = \frac{p_T}{|q|B} \tag{2.1}$$

In HEP experiments, the particles have typically a charge q of ± 1 or 0. With the known magnetic flux density B and the measured trajectory radius R, the momentum perpendicular to the magnetic field p_T can be calculated. The total momentum p is inclined at an angle θ to the primary beam and is derived from its perpendicular component:

$$p = \frac{p_T}{\sin\theta} \tag{2.2}$$

The origin of the particle trajectory is not known, but as typically many particles emerge from the primary or subsequent vertices, a point in which many trajectories cross, is likely to be a vertex.

According to energy-mass-momentum relation

$$E = \sqrt{(mc^2)^2 + p^2c^2} \tag{2.3}$$

or

$$m = \sqrt{\frac{E^2}{c^4} - \frac{p^2}{c^2}}, \tag{2.4}$$

the missing information to determine the rest mass of the particle is its total energy. It is measured in calorimeters by stopping them. The particles lose energy, which creates a large number of secondary particles, which can create subsequent particles as well, thus a tree-like structure of particles is created, the particle shower.

Charged particles create an electromagnetic shower. The two main mechanisms of energy loss are pair-production and Bremsstrahlung. There is a large number of possible calorimeter assemblies for collider experiments. Two approaches can be differentiated:
Homogeneous detectors use absorption material that at the same time generates the signal. Therefore, the entire detector volume is sensitive, meaning good energy resolution, but with limited spatial resolution. Examples are scintillating crystals with photomultiplier tubes, liquid noble gas detectors or Cherenkov detectors.
The second approach is sampling detectors. The working principle is the same as homogeneous detectors, but sensing material (active) and passive material are assembled in alternating patterns. The passive material is typically a high-Z (mass number) element, granting high energy absorption, but can not contribute to the signal. The active material generates the signal, but has a lower Z (e.g. scintillator or ionization chambers). Common arrangements of active and passive materials in sampling calorimeters are in alternating layers (sandwich-calorimeter) or bundles of active fibers in a passive material (spaghetti-calorimeter).

The energy of particles without charge is measured in hadronic calorimeters. The difference to electromagnetic calorimeters is that energy loss is not intermediated by the electromagnetic force, but by nuclear force.

Silicon based sensors play a leading role in particle tracking detectors with high spatial resolution. This chapter will briefly introduce non-silicon detectors, sketch the fundamentals of semiconductors in general and delve into silicon detectors, down to High-Voltage Complementary Metal-Oxide-Semiconductor Monolithic Active Pixel Sensors (HV-CMOS MAPS).

2.1 A shortlist of particle tracking detectors

The core of this thesis is the characterization of HV-CMOS sensors for HEP. The advantages of these sensors over other detector concepts, require mentioning the competitors and precursors [3, 4][1]:

Gaseous Ionization detector Traversing particles ionize the eponymous gas, the ions and electrons are dragged by an electric field to electrodes generating an electrical signal [6]. This principle is best known for its application in Geiger-Müller-counter for hand-held radiation detection. Many other detectors are based on this principle.

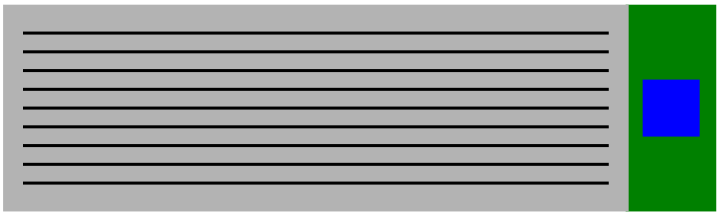
Wire Chamber Wire chambers are a form of gaseous detector with improved spatial resolution [7]. The conductive walls are used as cathode. Parallel, between the walls a large number of wire anodes is stretched. The particle trajectory is reconstructed by the evolution of the signal in these wires. Simple wire chambers have a high spatial resolution in only one direction, it is limited by the distance of the wires. Stacks of wire chambers with different orientation provide the missing information. More advanced approaches, like time projection chambers [8], deliver two-dimensional spacial and good time information.

Resistive Plate Chambers RPCs are assembled of alternating plates of high resistive material and gas-filled gaps [9]. The plates are biased with a high voltage, the gas is being ionized by traversing particles and the charge is dragged to the plates. It is detected capacitively by strips of conducting material on the other side of the resistive material.

Ring-Imaging Cherenkov Detector RICH detectors exploit the Cherenkov effect and deliver information not only on the location but also the speed of particles [10]. A particle emits Cherenkov light if it travels through a dielectric material with a larger speed than the phase-velocity of light in this material. The light is emitted in a speed-dependent angle to the moving direction of the particle, therefore on a plane sensor these photons are detected in the form of rings ($\cos\theta \propto 1/v$).

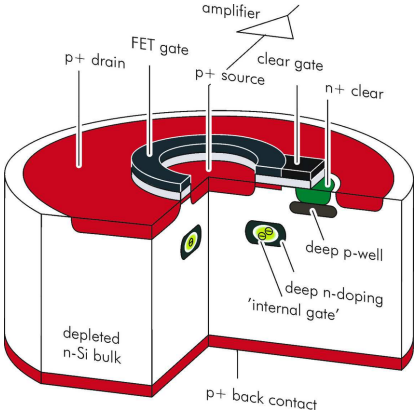
Semiconductor Detector Diodes formed by doped semiconductors are the sensing element of semiconductor detectors. Due to availability and price, silicon is most widely used in large-scale HEP experiments, but smaller experiments can be equipped with other semiconductor sensors as well. Traversing particles leave a trace of electron-hole pairs along their path, which are collected to create an electrical signal. These detectors can be produced in various forms:
Strip detectors are large pieces of silicon, up to wafer-scale, segmented by readout lines in long narrow strips (Figure 2.2a). Pixel sensors work in a similar way, but are segmented into smaller areas. The pixels are contacted individually (Figure 2.2b). The classical readout technique in both approaches requires one-to-one connections from each pixel or strip to a readout chip. Strip detectors have it located at one or more sides, pixel sensors have it connected flat on top of the sensor chip by bump bonds in a flip-chip process. This hybrid-pixel approach is widely used and is – due to the specialized technologies used for sensor and readout electronics – radiation tolerant, however, for the same reason, costly. Furthermore, the pixel size is limited by the size of bump bonds.
In both stripes and pixels, the generated charge can be collected by diffusion or drift. Collection by diffusion dominates if the passive sensor diode is left unbiased. It is less efficient and slower than collection by drift. The electric field required for drift is

[1] The cited text books give a great overview over detectors in particle physics and are recommended for further reading.



(a) Silicon strip sensors are a convenient option to fabricate large sensors at moderate costs. One element can cover several 100 cm^2. Spatial resolution is only good in one dimension. The readout ASIC (blue) is typically located at one or both ends of the sensor.

(b) Hybrid pixel sensors are composed of a sensor wafer (gray) for signal generation and readout chip(s) for signal processing (blue). The signal is transmitted by bump bonds (gold). Hybrid sensors are radiation hard and show good performance, but are expensive and require more material budget than other concepts.



(c) DEPFET sensors are a form of semiconductor sensors. They do not sense a generated charge or current directly, but the originally deposited charge is captured in a floating gate changing the characteristic of the FET's source-drain current. Active reset is required. (From [5])

Figure 2.2: Silicon strip sensors, hybrid sensors and DEPFET sensors are alternative technologies to HV-CMOS sensors for particle tracking.

induced by an external voltage, it can exceed 1 kV.
Instead of passive diodes as sensing elements, depleted p-channel field-effect transistor (DEPFET) sensors can be used (Figure 2.2c) [11]. Their advantage is that the signal is not deleted by readout, but remains stored in the sensor and can be read out repeatedly to reduce statistical uncertainties, until it is actively reset. Additionally, the MOSFET sensor structure has an intrinsic signal amplification.
Commercial complementary metal-oxide-semiconductor (CMOS) technology is an approach to combine sensor and readout electronics on the same piece of silicon, rendering connections between sensor and readout chip obsolete and allowing for a reduced material budget. However, full area efficiency and decent signal strength require modification of the standard process: Either the depleted sensor volume is to be enlarged by high resistive substrate (HR-CMOS) or by a high reverse bias voltage (HV-CMOS).

2.2 Fundamentals of silicon semiconductors

HV-CMOS sensors are based on silicon semiconductor technology. It is the most often used type of semiconductor, mainly in consumer electronics, and therefore its availability and price are attractive.

Silicon for use in electronics is produced as extremely pure mono-crystals, which are then sliced according to the crystal-orientation and further processed. Due to its extensive use, production is cheap and properties are well known.

2.2.1 Semiconductor fabrication

Silicon is a semiconductor, however this statement is only true for a mono-crystal. Silicon is produced from silicon-dioxide, which after solidification forms a poly-crystalline solid. The interfaces of the small crystals do not have deterministic electrical properties. For wafer production, the poly-silicon has to be converted to a mono-crystal and further has to be purified.

Figure 2.3: Mono-crystalline silicon is produced as cylinders, so-called boules, which can be cut into wafers. (From [12])

There are two methods to create a mono-crystalline boule (Figure 2.3), from which wafers can be cut [13]: In the Czochralski process, poly-silicon granulate is molten and a small

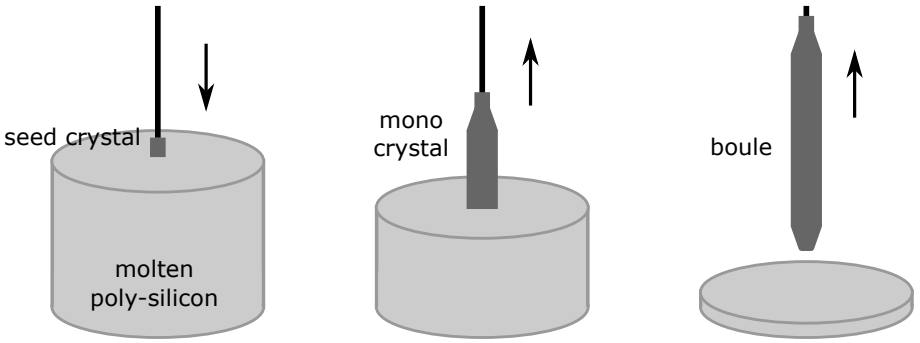


Figure 2.4: The principle of the Czochralski process of boule fabrication is to grow a large silicon mono-crystal from a small seed crystal, which is first submerged in molten silicon and then slowly retracted.

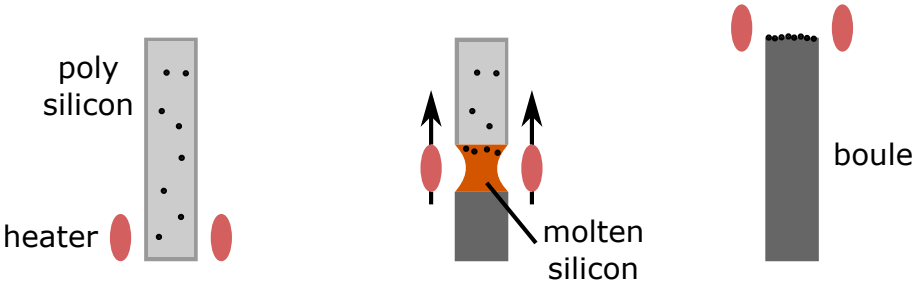


Figure 2.5: Zone melting refines a (poly-)silicon ingot by successive melting of a small part. After recrystallization, a mono-crystal is formed and impurities stay in the molten part, to be dumped in the last part of the boule.

seed crystal is brought in contact with the molten silicon. Then it is slowly and carefully pulled out again, so that the crystal grows at the surface of the molten silicon (Figure 2.4).

The second method is the Zone Melting process. It purifies a poly-crystalline boule and turns it into a mono-crystal, by melting only a disc-shaped part of the raw boule. The heating ring is slowly moved over the boule. The silicon recrystallizes and forms a mono-crystal, while impurities remain in the molten area and are dragged to one end of the boule (Figure 2.5). This method can be applied to already mono-crystalline boules for further purification.

Boules are then sliced and polished. Each of these slices is called wafer, ready for production of micro electronic devices.

2.2.2 Physics of semiconductors

The allowed states of electrons, thus the electronic behavior in a large crystal, can be determined by the time-independent Schrödinger equation [14]:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V\left(\vec{r}\right)\right)\Psi\left(\vec{r}\right) = E\Psi\left(\vec{r}\right) \tag{2.5}$$

where $V\left(\vec{r}\right)$ is the periodic potential by the crystal lattice with

$$V\left(\vec{r}\right) = V\left(\vec{r} + \vec{R}\right). \tag{2.6}$$

The potential periodicity $\vec{R}$ is defined by the crystal lattice.

The solutions to this equation are Bloch-waves [15]:

$$\Psi(\vec{r}) = e^{i\vec{k}\vec{r}} u_{\vec{k}}(\vec{r}) \tag{2.7}$$

The periodic nature of the problem is reflected by $u_{\vec{k}}(\vec{r})$, which has the same periodicity as the lattice.

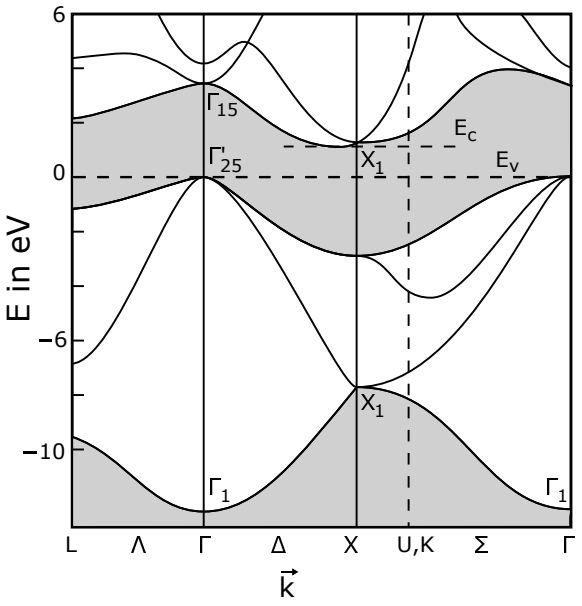


Figure 2.6: Band structure of silicon. The minimum of conduction band and valence band are not above each other which makes silicon an indirect semiconductor. The band gap is approximately 1.12 eV. (From[16], after [17])

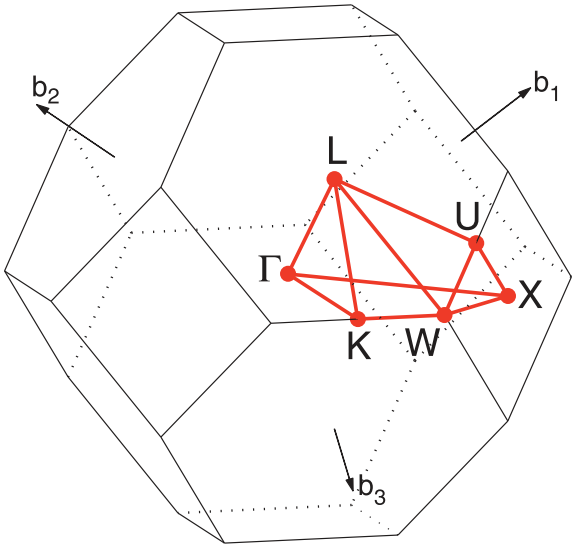


Figure 2.7: The Brillouin zone of a face centered cubic lattice, such as silicon. The points and axis of high symmetry are labeled. (From [18])

The dispersion relation of silicon is shown in figure 2.6, it shows possible energy states for electrons in the reciprocal momentum space, the $\vec{k}$-space. Quasi-continuous electron states are called band. The band with lowest energy, which is not fully occupied in ground state, is called conduction band, the next lower energy band is called valence band. The elementary cell in momentum space is the Brillouin zone, the equivalent of the Wigner-Seitz cell in normal space. Figure 2.7 shows the Brillouin zone of a face-centered cubic lattice,

e.g. of a silicon crystal. Points and edges of high symmetry are highlighted. As the crystal lattice is only symmetrical in certain directions, the electron states depend on the investigated orientation, the shown dispersion relation depicts the states along such edges of high symmetry. As silicon sensors and electronics are predominantly designed in two dimensions, the orientation of the silicon wafer, they are implemented on, is of high importance. The orientation indicated by a set of flat sides on the otherwise round wafers, which denote the crystal orientation by Miller indices [13]. Examples of widely used wafer orientations are {100} and {111}.

The shown dispersion relation of silicon shows two paramount properties of silicon: It has a band gap between conduction band energy E_C and valence band energy E_V and the minimum of the conduction band is not at the same value of $\vec{k}$ as the maximum of the valence band. Consequently, exciting an electron with the gap energy difference

$$E_G = E_C - E_V \tag{2.8}$$

is insufficient to bring an electron to the conduction band. Either, a larger energy has to be provided or a momentum has to be exchanged in addition (phonon). At absolute zero temperature, the valence band is fully occupied, while the conduction band is empty, because the Fermi-energy E_F is right in the middle of the only 1.12 eV wide band gap. Without thermal or other excitation, silicon behaves like an insulator.

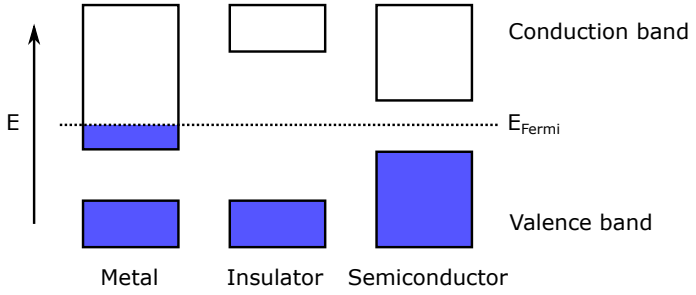


Figure 2.8: Whether a material behaves like a metal or insulator, depends on the Fermi energy. Metals have it in a band, insulators have it in a band gap. Semiconductors are insulators with a very small band gap. (From [19])

The different band configurations of metal, insulator and semiconductor are shown in figure 2.8. The Fermi level marks the boarder between occupied states (blue) and unoccupied states (white) at T = 0 K. If the Fermi level is located within an energy band of a material, it is a metal. Is it in a band gap, the material is an insulator or semiconductor. The discrimination between both is made by the size of the band gap: Materials with band gab larger than 4 eV are called insulators, with less are called semi conductors. Already a small additional energy input is sufficient to bring electrons from the valence band to the conduction band.

Usually semiconductors are operated at room temperature, in case of semiconductor detectors typically between −20°C and +50°C. Thermal excitation at these temperatures is in a semiconductor sufficient to bring electrons from the valence band to the conduction band. Both these electrons and the remaining vacancies in the valence band (holes), are free to move through the solid and can participate in a current. Thermal excitation is a statistical process, which leads to a certain fraction of electrons moving to the conduction band. The probability of a state with energy E being occupied, is described by the Fermi-Dirac distribution:

$$f(E) = \frac{1}{\exp\left(\frac{E-E_F}{k_B T}\right) + 1}, \tag{2.9}$$

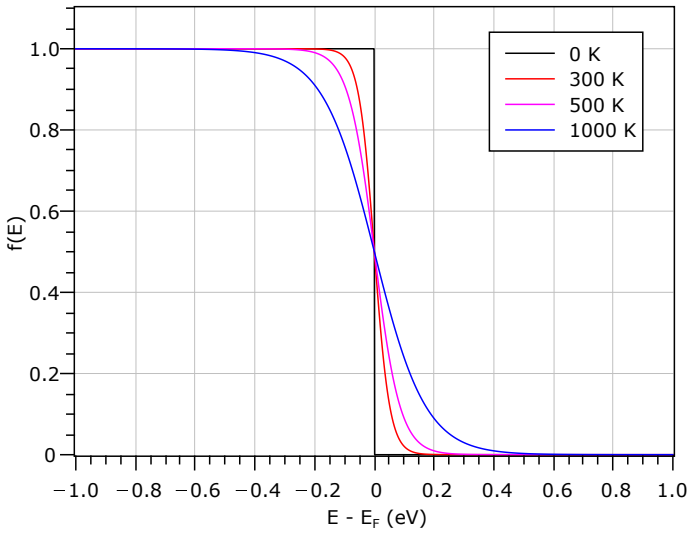


Figure 2.9: The Fermi distribution describes the fraction of occupied electron states in solids for a given temperature. (From [19])

where E_F is the energy of the Fermi level, k_B the Boltzmann constant and T the temperature. A plot around the Fermi energy of the Fermi distribution is shown in figure 2.9. At 0 K, it is a step function, all states left of the Fermi energy are occupied, all states right are empty. At higher temperature, some states to the left are unoccupied, which is interpreted as holes. They can contribute to a current. And some states to the right are occupied, which means free electrons that can contribute to a current, too.

2.2.3 Doped semiconductors

Pure semiconductors are bad conductors at room temperature, even with their small band gap, as thermal excitation allows only a small amount of free electrons and holes [13, 20]. According to Drude model, conductivity σ is:

$$\sigma = e \cdot n \cdot \mu \tag{2.10}$$

With elementary charge e, charge carrier density n and charge carrier mobility μ. The conductivity and electrical properties can be engineered by doping.

In pure silicon, the intrinsic carrier density is in the order of $10^{10}/\text{cm}^3$ at room temperature. Typical doping concentration is $10^{16}/\text{cm}^3$, so a million times higher.

Silicon atoms have four unpaired electrons, as silicon is in group IV of the periodic table of elements. Silicon forms a tetrahedral crystal with four bonds per atom. It is typically doped with elements of group III or group V. These elements are built in the crystal lattice without major distortions, but have one electron more or one electron less than silicon.

A missing electron of a group III atom causes a vacancy in a covalent bond (Figure 2.10a). This kind of vacancies cause additional energy states, close to the valence band. If an electron from the valence band fills it, a hole is left behind. Holes in the valence band are mobile. Bringing electron acceptors into the silicon lattice is called p-doping.

The analog effect occurs if a group V atom is introduced to the silicon crystal (Figure 2.10b). The un-paired electron is only weakly bound to its atom and can be easily brought to the conduction band from an additional state in the electron band structure close to the conduction band. The then free electron can participate in a current. Bringing electron donors into the silicon lattice is called n-doping.

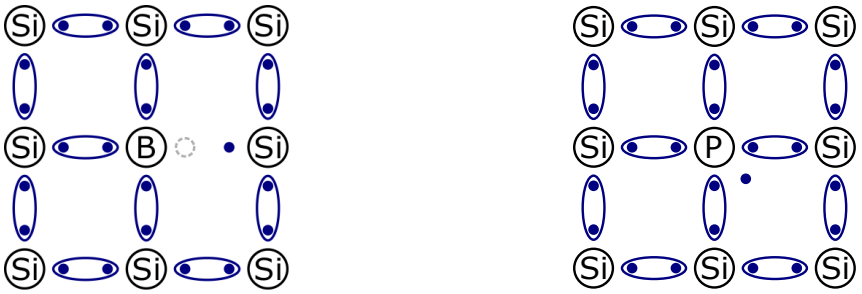


Figure 2.10: Examples for doped silicon in two dimensional representation. The boron atom (a) built into the lattice of the silicon crystal has only three outer electrons, therefore one covalent bond remains unpaired, which is interpreted as hole. The phosphorous atom (b) has five outer electrons, but the lattice offers only four partners to pair these electrons in covalent bonds. The left over electron can easily removed from its atom and participate in a current. (From [19])

The new energy stats affect the Fermi energy level of silicon: n-dopeding rises the Fermi level, p-doping decreases the Fermi level.

Wafers are produced with a low doping concentration of n- or p-type. The dopant is already brought into the silicon at the stage of boule manufacturing in low concentration. Typical effective doping concentrations range from $10^{15}/\mathrm{cm}^3$ to $10^{16}/\mathrm{cm}^3$, this results in a resistivity between $1\Omega\mathrm{cm}$ and $10\Omega\mathrm{cm}$. The medium and high doping concentrations needed for electronic components on the surface of the wafer are implemented by diffusion or ion implantation. The area of doping is usually restricted by photolitography.

2.2.4 The p-n junction

The simplest electronic component is a diode, formed by an n-doped and a p-doped area of silicon in contact. The doping atoms create new states, which change the relative Fermi levels. A solid object in equilibrium has a constant Fermi level. Therefore, as soon as the p- and the n-doped regions are in contact, an electron current from the n-doped area flows to the p-doped area and a hole current in the opposite direction, until the current is stopped by the building up electrical field [4, 13, 20].

Each area for itself has a surplus of either free electrons or free holes. According to Fick's law of diffusion

$$\vec{j}_{\mathrm{diffusion}} = -eD\nabla n \tag{2.11}$$

where D is the diffusion constant, e the elementary charge and n the charge carrier density, the carrier density in both parts are equalized. However, the left over ionized atoms build up an electrical field, which stops the diffusion eventually. Their electrical potential $\Phi(x)$ is defined by the Poisson equation:

$$\frac{\partial^2\phi}{\partial x^2} = -\frac{1}{\varepsilon_0\varepsilon_{\mathrm{Si}}}\varrho\,(x) \tag{2.12}$$

With the vacuum permittivity ε_0, the relative permittivity of silicon $\varepsilon_{\mathrm{Si}} = 11.68$ and the charge density $\varrho(x)$. In Schockley approximation the charge density is constant close to the interface of p- and n-doped region and 0 elsewhere. From Poisson equation in Schockley

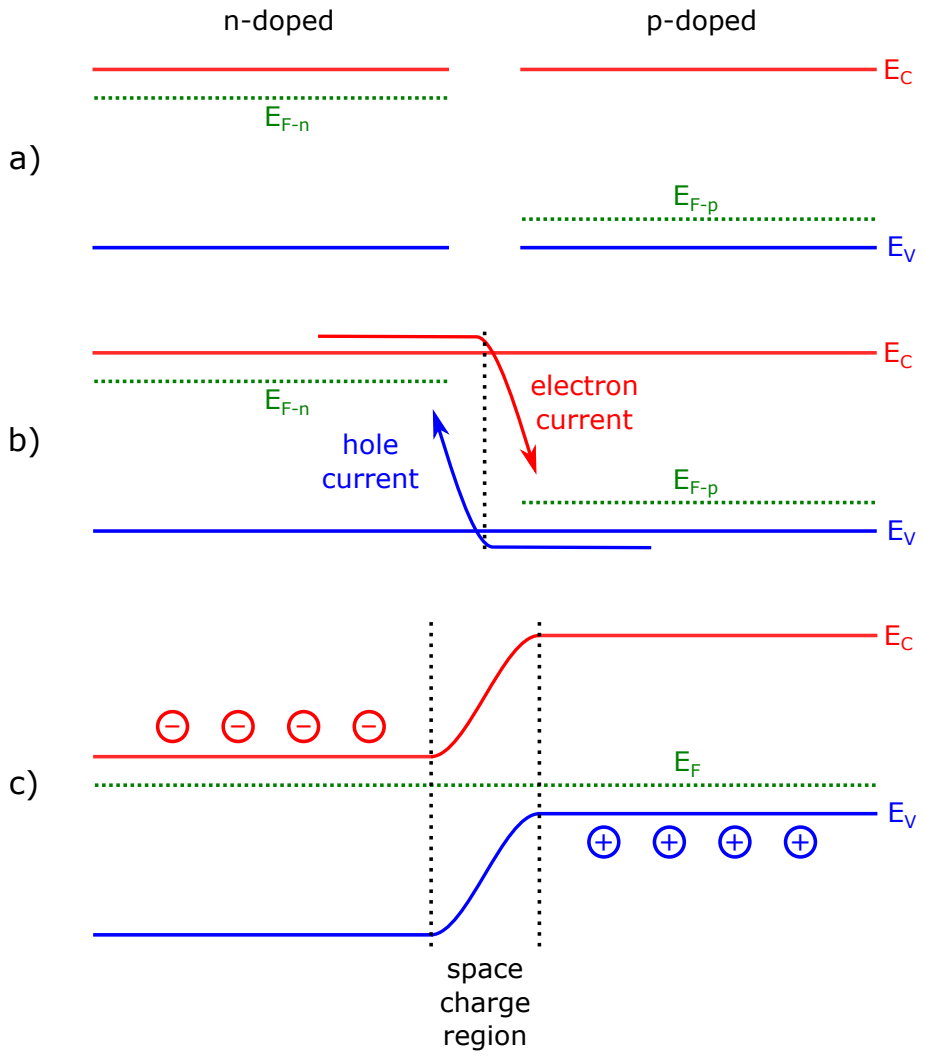


Figure 2.11: The band structure of a p-n junction. Doping introduces new states to silicon's band structure close to valence or conduction band (a). Once brought into contact, the majority charge carriers of each part diffuse into the other part (b). The process is slowed down by the electrical field, which is built up by the left behind dopant atoms, and is finally stopped when equilibrium is reached (c). In this state the bands are bent, so that the Fermi level of both parts is identical. At the interface of both parts an insulating space charge region without free charge carriers is formed. (After [4])

approximation, the width w of the space charge region can be calculated to[2]:

$$w = \left(\frac{2\varepsilon_0 \varepsilon_{\mathrm{Si}}}{e} U_D \frac{n_A + n_D}{n_A \cdot n_D} \right)^{1/2} \tag{2.13}$$

Where U_D is the diffusion voltage, and n_A is the density of acceptors in the p-doped region and n_D is the donator density in the n-doped region.

The diode is the sensing structure of semiconductor detectors. If traversing particles lose energy in the depleted volume, electron hole pairs are generated and separated by the electric field, so that they can not recombine as easily as they do in the non-depleted volume (see chapter 2.3). The separated charges are further collected in electrodes where a signal can be recorded. For strong signals, a thick space charge region is favorable. It can be influenced by the diffusion voltage and the density of charge carriers. To the diffusion

[2] Detailed deduction e.g. in [4]

voltage (typically ≈1 V) an external voltage U_{ext} can be added to increase the space charge region (reverse biasing):

$$w \propto \sqrt{U_D + U_{\text{ext}}} \approx \sqrt{U_{\text{ext}}} \tag{2.14}$$

Forward biasing reduces the depletion zone until conductance is possible. This behavior of blocking currents in one direction and letting them pass in the opposite direction, is widely used in electronic circuits, e.g to rectify alternating currents.

The diode of HV-CMOS sensors is formed by the p-doped bulk with a smaller but higher n-doped implant.

$$n_A << n_D \tag{2.15}$$

Consequently, the depleted volume is predominantly formed by the bulk and only a negligible part by the n-implant. After equation 2.2.4 is

$$w \propto \sqrt{\frac{n_A + n_D}{n_A \cdot n_D}} \propto \sqrt{\frac{1}{n_A}} \propto \sqrt{\rho} \tag{2.16}$$

with wafer resistivity ρ. Therefore, the usage of slightly more expensive wafers with higher resistivity can pay off[3].

2.2.5 Metal-oxide-semiconductor field-effect transistor

One strength of detectors based on HV-CMOS technology is that they do not only feature sensor diodes as sensing elements, but have at least signal amplification on the same piece of silicon. Some concepts feature a part of the readout chain, too. This can be driven as far as signal generation, amplification, digitization, storage and encoding of data to fulfill a standard serial data transmission protocol, are combined on the same piece of silicon, called die. The approach to combine as much functionality (analog and digital) as possible with the sensing diode, is called monolithic sensor [13, 20].

Both analog and digital electronics are based on transistors, in most cases Metal Oxide Semiconductor Field-Effect Transistors (MOSFET). Their working principle is illustrated in figure 2.12 on the example of an NMOS transistor. The basic idea is a voltage controlled resistor or voltage controlled current source, respectively. It is realized by two high n-doped implants in a p-doped substrate, the source and drain. Without applied voltages, the n-implants and the p-substrate form diodes, which blocks currents both from source to drain and from drain to source. The gate electrode is electrically insulated from the bulk by a layer of silicon dioxid. Substrate contact is established via a highly doped p-implant. Such tunnel contact provides a low resistive connection between metal and silicon.

In figure 2.12a, all aluminum contacts are at ground potential. Aluminum is the typical wiring metal on silicon chips (see Appendix A.6). The contact of aluminum and semiconductor is a the Schottky contact. The band bending on this interface leads to a substrate potential of approximately -0.75 V [23]. The area below the gate is therefore also depleted.

Applying a gate-source voltage U_{GS} greater than the threshold voltage U_{Th} forces electrons into the area below the gate (Figure 2.12b). The threshold voltage is defined as the gate-source voltage at which a conducting channel is formed below the gate, connecting source and drain. This is the case, if the density of free electrons reaches the level of free

[3] Note that NIEL damage by irradiation affects the bulk and changes the effective doping. Therefore, in experiments with high fluence, it is not reasonable to use very high resistive wafers ($> 2\,\text{k}\Omega\text{cm}$), as their advantage is quickly lost. [21]

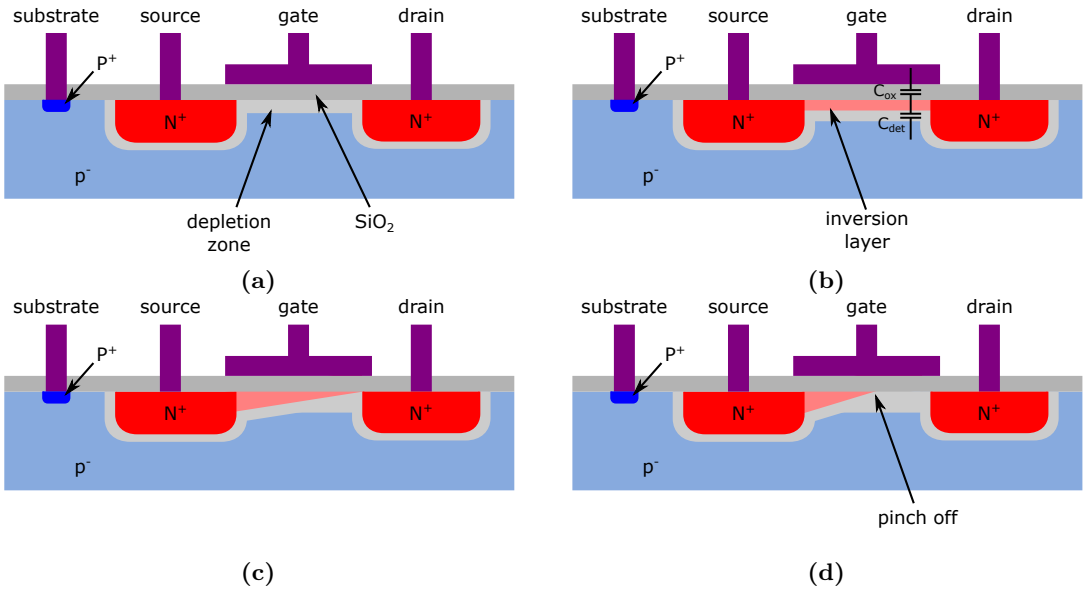


Figure 2.12: Schematic drawing of an NMOS transistor. Without applied voltages (a) and in linear mode with $U_{\mathrm{GS}} > 0$ (b). In (c), the transition from linear mode to saturation (d), is shown. (After [22])

holes in the undepleted p-type silicon. As the silicon below the gate is a p-doped area, where usually holes are the majority charge carriers, this channel is called inversion layer. This layer electrically connects drain and source.

The threshold voltage can be approximately estimated by capacitive voltage divider: The oxide-capacitance C_{ox}, between gate and inversion layer, and the depletion-capacitance C_{dep}, between inversion layer and substrate. Oxid thickness and substrate doping are usually chosen, so that $C_{\mathrm{ox}} = 2 \cdot C_{\mathrm{dep}}$. Together with relative permittivities of silicon dioxid ($\varepsilon_r = 3.9$) and silicon ($\varepsilon_r = 11.0$), the threshold voltage can be approximately calculated to $U_{\mathrm{Th}} = 0.5\,\mathrm{V}$.

At a given gate-source voltage U_{GS}, two operation modes are discriminated: For $U_{\mathrm{DS}} < U_{\mathrm{GS}} - U_{\mathrm{Th}}$, the n-channel provides an ohmic connection between source and drain. In this linear mode, is $I_{\mathrm{DS}} \propto U_{\mathrm{DS}}$.

With increasing U_{DS}, the free electrons of the channel are dragged towards the source. At $U_{\mathrm{DS}} = U_{\mathrm{GS}} - U_{\mathrm{Th}}$, the conducting channel is pinched off at the drain (Figure 2.12c). With further increasing U_{DS}, the point of pinch off moves towards the source. The shortened channel does not stop the current, but limits it to the saturation current I_{sat}: The increased voltage is partially compensated by the shorter channel and the current increases slower (Figure 2.12d). This mode is called saturation region.

The characteristic of a MOSFET is shown in figure 2.13. It depicts the source-drain IV-characteristic for fixed gate voltages. The source-drain voltage at which the inversion layer is pinched off, is indicated by the red line. The area left of it is called the linear region, even though good linearity is only given for small voltages. In the linear region, the MOSFET acts like a voltage controlled resistor.

The area right of the red line is the saturation region. The curves are not perfectly flat, but are still rising with a very small slope. In this regime, the MOSFET acts like a voltage controlled current source. The slope is called transconductance g_m, which calculates from

$$g_m = \frac{2 \cdot I_{\mathrm{DS}}}{U_{\mathrm{GS}} - U_{\mathrm{Th}}}. \tag{2.17}$$

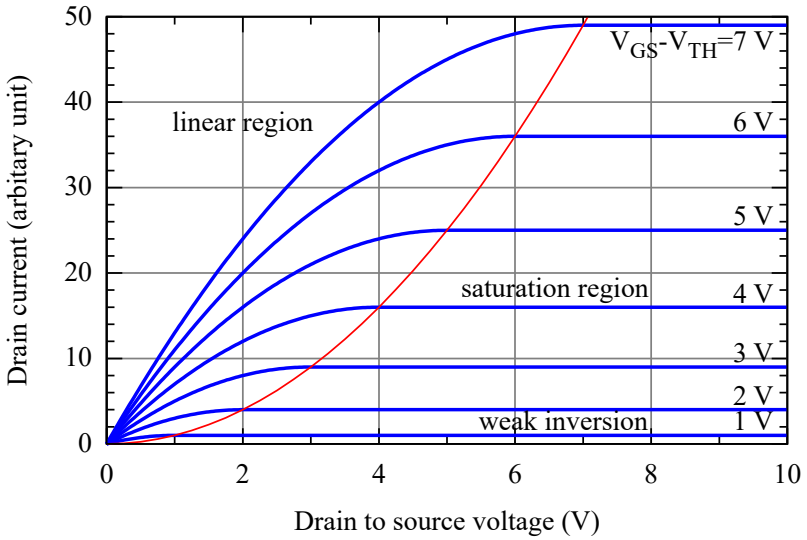


Figure 2.13: Drain-source characteristic of a MOSFET transistor for different gate voltages. (From [24], modified)

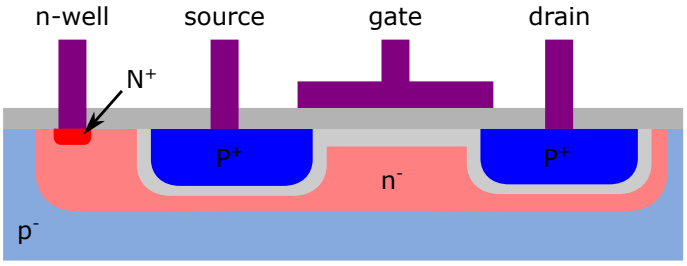


Figure 2.14: Schematic drawing of a PMOS-transistor.

So far only the NMOS transistor has been discussed. The second type, the PMOS transistor, is in many aspects just a mirrored twin of the NMOS transistor, but with important differences: The doping profile is exactly inverted to NMOS. Drain and source implants are of p-type surrounded by n-type silicone. In order to have NMOS and PMOS transistors on the same substrate, a deep n-well is added prior to PMOS implementation to provide a local n-type surrounding. As for NMOS, the inversion layer is controlled by the gate voltage, but with opposite sign. The source-drain current is sustained by holes. Their reduced mobility, compared to electrons of the NMOS, makes the PMOS transistor react slower to changes of the gate voltage.

2.3 Signal generation in silicon sensors

One advantage of HV-CMOS sensors over common CMOS sensors is the increased depletion zone by reverse biasing of the sensor diode. Incoming particles interact with silicon and generate electron hole pairs. Only charge carriers in the space charge region are dragged by drift to the collection electrode. Charges outside, can contribute to a signal as well, but for HEP experiments, they arrive too late at the electrodes, as they move by diffusion. Furthermore, is charge collection outside the electric field of the depletion zone inefficient, as electrons and holes are not actively separated from each other and therefore have an increased probability of recombination, or are trapped on the long diffusion path through the sensor by defects.

The mean energy required to generate a electron-hole pair in silicon is 3.6 eV, which was

found empirically. This seems rather surprising, as the band gap is only 1.2 eV. But as silicon is an indirect semiconductor, not only the energy of the band gap has to be provided, but also the energy to create a phonon to compensate for the momentum difference between conduction band minimum and valence band maximum. Such phonon can be caused by thermal excitation or by other incoming particles. Still direct transition from valence to conduction band is possible if sufficient energy is provided.

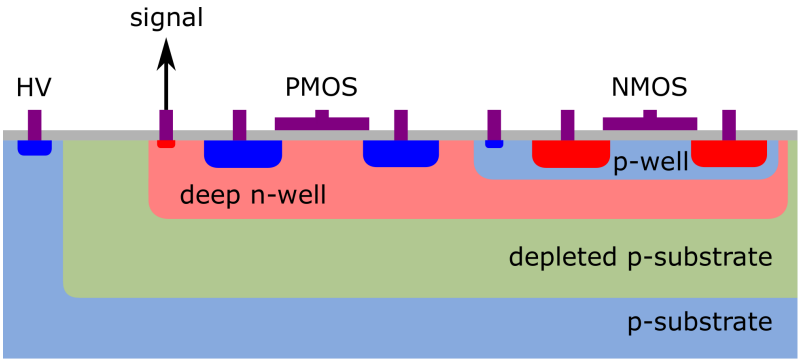


Figure 2.15: Schematic drawing of a pixel with smart diode design: The n-well, which forms the sensor diode together with the p-substrate, houses the electronics for signal amplification and -processing. The substrate is depleted by a high voltage (HV).

The increased lateral and transversal sensitivity is not the only advantage of HV-CMOS sensors over CMOS sensors: The concept of smart diode integrates signal amplification and processing inside a deep n-well which is also the charge collection electrode [1]. Figure 2.15 shows the schematic of a single smart diode. The p-type substrate and a deep n-well form the sensor diode. The thickness of the depletion zone is increased by reverse biasing of this diode via the substrate contact (HV). Due to differences in doping concentration, the depleted volume is predominantly p-substrate and only to a small extent deep n-well. The electron-signal is sensed at the deep n-well.

The transistors used in signal amplification and -processing are placed inside the deep n-well. PMOS transistors can be placed directly in the deep n-well, NMOS transistors require an additional shallow p-well.

2.3.1 Charge generation by absorption

The typical usecase for HV-CMOS sensors is particle tracking, not energy determination by absorption. However, the absorption of X-ray photons is an important tool for calibration and characterization of then sensor and the enclosed front-end electronics.

It is a proven way to generate a precisely known charge in the sensor diode and track the signal from generation to readout. The process of absorption is shown in figure 2.16. The absorbed photon transfers its entire energy to an electron of an atom. This excited electron moves a short distance[4] through the sensor, losing its energy to generate electron hole-pairs. If the photon is absorbed in the space charge region, all created electron-hole pairs contribute to the signal of a single pixel. Therefore, charge sharing is highly improbable and a pixel collects either all electrons of an event or none at all.

In this way, monochromatic X-rays can be used to feed a well known signal to the detector. A common source for X-rays are radioactive sources, e.g. ^{55}Fe (5.888 keV). Another possibility is to use the spectra of certain elements, which often have a dominant energy and some secondary peaks. The source of these spectra are target discs, usually made from a single element, which are illuminated by an X-ray tube with a wide energy range (see

[4]Travel distance depends on the deposited energy. For applications in this thesis, it is in the order of 1 µm.

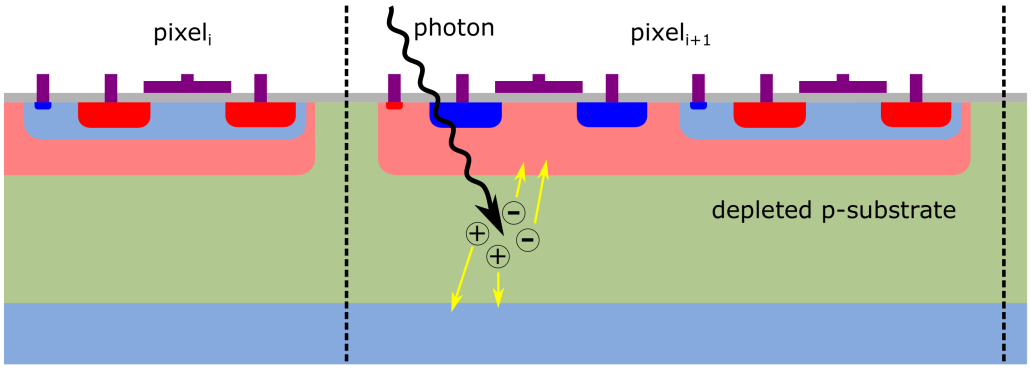


Figure 2.16: If a photon is absorbed by the sensing material of a pixel, its entire energy is available for electron-hole generation. This very localized process provides a well known signal in a single pixel.

Appendix B.2). The incoming X-rays excite the atoms of the target, which in turn emit the characteristic X-rays when returning to ground state.

2.3.2 Charged particle interaction with silicon

Semiconductor sensors can be used in particle physics, to track unknown charged particles on their trajectory through a magnetic field, to identify them. The energy of these particles is determined in calorimeters, outside the tracker.

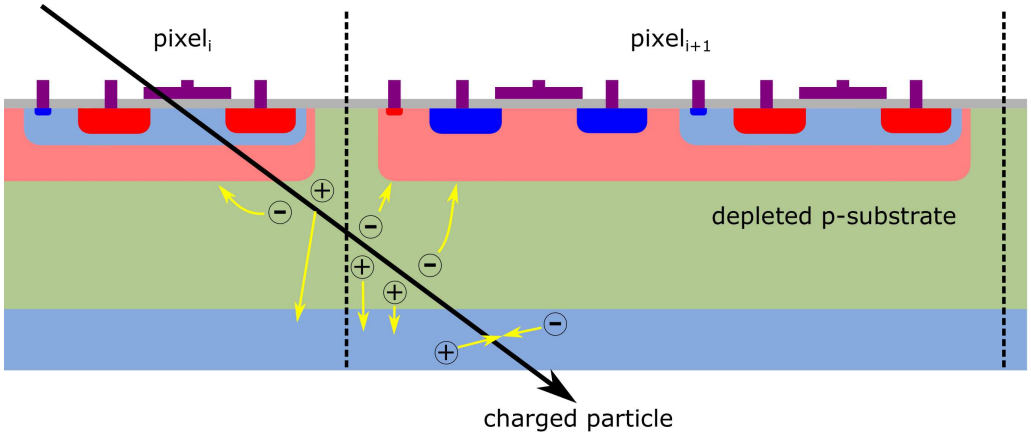


Figure 2.17: A charged particle moving through a silicon detector leaves a trace of electron-hole pairs. If generated in the space charge region, they are separated by the electrical field and electrons are collected by drift in the pixel electrode and contribute to its signal. The holes are collected by the substrate contact. Electron-hole pairs generated outside the space charge region are likely to recombine or diffuse and get trapped.

Particles of interest have enough energy to cross many sensor layers without being too much affected. However, charged particles interact with the electron hulls of the atoms in the sensor material by the long range Coulomb force along their trajectory. Figure 2.17 shows electron-hole production along a charged particle's path through the silicon sensor. Especially inclined trajectories leave the charge often in more than one pixel. This is called charge sharing, and can also happen for particles traveling perpendicular to the sensor.

The amount of energy transferred per traveled length depends on the particle's energy

(speed) and the cross section. It is described by the Bethe-Bloch formula [25, 26]:

$$-\frac{\mathrm{d}E}{\mathrm{d}x} = \frac{4\pi n z^2}{m_e c^2 \beta^2} \cdot \left(\frac{e^2}{4\pi\varepsilon_0}\right)^2 \cdot \left[\ln\left(\frac{2m_e c^2 \beta^2}{I \cdot (1-\beta^2)}\right) - \beta^2\right] \tag{2.18}$$

where v is the particles velocity, c is the speed of light, β is v/c, E is the particle's energy, x is the distance, e is the elementary charge, z is the number of e, ε_0 is the vacuum permittivity, n is the electron density, m_e is the electron mass and I is the mean excitation potential of the material. The mean excitation potential for silicon is experimentally found to be 173 eV.

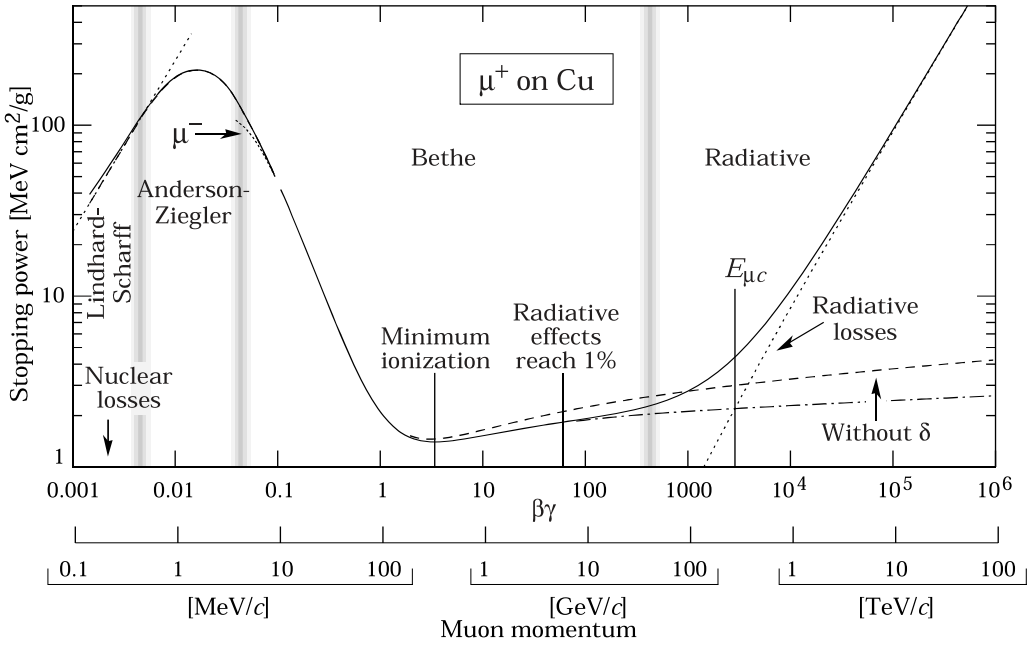


Figure 2.18: Stopping power of muons in copper as a function of muon momentum. The interaction of electrons in silicon is of the same shape. The minimum in the Bethe-dominated region is of special interest, as such minimum ionizing particles (MIPs) are numerously created in particle physics experiments and need to be detectable. (From [27])

Figure 2.18 shows the stopping power of muons in copper. The absolute values are different from those of particles in silicon, but the general behavior is the same: The energy loss of medium energetic particles is dominated by ionizing energy loss (labeled 'Bethe'). At very high energies radiative loss (Bremsstrahlung) becomes important and finally dominates over the energy loss by ionization. The so created photons most likely escape the detector, thus can not contributed to the signal. Lower energetic particles are not important for the applications in HEP.

Of special consideration for particle trackers is the global minimum of the energy loss function, the point of minimum ionization. This minimum is found in the range described by Bethe-Bloch. The minimum ionizing electron has an energy of approximately 1.5 MeV. The most probable number of electron-hole pairs generated per micrometer is 76 [4], the mean is 108.

These values imply an asymmetric distribution of dE/dx, which is of the form [28]:

$$f_{\mathrm{L}}(\lambda) = \frac{1}{\pi}\int_0^\infty e^{-t\ln t - \lambda t}\sin(\pi t)dt \tag{2.19}$$

Shifting and scaling of this Landau formula delivers the energy loss distribution for charged particles of a given energy traveling through a thin sensor.

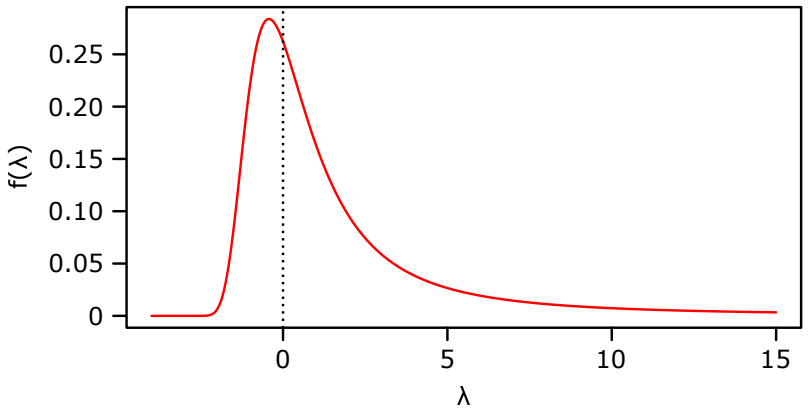


Figure 2.19: The Landau function describes the probability of energy transmission from a traversing charged particle of fixed energy to the surrounding medium. (After [29])

Figure 2.19 shows the plot of a Landau distribution. Note that the most probable value (MPV), the peak, is not the mean of this function. The tail of the function can be partially explained by δ-electrons, which have the potential to transmit a lot of energy from the original particle to the sensing material, if they travel in the plane of the sensor.

2.3.3 Particle scattering

A thick depletion zone in the sensor diode is beneficial, as it increases the charge signal. The amount of material in the path of the particle should be as small as possible, as more material means that the particle loses energy before it is measured in the calorimeter. Further this means more chance for scattering, which alters the direction of the particle and thereby reduces the momentum resolution measured in the detector by track reconstruction.

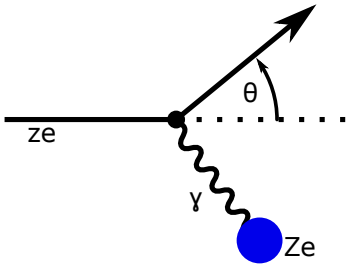


Figure 2.20: Scattering of an incoming particle with charge z on a nucleus with charge Z, is described by Rutherford scattering.

Scattering of charged particles in HEP is dominated by Coulomb-scattering on the electric field of a nucleus, not on the electron hull [3]. This process is described by Rutherford-scattering (Figure 2.20). The differential cross section according to Rutherford is given by:

$$\frac{d\sigma}{d\Omega} = \left(\frac{1}{4\pi\varepsilon_0}\frac{zZe^2}{4E_0}\right)^2 \frac{1}{\sin^4\left(\frac{\theta}{2}\right)} \tag{2.20}$$

Where z is the charge of the incoming particle, Z is the charge of the nucleus, e the elementary charge, E_0 the initial particle energy and θ is the deflection angle.

A particle traveling through a detector scatters multiple times. From the central limit theorem of statistics follows a Gaussian distribution of the angle of the emerging particles[5] [3]:

[5] The actual angular distribution has been calculated by Molière [30]. In Gaussian approximation, small angles are over-represented, large angles are under-represented.

$$f\left(\theta_{\text{plane}}\right) d\theta_{\text{plane}} = \frac{1}{\sqrt{2}\pi\theta_0} \exp\left(-\frac{\theta_{\text{plane}}^2}{2\theta_0^2}\right) d\theta_{\text{plane}} \tag{2.21}$$

This formula describes the projection to a plane of the angle between incoming and outgoing particle θ_{plane}. Its parameter θ_0 is approximated by the Highland equation [31] (with parametrization from [32]):

$$\theta_0 = \frac{13.6 \text{ MeV/c}}{p\beta} z \sqrt{\frac{x}{X_0}} \left(1 + 0.038 \ln \frac{x}{X_0}\right) \tag{2.22}$$

Besides the speed of light c, it depends on properties of the scattering particle (momentum p, charge z and velocity $\beta = v/c$) and on the material at which it scatters (thickness z, radiation length X_0). We see that the thickness of a detector has to be thin, in order to minimize the influence of scattering to the particles trajectory.

3. HV-CMOS sensors

The signal charge in HV-CMOS sensors is generated in the space charge region of the diode formed by a deep n-well in p-substrate (The generation process is described in chapter 2.3). The electron part of the signal charge is collected in each pixel. Elementary property of HV-CMOS sensors is the signal amplification within the perimeters of each pixel. However, details and further signal processing differ:

Hybrid assembly is the least ambitious approach. Only the sensing element of hybrid detectors is replaced by a less expensive HV-CMOS pixel array, which is connected via bump bonds to the readout chip [33]. The, due to reduced depletion voltage, smaller signals are amplified on sensor, every consecutive step in the readout chain is taken over from the conventional hybrid detector.

Capacitive Coupled Pixel Detector is similar to the hybrid approach, but instead of signal transmission via bump bonds, the bump pads are left unconnected [34]. The signal by HV-CMOS sensors is strong enough to be transmitted capacitively to the readout chip. Only minor or even no adjustments on the readout chip are required. Sensing and readout chip are connected with glue and a small number of bump bonds for power transmission. This approach allows for significant smaller pixel geometries, as the bumping technique is the limiting factor for pixel size in hybrid detectors.

Monolithic Active Pixel Sensors are the goal of HV-CMOS development [1]. Not only signal generation and amplification, but all steps of the readout chain are on the same die. HV-CMOS allows for powerful analog electronics next to complex digital logic. It seems feasible to cast the analog signals of a pixel matrix into a binary synchronous data stream with standard transmission protocol and send it to computer at a speed of over a gigabit per second.

HV-CMOS and HR-CMOS are two approaches to gain a sufficiently large depletion volume [35]. While HV-CMOS uses cheap standard substrate wafers and uses carefully designed guard ring structures to increases the depletion voltage before breakdown, HR-CMOS relies on more expensive substrate with higher resistivity rather than a high depletion voltage (see chapter 2.2.4). These two variants do not exclude each other and are often combined, but depending on the application, focus usually is laid on one.

The HR-CMOS approach is not to be confused with the commercial CMOS technology used in optical sensors. Like HV-CMOS sensors, HR-CMOS sensors strive for a 100% fill-factor, while optical sensors, e.g. in mobile devices, have a much lower fill-factor.

Location of digital circuitry is an issue in chip design: Digital electronics by definition switches only between high level and ground at a rapid rate. The so generated switching noise is prone to couple into the sensor diode or analog electronics, rendering them insensitive to actual hit signals. Two solutions with their own benefits and disadvantages have been developed: The 'analog islands in a digital sea' approach is

pursued in some readout chips of hybrid detectors (e.g. RD53A [36]). The sensitive analog components are clustered in islands and are protected from noise of the digital sea by guard rings. This concept is in principle transferable to the HV-CMOS sensors. The other approach is to cluster the digital circuitry at one side of the sensor chip, the digital periphery. This insensitive region is the only synchronous part of the chip, the pixel matrix operates asynchronously, which needs no noisy clock distribution over the matrix. Signal from pixels to periphery can be transmitted analog or via quasi digitized signals, which means their logic levels are closer to each other to minimize noise and cross talk. The sensors characterized in this thesis follow the latter approach.

Diode size is a yet open intensively discussed topic. The sensor diode can be either formed by a large n-well, which houses the electronics, with the substrate (large diode design), or it can be formed by a small n-implant, which is placed next to the electronics n-well (small diode design). The two variants are sometimes referred to as 'large fill factor' and 'small fill factor' design, even though these terms are misleading in the context of CMOS sensors. Both variants attempt 100% fill factor, meaning 100% of the pixel area is sensitive. The benefit of a small collection diode is a reduced sensor capacitance and therefore a larger voltage signal.
A small collection electrode in a large pixel requires careful engineering of the electrical field and limits the depletion voltage. The electrical field lines of a large n-well pixel are parallel to each other, those of a small n-implant pixel are in-homogeneous. Especially in a radiative environment, fabrication modifications are essential to reach and maintain full efficiency. This contradicts the primary concept of HV-CMOS sensors being a cost effective alternative to hybrid detectors.
This thesis focuses on large diode designs.

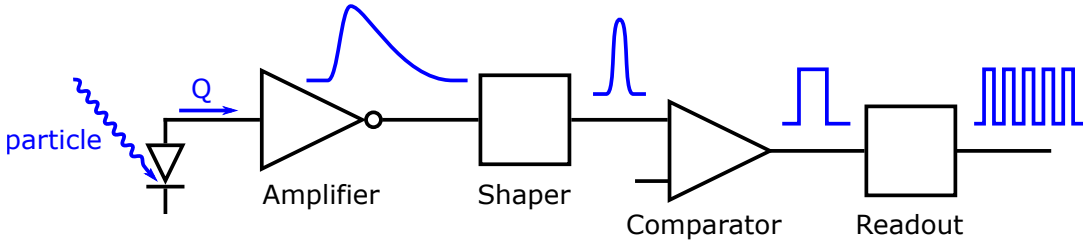


Figure 3.1: Typical signal path of a monolithic HV-CMOS sensor, from generation to readout. A charge signal is generated in the sensor diode and converted to a voltage signal by the charge sensitive amplifier. Optional is the shaping stage, which may change the pulse shape or detect edges. The comparator digitizes the signal and the readout stage wraps the provided information in a data set, which is multiplexed with the data of all pixels and finally sent out to an FPGA.

A simplified signal path of a monolithic active pixel sensor is shown in figure 3.1. The charge signal caused by a particle in the sensor diode, is converted to an analog voltage signal. Optional intermediate stages, such as edge detection or impedance conversion, shape the signal. One or more comparators digitize the signal, which is then used to create a data package. Such package can contain diverse information, e.g. hit location, hit time, but also analog information or others. In characterization setups all data packages of a sensor are multiplexed and sent as differential signal to an FPGA.

3.1 The HV-CMOS technology

The High Voltage Complementary Metal-Oxide-Semiconductor (HV-CMOS) technology has not been developed for sensors primarily. It is a technology platform for mixed signal processing that features both PMOS and NMOS transistors (CMOS). 'High voltage' in the

context of ASICs indicates that voltages above 5 V are possible. HV-CMOS technologies of all feature sizes are used in many fields, e.g. auto-motive, power management, motion-sensors, light-sensors, wireless data transmission and many more [37].

In HV-CMOS sensors, the high voltage is applied between p-substrate (negative) and n-well (positive) to enlarge the sensitive volume. All other functions are in the low voltage domain.

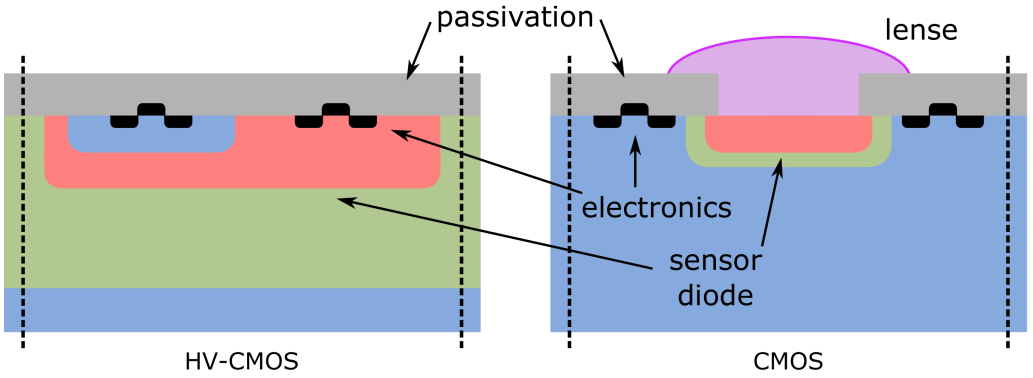


Figure 3.2: Comparison of HV-CMOS sensor (left) and optical CMOS sensor (right).

The HV-CMOS technology can be compared to, yet is substantially different from commercial CMOS sensors, used e.g. in mobile phone cameras. The cross-section of both is shown in figure 3.2. HV-CMOS sensors are to detect high energetic charged particles (not stopped by the sensor) or high energetic photons (sufficient penetration depth), thus have typically no need for a passivation opening and are yet sensitive on the area of an entire pixel (100% fill-factor).

CMOS sensors require a passivation opening to grant optical photons access to the sensitive volume of the sensor diode. The eponymous CMOS electronics is implemented to the sides of a pixel in the insensitive area (fill-factor $\ll 100\%$). The sensitive volume of HV-CMOS sensors is deeper, due to the additional depletion voltage applied to the sensor diode.

Charge collection in CMOS sensors is dominated by diffusion and therefore slower than in HV-CMOS sensors, where the electric field of the high depletion voltage collects charge carriers by drift faster and more efficiently.

The readout strategy of CMOS senors is typically the rolling-shutter principle: Consecutive readout of each pixel row, one by one. This is a good approach if the entire frame needs to be read out with limited time resolution. Virtually all pixels record events during a readout cycle ($\approx 100\%$ occupation).

Occupation of HV-CMOS sensors in high-energy physics is much lower, but a very high time-resolution and repetition rate is required, as single particles are to be tracked. Therefore, readout is implemented event-driven and zero-suppressed. Often an additional trigger defines time-frames of interest.

3.2 Signal amplification

The sensor diode provides a short current pulse, its integral is the signal charge. This current is typically very short and therefore often only the signal charge is used. A charge sensitive amplifier (CSA) converts the charge to a voltage signal and adjacent shaper stages modify or strengthen the signal.

The circuit diagram of a CSA is similar to an integrator, which is one of the most basic applications for an operational amplifier.

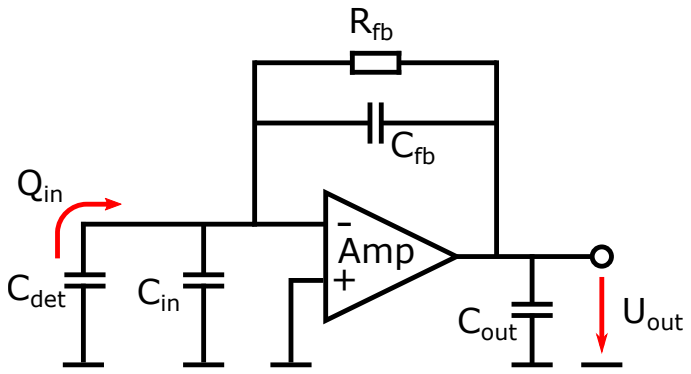


Figure 3.3: Simplified schematic of a charge sensitive amplifier.

Figure 3.3 shows the simplified configuration of a CSA with the most important components. Its output voltage can be described by:

$$U_{\text{out}} = -\frac{Q_{\text{in}}}{C_{\text{fb}} + \left(C_{\text{det}} + C_{\text{in}} + C_{\text{fb}}\right)/A} \tag{3.1}$$

where U_{out} is the output voltage, U_{fb} is the capacitance in the amplifier's feedback loop, A is the amplification, C_{det} is the detector capacitance and C_{in} the input capacitance of the amplifier. The feedback resistance defines how quickly, after a signal, the output voltage returns to its idle value. The shown layout features an (adjustable) resistor for continuous reset, but active reset is a possibility, too.

Besides the output voltage, the rise time of the amplifier is of special interest. The time between two events that have to be distinguishable in HEP experiments, can be much less than 100 ns. The rise time can be estimated by

$$\tau_{\text{rise}} = \frac{C_{\text{out}}C_{\text{fb}} + C_{\text{out}}C_{\text{det}} + C_{\text{fb}}C_{\text{det}}}{g_m C_{\text{fb}}} \tag{3.2}$$

C_{out} is the output capacitance and g_m the amplifier transconductance. Note that the rise time does not depend on the signal charge, but the output voltage does, though.

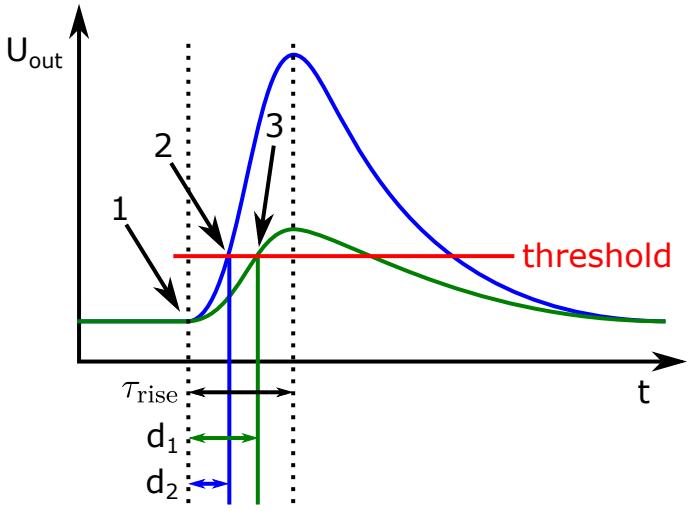


Figure 3.4: The time-walk effect appears whenever an analog pulse is digitized by comparing it to a fixed threshold level. It limits the time resolution to the rise time of the amplifier.

The constant rise time in combination with various pulse heights, leads to a time uncertainty

during digitization of the signal. A comparator compares the voltage signal to a fixed threshold (Figure 3.4 in red). Two signals starting in the same moment (1), but with different amplitude, cross the threshold at different times (2 and 3). The time-walk can be as large as the rise time: The smallest detectable signal touches the threshold at its peak time, while an infinite signal reaches the threshold already at starting point. Therefore is the rise time the lower limit for time resolution. In HV-CMOS sensors the rise time is typically between 20 ns and 100 ns, which is insufficient for many experiments. Measures to significantly improve time resolution will be presented.

3.3 Digitization

A differential amplifier digitizes the analog signal by comparing it to a threshold. A simple schematic is shown in figure 3.5. One input U_{in} is capacitively connected to the output of the CSA and the other is connected to a fixed voltage, the threshold voltage Th.

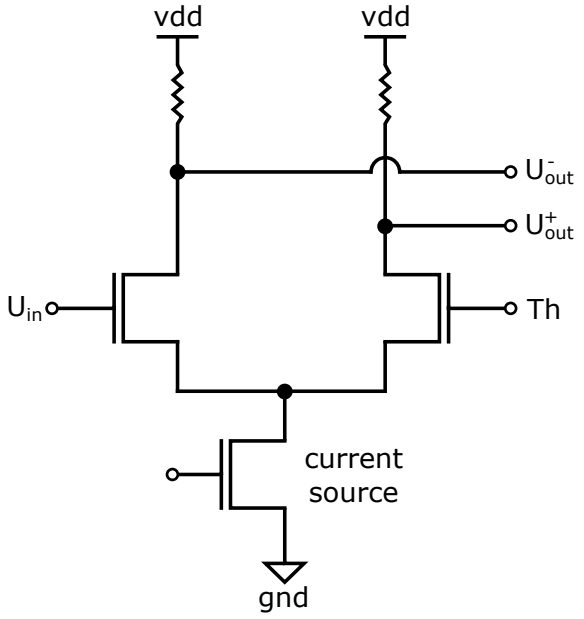


Figure 3.5: Digitization of analog signals is done by a comparator. It is formed by a differential amplifier, with one input being fixed to the threshold voltage.

In idle, U_{in} is kept at the baseline voltage (not shown) slightly above threshold level. The current through the differential amplifier is defined by a current source. Without signal, the current through each arm is identical. A positive signal at U_{in} increases the gate source voltage U_{gs} of the transistor, therefore its drain source current I_{ds} rises which leads to a current difference between both arms. The resulting difference in voltage drop at the resistors of each arm, is the output signal. Therefore, a positive input signal causes a positive deviation of output signal U^{+}_{out} and a negative deviation of output signal U^{-}_{out}. Sometimes only one of these outputs is used.

The resistors can be replaced by a current mirror, which is especially beneficial if only one output is used. And it is more convenient to implement on an ASIC. The configuration of a differential amplifier with one fixed and one dynamic input, is called comparator.

The output of the comparator is an asynchronous quasi digital signal[1], with a length that reflects the analog pulse length. It is used in the readout step to generate all necessary information: Depending on the signals origin pixel, an address is looked up and stored in

[1] Typically, it is not a rail-to-rail CMOS signal, nor are the edges very steep. Both can be achieved by adjacent stages.

the hit data package. Upon arrival, the state of a counter is stored as time reference, the time stamp. This counter is operated at a well-defined frequency and phase. More data may be collected in this state, but hit location and hit time are the essential information and should be included in every readout package.

Typically a group of pixels, e.g. a column, shares this end-of-column block. Many of them generate data packages, to be sent out to an FPGA. The readout managing block multiplexes these packages to avoid data collision. The data are then serialized and transmitted. The output is often designed as differential low voltage signals (LVDS). Differential lines have the advantage that noise, that might be picked up in long cables, affects both the positive and the negative line in the same way and therefore cancel each other out. Additionally, is the voltage drop in cables of no major concern, as a differential line transmits data as a current signal and not a voltage signal, such as single ended lines.

3.4 Noise sources

In the previous sections, the basics of signal generation from particles and propagation in (monolithic) HV-CMOS sensors have been discussed. An ideal sensor would detect even the tiniest deposited charge in the sensor diode and nothing in addition. However, in a real system, noise plays a significant role in all considerations: Increasing sensitivity of the system, makes it at the same time more prone to misinterpreting noise as signal. The signal-to-noise ratio (SNR) is a measure that puts the expected signal in relation to its uncertainties and perturbations, to evaluate the quality of a detector.

Of course, noise and SNR can be measured once a sensor system has been set up, but common noise sources are known and an estimation of noise in a system prior to implementation is possible. The noise sources in a detector system are caused by all electronic components. The following list comprises the most important noise sources of monolithic HV-CMOS detectors according to [3] and is not closed:

Shot Noise Despite the depletion layer, a (sensor) diode shows a small leakage current when being reverse biased, even before breakdown voltage. The generation process is a statistical process with 'countable' generations per time. Therefore, its probability distribution is described by the Poisson-distribution:

$$P_\lambda(n) = \frac{\lambda^n}{n!} e^{-\lambda} \tag{3.3}$$

with parameter $\lambda > 0$. Applied to shot noise it delivers the noise power spectral density $d\langle i^2 \rangle_{\text{shot}}$ at a current I_0:

$$d\langle i^2 \rangle_{\text{shot}} = 2eI_0 df \tag{3.4}$$

Shot noise appears where a current crosses a potential barrier. It is independent of the temperature.

Thermal Noise Thermal noise is caused by the thermal movement of free charge carriers. It occurs in components with and without current, other than shot noise, which requires a current. The current density fluctuation $d\langle i^2 \rangle_{\text{therm}}$ is given by:

$$d\langle i^2 \rangle_{\text{therm}} = \frac{1}{R} 4k_B T df \tag{3.5}$$

with the Boltzmann constant k_B, temperature T and resistance R. The resistance R is not necessary a resistor, it can be also the resistance of the inversion layer of a MOSFET. The thermal dependence allows for noise reduction by cooling. This form is also called Johnson-Nyquist noise.

Flicker Noise This type of noise has a spectral density proportional to the inverse frequency $1/f$ and is therefore also called $1/f$-noise. It appears in electronic systems and possibly has multiple causes, which makes a universal theoretical description difficult. In common is their behavior that leads to the eponymous $1/f$-behavior. In HV-CMOS detectors, it appears in the conducting inversion channel of MOSFETs. The oxide layer insulating gate and substrate can trap charges, which varies the conductivity in the channel, thus causes noise. The trapped charges have an exponential probability to be set free again.

Pick-up Noise As the name indicates, this noise is not necessarily a property of the investigated sensor system. Switching noise in the proximity of a characterization setup or electromagnetic waves in general can couple into the circuitry and cause noise. A specific kind of pick-up noise is cross talk. It occurs if one part of the sensor is affected by another, e.g. unwanted feedback from comparator to amplifier input or from one pixel to another.

Jitter Signal noise can cause uncertainties in signal recognition time, additionally to the previously discussed time-walk effect. Signal noise causes time uncertainty at the digitization stage. Noise on the comparator's inputs (signal or threshold) causes a deviation of threshold crossing time that can not be recovered afterwards. However, it can be reduced by increasing the rise time of the amplifier, as jitter is proportional to it, and by reduction of the input capacitance of the comparator, so that a deviation of the baseline leads quicker to a response of the comparator. An increased signal is beneficial for jitter too, but not always feasible.

3.5 Radiation damage in HV-CMOS detectors

HV-CMOS sensors detect high energetic charged particles, in other words hazardous radiation. Depending on the particle rate and the lifetime of an experiment, the detectors have to withstand a significant dose and fluence.

3.5.1 Bulk damage

Bulk damage predominantly affects the sensor diode. It is caused by non-ionizing energy loss (NIEL) of particles.

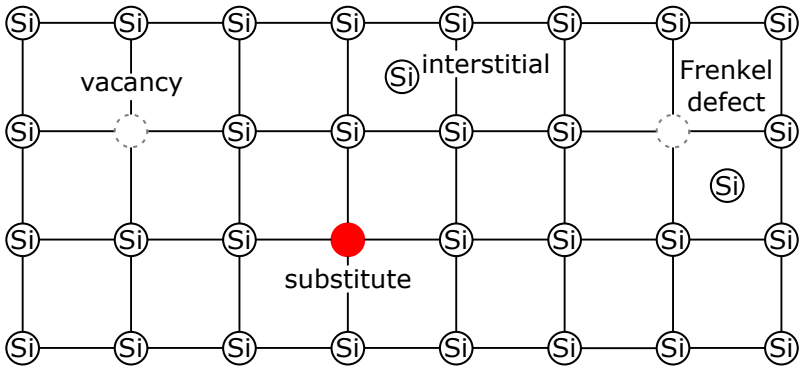


Figure 3.6: Non-ionizing energy transfer from particles to the crystal lattice causes defects in the lattice that can change the electrical properties of the semiconductor. The 2D sketch shows the most common defects: vacancy and interstitial atom. Their localized combination is a Frenkel defect. Sometimes particles are stopped in the lattice and substitute silicon in their place in the lattice or remain as interstitial atom.

The transferred energy from particle to lattice can be large enough to remove atoms from the lattice, leaving a vacancy behind. The free atoms can travel large distances in the material.

They come to rest somewhere in the lattice as interstitial atom. These defects change effective doping, resistivity and thereby the electric properties of the semiconductor [38]. If vacancy and interstitial atom are close to each other, they are called Frenkel defect. A two dimensional representation of the mentioned defects is shown in figure 3.6. Particles might be stopped and remain in the bulk as additional interstitial defects or fill lattice vacancies.

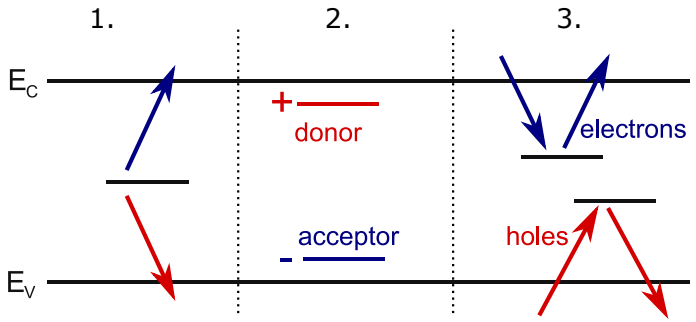


Figure 3.7: Lattice defects produce additional electron states. Their effect on the electrical properties of the semiconductor depends on their energy level: States in the middle of the band cause increased leakage current (1). States close to bands act as acceptor or donor and change the effective doping (2). Deep states trap generated charge long enough to prevent it from contributing to the signal (3). (After [4])

These defects cause electron states between valence and conduction band, which have a number of negative effects on the sensor diode. Figure 3.7 shows the most important effects of bulk defects [4]:

1. States close to the middle between valence and conduction band are responsible for the major part of the sensors leakage current. Shot noise is proportional to leakage current.
2. States close to valence or conduction band, are similar to volitional impurities and change the effective doping. Consequently, the depletion zone can change. Even type inversion of n-type substrate to p-type is possible.
3. Charge, which has been generated by particles, can be trapped in deep states long enough to withhold them from participation in the signal.

Some of these defects can heal over time, especially Frenkel defects. This process, called annealing, can be accelerated by moderate heating. Low temperature prohibits annealing.

3.5.2 Surface damage

While bulk damage affects signal generation in the sensor diode, ionizing damage affects the CMOS electronics. Charged particles leave a trace of electron hole pairs in the semiconductor. What is desired in the sensor region, can become a problem for the electronics: Charge carriers in the bulk are mobile and are dragged to the electrodes to be neutralized. Electron-hole pairs in the silicon oxide are not that free and remain longer in the oxide if they get separated and can not recombine. The electrons are more mobile and will eventually leave the oxide. The mobility of holes is limited, as they tunnel between rare defects, before eventually getting trapped in the Si-SiO^2 interface. Over time, the oxide gets positively charged. This changes the characteristics of MOSFETs, as the effective gate voltage is altered by the positive charge in the gate oxide. After ionizing irradiation, PMOS transistors require a higher voltage to start conduction. This is an issue, but can be respected in circuit design or corrected by a larger gate voltage. NMOS transistor, however, can not be switched off anymore, which can not be corrected, as negative gate voltages are not foreseen. An effective countermeasure is to keep the silicon oxide layer as thin as

5 nm, as this is a distance that holes can cross by tunneling [39]. This is only possible in some technologies and for limited voltages. Additionally to the threshold shift, the charged oxide causes an increased leakage current in either type of MOSFET.

The second type of surface damage caused by irradiation is the activation of interface traps. The interface between silicon oxide and pure silicon is not perfect, as the lattices do not match. The result is a large number of dangling bonds, the interface traps. During production of silicon devices, they are passivated by hydrogen atoms, which pair with the dangling bonds. However, they are not strongly bound to the lattice and are easy to be removed by irradiation, activating the interface traps. They cause increased threshold voltage and flicker noise.

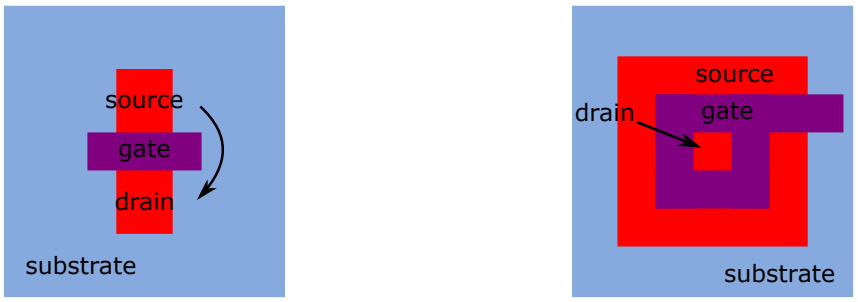


Figure 3.8: Layout of linear NMOS transistor (a) and annular NMOS transistor (b). While leakage current can bypass the gate of a linear NMOS via radiation induced oxide charge, this current is blocked by an annular gate.

PMOS transistors are considered radiation tolerant as ionizing radiation charges the gate oxide, which increases their threshold voltage. NMOS transistors on the other hand, suffer quickly from ionizing damage [19]. As the oxide thickness is fixed for a chosen process, other ways have to be found to make HV-CMOS MAPS sufficiently radiation tolerant. One way is to use PMOS transistors, but they can not replace all NMOS functionality. Sensitive circuits that depend on NMOS, can be equipped with annular, instead of linear transistors. Figure 3.8 shows the layout of a standard linear NMOS (a) and a radiation hard NMOS with annular gate (b). The linear transistor allows for radiation intensified leakage current (arrow). The annular gate blocks this path. However, these transistors are much larger and have a reduced resistance in linear region and a large transconductance in saturation region, as their width-to-length ratio can not be chose freely and is always large. Annular transistors are also called circular or enclosed-gate transistor.

Part II

H35Demo – a reticle size HV-CMOS sensor for ATLAS ITk

This section presents the H35Demo (Figure 3.9) detector ASIC and its characterization. It is the first full reticle size HV-CMOS sensor with monolithic readout developed for ATLAS ITk. Being the successor to a series of smaller sensors, it combines a number of carefully selected design features, which have been evaluated in the characterization of smaller test chips and comes with novel circuitry and approaches.

Chapter 4 outlines the target experiment of the H35Demo, the ATLAS ITk in the context of the High-Luminosity upgrade of the Large Hadron Collider at CERN.

Chapter 5 introduces the H35Demo and its most important design features. Besides the size, the H35Demo features monolithic readout and a novel digitization stage for time-walk suppression, the time-walk compensated comparator.

In chapter 6, the characterization setup for the H35Demo is discussed. It was developed by the author and consists of tailored hardware, software and firmware.

The largest chapter in this section is chapter 7: It summarizes the most important steps in the characterization of H35Demo by the author. Besides the overall functionality, special focus is put on signal variations of the pixels in the matrix and on temporal inhomogeneity effects. Compensation mechanisms for both are explored.

The Belle II experiment at KEK, Japan, is sketched in chapter 8. In the context of this experiment, three H35Demo sensors on a single piece of silicon have been produced, to evaluate the feasibility of a multi-reticle HV-CMOS sensor prototype on selectively thinned substrate, which allows for a reduction of material budget by omittance of support structures, even for very thin sensors. This H35Demo module is being presented in chapter 9.

The most important findings and their interpretation are summarized in chapter 10.

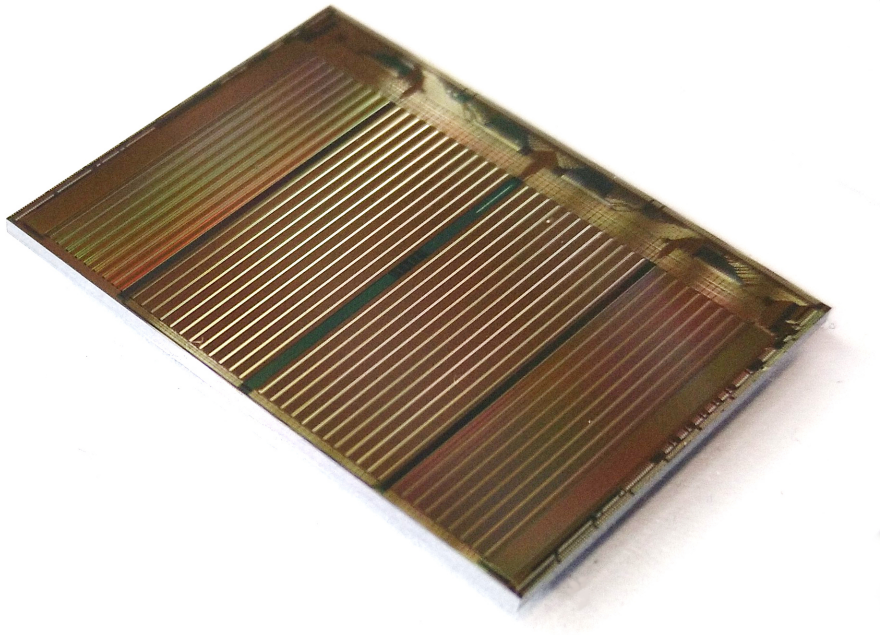


Figure 3.9: Photograph of H35Demo.

4. The Large Hadron Collider and its experiments

4.1 The Large Hadron Collider (LHC)

The Large Hadron Collider (LHC) is the largest particle accelerator built to date [40]. It is a storage ring of nearly 27 km length, designed to collide two beams running in opposite directions. The beams consist of protons or heavy-ions (e.g. lead), which collide at a center of mass energy of 14 TeV or even 1148 TeV. To reach these energies, protons are first accelerated in a linear accelerator (LINAC 2) and three circular pre-accelerators: Proton Synchroton Booster (PS BOOSTER), Proton Synchroton (PS) and Super Proton Synchroton (SPS). In heavy-ion mode, the first two acceleration stages are LINAC 3 and the Low Energy Ion Ring (LEIR). The later acceleration stages are shared with the proton mode.

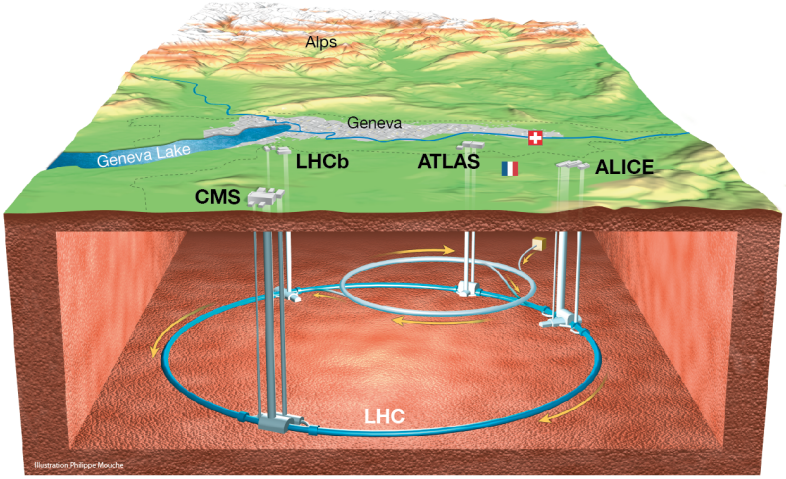


Figure 4.1: Artistic representation of the main parts of the Large Hadron Collider and its main Experiments in the vicinity of Geneva. (From [41])

LHC is located in the underground of Switzerland and France, close to Geneva, on the property of the European Organization for Nuclear Research (CERN). Figure 4.1 shows the LHC storage ring, which is upto 175 m under ground with several access tunnels.

The opposite running beams are brought to collision in four points of the storage ring. In these collision points, the main experiments are located: two general purpose detectors ATLAS [42] (A Toroidal LHC ApparatuS) and CMS [43] (Compact Muon Solenoid) and two smaller, specialized detector experiments ALICE [44] (A Large Ion Collider Experiment) and LHCb [45] (Large Hadron Collider beauty).

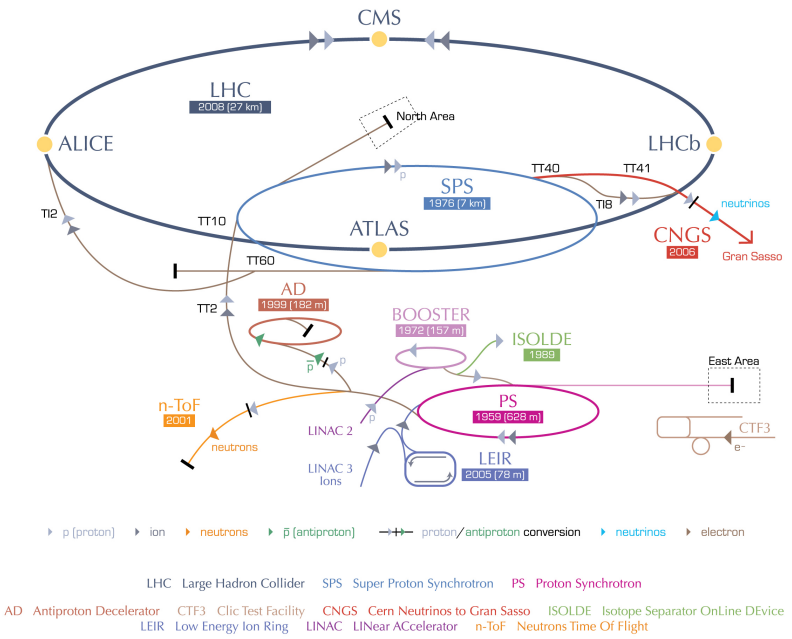


Figure 4.2: Schematic drawing of LHC. Most of the pre-accelerators are actually the main accelerators of previous experiments. (From [46])

A schematic drawing of the accelerator and its associated experiments is shown in figure 4.2. In this configuration, the LHC was operating from 2010 to 2018. The nominal instantaneous luminosity of 10^{34} cm^{-2}s^{-1} [47] was reached in 2016 and has been doubled, before the most recent shut down in 2018.

Luminosity L is – beside particle energy – the most important key number of an accelerator experiment. For LHC it is the number of particles per area and time, that actually interact:

$$L = \frac{1}{\sigma_{\mathrm{cs}}} \frac{\mathrm{d}N}{\mathrm{d}t} \tag{4.1}$$

Where σ_{cs} is the cross section of the interaction, N is the event rate and t is the time. By integration over run time, the integrated luminosity L_{int} is calculated:

$$L_{\mathrm{int}} = \int L \, \mathrm{d}t \tag{4.2}$$

The common unit for the integrated luminosity is inverse femto barn (fb^{-1}, 1 barn = 10^{-28}m^2).

The development of the instantaneous and integrated luminosity of LHC is shown in figure 4.3. The High-Luminosity upgrade after Long Shutdown 3 (LS3), is expected to increase the luminosity to $7.5 \cdot 10^{34}$ cm^{-2}s^{-1} (see chapter 4.2).

LHC operates in a pulsed mode, which means that particles are injected as bunches into the storage ring [47]. During a bunch crossing, the event rate is high, the time in between passes without event and is used to readout data from the experiments. In proton mode, the distance between two of the 2808 bunches, with about $1.5 \cdot 10^{11}$ protons each, is about 9.4 m. On average, 32 collisions happen per bunch crossing[1], this is called pile-up, as the individual events can not be disentangled by time stamping, but only by track reconstruction.

[1] On average Run1 had a pile-up of 25, Run2 had 32 and for short periods even a pile-up of 50 was recorded. [48]

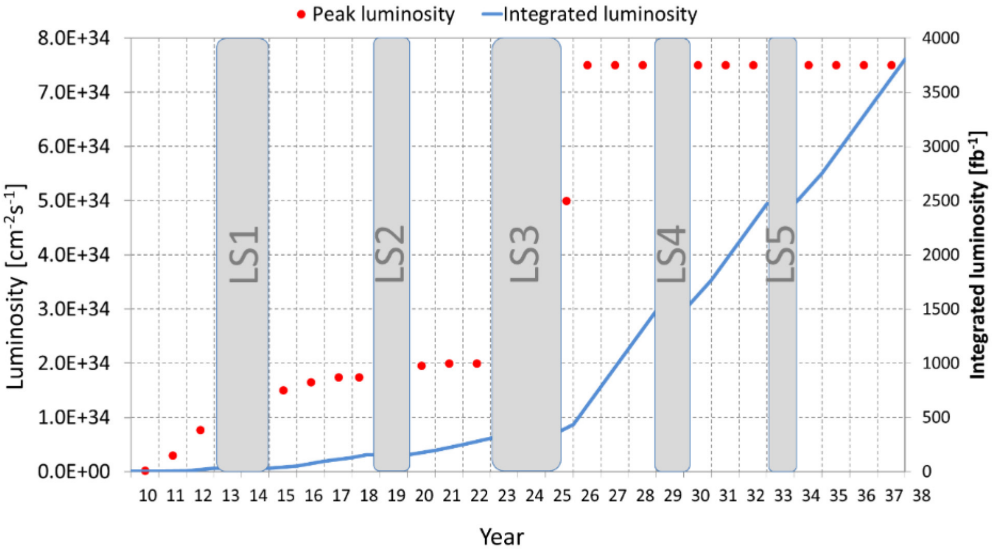


Figure 4.3: Instantaneous (red) and integrated luminosity (blue) as projected. The long shut downs (LS) are used for maintenance and upgrade of LHC and its experiments. LS3 will be used to implement the High-Luminosity upgrade. (From [47])

4.2 The High-Luminosity Upgrade of Large Hadron Collider

The main goal for the High-Luminosity (Hi-Lumi) upgrade of LHC is the increase of instantaneous luminosity, which at the same time increases the number of events that can be recorded per time. Thus, rare processes can be recorded in larger number and their statistical evidence is expected to rise. Without upgrade, the statistical gain of running the LHC and its experiments for another ten years past LS3, is only halving errors [47].

Prior to the upgrade, about 1000 fb^{-1} of integrated luminosity have been recorded. After the upgrade, this value can be more than tripled to almost 4000 fb^{-1} by 2037. This is achieved by increasing the instantaneous luminosity from $2 \cdot 10^{34}$ $\text{cm}^{-2}\text{s}^{-1}$ to $5 \cdot 10^{34}$ $\text{cm}^{-2}\text{s}^{-1}$ or even $7.5 \cdot 10^{34}$ $\text{cm}^{-2}\text{s}^{-1}$.

The Hi-Lumi upgrade will not only allow high-precision measurements of already known processes within the Standard Model [49], but will allow for the search for yet unknown particles, e.g. dark matter particles [50], new Higgs bosons [51] and in general physics beyond the standard model, such as supersymmetry.

The increase in luminosity is predominantly achieved by the change of two beam parameters: The number of protons per bunch is to be increased from $1.5 \cdot 10^{11}$ by almost 50% to $2.2 \cdot 10^{11}$. This is only possible to a certain extent, without reduction of the number of bunches, as the beam current is limited by the entire design of LHC (magnets, cooling and more).
The second parameter is the size of each bunch in the collision point. The emittance, which describes the beam size in position-momentum phase space, is to be reduced from 3.75 µm by one third to 2.5 µm. This is achieved by improvements in all stages of the acceleration process and the implementation of stronger and better focusing magnets at the interaction point. Additionally, the usage of crab-cavities is planned, which rotate the ellipsoidal bunches to reach maximum overlap in the collision point.

The bunch-crossing repetition rate is 40 MHz, which means a bunch spacing of 25 ns. In deviation of the baseline operational frequency, a mode with 50 ns bunch spacing and a higher number of protons per bunch ($3.5 \cdot 10^{11}$) is developed as an alternative in case of problems with the baseline approach.

The Hi-Lumi upgrade does not foresee a change of the 14 TeV beam energy.

The increased luminosity is not only a challenge for the LHC but also for its experiments, which in their current design are not suited to deal with the increased event rate. After the upgrade, not only the decay products of 27 piled-up events have to be recorded and read out per bunch crossing, but the products of 198 (25 ns bunch spacing mode) or even 454 (50 ns bunch spacing mode) piled-up events have to be dealt with. The required changes to ATLAS detector are being addressed in the next section, with focus on the inner tracker (ATLAS ITk). H35Demo and ATLASpix1 are developed within the HV-CMOS proposal for the outermost layer of ATLAS' inner tracker. HV-CMOSis expected to be able to deal with the increased event rate and event pile up, while the material budget, the costs and radiation tolerance is maintained.

4.3 The ATLAS detector

The ATLAS detector is one of two general-purpose experiments at LHC[2]. It features sub-detectors in order to identify all expected particles (with the exception of neutrinos), which are arranged similar to the skins of an onion symmetrically around the interaction point. The tracking detectors are within a magnetic field that bends the trajectory of charged particles. Together with the energy information gained in the calorimeters, the reconstructed trajectories are used to determine particle type, trace them back to a collision vertex and finally reconstruct the entire event.

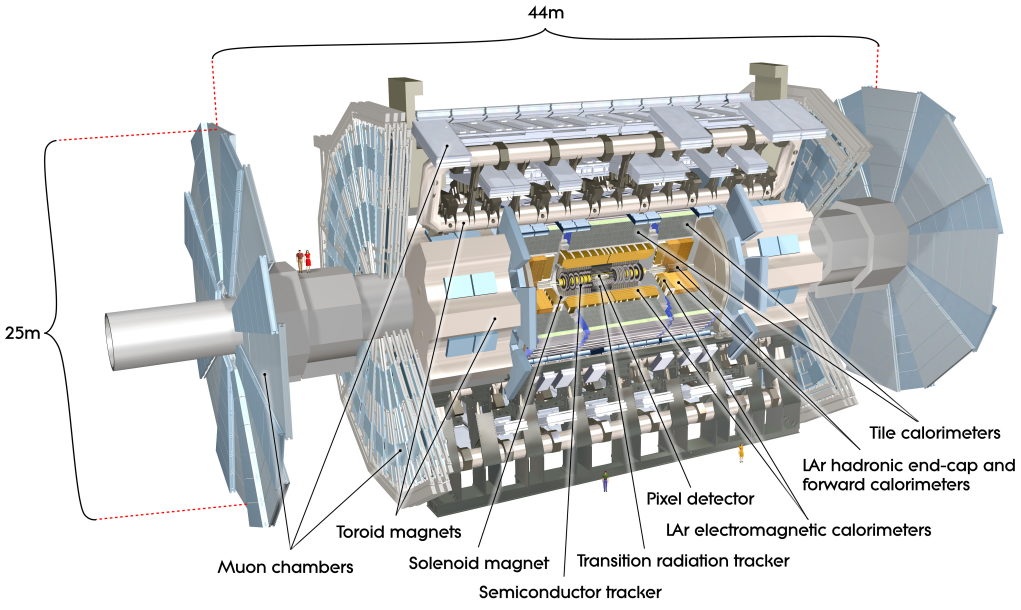


Figure 4.4: Schematic drawing of ATLAS Detector. For size comparison several people are displayed standing next and on the detector. (From [52])

Figure 4.4 shows a design drawing of ATLAS, with some components removed to allow for a view inside. For size comparison several persons are shown on and next to the 25 m tall detector. The main components, listed from the interaction point outward, are the inner detector for tracking, electromagnetic calorimeter and muon spectrometer. Additionally, the magnet system is partially inside the detector.

The inner detector of ATLAS is composed of several sensor layers (Figure 4.5) [54]. In general, the sensors closer to the center have to have a better spatial and time resolution

[2]The second general-purpose experiment CMS and the two other large experiments are shown in Appendix D.3

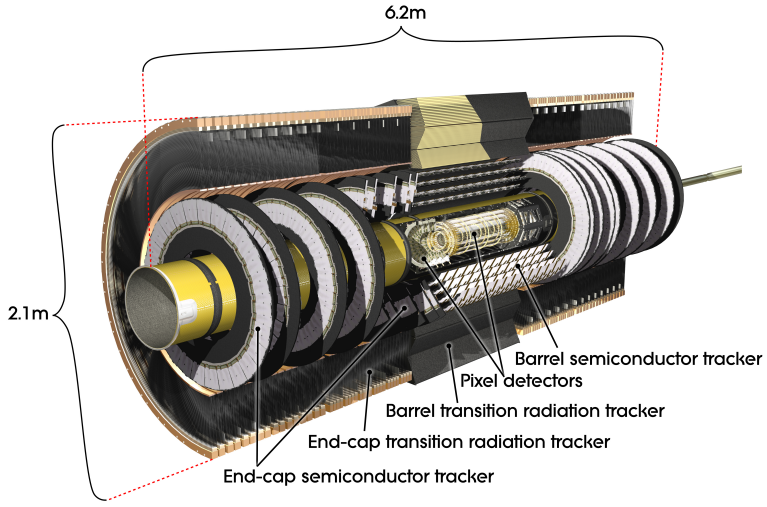


Figure 4.5: Schematic drawing of the ATLAS Inner Detector. (From [53])

to disentangle the traces of individual particles from each other. At the same time, these sensors have to withstand a larger lifetime dose and fluence, than sensors in outer layers, while the material budget is very limited to reduce scattering to the minimum. This limits the allowed heat dissipation as well, as cooling structures require space and add to the material budget.

The current baseline technology for the pixel detector is hybrid detectors (see chapter 2.1). They are arranged in a barrels around the interaction point with increasing radii: 50.5 mm, 88.5 mm and 122.5 mm. Additionally a fourth layer has been added in 2015 during LS1 at a radius of 34 mm [55]. These barrels are 'closed' by four end-cap disks per side in a distance of 400, 495, 580 and 650 mm from the interaction point. The pixels of the barrels and end-caps are 50×300 μm^2 large.

The pixel detector is embedded in the Semiconductor Tracker (SCT). Its silicon strips are 128 mm long and have a pitch of 80 µm. They are arranged similarly to the pixel layers in four barrels and nine end-cap disks per side.

The outermost component of the inner detector is the Transition Radiation Tracker (TRT). Transition radiation appears, if high energetic particles pass from a medium with a certain relative permittivity to a medium with a different permittivity. The TRT consists of xenon filled straws with 4 mm in diameter and up to 150 cm length. Between the straw and the sense wire in its center, an electrical potential drags free electrons out of the gas. Free electrons are created when charged particles or photons interact with and ionize the otherwise neutral gas. The 370 000 straws are densely packed, so that each particle track typically traverses 36 straws. The remaining space between the straws is filled with polymer fibers in the barrel and polymer sheets in the end-cap region to create transition radiation proportional to

$$\gamma = \frac{E}{mc^2}. \tag{4.3}$$

Lighter particles, e.g. electrons, deposit more energy in the polymer, than heavier particles of the same total energy, which is determined in the calorimeters. The deposited energy generates photons that contribute to the signal of the straw detectors nearby. Therefore, larger signals are identified as being of electron origin.

In order to identify particles, not only their trajectory but also their total energy is to be determined [57, 58]. For this purpose ATLAS has two calorimeters (Figure 4.6), the

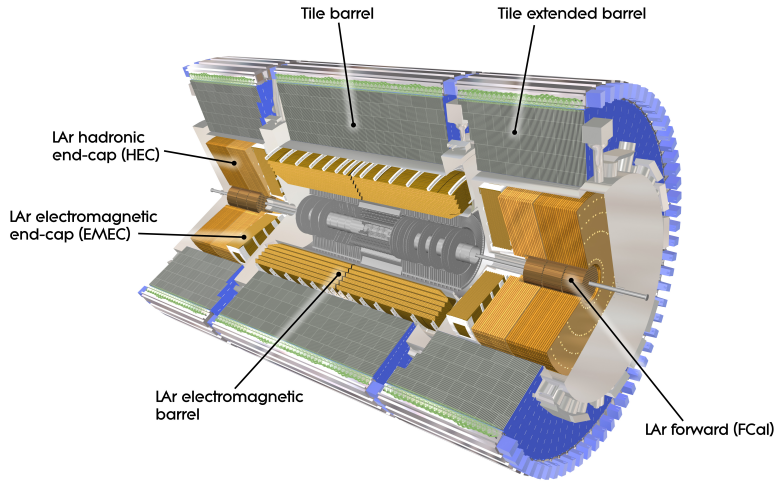


Figure 4.6: Schematic drawing of the ATLAS calorimeters. (From [56])

Electromagnetic (EMCal) and the Hadronic Calorimeter (HCal), arranged in two closed barrels around the inner detector and parts of the magnetic system. Both are sample calorimeters, which means that sensitive and absorbing volumes are placed alternating in the path of particles. The more layers a particle can penetrate, the more energy it originally had. The difference between EMCal and HCal is the used absorbing material: The inner EMCal uses lead, the outer HCal uses stainless steel. The sensing material is Liquid Argon (LAr), which has to be cooled down to 88 K. This is quite challenging in the proximity of components with significant heat dissipation.

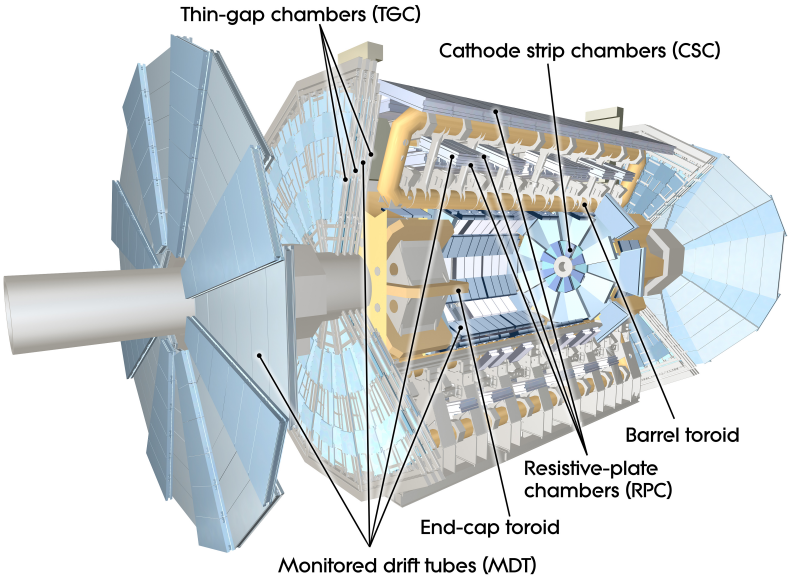


Figure 4.7: Schematic drawing of the ATLAS Muon Calorimeter. (From [59])

Virtually no particles make it past the calorimeter, with the exception of muons and neutrinos. Detection of neutrinos is not feasible, as they are only weakly interacting. Muons however, can be identified by the mere fact that their tracks do not end in the calorimeters, but continue outside. The Muon Spectrometer is the outermost layer of ATLAS [60], which at the same time makes it the largest component of ATLAS. It is within the magnetic field bending tracks of emerging muons for momentum and charge

measurement. The sensing elements are Monitored Drift Tubes (MDT), Cathode Strip Chambers (CSC) and additionally for triggering Resistive Plate Chambers (RPC) in the barrel region and Thin Gap Chambers (TGC) in the end-cap region.

As already mentioned is a strong magnetic field crucial for the function principle of ATLAS [61]. The magnetic field for the Inner Detector is generated by a solenoid magnet. It has a flux density of 2 T and is orientated in parallel to the beam line. The field for the Muon Spectrometer is generated by eight toroids and has a flux density of up to 1 T.

The luminosity of ATLAS is measured by four detectors in a distance of 17 m (LUCID) and 240 m (ALFA) to the interaction point around the beam pipe [62].

The three main detector components produce an enormous amount of data. Approximately 25 MB of raw data are recorded per bunch crossing [63]. At LHC's bunch crossing rate of 40 MHz, this sums up to 1 petabyte of data. Too much to be stored. Therefore, the data rate is reduced by zero suppression, which reduces the 25 MB per bunch crossing to 1.6 MB. Further reduction is achieved by triggering [64]. The ATLAS trigger system selects events of interest and thereby reduces the event rate from 40 MHz to 1 kHz. It uses coarse grain information of the calorimeters (level-1 hardware trigger) and online partial track reconstruction (high level software trigger). The data of the remaining events are actually stored for offline analysis.

4.4 The High-Luminosity Upgrade of ATLAS Inner Detector

The Hi-Lumi upgrade of the Large Hadron Collider (see chapter 4.2), requires an upgrade of the detectors and their components. However, the detectors are not only adapted to the higher luminosity, meaning higher event rate and radiation dose, but the upgrade is used to implement more substantial improvements. The Inner Detector needs to be replaced during LS3, as it will have reached its projected lifetime dose and fluence. The replacement is going to be the ATLAS Inner Tracker (ATLAS ITk) [65].

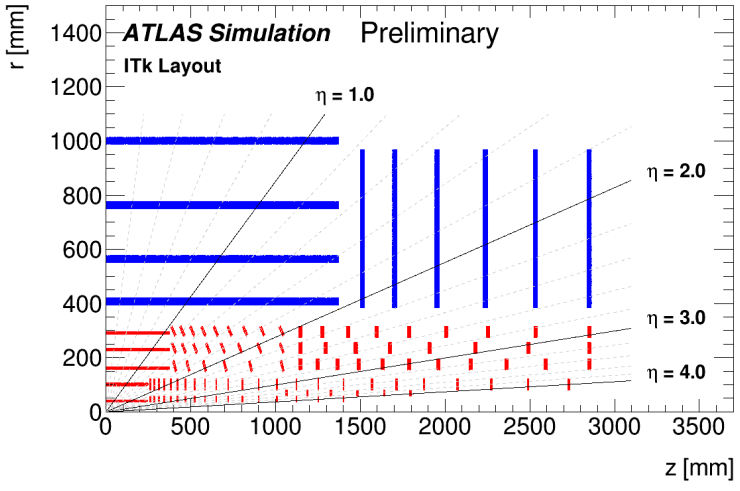


Figure 4.8: Schematic drawing of a quarter of ATLAS ITk as projected for the Hi-Lumi upgrade. The distance to the interaction point parallel to the beam line is labeled z, the distance in radial direction is labeled r. The five pixel layers are drawn in red, the strip layers in blue. ITk is going to be an all-silicon tracker. (From [65])

Figure 4.8 shows ATLAS ITk as currently projected. The number of pixel layers is going to be increased to five and all other sensing detector components are going to be replaced by silicon strips, as TRT would not be able to cope with increased occupancy.

The baseline technology for the pixel layers are hybrid-detectors which have already been used in ATLAS prior to the upgrade. HV-CMOS sensors are proposed for the implementation in ITk's outermost pixel layer. This layer has the same area as all inner layers combined. Therefore, one of the strengths of HV-CMOS sensors – the cost effectiveness – show to advantage, while radiation damage is limited [66], as the interaction point is at a moderate distance of 291 mm (barrel). H35Demo and ATLASpix1 are two sensors designed to demonstrate that HV-CMOS Monolithic Active Pixel Sensors are an option for this layer.

The requirements for sensors in the outermost pixel layer are still challenging [65, 55]:

Pixel geometry The pixel size has to shrink from 50×300 μm^2 to 50×50 μm^2 or 25×100 μm^2. This is no problem per se, but it increases the number of interconnections on a sensor and the data rate.

Efficiency A high detection efficiency of 99% has to be maintained even at the end of the detector's lifetime, while the noise rate per pixel can not exceed 40 Hz.

Time resolution and response time Additional to the plain detection efficiency, events have to be detected and processed within a moderate delay time ($\mathcal{O}(5\mu s)$) and need to be precisely appointed to a single bunch crossing. As the time between two bunch crossings is 25 ns, the time resolution of the sensor has to be[3]

$$\frac{25 \text{ ns}}{\sqrt{12}} = 7.2 \text{ ns} \tag{4.4}$$

Event and data rate The trigger rate for the outermost layer is simulated to be 4 MHz. The event data have to be stored on chip until a trigger decision is met and are then transmitted via a single gigabit link.

Radiation tolerance According to simulations, all components of the outermost layer have to deal with a lifetime dose of 1.6 to 3.5 MGy and a fluence of 28 to $38 \cdot 10^{14}\, n_{eq}/cm^2$.

Power consumption Power consumption is limited by cooling capability, which is $0.7\, W/cm^2$. This power has to be shared by all components, such as sensor ($0.1\, W/cm^2$) and readout chip ($0.5\, W/cm^2$). In case of a monolithic approach $0.8\, W/cm^2$ are available.

[3] The formula is deduced in Appendix A.5 for the resolution of a pixel and applies just as well to a time bin.

5. The H35Demo ASIC

The H35Demo was the first full reticle size monolithic sensor designed in HV-CMOS technology. It is a prototype for the outer most pixel layer of the inner tracker (ITk) of the ATLAS detector for the high-luminosity upgrade. It has been designed and its production was financed by KIT in close collaboration with the universities of Bern, Geneva and Liverpool, as well as IFAE Barcelona and CERN. This ASIC has been fabricated by *Austria Micro Systems AG* (*ams* AG)[1] in their 350 nm high-voltage process H35.

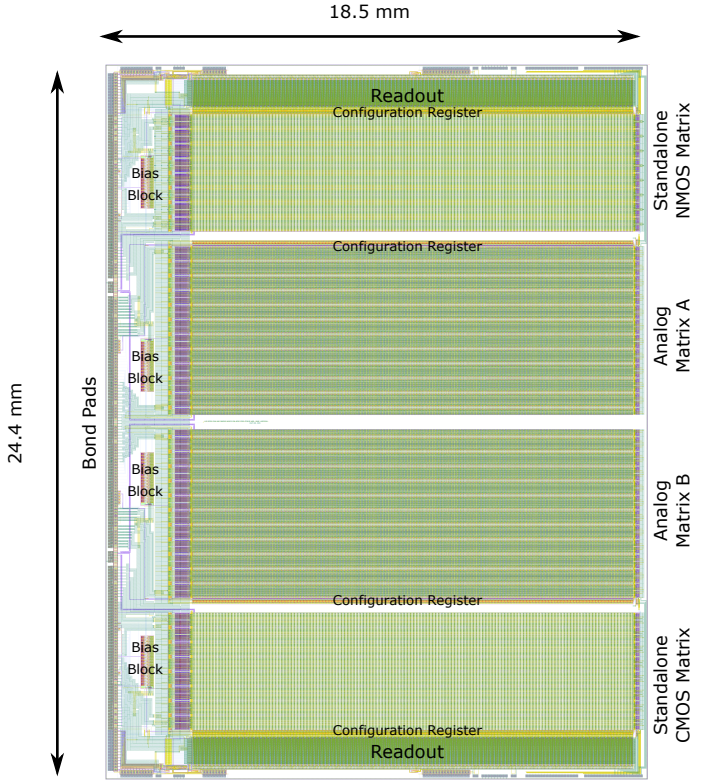


Figure 5.1: The layout of the H35Demo comprises of four independent matrices. Listing from top to bottom these are: a standalone matrix with in-pixel NMOS comparators, two analog matrices for CCPD readout and another standalone matrix with CMOS comparators in the digital periphery. Each matrix has submatrices with pixels of different flavor.

As the name H35Demo(-nstrator) indicates, this chip has been designed to demonstrate the capabilities of HV-CMOS sensors. The first goal is to prove that HV-CMOS-sensors work on the scale of a reticle, not only on the scale of small test chips ($\mathcal{O}\left(5 \times 5\ \mathrm{mm}^2\right)$). Secondly, the interconnection of already existing and in-use readout ASICs, e.g. FE-I4 [67], with HV-CMOS-sensors by capacitive coupling is to be demonstrated. The last functionality to

[1] `https://ams.com/`

be demonstrated is the monolithic or standalone readout. Basic monolithic readout has already been proven functional on the HVStripV1 [19, 68], however the monolithic readout scheme implemented in the H35Demo is much more sophisticated than previous versions.

The total area is 18.5×24.4 mm^2, which is the maximum area possible in *ams* H35 process. The chip has been produced on substrates with different resistivities ranging from the standard 20 Ωcm substrate, over 80 Ωcm and 200 Ωcm, to high resistive substrate with 1000 Ωcm. The H35Demo comprises four fully independent sections. They are not connected to each other, but only to bond pads. Their only connection is the substrate on which they have been produced. In case of a production or design fault in one section, the strict segregation protects the other sections from collateral harm.

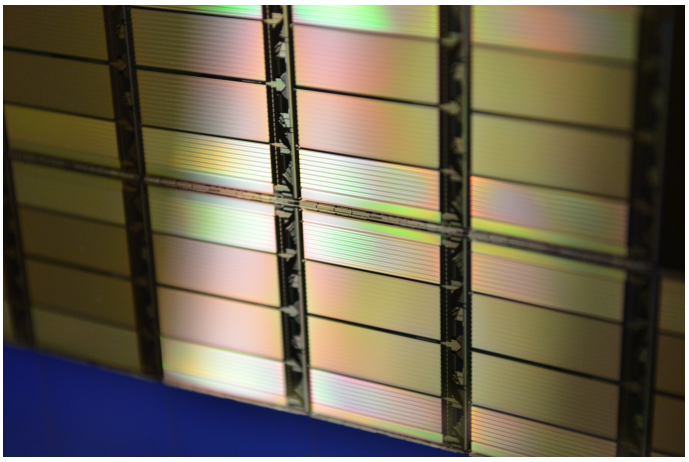


Figure 5.2: The foundry had the sensors produced on 8-inch wafers. The wafers arrived at KIT already diced, but stilled glued on a UV-release membrane. (Photo: R. Blanco)

The H35Demo was designed by a team led by Prof. Ivan Perić. The other main designers are Eva Vilella and Raimon Casanova. They signed the H35Demo (Figure 5.3).

Figure 5.3: The main designers signed the H35Demo between the two analog matrices.

5.1 Functional description

The four sections of the H35Demo are also referred to as matrices. In figure 5.1 the matrices from top to bottom are: Standalone NMOS Matrix, Analog Matrix A, Analog Matrix B and Standalone CMOS Matrix.

All four matrices have similar configurable bias blocks located on the left hand side of the chip. The bonding pads of all four matrices are at the left edge of the chip.

The major part of the chip's area is dedicated to the pixel matrices. The pixel size is 50×250 µm^2. Each pixel has one to three n-well electrodes which form the sensor diode in combination with the p-substrate. The in-pixel electronics is placed inside these n-wells. The front-end electronics is located inside each pixel and is the same for all pixel flavors of all matrices (Figure 5.4).

A charge signal is generated by a depleted diode and amplified by a charge sensitive amplifier with adjustable feedback. Adjacent to the amplifier is a source follower for impedance transformation and signal shaping. The output of the source follower is capacitively coupled to a comparator. It acts as inverter and second stage amplifier. The quasi digitized signal is

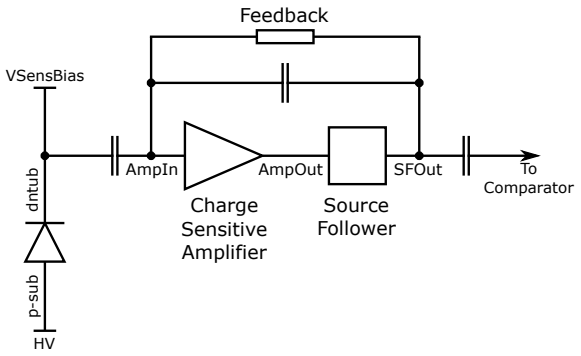


Figure 5.4: Simplified schematic of the front-end on H35Demo. The signal is generated by a reverse biased diode formed between p-substrate and n-well. The charge sensitive amplifier converts the charge into a voltage signal. Its feedback can be implemented with linear or enclosed transistors. The output is capacitively coupled to the adjacent electronics line. For schematic on transistor level see Appendix E.1.

then ready for transmission via monitor lines to the periphery or for capacitive transmission to a readout ASIC.

The baseline pixel layout is altered in some matrices and submatrices. It comprises the pixel diode, a charge sensitive amplifier with PMOS input transistor and regulated cascode, a source follower and a comparator. The feedback loop of the amplifier connects the output of the source follower to the input of the amplifier. The adjustable feedback electronics uses enclosed transistors. The simple comparator is located in-pixel. No deep p-tub (DPTUB) is used. Deviations of this baseline design will be addressed in the detailed descriptions of the four matrices below.

There are one to three monitor line output pads per matrix, which can be connected to single pixels or pixel groups by configuration for debug and detail analysis purposes. Additionally, the Standalone Matrices have the option to use fully monolithic readout. Another debug feature of each matrix is an AmpOut line and pad. It can be connected by configuration to the source follower output of all pixels of either column one or two (depending on the matrix).

The configuration of the chip is set via the configuration shift registers.

For laboratory-testing, the H35Demo has the option to connect single pixels to a injection line. Via this injection line, an externally generated electrical pulse can be guided to a specific pixel, where it injects a well-defined charge. This is very useful for debugging and calibration. Both the amount of the generated charge and its timing are precisely known and thus serves as reference.

All pixels of all matrices can be read out capacitively by gluing a readout ASIC onto them. In this mode only the front-end of H35Demo is used.

Besides the four matrices, there are small independent test structures. These test structures are used to test singular electrical properties of a design or production feature of the chip. Located at the very edge of the layout, pixel-like diodes can be used in edge TCT to determine the degradation of the sensor by NIEL damage introduced by irradiation.

5.1.1 The standalone NMOS Matrix

Although all four matrices have approximately the same area, the Standalone Matrices have a reduced number of pixels in favor of a digital periphery. Figure 5.5 shows the layout of the NMOS Standalone Matrix. The pixels are arranged in 16 rows and 300 columns. The

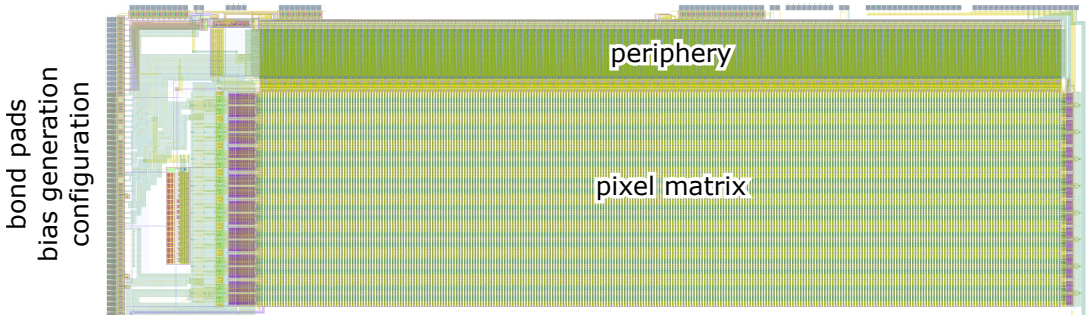


Figure 5.5: Layout of the first standalone matrix with an NMOS comparator in each pixel. Half of the pixels feature a time-walk compensated comparator. The pixel matrix consists of 16 rows and 300 columns. The digitized signals are send to the periphery where the signals are buffered and prepared for readout.

first 150 columns (from left to right) contain a simple NMOS comparator (Figure 5.6) and the second 150 columns contain a time-walk compensated NMOS comparator (Figure 5.7). The time-walk compensated comparator is explained in chapter 5.1.2. The usage of NMOS comparators, converting the analog signal into a rectangular signal with length carrying analog information, gave this matrix its name.

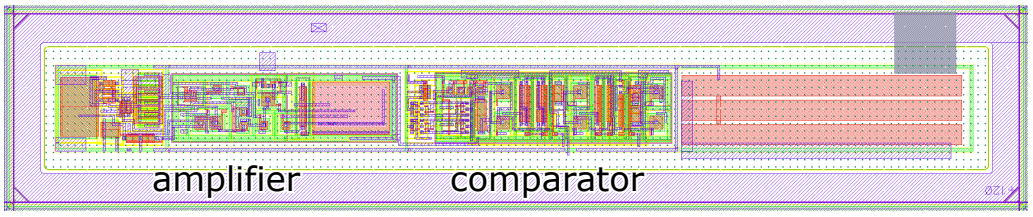


Figure 5.6: Layout of an NMOS Matrix pixel with normal comparator. There are three main parts (from left to right): amplifier, comparator and area without electronics. The whole area is used as collection electrode. The gray square on the top right is the bump-bond or CCPD pad.

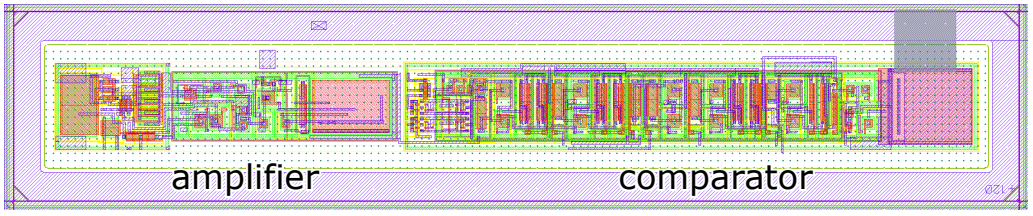


Figure 5.7: Layout of an NMOS Matrix pixel with time-walk compensated comparator. The area of the amplifier (left third) is the same. Compared to a normal comparator larger time-walk compensated comparator takes two thirds of the available space.

The term 'Standalone' indicates that no dedicated readout ASIC (e.g. FrontEnd-I3 [69]) is needed for full digital readout. This functionality has been implemented on-chip in the digital periphery.

The quasi digitized signal from a hit pixel is transmitted to the digital periphery. For each pixel, there is a Read Out Cell (ROC). A comparator converts the low voltage signal from the pixel into a CMOS signal. The output line and corresponding bond pad are called HitBus, which can be probed. All ROCs can be connected to this HitBus line via the configuration registers. For monolithic readout, the time of arrival is stored as 8-bit time stamp with 40 MHz or 80 MHz precision. Together with the pixel's row address (8-bit) this information is stored in a buffer cell.

Forty ROCs share an End-Of-Column (EOC) cell. The EOC scans the ROCs for hits by priority chain logic and reads them from the ROC buffer into the EOC buffer. There are 30 EOCs on the left hand side and 30 EOCs on the right hand side. The left and right EOCs are connected to two independent serializers in a daisy-chain fashion. There are two serializers for each half of the matrix. One for address information and one for timing information. The column address is not explicitly transmitted, and has to be recovered from the position of the EOC buffer in the readout daisy-chain. Each serializer feeds a differential line at 320 MHz. Hence, the information stored in the EOCs is send out to four differential readout lines: Address of the right half, time stamp of the right half, address of the left half, time stamp of the left half. One full readout cycle takes 1.5 µs for 40 EOCs, equal to 60 columns.

As the standalone readout operates synchronously at a high speed, it requires a control unit generating the control signals for all elements of the readout. This control unit is implemented as a finite state machine. The input clock for the state machine is provided by the FPGA.

5.1.2 Time-walk compensated comparator

For many experiments, time resolution is crucial. The intrinsic time resolution of low power HV-CMOS sensors is in the order of 100 ns (see chapter 3.2). This is not sufficient for many experiments. One way to improve time resolution is to replace the standard comparator with a time-walk compensated comparator. The idea behind this feature is to compensate for the time-walk effect by a comparator that reacts slowly to a large input signal and fast to a small input signal. There have been implementations of this idea already on previous chips, such as HVStripV1 and H35CCPDv2 [19]. On the H35Demo an improved version of the time-walk compensated comparator is implemented in the pixels of the right half of the NMOS Matrix.

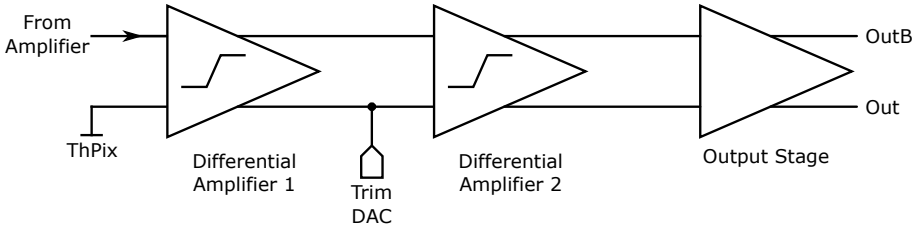


Figure 5.8: The normal comparator of the NMOS Matrix consists of two consecutive differential amplifiers and an output stage. The local effective threshold can be altered by a trimming circuit to compensate for behavioral variations. (Detailed schematic in Appendix E.1)

In general, the time-walk compensated comparator is similar to the normal comparator on the other half of the matrix (Figure 5.8). Both actually consist of two differential amplifiers. The signal from the amplifier *CROut* is used as positive input and the threshold *ThPix* is connected to the negative input. In case of the normal comparator, the differential output of the first stage is passed on to the inputs of the second stage. The effective value of the negative branch can be altered by the trimming circuit. Finally, a switchable output stage connects the resulting signal to the output line and CCPD pad.

The time-walk compensated comparator works slightly different (Figure 5.9). The input to the second stage discriminator is not directly the first stage, but the charge status of a capacitor *C*. If there is no signal, the voltage level of this capacitance is determined by the bias voltages *nVPlusTWPix* and *VNTwDown*. It is constantly charged via transistor *N24* and constantly discharged by *N22* and *N25*. When a signal occurres on *CROut*, it is

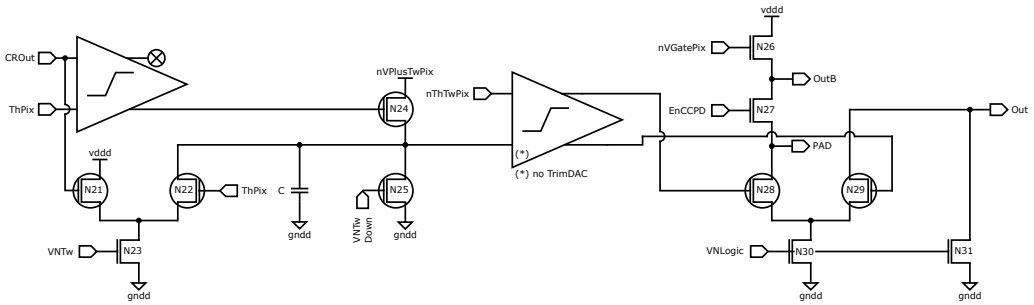


Figure 5.9: A signal in the the time-walk compensated comparator starts the discharge of a capacity. The speed of this discharging is determined by the signal height. A small signal speeds up the discharging and a large signal slows it down. So the behavior is inverse to the time-walk effect, thus compensates for it. Once the voltage level has dropped sufficiently, the second discriminator is triggered. (Schematic from [70], modified)

not only compared in the first differential amplifier to a given threshold, but is also used in an additional circuit. The output of the first differential amplifier cuts off the loading current to C at *N24* and its voltage level starts to drop. How fast the voltage is dropping at C, is determined by the extra circuit. In case of a large signal, the constant current set by *VNTw* at *N23*, flows mainly through *N21* and not through *N22*. This means that C is slowly discharged. For a small input signal, which allows only a reduced flow through *N21*, C discharges fast via *N22*. Once the voltage of C has dropped below *nThTwPix*, the second discriminator is triggered. This means that the temporal behavior of the second stage is the opposite of the time-walk effect, thus compensates for it. The output of the second comparator is connected to an output stage, just as the normal comparator.

5.1.3 The Analog Matrices

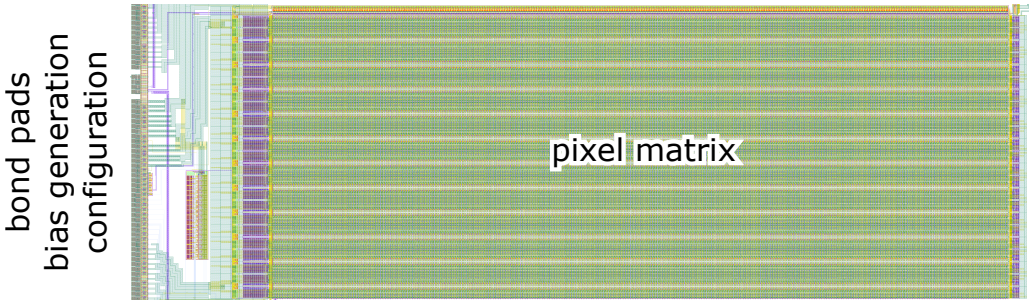


Figure 5.10: Analog Matrix A has no periphery for readout. The major part of the chip is the pixel matrix with 23 rows and 300 columns. The bond pads are located on the left edge of the chip. The space in-between is used for configuration circuits, the bias block and signal routing.

The layout of Analog Matrix A is shown in figure 5.10 and the layout of Analog Matrix B in figure 5.11. Most space of both is occupied by pixels (green), which are arranged in a matrix with 300 columns and 23 rows. Only a small fraction is used for utilities, such as configuration shift register (purple and orange), bias block (square structure in red and green) and pads (at left edge).

There is no standalone readout implemented. For full digital matrix readout a readout chip has to be glued onto the H35Demo.

Analog Matrix B is flipped vertically with respect to Analog Matrix A. The difference between the analog matrices is the pixel flavor:

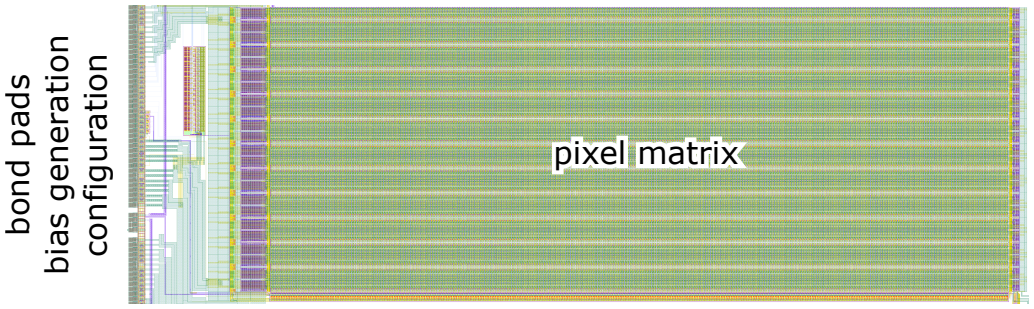


Figure 5.11: Analog Matrix B is mirrored vertically with respect to Analog Matrix A. The overall structure is the same, the difference is the pixel flavor.

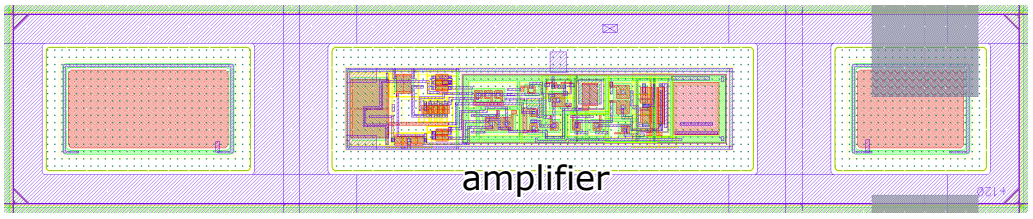


Figure 5.12: The pixels of the analog matrices have a size of 50×250 µm^2. They are divided into three segments. The left and right segments are pure collection electrodes, the middle one contains the pixel electronics: a charge sensitive amplifier. The gray box on the right is the bump-bond and CCPD pad.

Analog Matrix A is divided into three sections of 100 columns times 23 rows each. The first part has an extra DPTUB. The second part uses the baseline pixel design and the third part has linear transistors in the feedback loop instead of enclosed transistors. Linear NMOS transistors are not radiation tolerant, as enclosed ones. However, their gain is higher and they occupy less space.

Analog Matrix B is also divided into three sections. But the flavors of the pixels are partially different. The first group is the only one in this matrix with an extra DPTUB. Group two is realized in baseline design. The last group features pixels of higher speed on the expense of lower gain. This is achieved by an additional capacitance between the output of the source follower and DNTUB.

In contrast to the Standalone Matrices, the analog matrices have three monitor lines, which can be operated in parallel. This allows independent simultaneous monitoring of the analog output of three horizontally neighboring pixels.

5.1.4 The standalone CMOS Matrix

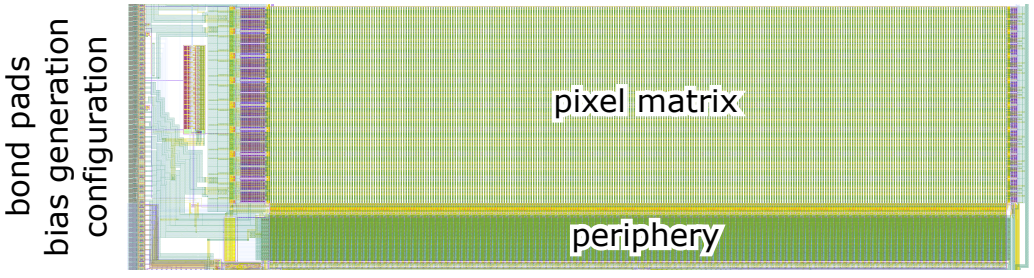


Figure 5.13: The layout of the Standalone CMOS Matrix is vertically mirrored with respect to the Standalone NMOS Matrix. In this part of the H35Demo, CMOS comparators are used in the periphery to digitize the analog signals coming from the pixels.

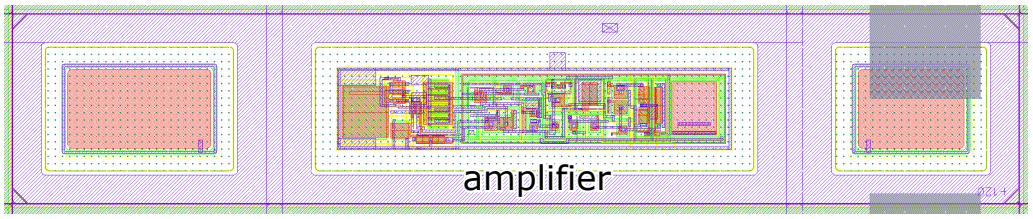


Figure 5.14: The layout of a CMOS pixel is similar to the layout of an analog pixel. All components for digital readout are moved to the periphery. In order to drive the analog signal from pixel to periphery, a second stage amplifier is implemented.

Like the Standalone NMOS Matrix, the Standalone CMOS Matrix consists of a pixel matrix with 16 rows and 300 columns. The major difference is that instead of an in-pixel NMOS comparator, a CMOS comparator located in the periphery is used (Figure 5.15).

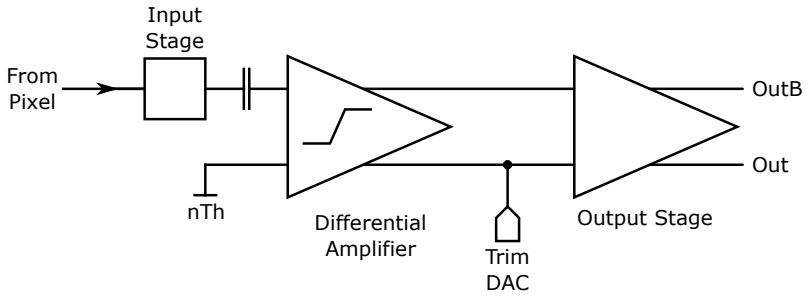


Figure 5.15: The input from the pixels to the CMOS comparator in the periphery of the CMOS Matrix is AC-coupled to a CMOS comparator. The effective threshold can be adjusted by the trimming circuit. (Detailed schematic in Appendix E.1)

Instead of the comparator, each pixel has a second stage amplifier, in order to drive the amplifier's signal to the comparator in the up to 5 mm remote periphery.The monolithic readout works the same way as in the NMOS Matrix. All pixels have a CCPD readout pad for bump bonding or capacitive readout to a readout chip. Each pixel can be connected to the HitBus or monitor line. Further, the signals of the first amplifiers in the second column can be connected to a bonding pad.

5.1.5 The bias block

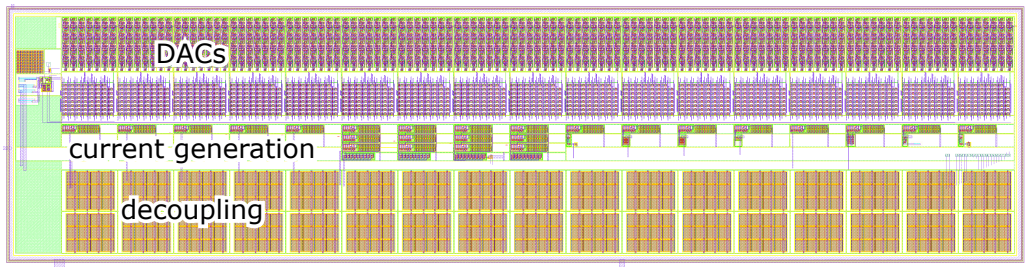


Figure 5.16: The area occupied by the bias block is 465×1800 µm^2 large. Each of these blocks generates 17 bias voltages for one matrix. The DACs can be found in the top third of the layout. The structure in the middle is for current generation. The decoupling capacitors occupy the bottom third. This layout is displayed rotated clockwise by 90° with respect to the H35Demo's layout.

There are four bias blocks on the H35Demo design, one for each matrix (Figure 5.16). Each comprises 17 configurable voltage generators, digital to analog converters (DACs), with 6-bit precision. The generated voltages are used as input to current sources to allow significant current loads. The value of each DAC is stored in a shift register.

5.2 Configuration

The configuration of the H35Demo is based on a long shift register for every matrix. Each bit is realized as a two-phase clock flip flop with attached latch. The shift register is segmented. Each segment is close to the feature it configures.

The first segment, closest to the input pad, is the DAC register. Each DAC register has 7 bits. One of these bits is a spare bit. Some of these spare bits are used for additional configuration.

The second segment is the horizontal control register of the pixel matrix. This register holds 4 bits per row. The first bit marks a row for test signal injection. The second marks a row for connection to the monitor readout line. In order to actually connect a pixel to the injection or monitor line, it needs to be marked both in horizontal and vertical control register. The third bit connects the amplifier of the second pixel of a row to the AmpOut pad. No selection in vertical control register is necessary. The fourth bit is used to load the pixel RAM. It can be interpreted as *write enable* signal. As the analog matrices do not have a RAM, this bit is not used in them.

The third segment is the vertical control register. It is the second part of the matrix configuration, responsible for the columns. Just like the horizontal control register it comprises 4 bits per column. The first one connects the injection line to the input of the output buffer (source follower). This is a characterization feature. Injecting pulses directly into the source follower and measuring its response without additional electronics allows direct calibration of this circuit. The second bit is used to connect the pixels output to the monitor line. The third bit connects the injection input line to the amplifier input of a column. In case of a Standalone Matrix, the third and the fourth bits are used as payload when writing the RAM. In Standalone Matrices bit four is not used.

The fourth segment of the configuration shift register configures the End-Of-Column blocks. Analog Matrices do not have EOC, thus it exists only in the standalone matrices. It is 2 bits per readout column long, the first is activating the HitBus and the second is a *write enable* signal for writing the digital RAM. The values of the first two bits of the column register are interpreted as payload.

6. Characterization setup for H35Demo

Once the H35Demo design has been finished and was submitted to the foundry, the work on a dedicated characterization setup started. Even though extensive simulations and careful design promise good functionality, the characterization of sensors needs to be done on the actual chip. A characterization setup for HV-CMOS sensors comprises three main parts: A printed circuit board (PCB) carrying the chip, an FPGA with an adapted firmware and a computer with matching software.

The H35Demo is mounted and wire bonded on a carrier PCB. It provides power supply, voltage supply (from daughter boards) and supplemental components. Additionally, signal lines connect the H35Demo to an FPGA board. The FPGA configures, and drives the sensor as well as it receives readout data from it. The computer software establishes a connection to the sensor, via the FPGA and controls all functions of the setup.

6.1 Hardware

The carrier PCB has four layers, the inner layers for power supply, the outer layers for signal routing. It has been designed to support not only the H35Demo, but also a readout ASIC glued on top of it, the FE-I4. For use with the NexysVideo[1] it has a FMC-LPC connector. This FPGA development board is used by KIT-ADL. The same signals are routed to the redundant KEL connector, for usage of the General Purpose AnalogCard (GPAC)[2] system of Bonn University.

The H35Demo requires a number of additional bias voltages. Some of them need to be adjusted quickly and frequently during operation, others are set to a fixed value. The latter are defined by a 50 MΩ trimmer between 3.3 V and ground. The ones that need to be set repeatedly for measurement to different, well-defined values, are generated by voltage DACs and adjacent amplifiers. As these components with a reasonable precision are rather expensive, they have been outsourced to two daughter boards. The original version was designed by R. Blanco, later replaced due to noise issues, by the VoltageBoards developed for the Multi-purpose Adapter Board (see chapter 14.1.1). For generation of test signals, an injection circuit is placed on the carrier PCB.

Before the NexysVideo has been introduced, the Uxibo FPGA board[3] (Spartan IIe[4]) was used. It is known to cause a lot of noise on the ground potential and transfer it to the connected characterization system. This issue was tackled by implementing differential line receivers on the PCB converting LVDS signals to single ended CMOS signals, thus no common ground is required. Most input signals of the H35Demo and the used DACs expect this signal standard. Differential data transmission does not need a common ground

[1] https://reference.digilentinc.com/reference/programmable-logic/nexys-video/start
[2] https://basil.readthedocs.io/en/latest/hardware.html
[3] https://uxibo.de
[4] https://www.xilinx.com/support/documentation/data_sheets/ds077.pdf

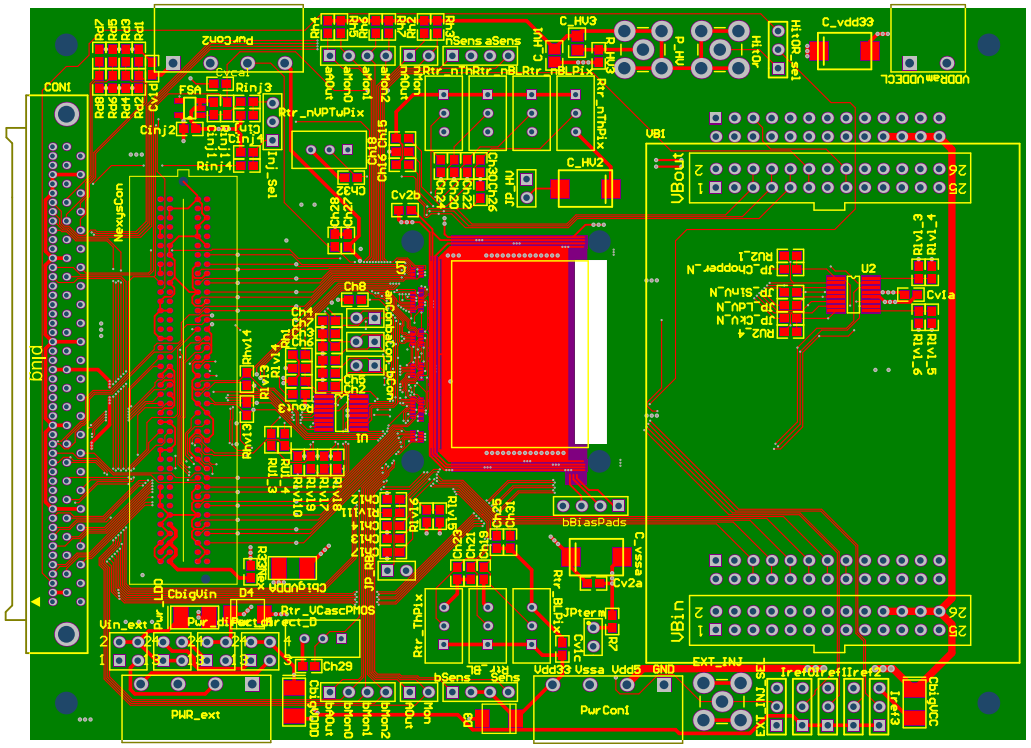


Figure 6.1: Top layout of the H35Demo carrier PCB. The H35Demo shall be glued on the top-copper square in the middle (red), bonded to the left. An FE-I4 can be glued/bump-bonded face down on-top: Its bonds go through the hole to the bottom side. One VoltageBoard can be connected to the box headers on the right. The connectors to the FPGA (KEL or FMC) are located on the left. Upper and lower edge of the PCB are used for power connections and signal probing.

like CMOS, hence noise proliferation can be prohibited. However, it was found that the NexysVideo boards do not cause that kind of noise.

The H35Demo requires several voltage sources with significant current. The respective bond pads are routed to power plugs, to be connected to analog power supplies. Additionally, the depletion voltage is routed to an SMA connector. From there, it can be connected with a shielded cable to an SMU with high voltage capability.

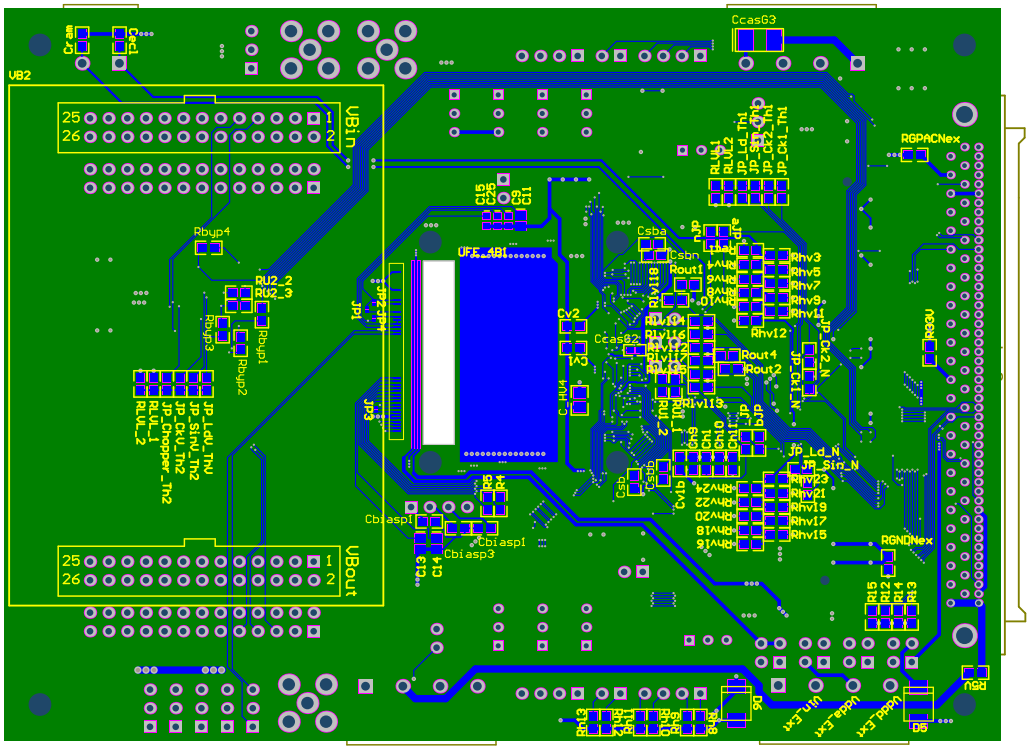


Figure 6.2: Bottom layout of the H35Demo carrier PCB. The bonds of an optional FE-I4 readout ASIC can be guided through the hole in the middle to bond pads on the left. On the far left side a second VoltageBoard can be plugged into box headers.

6.2 Firmware

The device under test (DUT) and all configurable parts of the PCB are controlled and managed by firmware. The firmware runs on an Artrix-7 FPGA, which is mounted on the evaluation board NexysVideo from Digilent. The firmware is written in Verilog using the Vivado IDE Version 2017.2[5]. The communication between computer and FPGA is established through a virtual COM-port via USB. The USB connection on the NexysVideo is provided by an FTDI-chip[6], which exchanges data between USB and UART. The FTDI chip can buffer a limited number of 8-bit data words for each direction.

The simplified block diagram in figure 6.3 shows the working principle of the firmware. Only the *FastReadout* module is H35Demo-specific, all other modules are reusable. Data received from the computer are stored in the FTDI chip and have to be fetched by the FPGA. A Finite State Machine (*FTDI FSM*) has been designed, according to the specifications of the FTDI chip, to handle the data flow. The FSM stores the data into an FPGA FIFO. The *Order Sorter* picks the 1-byte data words from there and interprets them as order-header, -address, -length or -payload for the *Register File*. The underlying protocol has been specifically designed to suite the expected requirements of a characterization setup for large monolithic sensors like the H35Demo. The protocol itself is explained in the next section.

The *Register File* is an accumulation of individually addressable registers. Each register has a dedicated purpose. All FPGA modules receive their settings from these registers within the *Register File*. But also the slow control of the H35Demo and the VoltageBoards are generated here.

[5]https://www.xilinx.com/products/design-tools/vivado.html
[6]`https://www.ftdichip.com/Products/ICs/FT2232H.html`

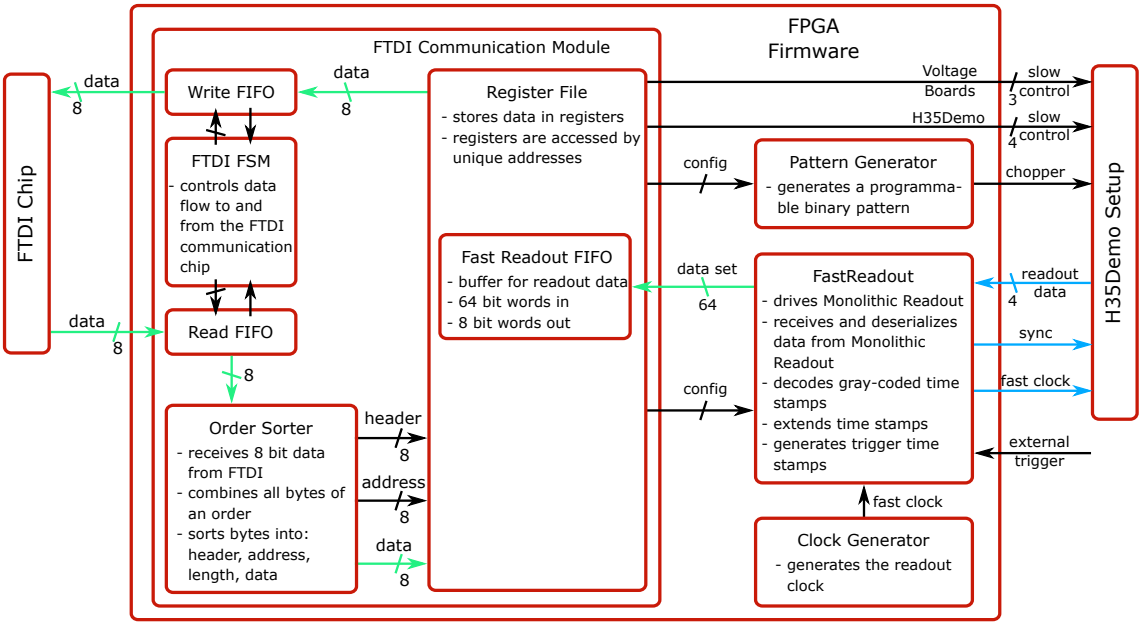


Figure 6.3: Block diagram of the firmware for the H35Demo characterization setup. The only sensor-specific module is the fast-readout module, all other modules are universal. The fast-readout module generates signals for the H35Demo and receives data from its Standalone Matrices. The FTDI communication module manages data transfer between computer and FPGA.

6.2.1 Slow control

While modules can simply read the information stored in their respective register, data transfer to hardware components works differently. The configuration of both H35Demo and VoltageBoards is held in storage registers inside the chip or DACs, which are loaded from a shift register. These shift registers are written by single ended slow control lines (CMOS). The sensor's shift register has a data line (Sin) a load signal (Ld) and a two-phased clock (Ck1, Ck2). The VoltageBoards have only one clock. These configuration signals are connected to registers in the *Register File*. For each bit that is to be written into the H35Demo configuration register, the respective registers in the *Register File* are written five times. At first, the data bit d is set, then the two clocks are pulsed after each other. After the configuration has been completed, the load pulse is sent to transfer the data from the shift register to the permanent storage register.

Consequently, the amount of data transmitted from computer to FPGA is much larger

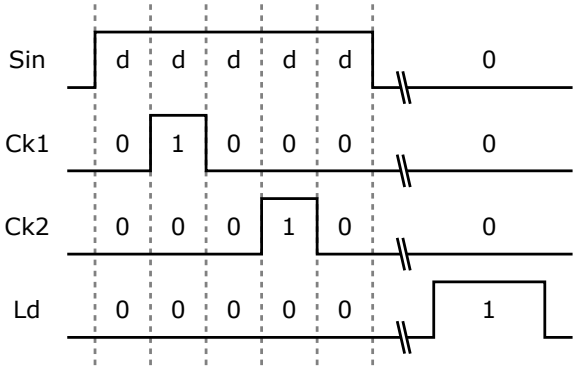


Figure 6.4: The configuration of the H35Demo is based on a two-phase clock shift register. Both data (Sin) and auxiliary signals (Ck1, Ck2, Ld) are generated by software.

than the configuration to be sent from FPGA to H35Demo. For setting one bit in the shift register, five parallel words are needed to be sent from the FPGA, each is sent from the computer as 8-bit word. Transfer speed is slowed down to avoid configuration errors by sending each word n_{div}-times. This clock divider n_{div} is in the order of ten: Therefore, $5 \cdot 8 \cdot n_{div}$ bits have to be sent from the computer to write a single bit into the configuration shift register of the H35Demo.

In principle, the auxiliary signals could be generated in the FPGA as well, however it is much more convenient to have the clock and load generation easily accessible in software. The increased data load on the connection to the computer has no effect on configuration speed, as the used shift registers are slow and the transmission speed to the sensor has to be curbed to work properly.

6.2.2 Pattern generator

The *Pattern Generator*[7] module was taken over from previous projects with the Uxibo Board with only minor changes. Its output is an arbitrary binary pattern, the chopper signal. It is used as input to a charge injection circuit. Important parameters are the pulse length, clocking speed, duty cycle and configuration bits. All parameters can be set by the control software with intermediate register file.

6.2.3 Fast readout module

The *Fast Readout Module* manages the functions necessary to run the Standalone Readout of the NMOS and CMOS Matrix on the H35Demo. Its configuration is handled by the *Register File*.

The state machine in the standalone periphery of the H35Demo is driven by a fast clock. This clock is generated on the FPGA by a PLL from the slower system clock. When the fast-readout is activated, the fast clock is passed on to the sensor as differential signal. In order to synchronize the state machines on the sensor and on the FPGA, a synchronization signal is generated periodically.

The hit data from the H35Demo are transmitted serially via four differential readout lines. An address line and a time stamp line for each left and right half of a Standalone Matrix. First of all, the serial data are stored in parallel registers. The pixels are arranged in 60 groups of 40 pixels. The hit data from the groups are sent one after the other, one data set of each group in each block cycle. After 60 block cycles, one data set of each group has been read out. If a group has not recorded a hit from the pixel matrix, an empty hit is sent.

The hit data consists of an 8-bit address of the hit pixel and an 8-bit time stamp. The address space covers only one group (40 pixels). The information which group has been hit, has to be reconstructed from the position, in the data stream.

The time stamp is Gray-encoded and 8 bits wide. Right after being received, it is decoded. As 8-bit is insufficient for most analyses, it is extended by another 8-bit counter in the *FastReadout* module. Between time stamp generation on the sensor and addition of the extended time stamp up to 500 fastclock cycles can elapse. Therefore, the extended time stamp is corrected for this uncertainty.

The last piece of event information collected is the trigger time stamp. An external trigger can be connected to one of the FPGA's PMOD connectors. When a trigger is received, the time of arrival is stored.

The total hit data set contains 64 bits:

[7] The Pattern Generator was designed by M. Koniczek, University of Mannheim.

Name	Label	Description	Bits
block index	ROBin	According to the time of arrival, the sending readout block is identified.	8
trigger time stamp	TrigTS	The trigger time stamp is temporarily stored, when a trigger was received. It is only read out, if a hit from the H35Demo is received before a new trigger pulse has occurred.	16
extended time stamp	ExtTS	The time stamp counter in the H35Demo is mirrored in the FPGA, but extended to 16 bits. Synchronous operation is ensured by the synchronization signal. The upper 8 bits are used here.	8
time stamp left	TSL	The time stamp received from the left half of the monitored matrix.	8
address left	ADDRL	The pixel address of the left matrix.	8
time stamp right	TSR	The time stamp received from the right half of the monitored matrix.	8
address right	ADDRR	The pixel address of the right matrix.	8

As the left and the right side of the matrix operate independently, a 64-bit hit word can contain a hit in the left half or the right half or in both the left and the right half. It is also possible, that in one data set no hit is recorded at all. In this case, the data set is discarded (zero suppression).

If a data set is complete and not empty, it will be sent out of the *FastReadout* module and stored in a dedicated FIFO inside the *Register File*. The input of this FIFO is 64 bit wide, the output is 8 bit wide to comply with the data transmission protocol.

6.2.4 Reading data from the FPGA

The FPGA does not automatically stream data out to the computer. For data transmission in this direction, a read request has to be sent. This request contains a *Register File* address and the number of bytes to be read from it. A read request to a configuration register has typically length 1. Reading from the *Fast Readout FIFO* is only reasonable when reading blocks of 8 bytes (one entire data set). Several thousand data sets can be buffered and requested at once.

The requested bytes are transferred to the *Write FIFO* and from there passed on to the *FTDI chip*. This transfer is managed by the same *FTDI FSM* that handles the other data direction to avoid data collision.

6.3 Software

Most high level functions are managed by a computer. The control software is written in C++ and makes use of the *Qt* framework[8]. Figure 6.5 displays a simplified block diagram of the software used to control the H35Demo and its setup.

[8] https://www.qt.io/

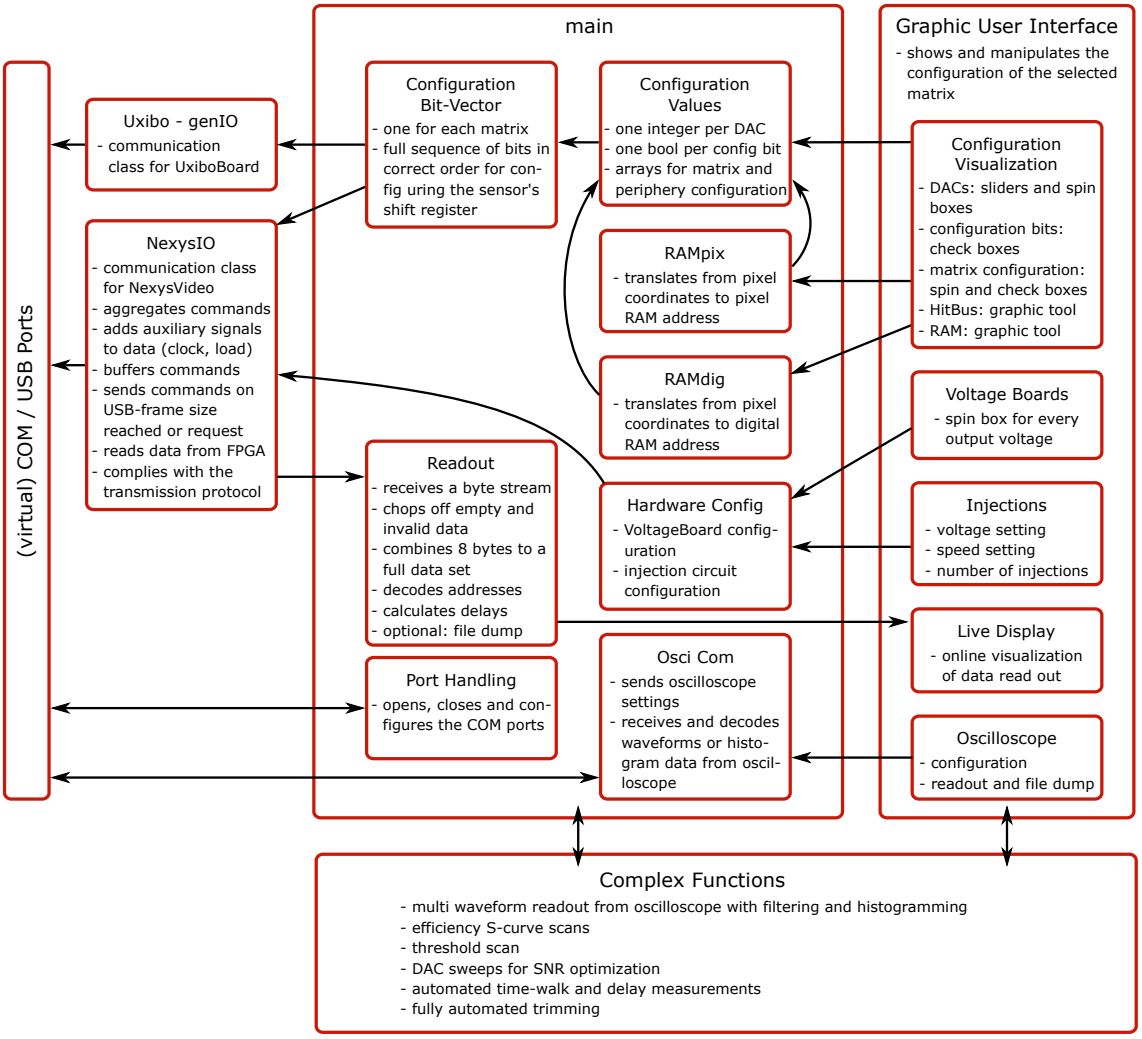


Figure 6.5: Block diagram of the computer software. The communication functions are drawn on the left side. The middle block shows the mediating functions between binary and encoded data from and to the H35Demo and their readable pendants. The box most to the right comprises the visualization of the major functions. Some measurements with compound and/or automated functions are listed in the bottom box.

6.3.1 FPGA communication

The connection to the FPGA is established via USB by virtualized COM ports. Other COM ports might be used to communicate with an oscilloscope or power supplies. The connections are automatically handled by the software, but can be overwritten manually on the graphic user interface (GUI).

The communication with the FPGA is based on a custom protocol. A variable command length on the cost of a small header to each command allows writing of a single register on the FPGA or several complete configurations of the H35Demo in one single command.

The communication protocol utilizes words of 1 byte length. This size was chosen with respect to the FTDI chip used on the NexysVideo, which handles 1 byte at a time. The actual target of all commands is the *Register File* in the FPGA. Each register has a dedicated purpose and can be both written and read from the computer.

The following table describes the transmission protocol for a data transfer between computer and FPGA:

number of bytes	purpose	description
1	header	Indicates whether the following command is write or a read command.
1	address	The address in the register file from which is read or to which is written.
2	length	The length of the data stream to be sent or received. The maximum length is restricted to 64 000 bytes.
< *length* >	data	Only write commands have the data part. The amount of data sent is indicated by the previous length bytes. A read request ends here, the FPGA will now start to send the requested data.

The first byte indicates only if the following command is a read or a write command. The second byte indicates the address in the *RegisterFile* on the FPGA from which is to be read or to which is to be written. Two bytes give the amount of data to be read or written. A read command ends here and receiving data from the FPGA starts. In case of a write command, the payload is sent now.

This protocol is very efficient for long read and write commands. Up to 64 000 bytes can be sent or received without additional overhead. This is exploited for the quite large configuration process[9]. Read commands benefit equally. When the H35Demo is illuminated with a source, the hit data have to be readout continuously and with potentially high load. With this protocol, the reading process only has to be interrupted after every 64k byte, meaning more than 8k data sets, by the next read request.

6.3.2 Configuration

The H35Demo is configured by concatenated shift registers. Each matrix has its own shift registers. The configuration is prepared as a whole on the computer, represented in a ≈ 1450 bit long vector. Each bit has a dedicated meaning or function.

However for a convenient way of modification, the configuration bits have to be combined according to their meaning. For better readability of the code, the single values are stored in variables with a meaningful label. For example, the 16 bits of each voltage DAC are stored in an integer labeled with the DACs name. This is done in the back-end of the software. The content of RAM is stored in arrays, which are read and converted during the process of sending the configuration, because their values are not statically map-able to the configuration bit-vector.

The configuration of the VoltageBoards and the injection circuit is handled in the same way, the voltages and the injection properties are stored and translated into a binary pattern on request.

6.3.3 Laboratory tool integration

Some measurements involve laboratory tools. Power supplies are always needed to power a characterization setup, but only some measurement sequences require them to change their setting during a measurement. Manually adjusting them between two measurement steps

[9] Configuration example for an Analog Matrix: 1411 configuration bits plus 5 load cycles, times clock divider 8 ≈ 11288 byte. The Standalone Matrices have an over 50 bit longer configuration. And to write not only the configuration but also the RAMs, this configuration has to be sent 2{16(matrix RAM) + 40(periphery RAM)}+1(normal configuration) times. That means over 826 000 bytes in several packages.

is not convenient. The need for automation is even more evident for some oscilloscope measurements. When a large number of waveforms has to be read out, the measurement time is very small, but the number of repetitions is large. Such measurement can only be done automated, therefore the integration of the laboratory tools into the setup control software is indicated. In this way, large measurement campaigns can be planned and are automatically conducted. The respective functions use the NI-VISA[10] library to configure and read out the oscilloscope. The received data, such as waveforms or histograms, is usually directly stored on the hard drive for offline analysis, but some basic functions of data analysis are implemented and can already be applied during the measurement.

6.3.4 Fast readout

The Standalone Matrices send their encoded data to the FPGA, where further data are added and chopped into 1-byte transmission packages. The software receives these packages and has to recombine them to an event data set. Invalid or empty data sets are discarded. Valid data sets are further processed. Either they are directly stored without decoding for maximum readout performance. Decoding has to be done offline. Or the data sets are decoded online. This includes translation from readout block index and pixel address to column and row position and calculation of the time difference between trigger time stamp and event time stamp. This decoded data are stored in a file.

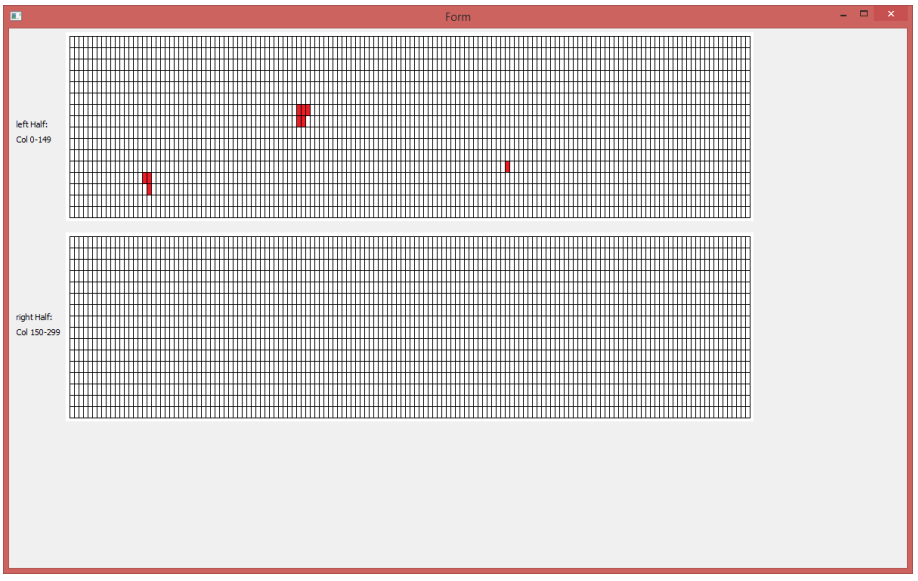


Figure 6.6: Screen shot of the live display function. To fit to common forms of computer screens, the matrix is divided in two and the pixel aspect ratio is not to scale. Pixels that have been hit, are dyed in red. After a fixed period of time, the color is reset.

For online compliance, the decoded data can be displayed quasi live as in figure 6.6. The visualization is the pixel grid of the H35Demo. Hit pixels are painted red for a certain duration. This feature is an instant feedback tool for the functionality of the chip. And especially for source- or beam-measurements, it is a handy tool to verify alignment.

6.3.5 Graphic user interface

In the previous sections, the interfaces between the H35Demo and the FPGA, between the FPGA and the Computer and between configuration and bit-vector have been discussed. The last important interface for the H35Demo setup is the Graphic User Interface which offers a convenient way of manipulating the settings of the characterization setup.

[10] https://www.ni.com/de-de/support/downloads/drivers/download.ni-visa.html

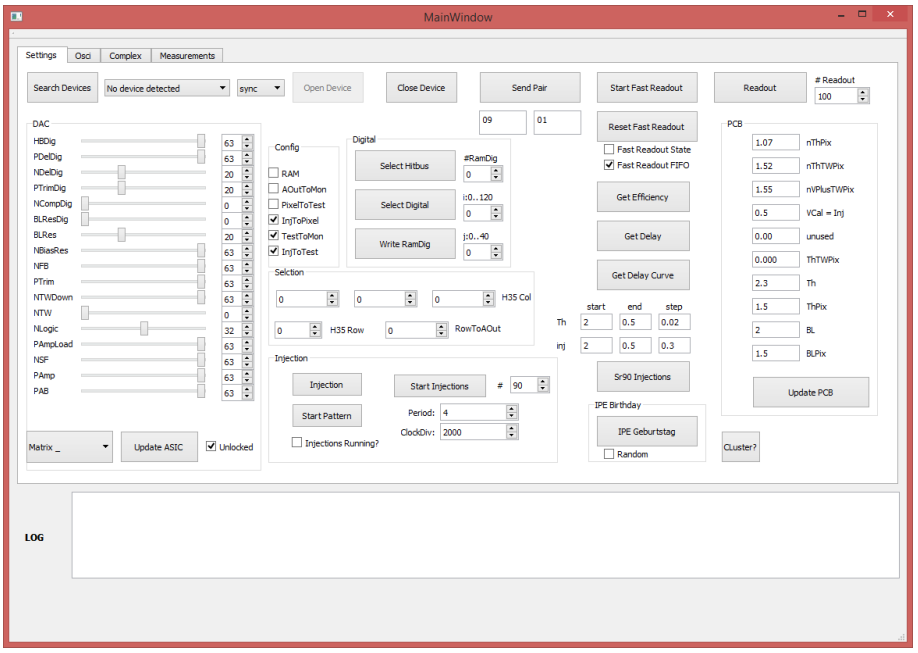


Figure 6.7: Screen shot of the control software. In this tab, the connection to the rest of the setup can be established and most chip settings can be modified here.

Figure 6.7 shows the main tab of the configuration software. The most basic chip configuration can be found here. The functions to establish a connection to the FPGA board and other laboratory tools is located on the top left corner. The area below is dedicated to manipulation of the configuration of the selected matrix (bottom left). From left to right are sliders and scroll boxes for DAC settings, configuration bit check boxes, direct chip RAM access and pixel matrix configuration. Graphic input assistants for HitBus and digital RAM can be launched here. Furthermore, on the far right the settings of the VoltageBoards can be changed. In the middle, right above the log, is a box to configure the injection circuit. Buttons for several often used composed functions are placed in the top right area. One button triggers the readout of a defined amount of data, another one starts continuous readout of hit data from the standalone matrices. Finally, with one button each, measurements of efficiency and delay for a selected pixel can be started using the provided settings.

The second tab manages all functions of a connected oscilloscope. This includes settings of the oscilloscope, as well as the readout of data (usually histograms or waveforms) to the computer for further analysis.

The tab *Complex* comprises automated measurements, which require a larger number of steps and orders. Measuring the analog signal of every pixel in a matrix with the oscilloscope is an example. Another example is the determination of signal-to-noise ratio of a pixel for several settings of a DAC. Last but not least, trigger handling is done here.

The last tab holds functions which are regularly used during operation. A screen shot is shown in figure 6.8. The first column comprises all functions necessary for S-curve measurements. Starting from single pixel efficiency S-curves to fully automated trimming of the whole matrix.
The second column contains a full matrix threshold scan with different settings. Optionally, the effect of the trimming value is investigated.
The column *RamPix* loads, sets and saves the RAM settings found by trimming.
The fourth column is dedicated to time-walk and delay measurements for single pixels or an entire matrix.

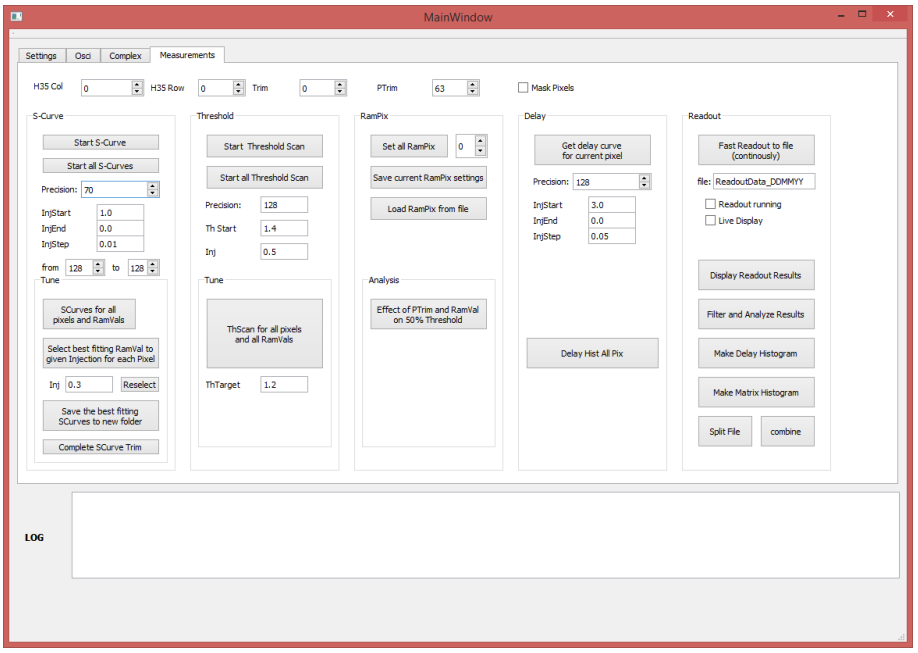


Figure 6.8: Screen shot of the measurement tab. Multi-pixel measurements that require several configuration steps are located here. An example is the full matrix efficiency S-curve measurement.

The last column holds not only a button to continuously read out data from FPGA to file and start a live display of the events, but also has some functions to handle and analyze readout data.

7. Characterization of H35Demo

The H35Demo is an ASIC designed to serve as monolithic silicon pixel sensor chip in the outer layers of ATLAS ITk detector of CERNs LHC. The characterization has to reveal if the *ams* H35 technology can meet all crucial requirements (see chapter 4.4). If a requirement is not yet met, possible improvements are to be identified.

7.1 Calibration

Generally speaking, this section will study how the H35Demo responds to well-defined signals. The signal response can be probed at three (Analog Matrices) or four (Digital Matrices): Charge sensitive amplifier output (S1 or AmpOut), second stage amplifier output (S2 or Monitor), comparator output (HitBus) and the digitized hit information in digital matrices (Fast Readout).

These outputs will be examined by sending known signals, e.g. X-rays from radioactive sources or X-ray tube, or electrical charge injections, into the chip. The latter need calibration themselves, but are the most convenient way of characterization as they are always available, are not harmful to humans or equipment and the point in time when an injection is triggered is known and does not require additional measurement.

7.1.1 Calibration of the injection circuit

The calibration of the H35Demo has been conducted on an 80 Ωcm sample of the H35Demo without post-processing. However, some results are expected to be universal. The first goal of the calibration is to relate the injection voltage to a charge generated by charged particles or photons in the pixel diode. Once this relation is known, most experiments can be done with the injection circuit and without hazardous radioactive sources.

In this measurement, the chip is illuminated with X-rays from an iron-55 source. The X-rays emitted from this source have a well-defined energy of 5.9 keV. When such photon is stopped in silicon, 1639 electron-hole pairs are created. The electrons are separated from the holes by an electric field. The charge signal is converted to a voltage signal and amplified by a charge sensitive amplifier (CSA). The output of the amplifier is monitored by an oscilloscope and the maximum value of each signal pulse is filled into a histogram (Figure 7.1). Even though the X-ray energy is well-defined, we expect a Gaussian distribution due to uncertainties in charge generation and in charge collection. The largest contribution to the uncertainty however, is the noise of the amplifier. A Gaussian fit is applied to the histogram data. The values obtained by fitting are:

$$U_{\mathrm{AmpOut}}\left(^{55}\mathrm{Fe}\right) = 79.7\ \mathrm{mV} \pm 6.4\ \mathrm{mV} \tag{7.1}$$

Note that this result is only valid for one pixel on one chip with the exact settings of the measurement.

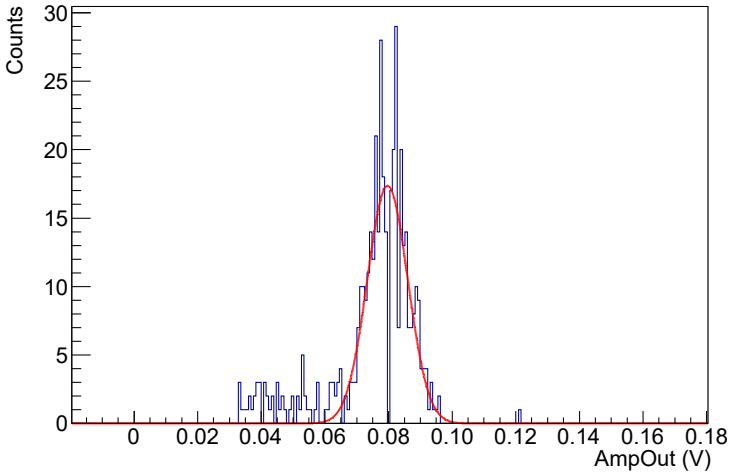
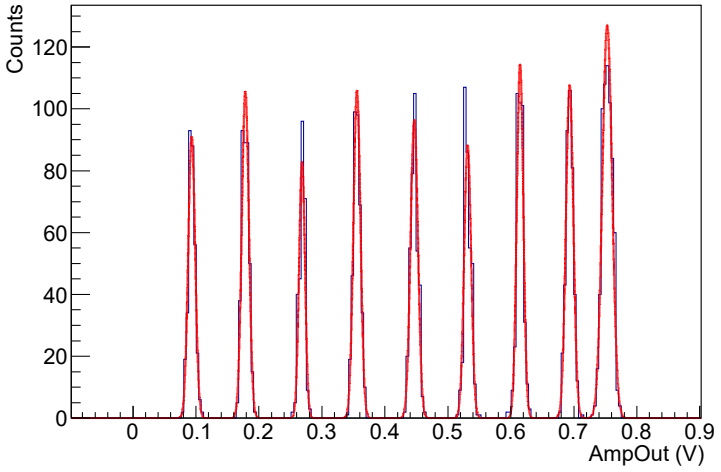


Figure 7.1: The signal of the first stage amplifier is filled into a histogram and a Gaussian fit is applied to the peak. The background noise is excluded from the fit.

Instead of charges generated by X-rays, the charges are now generated by negative voltage pulses. These pulses are generated by the injection circuit. Figure 7.2 shows the combined histogram of injections with different voltage. The used injection circuit is implemented on the carrier PCB and creates steep negative voltage steps between an upper voltage and ground. The upper voltage is defined by DAC which can be written by the computer via FPGA. The effective range of injections is about 0.1 V up to the supply voltage -0.4 V.

The injection is triggered by an analog switch, controlled by a pattern generator on the FPGA. Speed, number of injections, duty cycle and other properties can be set by the computer. Therefore, not only signal height but also timing properties of the device under test can be investigated.

The injection line is fed into the H35Demo. By configuration, it can be connected to the input of the amplifier of one or more pixels of the DUT.



Injection (V)	AmpOut (V) μ	σ
0.3	0.0933	0.0055
0.6	0.1784	0.0056
0.9	0.2684	0.0046
1.2	0.3556	0.0055
1.5	0.4466	0.0058
1.8	0.5314	0.0055
2.1	0.6146	0.0051
2.4	0.6933	0.0059
2.7	0.7530	0.0077

Figure 7.2: Test signals of different voltages are generated and the amplifier's response is monitored. A Gaussian fit is applied to each peak in the histogram.

The value found from figure 7.1 in equation 7.1 is compared to the respective values extracted from figure 7.2. The result is used to determine the number of charge carriers Q

generated by injection pulses:

$$Q\left(U_{\text{Amp}}\right) = 20746 \ \frac{e^-}{\text{V}} \cdot U_{\text{AmpOut}} \tag{7.2}$$

$$U_{\text{AmpOut}}\left({}^{55}\text{Fe}\right) = U_{\text{AmpOut}}\left(1639\, e^-\right) = 79.7 \text{ mV} \tag{7.3}$$

$$Q\left(U_{\text{Inj}}\right) = 5883 \ \frac{e^-}{\text{V}} \cdot U_{\text{Inj}} \tag{7.4}$$

As the pixel electronics design before the amplifier and the injection circuit are the same for all matrices and pixel flavors on the H35Demo, this result is assumed universal.

At this point the baseline noise is also determined by histogramming the untriggered baseline with an oscilloscope (Figure 7.3). For unirradiated sensors the shape of this noise can be assumed Gaussian. The fit delivers a noise sigma of 3.8 mV. The charge equivalent noise (ENC) is 78 electrons. For ^{55}Fe-signals we have therefore a signal-to-noise (SNR) ratio of 21.

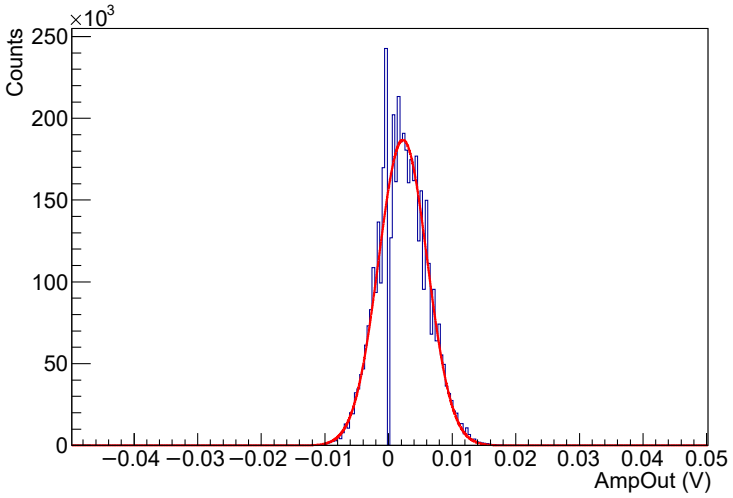


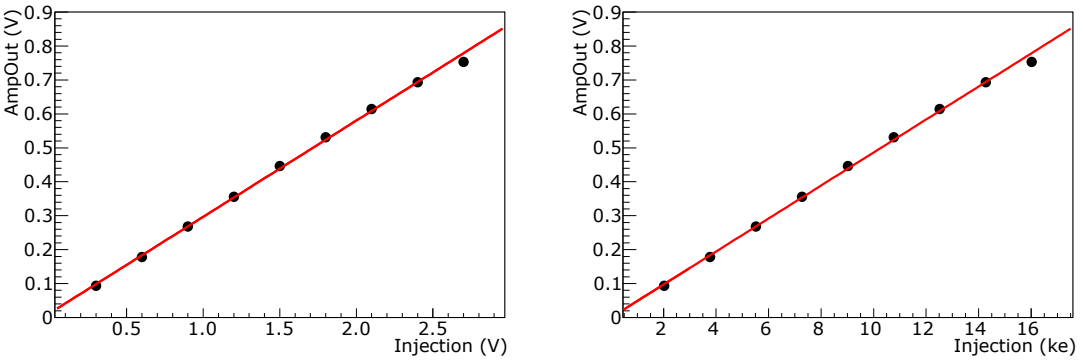
Figure 7.3: A histogram of the AmpOut baseline. The missing bin at 0 and the doubled bin left to it are due to an oscilloscope glitch. The baseline noise measured in this way is 78 ENC.

Note that this result was obtained using the analog output of the first amplifier. It does not represent the internal noise of the H35Demo, because the connection from amplifier to pad is not optimized. For a precise and representative noise measurement, the standalone readout or a readout ASIC have to be used to measure S-curves from which the internal noise can be extracted. However, the noise of AmpOut can be used as relative measure, e.g. to compare pixels of different flavor or different configurations of the chip.

7.1.2 Calibration of the amplifier

In figure 7.4 the mean value of the Gaussian fits are used to obtain AmpOut as a function of injection voltage. The sigma value of the Gaussians is used as error bars, however, they are hardly visible, because they are smaller than the markers. As anticipated, the relation is a linear function in the relevant range. The measurement range is limited by the injection circuit on the carrier PCB. The injection circuit is explained in more detail in chapter 14.6.

$$U_{\text{AmpOut}}\left(U_{\text{Inj}}\right) = 0.2836 \cdot U_{\text{Inj}} \tag{7.5}$$

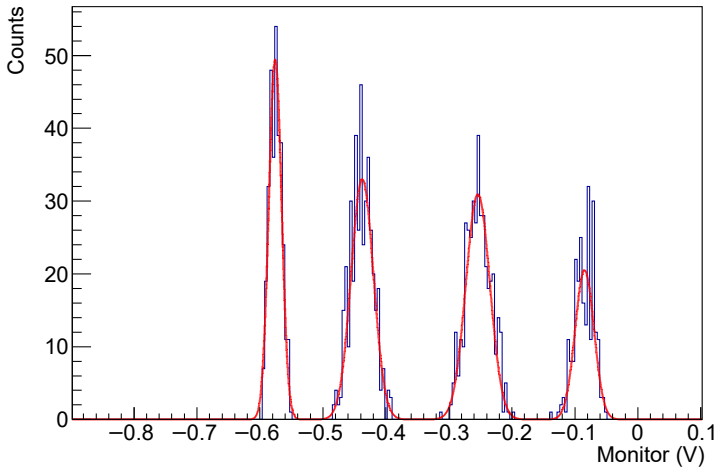


(a) **(b)**

Figure 7.4: The plots show the signal of the first amplifier (AmpOut) as a function of injection voltage (a) and injected charge (b). The values and their standard deviation are extracted from figure 7.2.

The second stage amplifier is, just as the first stage, an inverting amplifier. It is required to strengthen the signal and transmit it to a readout ASIC (via CCPD or bump bonding, applies to all matrices) or to transmit it to the digital periphery (standalone matrices).

In contrast to the first stage amplifiers, of which only the ones in the second pixel column in each matrix can be connected to a pad, all pixels of the second stage can be connected to monitor lines. Analog matrices have three monitor lines, standalone matrices have one. The response of this amplifier to injections is histogrammed in figure 7.5. Due to the additional amplification, saturation of the amplifier is reached at about 1.5 V or 9000 e.



Injection (V)	Monitor (V) μ	σ
0.3	-0.0846	0.0150
0.6	-0.2540	0.0190
0.9	-0.4375	0.0177
1.2	-0.5758	0.0101

Figure 7.5: Histogram of injections with different voltage after the second stage amplifier. Saturation is already reached at about 1.4 V injections.

The positions of the injection histograms are displayed in figure 7.6 and a linear fit is applied:

$$U_{\text{Monitor}}\left(U_{\text{Inj}}\right) = -0.5450 \cdot U_{\text{Inj}} \tag{7.6}$$

$$U_{\text{Monitor}}\left(Q_{\text{Inj}}\right) = -0.0939 \ \frac{\text{V}}{\text{k}e^-} \cdot Q_{\text{Inj}} \tag{7.7}$$

The output signal of the first amplifier is at the same time the input signal of the second amplifier. This can be exploited to determine the voltage gain of the second stage amplifier.

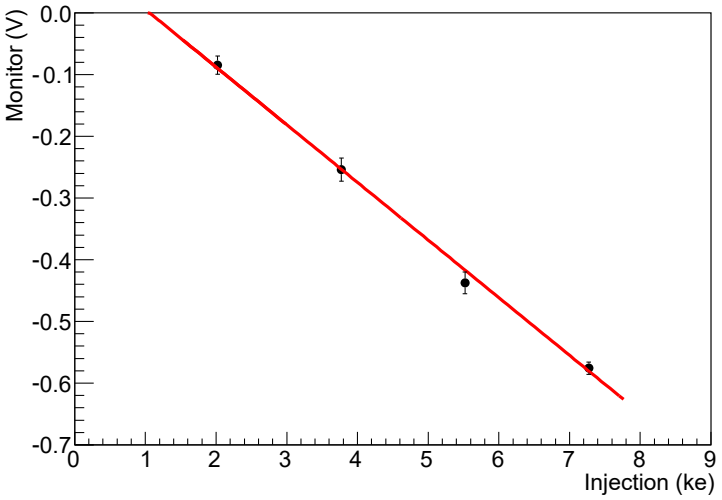
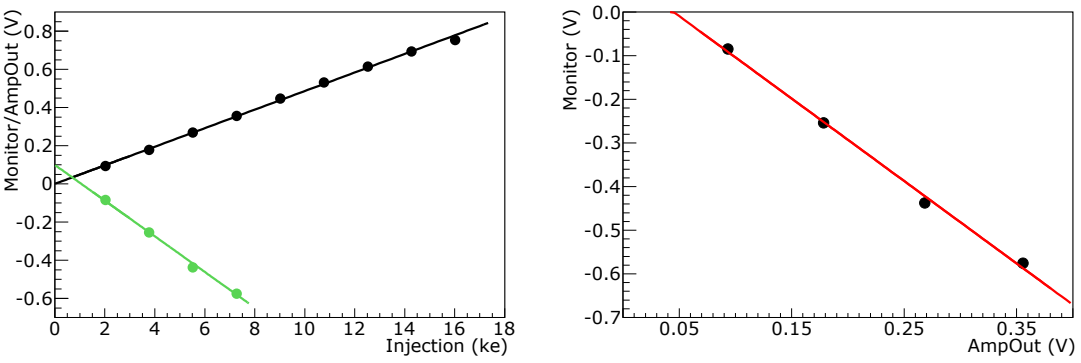


Figure 7.6: The analog response of the second stage amplifier to injections, measured on the monitor line. Signals of more than 8 ke saturate the amplifier.

Figure 7.7a shows the calibration lines of both the first stage and the second stage in one plot. As the second stage reaches saturation above 8 ke, the four points with smallest signal can be used to determine the amplification. In figure 7.7b the signals of the second stage are plotted as a function of the first stage. The amplification is assumed linear in this range and the fit result delivers for the voltage gain A:

$$A = -1.889 \pm 0.073 \tag{7.8}$$



(a) **(b)**

Figure 7.7: The combined calibrations can be used to determine the amplification of the second stage amplifier.
(a) Calibration of first amplifier in black, of second amplifier in green.
(b) Output signal of the second amplifier as a function of the output signal of the first amplifier for the same injected signal. The slope equals the amplification of the second stage amplifier.

7.1.3 Calibration of the pixel diode

The same settings as before are used for a measurement of charged particles originating from a ^{90}Sr source. The β^--rays from ^{90}Sr have an energy of 0.546 MeV. A depletion voltage of 80 V has been applied. Figure 7.8 shows the histogram which has been obtained in the same way as the histogram in figure 7.1.

The histogram is not only composed of a Landau-distribution (peak on the right), which is expected from energy loss of charged particles in matter, but also of a second, a background peak (left). The background peak is not be confused with a peak by baseline noise. This low-energetic peak occurs due to the way this histogram was obtained. For a single pixel measurement, the active area is rather small and therefore the probability of charge sharing between neighboring pixels is high. Whenever the monitored pixel receives charges from a shared event, it contributes to the background peak. Another source of events in this peak are electrons which have not traveled directly from decay to detector, but lost energy on the way by scattering. They could have been excluded by a trigger system. This background is approximated by a Gaussian distribution around 0 V, which is cut off by oscilloscope settings below 0.2 V.

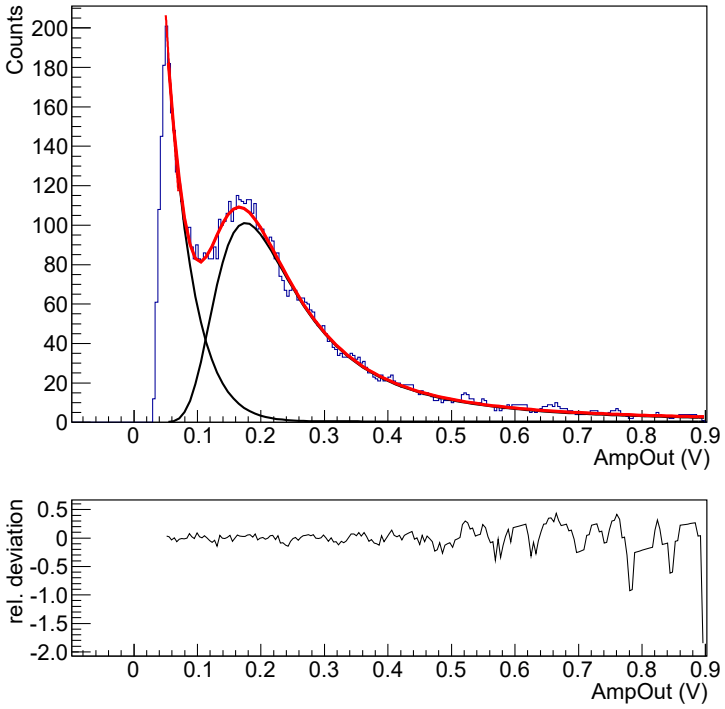


Figure 7.8: Histogram of ^{90}Sr-signals at the first amplifier. The fits of the background peak and the Landau-Gaussian are displayed in black, the envelope in red. The relative deviation of fit and data points is shown below.

In order to extract the actual signal, all features have to be fit together. However, the Landau-distribution itself is convoluted with another Gaussian. This Gaussian is the noise peak seen in chapter 7.1.1 (Figure 7.3). Figure 7.1 displays the envelope fit result in red and the partial fits of the background (left) and the Landau-Gaussian (right) in black. The most-probable-value (MPV) of the Landau distribution is 0.175 V or 3600 e.

7.2 Signal-to-Noise Ratio

The signal-to-noise ratio (SNR) of the detector is important for all experiments. It may be defined differently for different experiments. In case of the H35Demo, aiming at the ATLAS experiment, a MIP is often referred to as (the smallest) 'signal' and the analog noise is obtained by a baseline histogram. However, not always this definition of SNR is used, especially when not an absolute value, but a relative SNR is sufficient. In that case the signal is set to be a certain number of electrons generated by injection. The noise might still be the baseline noise or the variance of the measured signal.

This relative SNR has been used to compare the different pixel styles of the Analog Matrices A and B with a total of six different pixel flavors. Settings were chosen so that analog power consumption is the same for both matrices. This requires a setting of the main analog bias current (VNAmp) in Matrix A which is about 10% higher than in Matrix B.

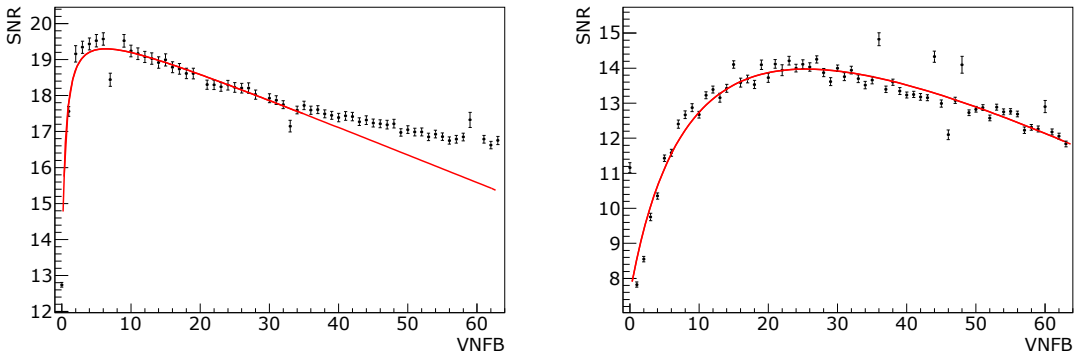
7.2.1 SNR as function of feedback

Each Matrix A and Matrix B comprise three pixel flavors each, of which only the first flavor can be investigated using AmpOut (see chapter 5.1.3). The analog behavior of the other flavors can be compared using the monitor lines. Figure 7.9 illustrates the effect of the feedback setting on signal-to-noise ratio. The AmpOut response to a 0.5 V injection (≈2900 e) was used as signal and its variation as noise. An amplifier with stronger feedback has a reduced noise, but also a smaller output signal [71]. Thus, SNR can be optimized with respect to *VNFB*.

The behavior around the optimal *VNFB* setting can be approximated by a function of the form

$$y(x) = a + bx + \frac{c}{x - d}. \tag{7.9}$$

The best SNR of Matrix A in this measurement was 19 for a *VNFB* setting of 6. In Matrix B, the best SNR is lower (14) and requires a higher setting of *VNFB* (25). As the pixel design is the same, the higher setting of the main bias DAC has been identified as the origin of this difference. It scales *VNFB*, resulting in a left shift shown in figure 7.9b relative to the plot in figure 7.9a. Furthermore it increases the amplification, which leads to a larger SNR.



(a) Matrix A **(b)** Matrix B

Figure 7.9: The plots show signal-to-noise ratio as function of the feedback setting (VNFB). The strength of the amplifier's feedback determined by VNFB has a large effect on SNR. The optimal SNR is found by fit. The main bias current of Matrix A (a) was higher than in Matrix B (b). This increases the SNR and left shifts the SNR curve.

7.2.2 SNR of the second stage amplifier

In contrast to chapter 7.2.1, the SNR of the second stage amplifier is not only strongly dependent on the feedback setting of the first amplifier, but also on the setting of the baseline resistance DAC *BLRES*. The reason is that the output of the first amplifier is coupled capacitively to the input of the second. The baseline of this input is defined by an external voltage. However, the coupling strength between the baseline voltage and the amplifier's input is defined by *BLRES*.

Consequently, two parameters have to be swept for the SNR analysis of the second stage amplifier. Figure 7.10 shows SNR as function of *VNFB* for several values of *BLRES*. In order to find the best value of *VNFB* for a fixed value of *BLRES*, one can simply pick the highest value in this graph. The best setting for *BLRES* is found by comparing the highest SNRs from each *VNFB* sweep (Figure 7.11a).

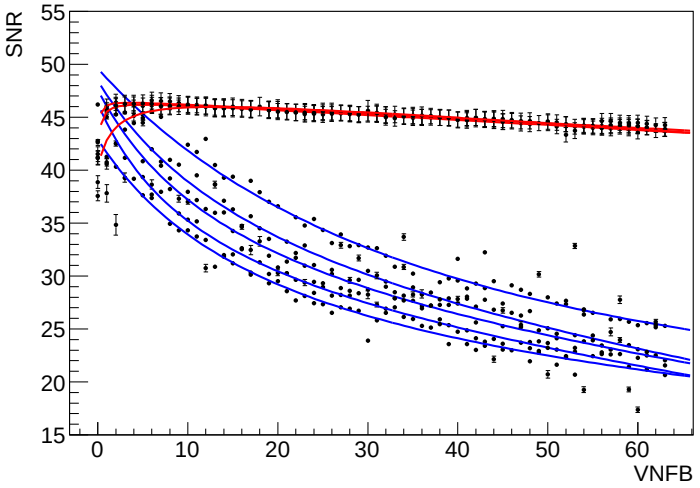
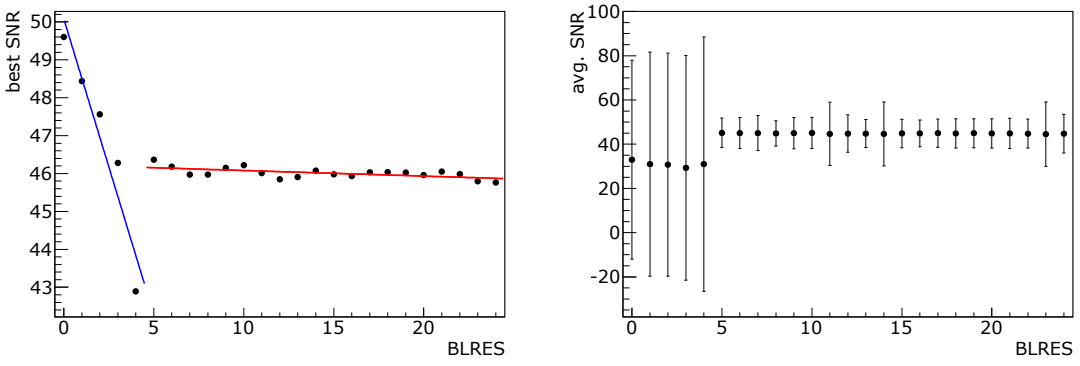


Figure 7.10: The signal-to-noise ratio after the second stage amplifier as a function of feedback setting. Each line corresponds to a different setting of the baseline resistance. In blue, the lines for $BLRES = \{0...4\}$, from high to low SNR values; in red $BLRES > 4$, from low to high. Excluding the blue lines, the best SNR found is > 46.

However, when looking at the shape of the SNR(*VNFB*) curves, it is apparent that there are two types: One type has shapes similar to the curves of the first amplifier in figure 7.9 (red). The other type shows a $1/x$ behavior (blue). The lines in blue were obtained with $BLRES = \{0...4\}$, the lines in red are very similar to each other for all $BLRES > 4$.



(a) **(b)**

Figure 7.11: Further considerations regarding the optimal *BLRES*-setting:
(a) The best SNR obtained for *VNFB*-scan at given *BLRES*. Low settings are sensitive towards variations of *BLRES* (blue), the SNR of higher *BLRES* settings is rather unaffected by small changes (red).
(b) Mean and standard deviation of a *VNFB*-scan at given *BLRES*. The SNR is much more stable for $BLRES \geq 5$.

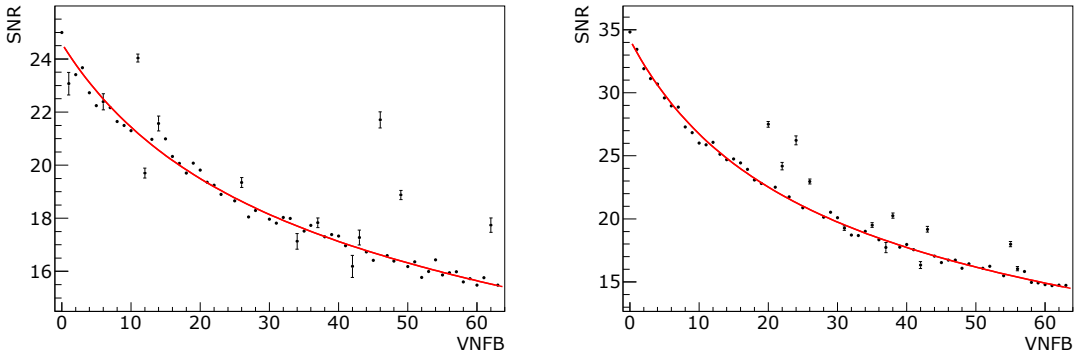
The best SNR of the blue curves are also the best for all possible *BLRES*. However, the stability of these values is not given. A small variation in experimental condition, e.g. a minor change in supply voltage or irradiation dose, has a large negative impact on the SNR,

when a low *BLRES* has been chosen. For this reason it is advisable to choose a point from the red lines with $BLRES > 5$. Above a rather low setting of *VNFB*, the SNR is nearly constant, thus independent of minor changes of the experimental condition. This behavior is illustrated in figure 7.11b by plotting the average SNR with its standard deviation of a given *BLRES* for all *VNFB*. For the baseline resistance set between 0 and 4, not only the average SNR is lower, but also their spread is up to 10 times higher, compared to the others.

Even though a very low coupling between amplifier input and baseline voltage can result in a good signal-to-noise ratio, it is advisable to pick the baseline resistance larger than 5. The setting of the amplifier's feedback may be set much less carefully in this case. Reasonable feedback settings range from 5 to 10 for an unirradiated sensor, without greater change when going to higher values. Sticking to this restriction the SNR in this measurement is always found around 46.

7.2.3 SNR of linear feedback transistor and enclosed transistor

As seen in the previous sections, the properties of the amplifier is strongly dependent on the feedback. Variation of different settings changed the SNR by more than an order of magnitude. However, the feedback transistor type has a huge impact on the SNR, too [19]. The third pixel group (columns 200 - 299) of Matrix A has linear NMOS transistors in the feedback circuit, instead of the baseline version with enclosed transistors (columns 100 - 199). Figure 7.12 shows the SNR of the second stage amplifier of a pixel with enclosed feedback transistors (a) and linear feedback transistors (b). The benefit of enclosed transistors is the radiation tolerance. This comes on the price of reduced gain. Pixels with linear transistors in the feedback on the unirradiated chip have roughly 50% increased SNR.

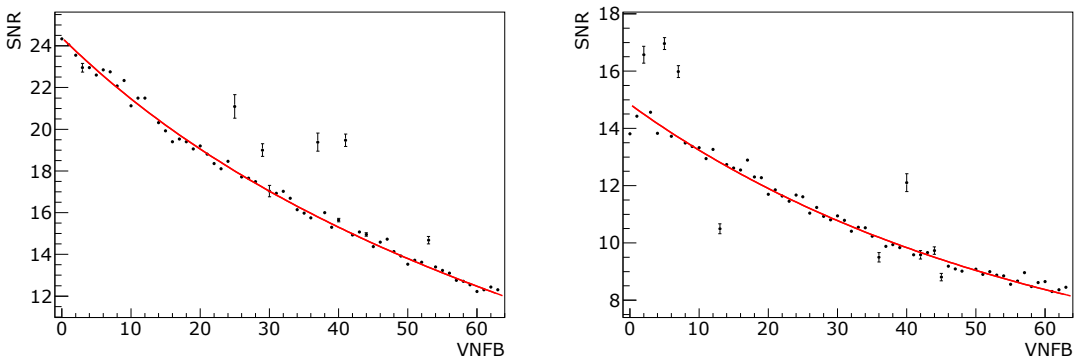


(a) enclosed feedback transistors **(b)** linear feedback transistors

Figure 7.12: In the feedback circuit of the first amplifier, enclosed or linear transistors can be used. The plots show SNR as a function of amplifier feedback measured on a pixel with enclosed transistor (a) and with linear transistor (b) in the feedback loop.

7.2.4 SNR of high gain and normal gain amplifier

As described in chapter 5.1.3, the third pixel group of Analog Matrix B has a lower gain in favor of faster response time. This is achieved by adding a capacitor between the source follower's output and the deep n-tub [70]. Figure 7.13 compares the signal-to-noise ratio of a pixel with baseline layout to a faster pixel with lower gain. As the faster layout does not reduce noise as it reduces the gain, the SNR is significantly reduced. With the chosen settings, the reduction is about $1/3$.



(a) baseline layout **(b)** faster layout

Figure 7.13: One way to reduce the time-walk effect is to increase the speed of the amplifier. This can be done by adding a capacitance between the source follower's output and the deep n-tub, on the cost of a reduced gain. The SNR behavior of a pixel with baseline design is shown in (a). The faster amplifier in (b) has an about 1/3 reduced SNR.

7.3 Charged particle detection

Unlike photons, charged particles do not generate electron-hole pairs in a localized spot but along a trajectory. When absorbed, photons release their entire energy at once. Charged particles are usually not absorbed, but interact with the sensor material during their flight, dissipating energy. Some of these interactions have long range effects of several hundred nanometers, therefore the hit signal of a charged particle is not necessarily limited to a single pixel event. An event with more than one pixel is called cluster event.

The size of the cluster depends on many factors. The volume in which the charge is generated depends on the sensor type, the particle type and its energy. In HV-CMOS-sensors, the charge carriers from within the depleted volume are collected by drift at the pixel electrodes. Charge carriers from outside the depleted volume do not contribute to the signal. They either recombine or are not collected in time (diffusion is slow). Therefore, the recorded cluster size depends also on the pixel geometry, its electrodes and the depletion voltage.

The active volume under each pixel can be assumed to have a box-like shape. Thus, the inclination of the incoming particle has a large effect on the cluster size. A particle moving perpendicular to the sensor may hit only one or few pixels, while a particle traveling in the sensor plane will intersect the active volume of many pixels.

High energetic particles may even generate secondary electrons in the detector with enough energy to travel long distances. If this secondary electron travels in the sensor plane, it can cause signals in a very large number of pixels. Such high-energetic secondary electrons are called Delta-electrons.

For many considerations, the Minimum Ionizing Particle (MIP) is of special interest (see Bloch formula 2.3.2). As the name suggests, such particle has the energy leaving the smallest possible signal in a silicon detector. The smaller the charge signal is, the harder it is to distinguish it from noise. Additionally, smaller signals have a larger time-walk. The energy of a minimum ionizing electron is ≈1.5 MeV. Such electrons can be provided by an accelerator or by radioactive decays. As accelerators are not always at hand, Strontium-90 is used as sources of MIPs, even though the decay energy is 0.546 MeV [72]. The energy deposited by ^{90}Sr-electrons is approximately 40% higher than the deposition of a true MIP [73]. However, ^{90}Sr is a good choice, because it is easily available, has a long half time and none of the secondary decays have energies close to the energy of the primary decay [74].

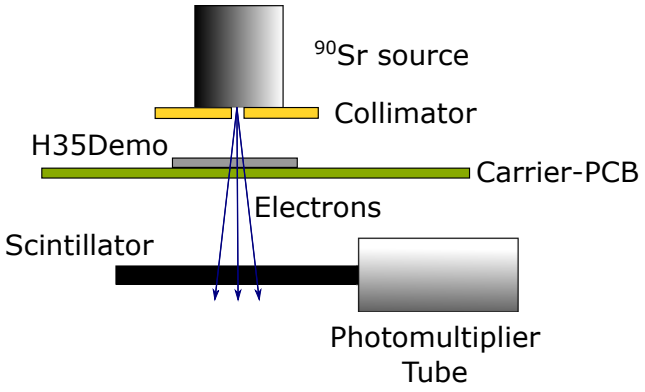


Figure 7.14: ^{90}Sr is a handy source of charged particles close to MIP energy. The encapsuled source is placed above the sensor with optional collimation. The emitted electrons pass through the sensor and a scintillator for precise time reference. The data from the sensor and the photomultiplier tube are collected by the FPGA.

The sketch of a typical source measurement setup is shown in figure 7.14. The used ^{90}Sr source is encapsulated in acrylic glass cylinder for save handling. During measurement it is stored in a brass housing with an exit window, pointing at the measured sensor. If the measurement requests a precise time reference, additionally a scintillator with photomultiplier tube is added to the setup. The scintillator can be placed behind the sensor, if the sensor has been thinned and is supported by a PCB with a cut-out under the active area. If these premises are not given, the scintillator has to be placed in front of the sensor, because virtually all electrons are stopped by a thick sensor and PCB.

Even though the rate of the used ^{90}Sr source is moderate compared to some others sources, the effective particle rate hitting the H35Demo is significant considering its large size. This rate can not be handled by analog readout, except for single pixel measurements. In order to read out the entire matrix simultaneously the standalone readout has to be used.

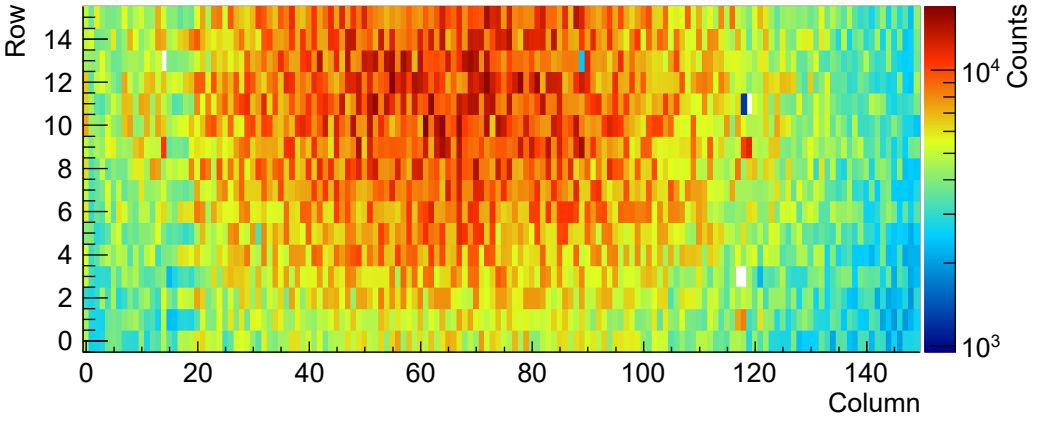
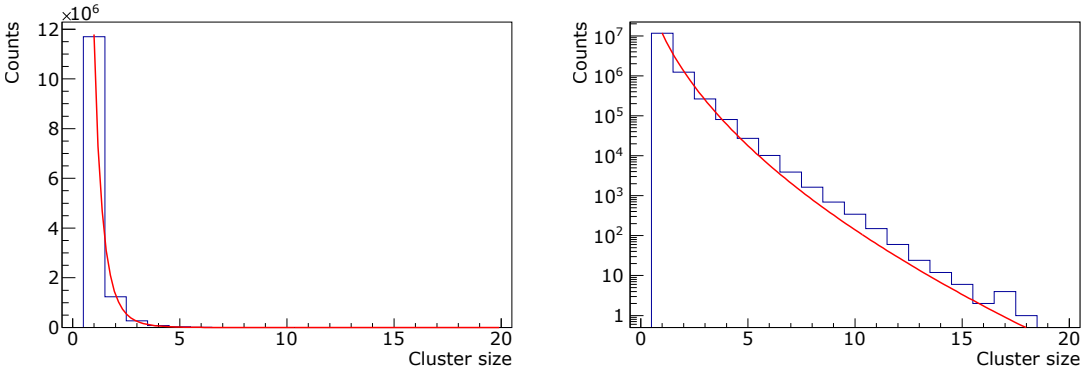


Figure 7.15: Hit map recorded by the NMOS Matrix, illuminated by a collimated ^{90}Sr source.

Figure 7.15 displays the hit map of the NMOS Standalone Matrix which has been illuminated with ^{90}Sr-decay electrons from a collimated source[1]. The ^{90}Sr source was placed above the H35Demo, with the aperture pointing at the left half of the NMOS Standalone Matrix (without time-walk compensation).

The hit data from the previous measurement comes with 8-bit time stamps. Each pixel hit

[1] A similar measurement, but without collimation is shown in Appendix E.6.



(a) **(b)**

Figure 7.16: The histograms show how often each cluster size was detected with linear (a) and logarithmic scale (b). The vast majority of events is single pixel hits. Clusters with size two and three are notable populated, too. Larger clusters can only be seen in a logarithmic plot. The average cluster size is 1.7 pixels per event.

has its own, which is extended in the FPGA to a total of 16 bits. The collected data have been analyzed using the spatial and timing data to get information about the cluster size single particles have left in the detector.

As the particles hit the sensor mainly perpendicular, the shape of clusters is assumed circular. The constant data stream from the H35Demo is zero suppressed in the FPGA. The event rate is low compared to the readout speed. Therefore, only a small number of consecutive data sets have to be considered in the search for clusters. To decide if a hit pixel is part of a cluster, two conditions have to be fulfilled: The spatial distance to the forming cluster has to be 250 µm or less and the time difference between any hit of the forming cluster and the candidate has to be smaller than 10 time stamps (12.5 MHz). Finally, the cluster size is filled into a histogram (Figure 7.16). The vast majority of events is in single pixel clusters, but a significant number is found in cluster of size two and three as well. Larger clusters occur only sporadically, therefore they can be seen only in the logarithmic plot. The best agreement with the collected cluster data shows the equation of the form:

$$f(x) = a \cdot \exp\left(-b\left(\sqrt{x} - c\right)\right) \tag{7.10}$$

Where a is a scaling factor, b is the inverse decay constant with unit 1/m and c is an x-offset. The cluster size x appears under the square root. That is reasonable, because the cluster size depends on the interaction range with unit m of the incoming particle with the sensor material. The cluster size x however is an area with unit m². The mean cluster size in this measurement is 1.7 pixels per cluster.

7.4 Mismatch

In the production of silicon chips statistical methods are used (see chapter 2.2.1). Some of these are very sensitive to small parameter variations. For example, photolithography of edgy structures is non-trivial and the making of masks is a complex science itself. Another example is that most etching techniques do not carve structures with a uniform speed and high aspect ratio, but etch some structures faster than others with limited aspect ratio.

Therefore, a transistor located in one region might have different properties than a transistor in another region of the chip, even though they share the same layout. In other words: their behavior does not match [75, 76].

There are no dedicated structures on the H35Demo to investigate this mismatch. However, the behavior of single pixels can be measured and compared to each other. The result may be interpreted as the sum of all mismatch effects a signal, which is assumed constant, has picked up on its path from generation to monitoring.

The effect of mismatch in the NMOS Matrix is investigated in this section. It is measured as variation of the detection threshold of efficiency S-curves. In order to be comparable to measurements in the Analog Matrices, all values are compared to the signal of ^{55}Fe X-rays.

A similar measurement on the Analog Matrices using the ^{55}Fe decay signal as reference, is presented in Appendix E.3. The major difference between both is that the measurement of the NMOS Matrix is based on electrical charge injections and the Analog Matrix measurements utilize X-rays. The first, excludes variation effects of the sensor diode, but includes variations of the test signal circuitry, while the latter has no digitization stage, thus does not include its influence.

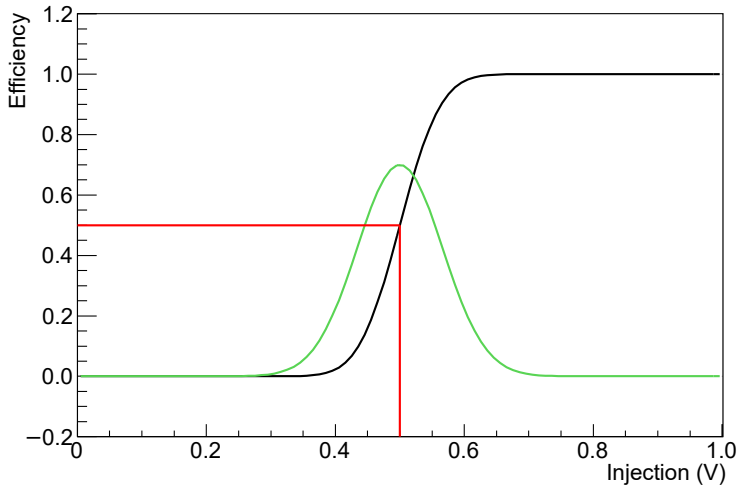


Figure 7.17: Characterization of standalone sensor chips is often based on efficiency S-curves (black). They are obtained by measuring the efficiency over the injected charge of a single pixel. For analysis the inflection point (red lines) is used. It marks the point of 50% efficiency and is called detection threshold.
The S-curve has no step-like shape due to noise. Hence, the width of the S is a measure for noise in the system. For some studies, the S-curves derivative is useful, it is a Gaussian (green).

For mismatch analysis of the NMOS Standalone Matrix, S-curves for every pixel is measured. All settings, especially the threshold of the comparator, remain fixed. Test signals of a certain charge are injected into a pixel n-times and the detection efficiency is calculated from the data read out from the standalone readout. The injected signal starts with 1 V and is then lowered until the detection efficiency reaches zero. The result is an S-curve as shown in figure 7.17. If the measured system was without noise, a step-function would be found instead of a smooth S-curve. Noise introduces a detection uncertainty, thus smears out the step-function, turning it into an error-function. The derivative of such function is a Gaussian. Its parameter σ is a measure of noise or width of the S-curve.

The inflection point of the S-curve marks the point of 50% efficiency. In the Gaussian representation of the S-curve, it is equal to the peak position μ. This 50% detection efficiency point is called detection threshold. It is not to be confused with the threshold of the comparator.

All pixels of a perfect sensor would deliver the same S-curve. In a noise-free system these identical S-curves were step functions. In a real sensor, the S-curves are shifted relative to each other and the transition width (width of the S-curve) is not uniform.

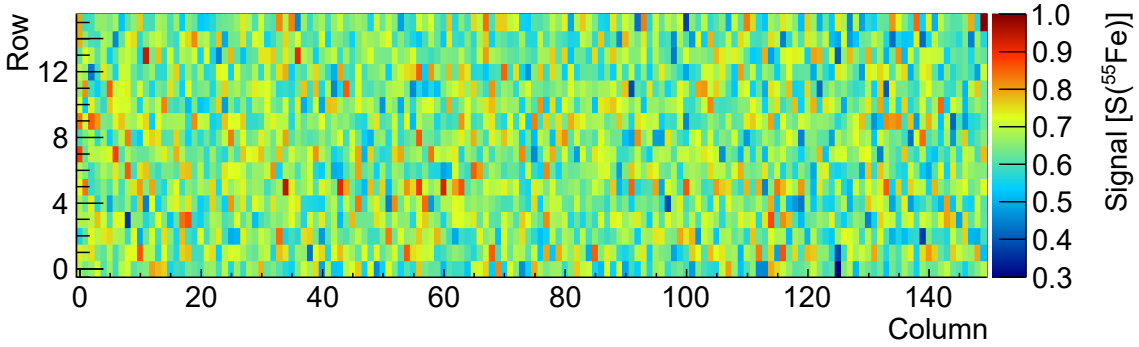


Figure 7.18: Color coded detection threshold measured via standalone readout. It is a measure of mismatch in the NMOS Matrix. For better comparability, the detection thresholds are given relative to the signal of ^{55}Fe.

For figure 7.18 the detection threshold of every NMOS pixel with standard comparator was measured. For better comparability, the 2D plot shows the inverse of the detection threshold normalized to the signal of ^{55}Fe:

$$\text{Signal} = \frac{1}{U_{\text{Detection Threshold}}} \cdot \frac{Q_{^{55}\text{Fe}}}{Q_{1\,\text{V Injection}}}$$

$$\text{Signal} = \frac{1}{U_{\text{Detection Threshold}}} \cdot \frac{1639\,e^-}{5883\,e^-/\text{V}}$$

$$\text{Signal} = \frac{0.28\,\text{V}}{U_{\text{Detection Threshold}}} \tag{7.11}$$

Nevertheless, the values are not comparable 1:1. The plots for the Analog Matrices show the analog response to a well-defined signal, whereas the NMOS Matrix plots show a variable signal with well-defined efficiency. Despite the differences the relative behavior can be analyzed.

It appears that the comparator is more sensitive to mismatch or supply voltage variations, than the amplification stages, as the variations of the Analog Matrices are smaller than those of the NMOS Matrix. The response of pixels is not necessarily randomly distributed over the matrix, but often a tendency along supply- or signal lines can be spottet. This behavior is investigated in Appendix E.2.

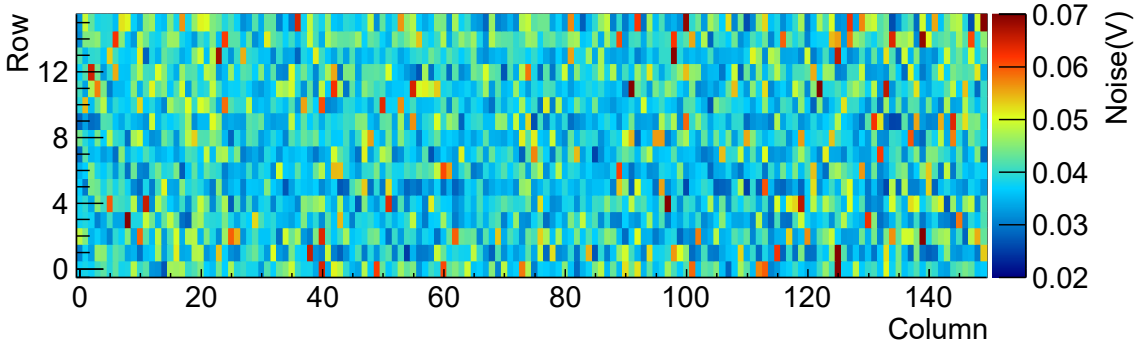


Figure 7.19: The noise of the NMOS Matrix is extracted from the measured S-curves.

The transition width of the S-curve of each pixel is displayed in figure 7.19. Just as in the detection threshold plot, no structure is visible. The fluctuation between single pixels is much larger than a possible change from column to column or from one row to the other.

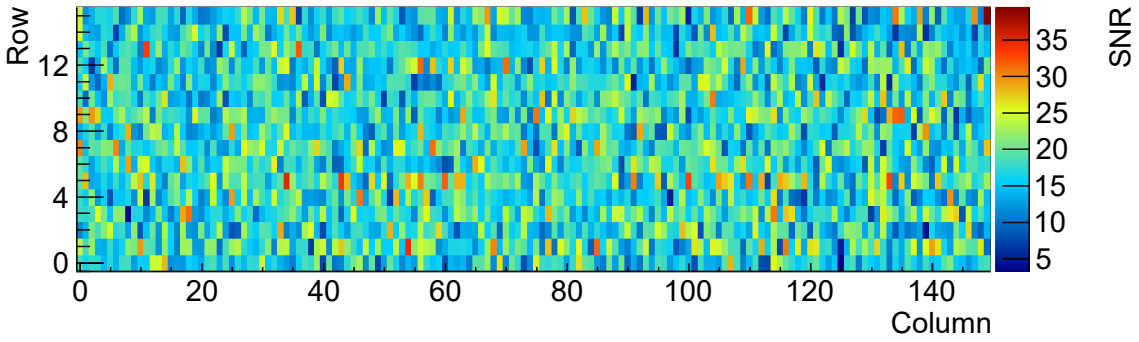


Figure 7.20: The SNR of the NMOS Matrix. It has been calculated from S-curves, using the definition 7.11 as signal and the transition width as noise. The random spread of the SNR values is large. Most are between 5 and 30.

Both the signal and the noise map of the left half of the NMOS Matrix were homogeneous. Consequently, the SNR distribution in figure 7.20 is homogeneous, too. SNR was calculated using the signal definition from equation 7.11.

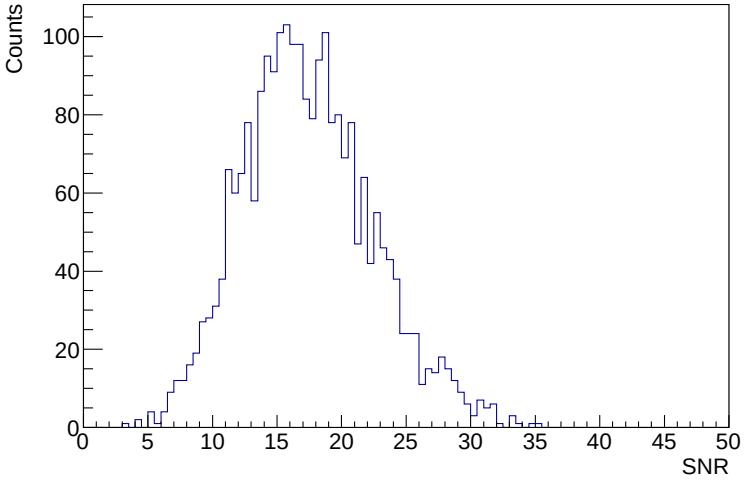


Figure 7.21: A histogram filled with the SNR values of each pixel in the NMOS Matrix without time-walk compensation. The distribution is wider than in the Analog Matrices, but the mean SNR is higher as well.

Figure 7.21 shows a SNR histogram of the whole matrix. The mean SNR is 17.4 with standard deviation of 5.0. The spread of SNR values in the NMOS Matrix is significantly larger than the spread in the Analog Matrices, and the mean SNR is larger, too. The analog part layout of the pixels is very similar in all three matrices. Therefore, it seems reasonable to identify the variations of the digitization stage as source of the widened spread. For particle detection in real physics experiments, the behavior of the pixels has to be unified.

7.5 Trimming

It is not only important for HV-CMOS detectors to be sensitive to even small signals, but also to have a uniform behavior of all pixels over the whole matrix. As seen in the previous section, mismatch and other effects can have a large impact on the uniformity of the sensor. This was illustrated by S-curve scans of all pixels (see chapter 7.4).

To correct for the relative shift of the S-curves, a trimming circuit is implemented on the H35Demo. The comparator of each pixel receives a global threshold from the bias block. This global threshold can be adjusted locally in each pixel by a 2-bit RAM setting. This RAM setting can take the values $N_{\mathrm{RAM}} = \{0...3\}$. The first two values decrease the threshold, the second two increase it. Another global 6-bit DAC, *PTrim*, defines the scaling $s(\mathrm{PTrim})$ of the local adjustment. All together, the local threshold Th_i of a pixel i can be described by

$$Th_i = Th_{\mathrm{global}} + (s\,(PTrim) \cdot (1.5 - N_{\mathrm{RAM}})) \tag{7.12}$$

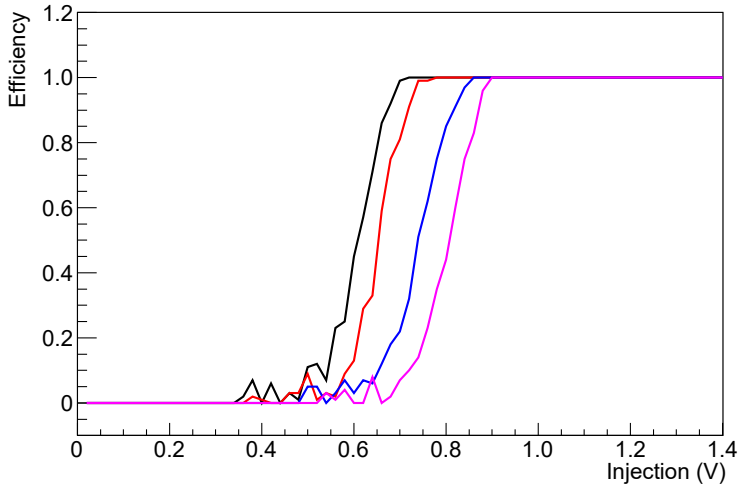


Figure 7.22: Trimming is based on adjusting the local threshold of each pixel so that all have the same detection threshold. This is achieved by writing the 2-bit trim value {0...3} into the RAM of each pixel, which shifts the local threshold up (0 black, 1 red) or down (2 blue, 3 magenta).

Figure 7.22 shows four S-curves measured from the same pixel with all four possible RAM-values. In this case the total detection threshold shift from the lowest to the highest setting is about 0.2 V. The data for these S-curves have been taken with rather small injection steps and high number of injections per step to obtain a smooth line. For the actual trimming procedure this is not feasible. As four S-curves per pixel are necessary, this process is very time consuming for a pixel matrix with a high number of pixels. In order to reduce the required trimming time to an adequate level, many things have to be optimized. These include a reduction of steps and injections per step, but also optimization of the procedure itself, e.g. measuring an entire row at once and improving the firmware and software for high data rate and efficient fitting.

The effect of the second parameter, *PTrim* is illustrated in figure 7.23. Just like in the previous figure, S-curves have been measured with all four possible RAM settings and the detection threshold has been determined by fitting an error function. This measurement was repeated 64 times, for each possible setting of *PTrim*. The resulting plot looks like

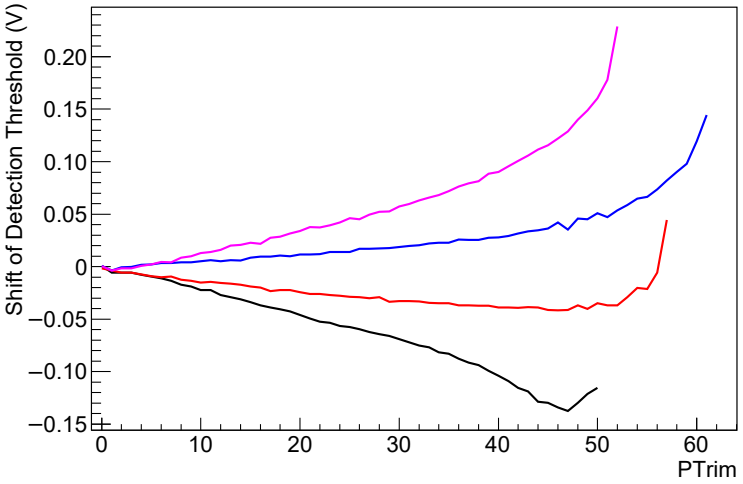


Figure 7.23: Detection threshold as a function of *PTrim*-DAC (matrix setting) for all four possible RAM settings (pixel setting). Above 40, the circuit starts to malfunction.

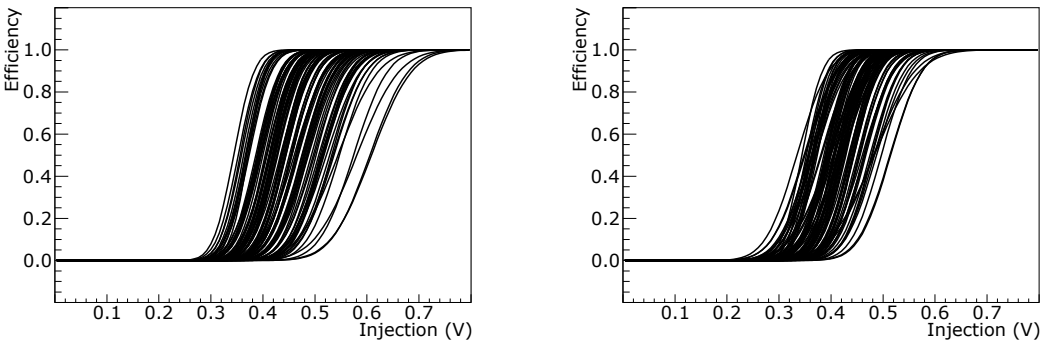
a fan for small values of *PTrim*. However, above ≈ 40, the linear behavior turns into an exponential characteristic, for all RAM values. The measurement is not possible anymore for values close to the maximum.

7.5.1 Trimming the standalone NMOS Matrix

The previous measurements demonstrated the principles and mechanisms of trimming the H35Demo. The following section shall illustrate the actual procedure of trimming half the NMOS Matrix without time-walk compensation.

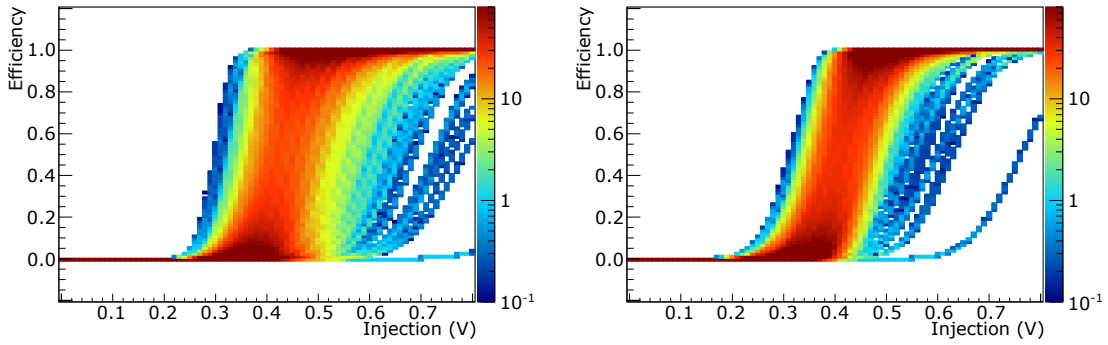
At first the value of *PTrim* is to be determined. This is done by running a full matrix S-curve scan with *PTrim* set to 0, to measure the primary spread of the detection thresholds without local adjustment of the global comparator threshold (Figure 7.23). A selection of such full matrix S-curve measurement is shown in figure 7.24a.

PTrim should be set to a value, at which both the S-curve with the highest and the lowest detection threshold can be shifted to the same value in the middle. However, it has turned out that the scaling of *PTrim* is also not uniform across the matrix. Therefore, *PTrim* is set to the value, which allows to shift the highest detection threshold onto the lowest and vice versa.



(a) **(b)**

Figure 7.24: The plots show S-curves of 100 pixels before (a) and after (b) trimming. The trimming target is 0.42 V and the trim DAC is set to 25.



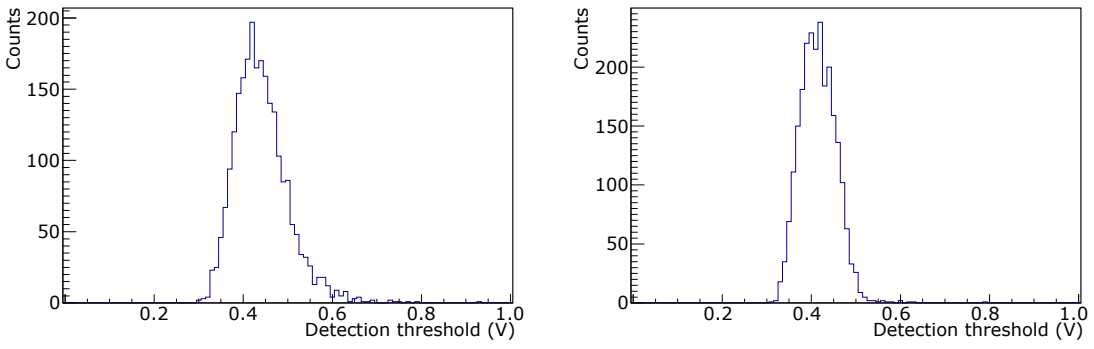
(a) **(b)**

Figure 7.25: The large number of pixels renders it impossible to display all S-curves at once. This S-curve density plot gives an impression on the spread of S-curves before (a) and after trimming (b).

With *PTrim* being fixed to a good value, S-curves of the whole matrix are taken again four times for each possible trim RAM setting. In general a binary search is preferable to this brute force approach, but for a search space of only two bits and an entire row being measured at once, no time improvement can be gained. To the contrary, a full data set with S-curves of all pixels and all possible RAM values, allows interesting insights in the trimming process and performance.

From the measured S-curves an average detection threshold has to be picked. This detection threshold is the value to which the S-curves of all pixels should be shifted. It is also referred to as the target threshold. Next, of the four possible S-curves of a pixel, the one with detection threshold closest to the target threshold is selected. This is repeated for every single pixel and the respective RAM values are stored and sent to the chip.

Finally, an S-curve scan of the whole matrix with the found RAM settings is conducted to verify the trimming success. Figure 7.24b shows the S-curves of the same pixels, but this time with the determined RAM values. S-curve comparison of an entire matrix is shown in figure 7.25. The color-code indicates how many S-curves intersect a bin and how long this intersection is.



(a) **(b)**

Figure 7.26: Histograms of the detection thresholds of all pixels in the NMOS Matrix without time-walk compensation before (a) and after (b) trimming. The comparably high setting of the trim DAC to 25, results in smooth Gaussian shape without additional features, which indicates successful trimming. However, the total spread remains rather large.

After trimming the global threshold can be decreased until the noise floor is reached, if a

very low comparator threshold is requested. Usually the effective threshold level, and thus the sensitivity of the sensor, can be substantially lowered after trimming. Before, it has to be set higher to avoid the noise floor of every pixel.

The spread of the S-curves, and consequently the success of the trimming, is best illustrated by filling the detection threshold of every pixel into a histogram. Figure 7.26 shows such histograms before (left) and after trimming (right). The standard deviation of the untrimmed S-curves is 61 mV, after trimming the standard deviation is reduced to 40 mV in this example. This is a reduction by over 34%.

On first sight, this is not a great value when translated to $\approx 230\,\mathrm{e}$. However, the premises have to be taken into account, which prevented this method of achieving a higher reduction, but had other benefits:

- It was requested to have not a single S-curve left untrimmed. Before trimming some S-curves were below noise floor, or above the S-curve range. Both render the S-curve immeasurable. Afterwards every single pixel was successfully measured with a detection threshold within the expected range.
  The price to pay for this success is a large value of *PTrim*. A large scaling factor on binary trim values naturally renders the result coarser than a smaller value.
- Each comparator has only two trim bits, meaning trim values ranging from 0 to 3. The trimming circuit is needed for each pixel, therefore the area scales with the number of pixels. The implementation used on the H35Demo is area saving, but does not allow a fine granularity, which limits the final trimming result.
- The H35Demo has no option to mask pixels. If some pixels, which perform way out of specifications (usually $\ll 1/1000$) were masked, *PTrim* could be greatly reduced and a higher granularity would improve the trimming result. Excluding these pixels from trimming is not an option, because they would cause a high noise rate on the readout, preventing actual signals from being transmitted.

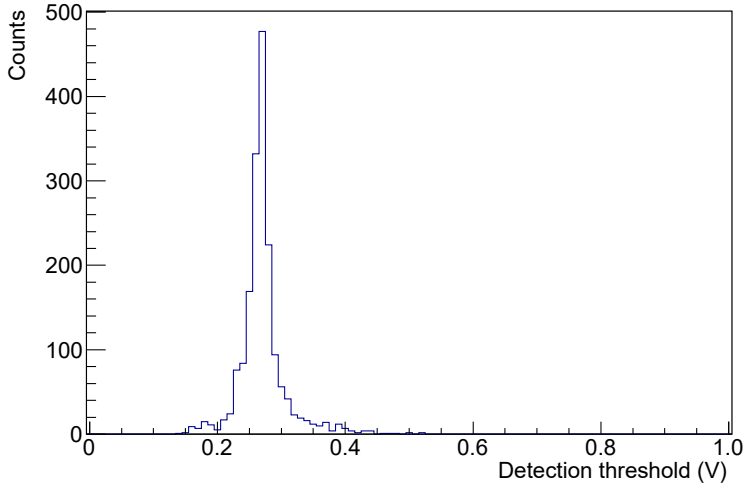


Figure 7.27: Histogram of detection thresholds after trimming with the trim DAC manually set to 15. The Gaussian is much more narrow, but rests on a plateau of entries from, with this setting, untrimmable pixels.

Figure 7.27 shows the detection threshold histogram of the trimmed matrix with a smaller *PTrim* than before. The distribution appears more narrow, but its sigma of 37 mV is only a little bit below the sigma of the previous trimming result. As anticipated is the peak in the middle of the histogram sharper than before, but is resting on a wider plateau of

detection thresholds. The entries of this plateau are formed by S-curves which were far away from the trimming target threshold and, due to the small *PTrim* value, could not be shifted far enough. Without this plateau, the trimming would have been a full success with a very small remaining sigma of less than 15 mV (about 85 e). Given the limitations of the trimming mechanisms on the H35Demo the first result is preferable for stable operation and a uniformly working pixel matrix.

As mentioned before, the large data set obtained by measuring S-curves of all pixels and all possible trim RAM values can be exploited for further analysis of the trimming circuitry of the H35Demo. Actually, picking the trimming's target threshold is not as straight forward as described before. A way to determine the optimal target threshold is described in Appendix E.7.

7.6 Time-walk effect

The time resolution of HV-CMOS sensors is dominated by the time-walk effect $\mathcal{O}(100\text{ ns})$. This effect occurs when digitizing the analog signal and arises from the finite speed of the charge sensitive amplifier.

Compared to the time-walk effect, the uncertainty in charge collection time is negligible. The electric field in the depleted active sensor volume drags the charge carriers quickly to the electrodes (charge collection by drift). CMOS sensors without depleted sensor volume collect charge carriers slowly by diffusion, which causes a significant contribution to the total time uncertainty.

The expected time-walk can be estimated by analyzing the amplifier's analog response to charged particles. An oscilloscope with high repetition rate monitors AmpOut. Note that AmpOut is not identical with the pixel's comparator input. While the comparator receives the amplifier's signal directly, the signal to AmpOut has to pass an analog buffer. Therefore, the result from measurements on AmpOut is not one to one transferable to the internal digitization process in the H35Demo. The time reference is provided by a scintillator. Its signal is fed into the same oscilloscope. On trigger the waveforms are automatically transmitted to a computer for further analysis.

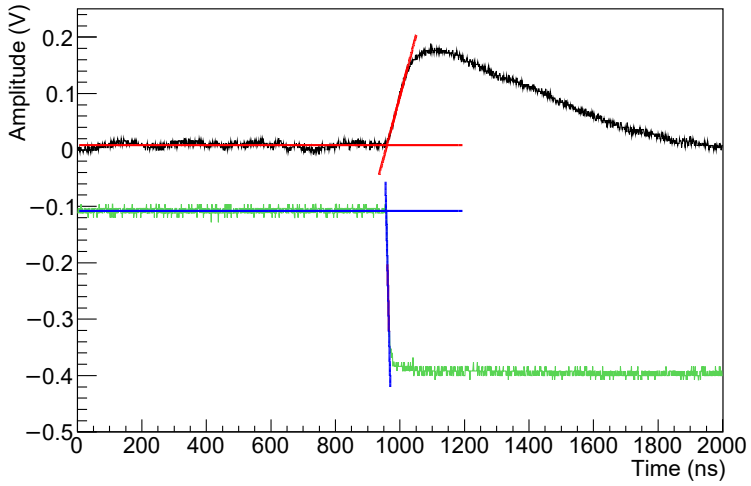
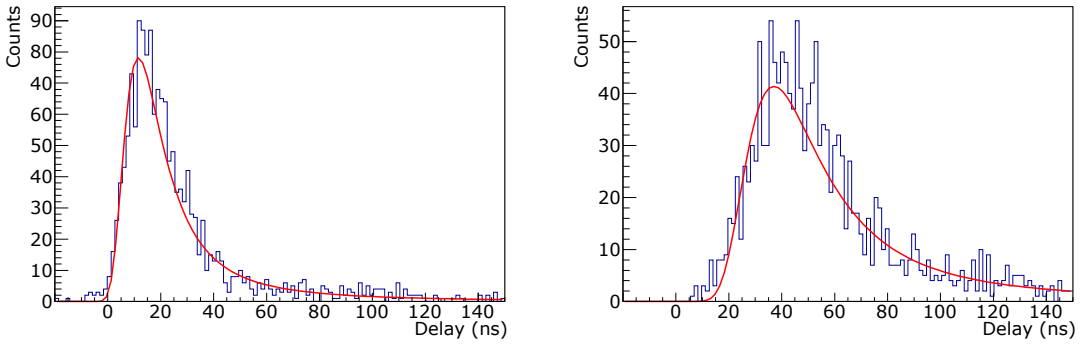


Figure 7.28: Low energetic electrons from a ^{90}Sr source pass through a H35Demo chip and a scintillator. The scintiallor signal (green) is used as time reference for the analysis of the H35Demo's signal (black). The baseline of both waveforms is averaged and a fit is applied to both slopes (red and blue).

A large fraction of the events have either a signal recorded on the amplifier waveform or on the scintillator waveform. These events will be excluded from the following considerations.

Figure 7.28 shows a typical example of an event from this measurement. The black line is the voltage level of the amplifier, green is the signal from the scintillator. A linear fit is applied to the negative trigger slope and is then intersected with the averaged baseline (both blue). This intersection point is used as time reference for the analysis of the amplifier's waveform.

A constant fit is applied to the baseline of the signal (red). This is necessary to get rid of a constant offset which can be introduced by the AC-coupling of the oscilloscope. The rise time is approximately 150 ns with the used settings. This is much longer than the few nano seconds rise time of the scintillator signal. Further, it is not only of interest to measure the point in time, when the signal started (Intersection point of the red lines), but also the point in time, when the signal crosses an arbitrarily set threshold. Therefore, for each threshold setting, the rising edge has been fit by finding the first point of the waveform being larger than the threshold and using few points before and after it for a linear regression for better stability.



(a) 20 mV threshold **(b)** 80 mV threshold

Figure 7.29: The delay between scintillator signal and threshold crossing for a given threshold above baseline. Small thresholds result in a narrow Landau-like distribution with low mean (a). For higher thresholds, the histogram is shifted to larger delays and widens (b).

Figure 7.29 shows the delay histogram from the same data set with a virtual threshold of 20 mV (7.29a) and 80 mV (7.29b). A small threshold results in an absolute small mean delay and a small standard deviation. The three times higher threshold shifts the mean delay to higher values an widens the spread. Consequently the best time resolution is a threshold of virtually zero. However, this is only true for a noiseless system. This measurement represents in this aspect an ideal system, as only oscilloscope snapshots were taken into account, which contained a verified signal. From these signals, attributes were extracted by fits. Naturally, such post-processing is not possible online for a vast number of pixels or channels.

The shape of the delay histogram is similar to that of a Landau-distribution. However, it is actually the convolution of a Landau- (charge deposition), a Gaussian (white noise) and a function for the time-walk effect (exponential-like).

The influence of the threshold on the delay was examined for thresholds between 0 and 0.4 V above baseline. Figure 7.30 visualizes data as two dimensional histogram. Low thresholds grant a sharp peak with low absolute delay, for higher thresholds the peak is shifted to higher delays and smears out. The mean delay between scintillator and signal starting point is below 18 ns (Figure 7.31). Increasing the threshold increases the delay proportionally

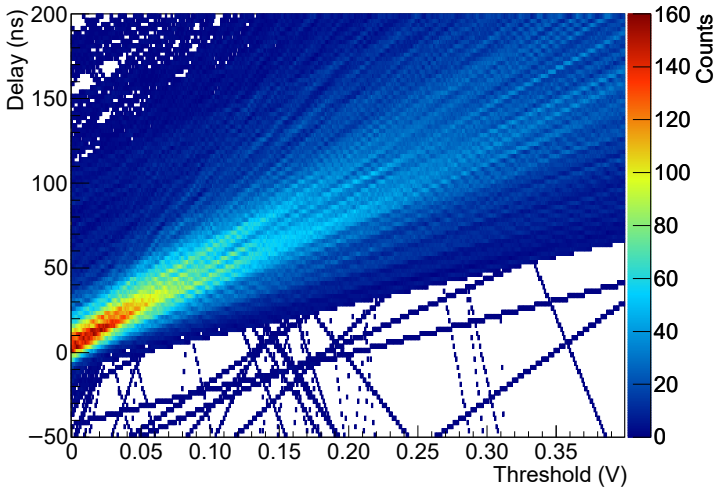


Figure 7.30: Two dimensional visualization of the analyzed data. The plot shows signal detection delay as a function of threshold. It was calculated on measured analog waveforms.
Small thresholds result in a small mean delay and small spread (area with high occupation). Going to larger thresholds increases the absolute delay (area with high occupation moves towards higher delays) and the time resolution (spread of entries) deteriorates.

until a threshold $\approx 0.1\,\mathrm{V}$. For even higher thresholds, the delay keeps increasing, but with reduced slope.

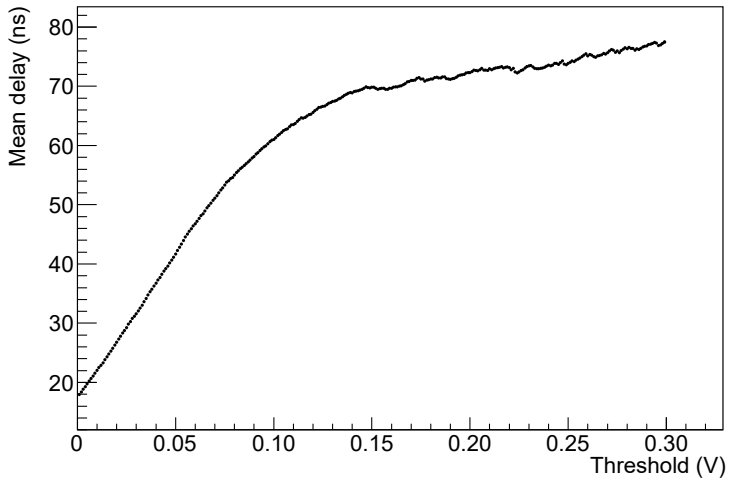


Figure 7.31: The mean delay between scintillator signal and threshold crossing as a function of threshold voltage. It rises linear at first, and flattens for higher values.

So far only the effect of threshold on delay was discussed. At the same time, an increased threshold does not only reduces the noise rate, but also might reduce the efficiency, as small signals do not cross the threshold level anymore.

Due to the described analysis method, the noise reduction effect could not be determined, but the reduction of signals exceeding the threshold could be monitored. In figure 7.32, the efficiency is plotted over simulated threshold. As the original signals originate from low energetic electrons, their energy deposition follows a Landau distribution. Thus, signals of various voltages were recorded. Above 80 mV threshold level, the efficiency starts to drop. Detection level of 50% is reached at a threshold of 160 mV. A threshold above 400 mV

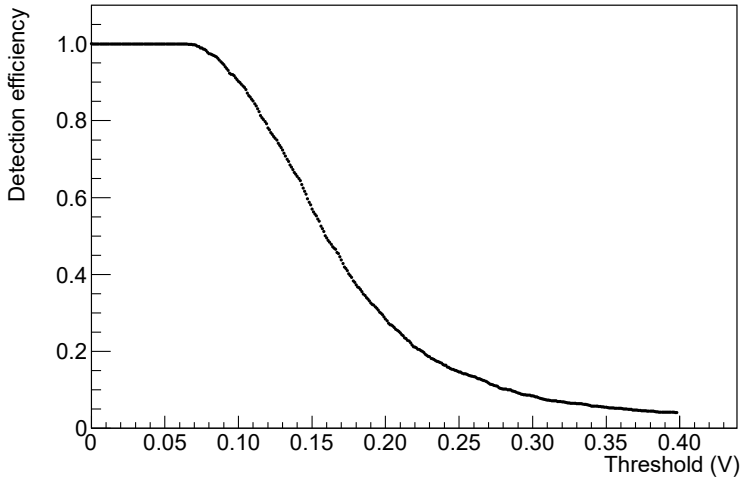
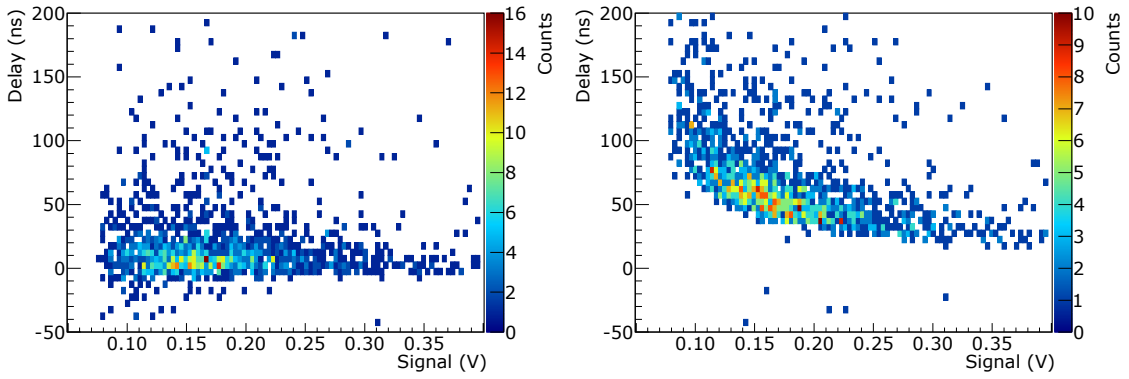


Figure 7.32: Detection efficiency as a function of threshold voltage. The threshold setting has not only an effect on timing, but also on the efficiency. Only strong signals can be detected for higher thresholds. The reason that a threshold greater than zero is needed is not part of this analysis: noise.

would discard most signals.



(a) 0 mV threshold **(b)** 100 mV threshold

Figure 7.33: The two dimensional histograms show delay over signal for a fixed threshold.
(a) The threshold is set to 0 mV. The delay depends only on the signal.
(b) A rather high threshold is 100 mV. Weak signals have a delay of around 100 ns. Stronger signals reduce the delay exponentially to a residual value around 30 ns.

The time-walk of a sensor can be displayed as two dimensional histogram with signal on the x-axis and delay on the y-axis (Figure 7.33). Each histogram entry represents a single event with its individual energy deposition. Therefore the spread is large, nevertheless accumulations in certain areas can be spotted.

With a virtual threshold equal to zero, the delay is rather independent of the signal (Figure 7.33a). An increasing threshold starts to tilt the distribution (Figure 7.33b). Weak signals have a high delay of around 100 ns. Towards stronger signals, the delay drops exponentially until it reaches a plateau of about 20 ns at 0.25 V signals.

Figure 7.34 shows the same plot for a realistic setting of the threshold. The time-walk is roughly 50 ns.

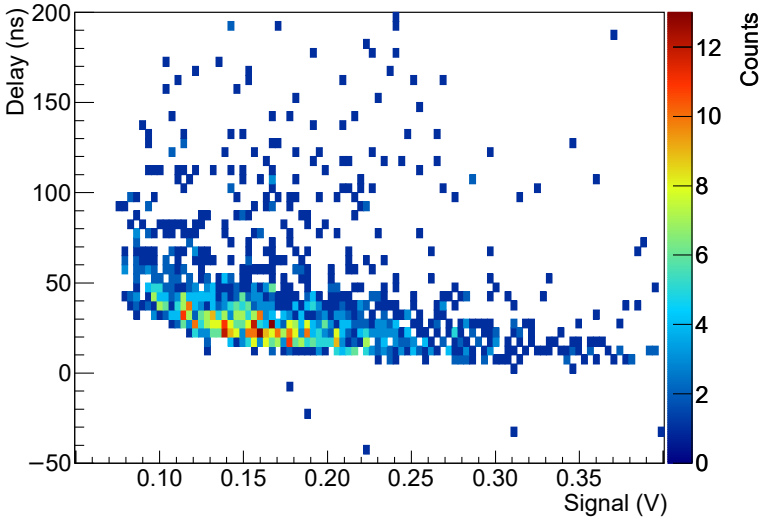


Figure 7.34: Two dimensional plot of signal delay over signal at a threshold of 40 mV. This is a good value for low noise, high efficiency and low time-walk. It is ≈ 50 ns.

7.7 Time-walk compensated comparator

The time-walk compensated comparator is designed to react slowly to a high signal and fast to a small signal. As the amplifier's temporal behavior is just inverted, the time-walk is compensated. It is evident that the timing properties of both components have to match, otherwise the time-walk is not fully compensated or over compensated. The working principle is based on discharging a capacitance. The voltage level of the capacitance is compared to the threshold voltage *nThTwPix* in a comparator. Details are explained in chapter 5.1.2.

Depending on the primary signal strength, the process of discharging has a different speed.

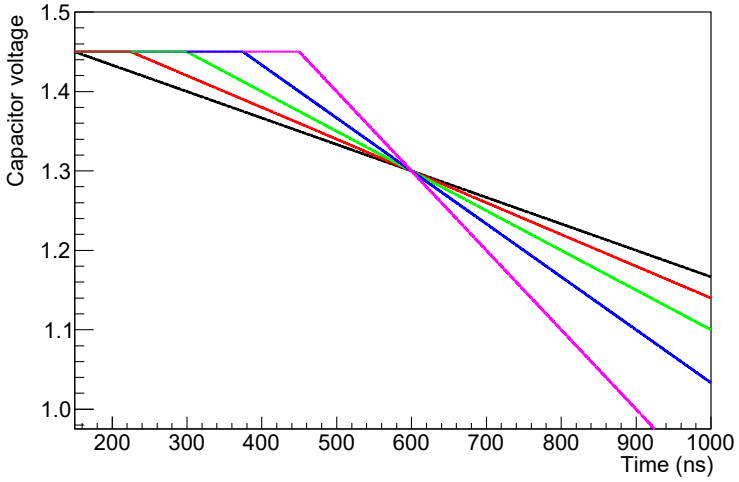


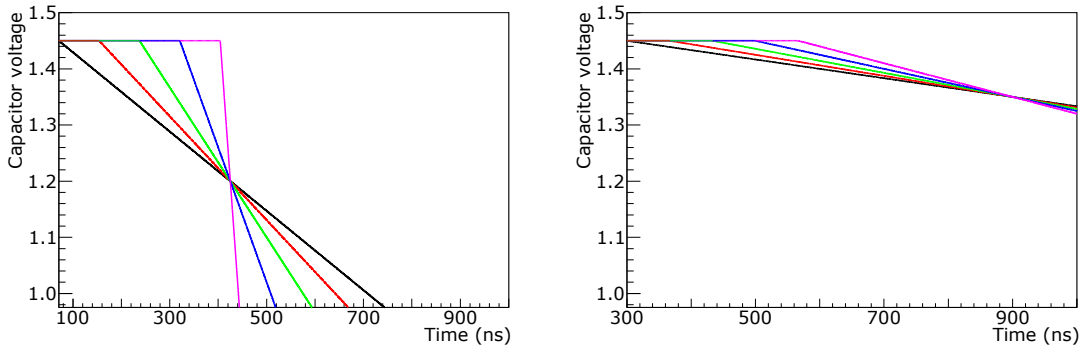
Figure 7.35: The output of the amplifier is not directly connected to the comparator, but is used to discharge a capacitance. The discharging of this capacitance C is sketched over time. It starts early for large signals (black) and late for small signals (magenta), due to the time-walk effect. The discharging speed however, is slow for large signals and fast for small signals. For good settings, a point can be found, in which all discharging curves, independent of the signal, intersect. The threshold of the comparator is set into this intersection point.

Therefore, plotting voltage over time for different signals should result in lines intersecting in a single point (Figure 7.35). Threshold voltage *nThTwPix*, also referred to as *Th2*, has to be set in this intersection point.

Due to production mismatch, the temporal behavior of the time-walk compensated comparator can not be perfectly matched to the amplifier's by design. Thus, three settings require fine tuning:

- The idle load of the capacitance is defined by the pull-up voltage *nVPlusTWPix*.
- Its antagonist is DAC *VNTwDown*, which acts as idle pull-down to ground. Without signal and depending on the actual settings the inflow to and outflow from the capacitor reach equality. The voltage is between *nVPlusTWPix* and ground.
- In case of a signal, the capacitor receives an additional outflow channel via a transistor, biased by the DAC *VNTw*.

Together they define the starting voltage of the discharging curve and its slope in case of an event. As the discharging curves are not infinite, the intersection point has to be in the range of *Th2* ($\approx 1...2\,\mathrm{V}$) and the slopes can not be too steep nor too flat. A too steep slope results in a strong sensitivity towards variations of *Th2* or even the lack of an intersection point (Figure 7.36a). A too flat slope increases the total delay time of the detector and the dead time of the hit pixel (Figure 7.36b).



(a) **(b)**

Figure 7.36: Time-walk compensation is only given, if the intersection point has a voltage in the operational range of the comparator. There are several settings which can be adjusted to achieve this. However, the slope of the discharging curves has to be considered. A too steep curve (a) creates a strong sensitivity of the remaining time-walk on the threshold voltage. If discharging is too slow (b), the absolute delay can get very large.

7.7.1 Time-walk measurement with test signals

The working principle of the time-walk compensated comparator is to be characterized. At first the discharge curves need to be measured for different primary signals. The above mentioned variable DAC settings and external voltages have to be optimized for good time resolution.

However, direct measurement of the discharging curves is not possible, as there is no way to access the voltage of the capacitor. It is possible though, to measure the delay between the injection of a test signal and the end of its digitization (time stamp generation). For fixed settings, including the injected charge, the delay can be measured as a function of *Th2*. Each measurement provides one tuple (voltage, delay). With sufficient statistics, a smooth line is obtained. The measurement is then repeated for a different injected charge. The result of this measurement is shown in figure 7.37.

The *delay* is defined as

$$\text{Delay} = \text{TriggerTimestamp} - \text{HitTimestamp} \tag{7.13}$$

The used unit is 40 ns. The trigger time stamp is generated by the FPGA, the hit time stamp is generated by the H35Demo. Both counters run with the same clock but with a well-defined shift.

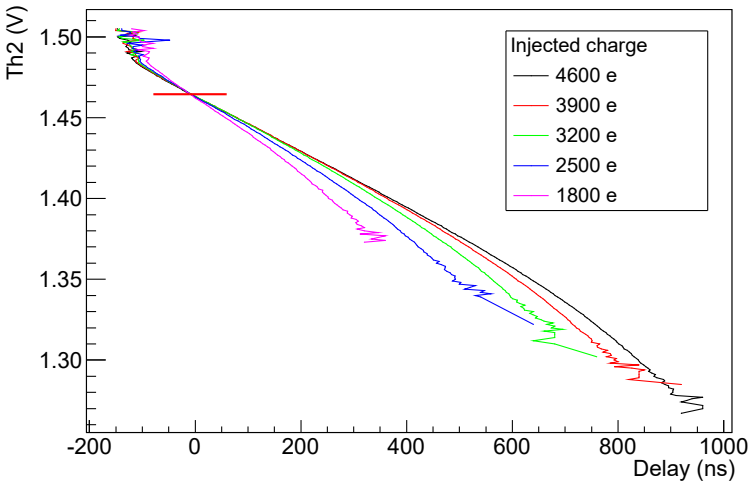


Figure 7.37: Each discharging curves in this plot has been measured by repeatedly injecting charges and varying the threshold while measuring the total delay between injection and read out. Each color represents a different injected charge. The point of least time-walk was found by fitting the data points of each color and intersecting them. The threshold of the intersection point is ≈ 1.46 V.

The injected signals range from 1800 e, which is about 1/2 the signal of a MIP, up to 4600 e. The change in curve shape is large for small signals and small for larger signals, therefore signals larger than 4600 e do not have to be considered.

If threshold *Th2* is too large or too small, the efficiency drops and a measurement is not possible anymore. This effect can be seen on the upper and lower end of the plot, where the discharge curves start to be spiky. Only few injected signals are detected without precise time measurement.

The best possible *Th2* in this area is determined by applying linear fits around the intersection point to each curve. Those fits are then intersected with each other and the mean of all is taken as point of no time-walk. In the example shown in figure 7.37, this point is reached at *Th2* = 1.4645 V.

7.7.2 Time-walk measurement with a strontium-90 source

In the previous measurement, test signal injections have been used to find good settings for the time-walk compensated comparator. Now, these settings are going to be used to measure the effect on timing in pixels with and without time-walk compensation. ^{90}Sr is used as signal source. The energy of the emitted β-rays is sufficiently close to the energy of minimum ionizing particles to represent a worst-case scenario with respect to time-walk.

The collimated ^{90}Sr source is placed above the H35Demo. A scintillator is placed below as time reference (Figure 7.14)[2].

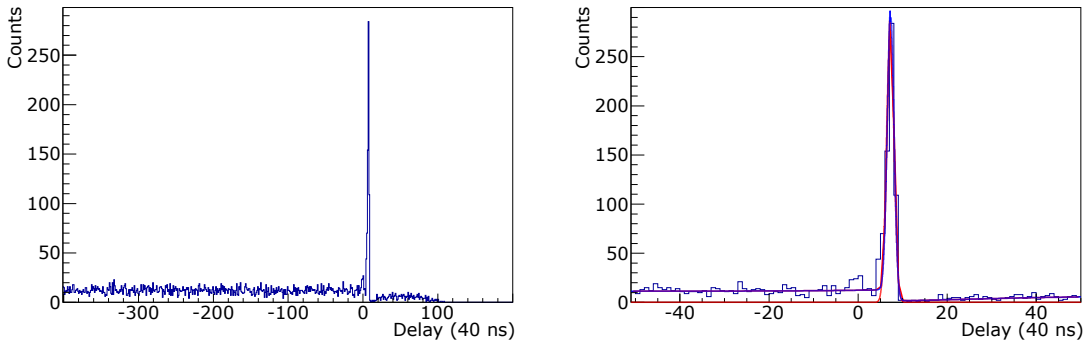
[2]The custom comparator PCB is shown in Appendix E.8.

The monolithic readout is used to collect data from the NMOS Matrix. Scintillator information is stored as second time stamp. This data are read out to a computer and the difference of the two time stamps is calculated. The result is filled into a histogram. Due to the size of the time stamps, the range of possible values is ±65000 steps, with each step being 40 ns long. This range is large compared to the expected delay.

Figure 7.38 shows a segment of the obtained histogram. It was measured on a pixel in the left half of the NMOS Matrix with a normal comparator. There is a distinctive peak on nearly uniform background.

This background occurs, because the scintillator is large compared to the measured pixel. Therefore, it is likely that the time stamp of an electron that causes a signal in the pixel, is set in relation to the scintillator time stamp of a different electron which has not passed the measured pixel. The background peak in positive delay direction is about 100 time steps wide. Its origin is not known but most likely a similar effect. For analysis, the background at negative delay is assumed constant, the background in positive direction Gaussian. The signal peak is fit with a Gaussian function. The total fit is shown in blue, each component in red.

The delay peak of this pixel is five bins wide and has a standard deviation of 30 ns.



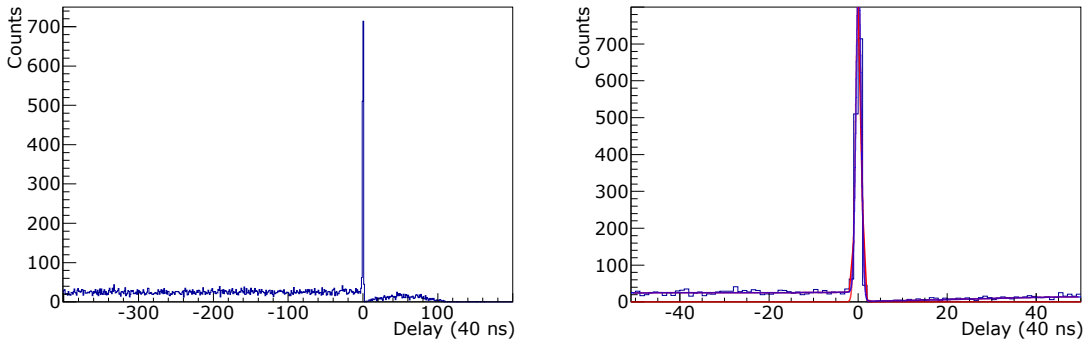
(a) **(b)**

Figure 7.38: The histograms show the delay between the time stamp generated by the H35Demo and the time stamp generated by the FPGA from a scintillator signal. The pixel is located in the left half of the NMOS Matrix and has no time-walk compensation. The measured particles are electrons from a ^{90}Sr source. The background originates from random coincidences. For fitting the data, the background towards negative delays was assumed constant, background towards larger delays was assumed Gaussian. The peak itself is Gaussian. The standard deviation of the signal peak is 30 ns.

A pixel on the right side of the H35Demo's NMOS Matrix with time-walk compensated comparator was measured in the same way and the obtained data were analyzed likewise. Figure 7.39 shows the resulting histogram. The previously described fit was applied to this histogram, too.

The delay peak of a pixel with time-walk compensated comparator is two bins wide and has a standard deviation of 23 ns.

We observe a relative shift between the delay peaks of the normal comparator and the time-walk compensated comparator. The time stamp of the pixel with time-walk compensation is on average generated 288 ns later, than the time stamp of a pixel without time-walk compensation. This additional absolute delay is the price to pay for a better time resolution with the time-walk compensated comparator on the H35Demo.



(a) (b)

Figure 7.39: Delay histogram of a pixel with time-walk compensation. It was obtained and analyzed just as the histogram presented before. The standard deviation of the delay is 23 ns. However, the absolute delay is larger than without time-walk compensation.

7.7.3 Landau-distributed test signals

The previous measurement requires a lot of hardware, a ^{90}Sr source and much time. An alternative was necessary. Instead of using particles from a source and measuring the time difference between the scintillator time stamp and the sensor time stamp, test signal injection can be used. These test-signals are Landau-distributed injections in random order and with random timing (Figure 7.40). Each set of injections consists of one thousand injections, the injection voltage is chosen randomly from the Landau-distribution. One thousand injections per set are sent. The delay is averaged within such set for sub-time-stamp resolution.

The Landau-distribution was obtained by fitting a histogram from an actual source-measurement, with depletion voltage set to 0. Without additional depletion voltage, the signals are small and represent the worst case time-walk scenario. The most-probable value of this Landau-function is found to be 1650 e and is cut off at 7000 e. Injection have been calibrated to generate the requested number of electrons.

In this way, a measurement on a pixel with normal comparator is conducted. Figure 7.41 shows delay over injected charge. Very low injections can not be measured properly, because

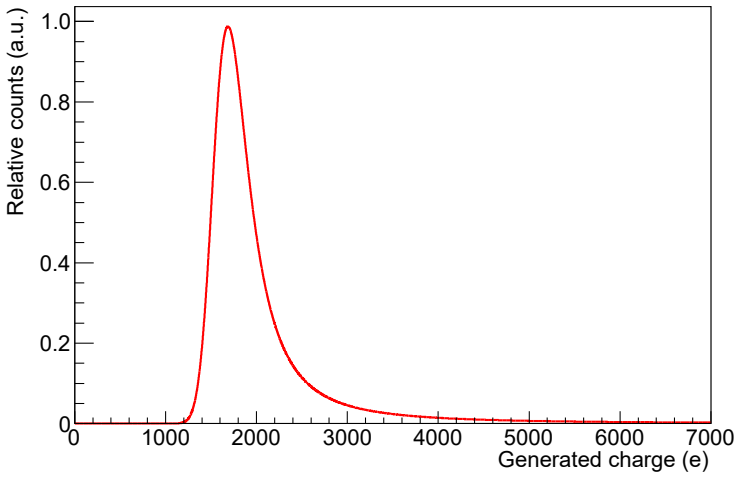
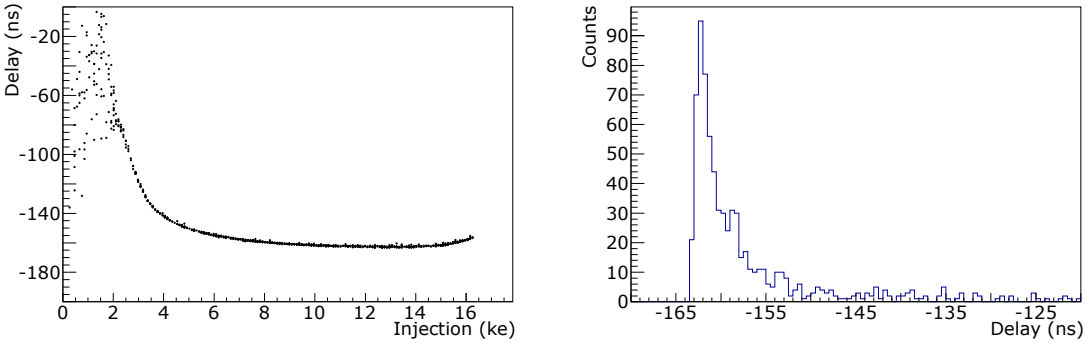


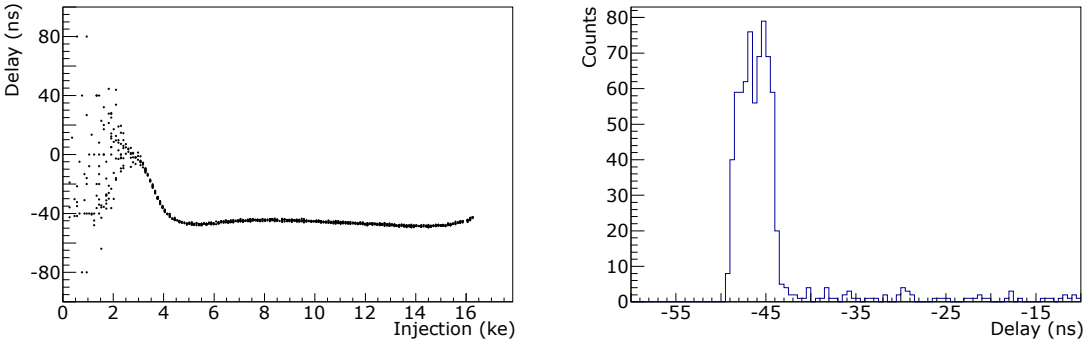
Figure 7.40: The plot shows the Landau-distribution used for generation of test signal injections. The most probable value is set to 1650 electrons. This value originates from measurements without additional depletion voltage and therefore smallest possible signals.

the detection efficiency is low. Consecutively, for further analysis only measurement points above detection level (>50% efficiency) are used. For better comparability, these points are filled into a histogram. The entries in the histogram filled with data measured on a pixel with normal comparator, are arranged alone the typical exponential decay, according to the time-walk effect. The standard deviation of the histogram without time-walk compensation is 8 ns.



(a) **(b)**

Figure 7.41: Landau-distributed test signals have been injected into an NMOS pixel without time-walk compensation. The left plot (a) shows delay over injected charge, the right plot (b) shows a histogram made from these data above threshold level. It shows a small delay for signals above a certain charge, with a large tail from signals below ≈ 4 ke. The standard deviation of the delay is 8 ns.



(a) **(b)**

Figure 7.42: The same plots as in figure 7.41, but for a pixel with time-walk compensation. The delay histogram (b) has a more box-like shape without tail. The standard deviation of the delay is 4 ns.

This is in contrast to the shape of the histogram from the time-walk compensated comparator measurement in figure 7.42. It has a rather box-like shape which is in line with the expectations. The standard deviation is 4 ns, this is an improvement of 50%.

8. The Belle II experiment

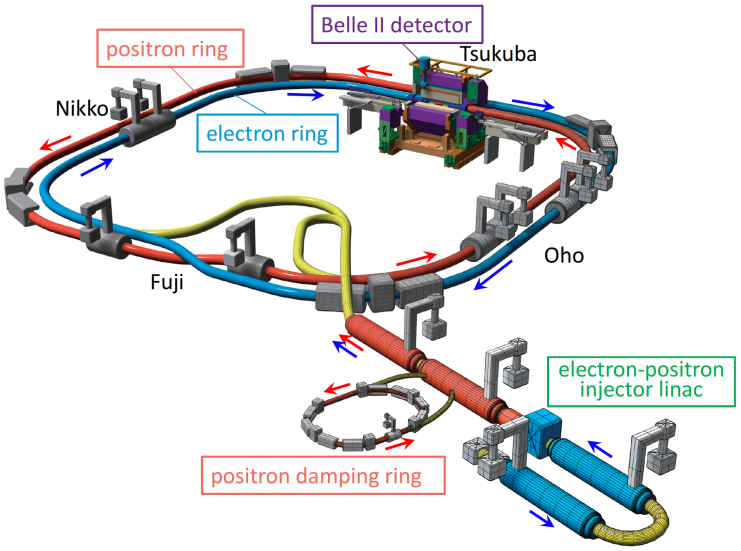


Figure 8.1: Schematic drawing of SuperKEKB with Belle II detector in the interaction point. (From [77])

KEKB and its upgraded version SuperKEKB is an electron-positron accelerator, operated by KEK, the High-Energy Accelerator Research Organisation, in Tsukuba, Japan [77] (Figure 8.1). Compared to LHC, the energies are moderate and the beam energies are not identical: The electron beam is operated at 7 GeV and the positron beam at 4 GeV. The center-of-mass energy is 10.38 GeV, the resonance energy of the Υ(4S), a $b\bar{b}$-meson. Therefore, SuperKEKB is often referred to as a b-factory, the B in its name reflects this. The beams are brought to collision in the Belle II detector with an instantaneous luminosity of up to $2.4 \cdot 10^{34}$ $cm^{-2}s^{-1}$ in mid 2020, which is world record [78]. 2503 bunches are stored in the accelerator ring, each has a length of 5 or 6 mm. The bunch spacing is 98 ns.

Belle II has a similar overall structure as ATLAS or CMS at LHC, but is much smaller (see chaper 4.1) [80]: For tracking, a pixel detector (PXD) based on DePFET sensors (see chapter 2.1) [81], a silicon strip vertex detector (SVD) and cell drift chambers (CDC) are implemented in close proximity to the interaction point. Due to their small distance to the beam (14 mm radius and 22 mm radius), the inner layers of PXD have to deal with a large amount of beam-induced background. To keep event pileup at a reasonable level, the readout frames are 20 µs long (integration time). Particle identification is executed by Cherenkov radiation detectors TOP and ARICH (see chapter 2.1), the magnetic field of up to 1.5 T for trajectory bending is provided by superconducting solenoid magnets. Calorimetry is the outermost component with RPCs (see chapter 2.1) as sensing elements. A schematic drawing of Belle II is shown in figure 8.2.

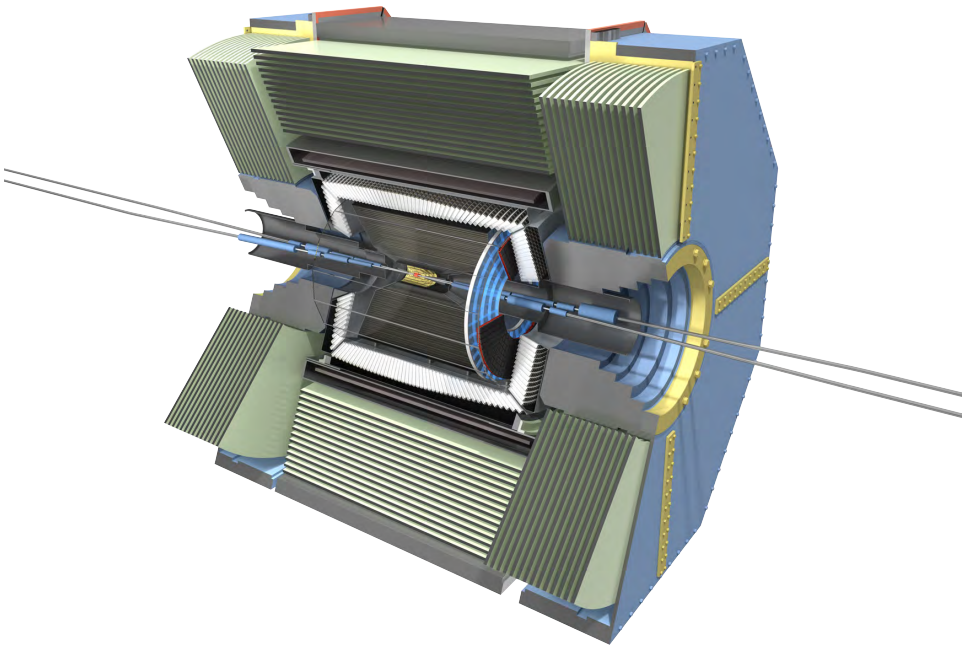


Figure 8.2: Schematic drawing of the Belle II detector. (From [79])

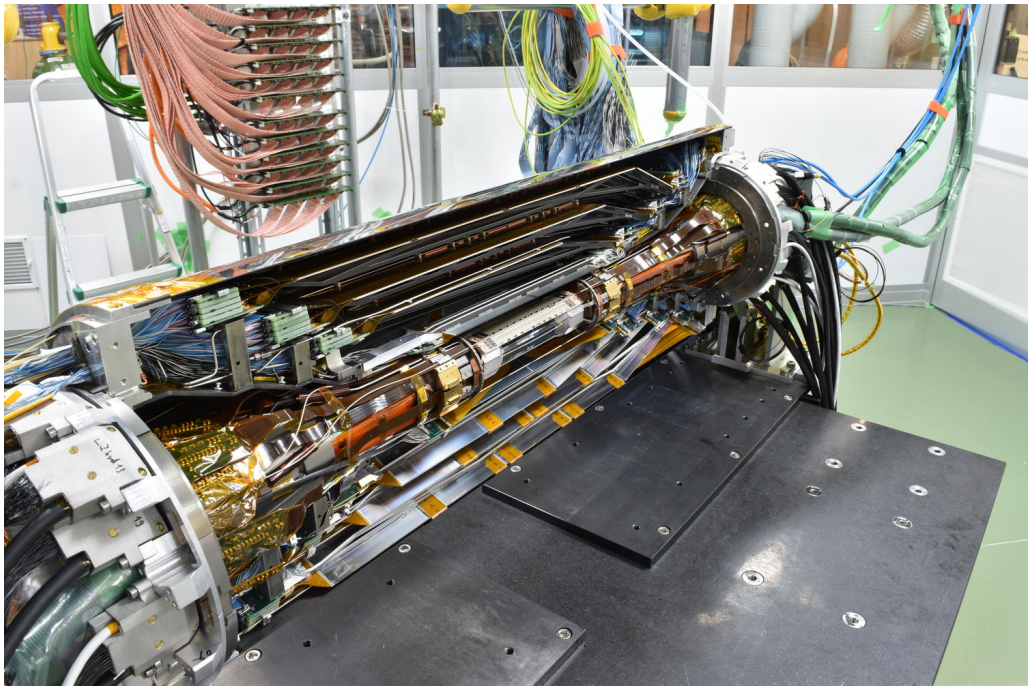


Figure 8.3: Picture of Belle II vertex detector VXD. (From [82])

The vertex detector of Belle II (VXD, combination of PXD and SVD) has been designed with minimal material budget (Figure 8.3). This is achieved by omitting most mechanical structures in the region close to the interaction point. The sensors have to be self-stable and are held in place by end-caps. This formulates the need for long monolithic sensors with minimum power consumption. The pixel size of PXD are between 50×55 μm^2 and 50×85 μm^2.

For a future upgrade of the pixel detector, we have proposed HV-CMOS sensors [83]. Construction of some 10 cm long self-stable HV-CMOS sensors at minimal thickness is not intrinsically possible as reticle size is in the order of 2×2 cm^2. However, it seems possible to leave several reticles undiced as one piece of silicon. Anyway, a 10 cm long sensor of standard thickness (≈ 150 μm) needs mechanical support to not bend or break from its own weight, while a unthinned sensor (≈ 700 μm) has the required rigidity, but adds unnecessary silicon to the material budget. This issue is tackled by selective thinning: only the middle area of the sensor is thinned, while a frame of thicker silicon provides the required stiffness.

Both mentioned techniques are not new, but have not been used on HV-CMOS sensors, therefore their suitability has to be proven, while unexpected events by thinning or crosstalk from one reticle to the other has to be excluded.

9. H35Demo module for Belle II

9.1 Selectively thinned sensor module

In order to prove the suitability of selectively thinned HV-CMOS sensor with more than one reticle per die, an undiced H35Demo wafer has been sent to Fraunhofer IZM[1] for post-processing. The wafer was thinned down to 450 µm and cut into modules of one to three sensors per module. Afterwards, each module has been further thinned in the central region only, down to 100 µm using deep-reactive ion etching.

Figure 9.1: Backside of a 7.5 cm long, selectively thinned H35Demo module.

Figure 9.1 shows a photograph of a face-down module composed of three H35Demo sensors. The thicker frame is only at the edges of the module, not on the edges of the individual sensors. The module is approximately 7.5 cm long and 2 cm wide, which is in the order of the sensor modules required for Belle II VXD.

To evaluate selective thinning of the modules, one of them has been cast in resin, cut and polished. The cross section was investigated by microscope, a picture is shown in figure 9.2. Deep-reactive ion etching allows for silicon milling with high aspect ratio (ratio between carved trench depth and width). In this special case, this means nearly 90° angles are possible. The ground of the etched area is plane already at a distance of about 300 µm from the edging corner. Typical for wet etching is over etching under the cliff, similar to a cave. The used combination of chemical and physical etching minimizes this effect.

[1] https://www.izm.fraunhofer.de/

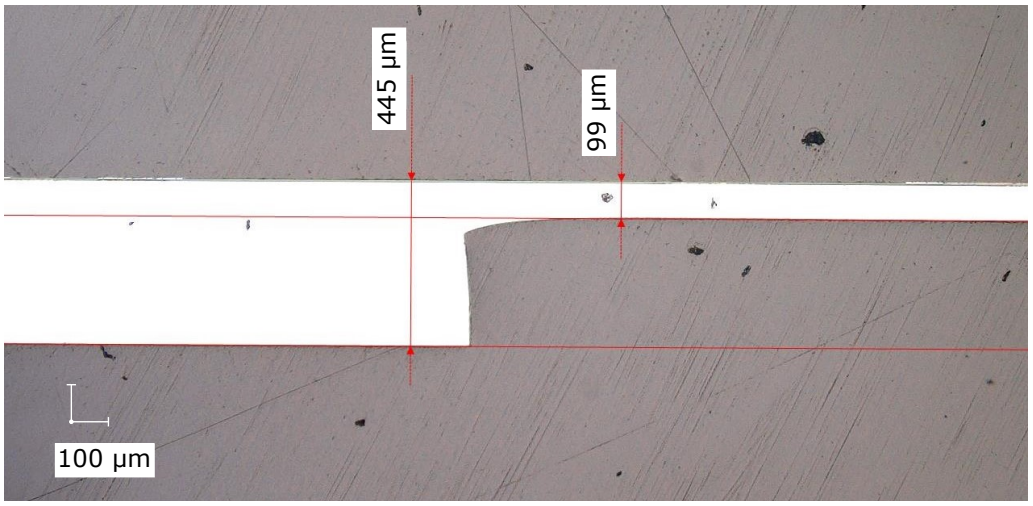


Figure 9.2: Microscopic view of the cross section of a selectively thinned H35Demo module. The thicker frame (left) provides rigidity, the thinned inner area (right) grants minimal scattering potential and energy loss for traversing particles.

9.2 Setup for H35Demo module

The characterization setup for the H35Demo Modul has been developed by J. Bader within a Master Thesis under the author's supervision [84] (Figure 9.3).

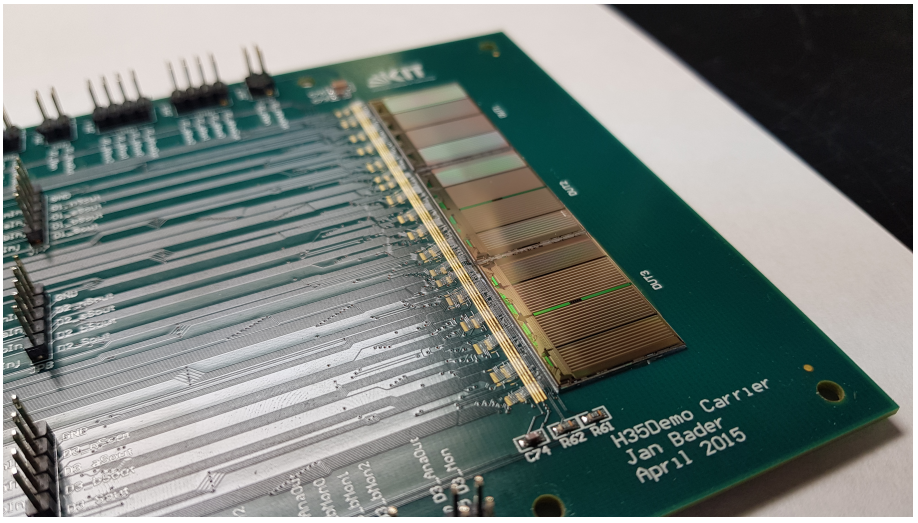


Figure 9.3: Photograph of the carrier circuit board for a H35Demo module. Up to three H35Demo sensors can be operated in parallel. (From [84])

Its hard- and firmware is based on the GECCO system (see chapter 16) and the single chip characterization setup for H35Demo (see chapter 6), its software is derived from GECCO telescope software (see chapter 16.3). The setup supports the operation of up to three H35Demo sensors, with four matrices each, in parallel.

9.3 Proof of principle for H35Demo module for Belle II

H35Demo is in all four matrices in every sense a fully functional design. Therefore, proving functionality of a selectively thinned H35Demo module, requires only verification of functionality of a single pixel on each of the three sensors.

The setup has been built and commissioned during the Master Thesis of J. Bader. All

twelve matrices are configurable and show responses, on analog outputs (Analog Matrix A and B) as well as on the monolithic output (CMOS Matrix) [84].

In the NMOS matrices, only the first stage amplifier is operational, as one supply voltage required for the second stage amplifier is, due to a PCB design flaw, too weak. The monolithic readout of the NMOS matrix works, but shows either no hits, or noise hits for a very low threshold.

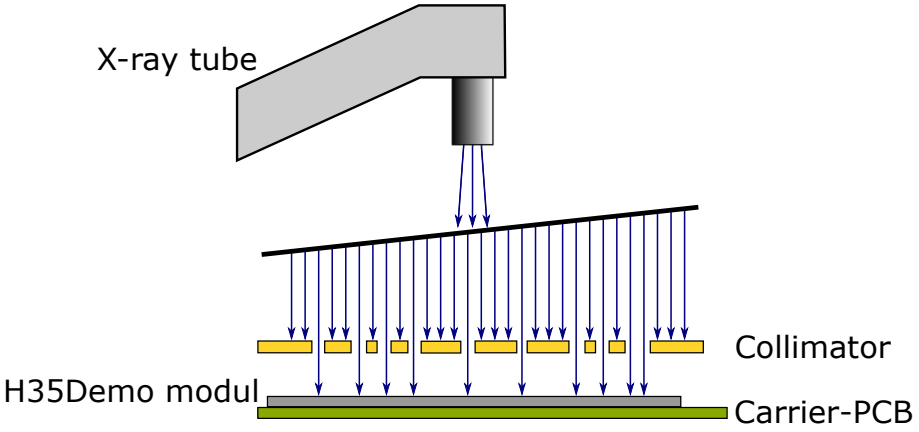


Figure 9.4: Schematic drawing of the H35Demo setup with collimator, being illuminated by an X-ray tube.

The first functionality tests have been conducted per matrix with electrical test signals and localized X-rays from a small ^{55}Fe source. For a whole module test, beam time in an X-ray tube (see Appendix B.2) was booked, which allows for illumination of a larger area at a significant rate (Figure 9.4).

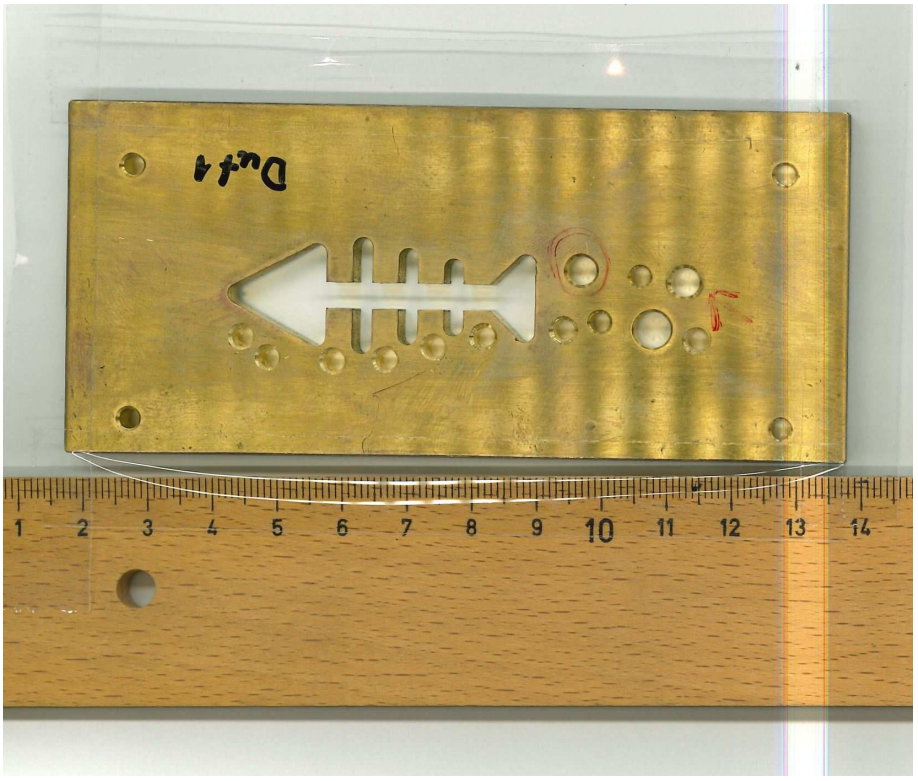


Figure 9.5: Photo of the collimator used in this measurement to screen some areas of the module from X-ray illumination.

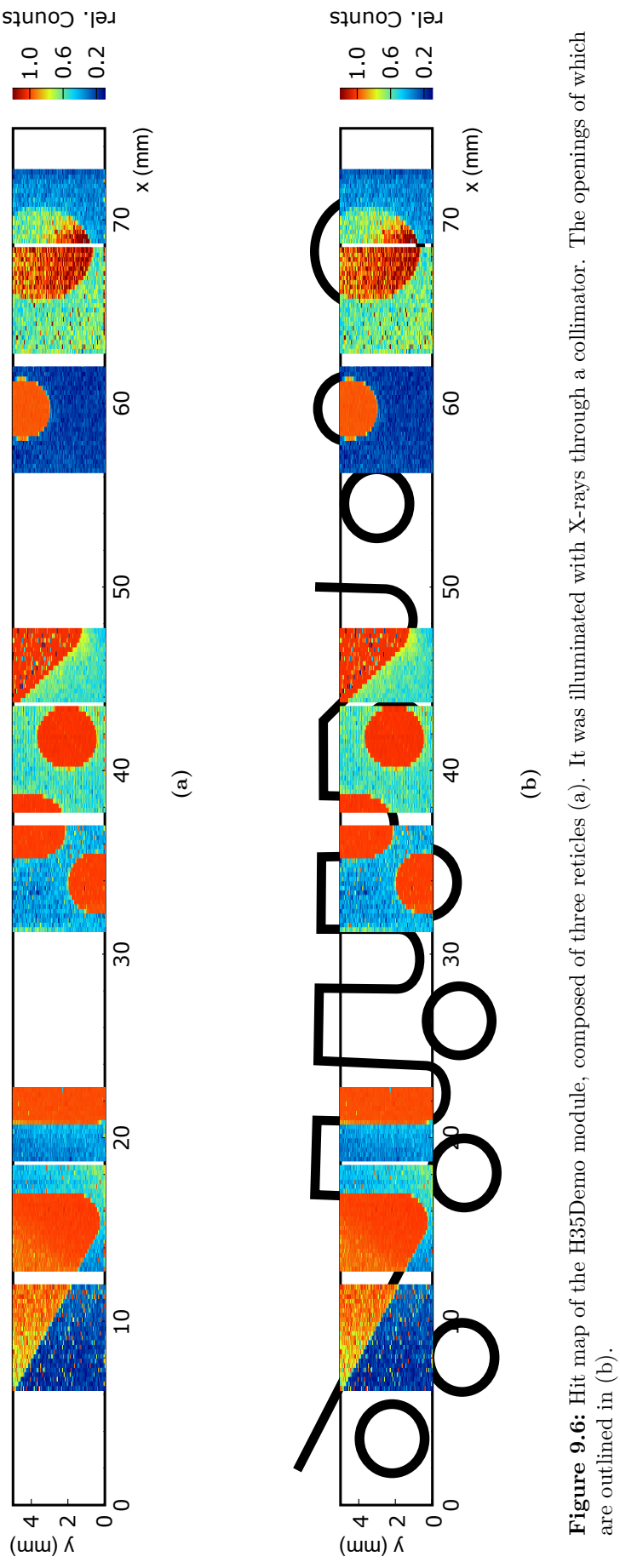


Figure 9.6: Hit map of the H35Demo module, composed of three reticles (a). It was illuminated with X-rays through a collimator. The openings of which are outlined in (b).

A brass-collimator screens the module partially from the incoming X-rays. Without collimator, the picture would be quite homogeneous and noise could not be distinguished from signals. The 5 mm brass shields most X-rays, thus the hit rate in shielded and non-shielded areas is significantly different. Figure 9.5 shows a photo of the used shielding. During the measurement, the matrices are configured and the second stage amplifiers of each individual pixel are in turn connected to an oscilloscope for several seconds, which counts the signals.

The normalized hit rate of the first 100 columns of each measured matrix is plotted in figure 9.6a and set in relation to the outlines of the collimator in figure 9.6b. The white area between the measured matrices refers to the missing NMOS matrices, the some 10 µm wide gap between the matrices of one reticle and the few 100 µm between reticles.

This simple measurement proves that a selectively thinned HV-CMOS module of three reticles operates in the same way as individual dies.

10. H35Demo summary

The H35Demo is a milestone on the path towards the first detector equipped with HV-CMOS sensors in high-energy physics experiments. It is not only a full reticle size sensor but has also successfully demonstrated functionality of several important technology options and key features for future detector systems:

- A reliable front-end that combines high signal-to-noise ratio with fast response time at a promising power consumption. This is achieved by a smart segmented-pixel diode design and a multi-stage amplifier layout.
- Digitization of analog signals in the periphery reduces the noise floor, as no fast alternating, digital signals run across the pixel matrix.
- Capacitive readout of individual pixels or groups of pixels is an intermediate step on the way towards fully-monolithic sensors, as it combines well-proven readout ASICs with cutting-edge HV-CMOS technology.
- Digitization and serialization of the collected data, is the key property of monolithic sensors. Two out of four matrices of H35Demo feature the necessary circuits in their periphery. The area of the periphery is insensitive and in the design of a detector, this has to be respected by smart geometric arrangement of the sensor modules.
- A smart variant of the comparator used for digitization compensates for the time-walk effect. The time of threshold crossing depends on the signal: Large signals reach threshold faster than small signals. The time-walk compensated comparator reacts fast to small signals and slow to large signals, thus compensates the time-walk effect.
- The trimming circuit adjusts the threshold of every comparator individually, thus tackles the effect of production variances on the homogeneity of the sensor.

As part of this thesis, a characterization setup for H35Demo has been developed from scratch. It comprises a sensor-specific carrier board, versatile FPGA firmware modules for data transfer and a control software with graphic user interface and automation of complex measurement procedures.

The characterization of H35Demo has been presented in the previous chapters. The most important results and their interpretations are summed up below:

- The linearity of the front-end has been demonstrated by electrical test signals and X-ray illumination for both amplifier stages. The deviation from linearity is below 2% for the first amplifier and below 4% for the second amplifier. The effect of design variations and bias DAC settings has been studied. A general finding is that increased amplifier speed has to be paid by reduced signal-to-noise ratio or increased power consumption.

- The effect of production mismatch has been determined by S-curve measurement (NMOS Matrix) or by compilation of independent X-ray spectra for every pixel (Analog Matrices). The variations of signal and noise have been analyzed. The compensation mechanism of local threshold adjustment, the trimming circuit, is able to reduce the sensitivity distribution in the NMOS Matrix by 40%.

- Charged particle detection has been evaluated on the example of strontium-90 decay electrons. A scintillator was used to put an external time stamp on events, which allows timing evaluation. The time-walk effect has been theoretically investigated on analog signal waveforms. The time uncertainties of standard comparator and the sophisticated time-walk compensated comparator were determined. The time-walk compensated comparator was able to reach a time uncertainty of 23 ns with decay electrons and 4 ns with Landau-distributed test signals. This meets the requirements of ATLAS ITk after high-luminosity upgrade.

  Furthermore, the gathered data have been searched for changes of signal delay depending on pixel position in the pixel matrix. Correction for such effects on chip is not implemented and also not advisable, as it would increase the size of the insensitive periphery. A better solution is a look-up table on the readout FPGA, which can be fed with calibration data prior to the start of a measurement.

- Charged particles interact with the sensor material, generating a charge cloud when passing through it. The sensitive volume is only the depleted zone of the sensor diode. It is a few 10 µm thick (actual value depends on substrate resistivity and depletion voltage). Not much, compared to the 100 µm (thinned) to 700 µm (not thinned) thickness of the chip.
  The lateral size of the charge cloud caused by passing charged particles has been estimated by running a custom clustering algorithm on the collected data. The average cluster size was found to be 1.7 pixels ($\hat{=}$ 21 250 µm^2), but with a tail to up to 20 pixels ($\hat{=}$ 250 000 µm^2). This is an important information when it comes to tailoring an HV-CMOS sensor for a high-energy physics experiment, as the designer needs to calculate the expected hit rate from simulated event rate.

- Three H35Demo sensors were left undiced in order to prove the suitability of HV-CMOS sensors for multi-reticle modules. Such large pieces of silicon – in this case 2×7.5 cm^2 – can be implemented in detectors without additional support structures for minimal material budget. Sufficient rigidity of thinned sensors is provided by a silicon frame, which is left behind during thinning by deep-reactive ion etching. The feasibility of this concept has been demonstrated.

The characterization of the H35Demo revealed a lot of expected and unexpected information. These findings are the basis to design its successors, ATLASpix1 and MuPix8.

Part III

ATLASpix1 and MuPix8 – monolithic HV-CMOS sensors for particle physics experiments

In this section of the thesis, the major sensors of a combined ASIC production in *ams* aH18 technology are discussed. ATLASpix1 and MuPix8 share many design features but differ in some key properties, as they aim for different experiments: ATLASpix1 is a demonstrator chip for ATLAS ITk's outermost pixel layer and takes about half of the available reticle area. MuPix8 covers the other half of the reticle and is designed as demonstrator for the Mu3e experiment. Both sensors are very close to a final design. If the characterization can proof all functions fully operational, each can be scaled up to full reticle size. ATLASpix1 (Figure 10.1) and MuPix8 are considered the last large-scale test chip iteration prior to prototype construction for the respective experiments.

Chapter 12 introduces the Mu3e experiment. ATLASpix1 aims for the ATLAS experiment, which has already been described in the context of the H35Demo in chapter 4.

The designs of ATLASpix1 and MuPix8 sensors, are described in chapter 11 and chapter 13.

In preparation of the characterization, a new measurement environment has been developed by the author. The Multi-purpose Adapter Board Setup (MAB) combines sensor-independent hardware, software and firmware for rapid characterization and is the topic of chapter 14. Its first application was the characterization of the ATLASpix1 and MuPix8 sensors.

Their characterization is presented in chapter 15. Special attention is directed to effects that occur on large sensors and full-matrix measurements.

The successor of the MAB is the GEneric Configuration and COntrol Setup (GECCO). It is described in chapter 16 and its first application, a beam telescope based on ATLASpix1 layers is presented.

Chapter 17 explores the possibility to use HV-CMOS sensors in the monitoring of medical beams for heavy-ion irradiation therapy.

The results of this section are summed up and discussed in chapter 18.

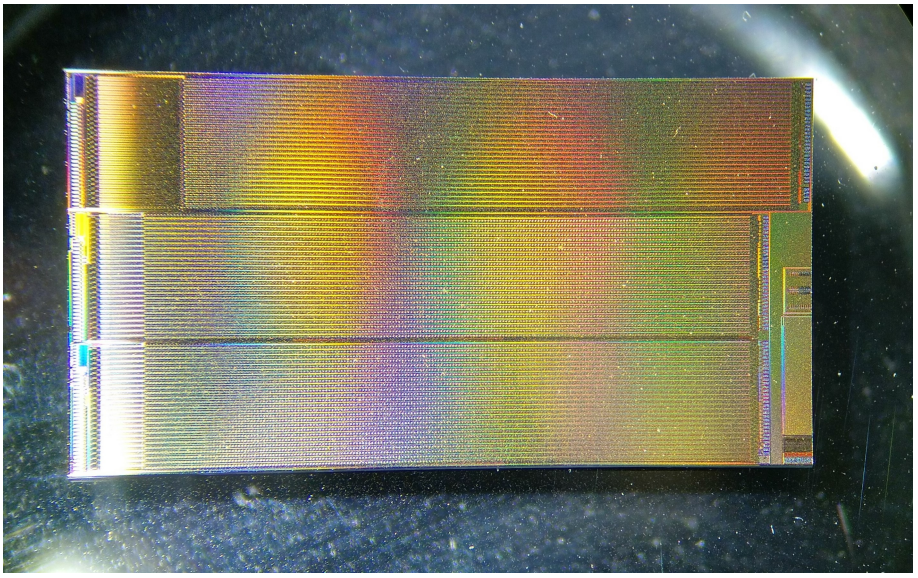


Figure 10.1: Photograph of ATLASpix1 sensor.

11. The ATLASpix1 ASIC

ATLASpix1 can be seen as one of two successors to H35Demo. However, the technology has been altered to the *ams* HV-CMOS aH18 process. The most significant change is the reduction of the minimum feature size from 350 nm to 180 nm. ATLASpix1 was part of a larger, joint submission with MuPix8 and a number of smaller chips. Its layout is shown in figure 11.1.

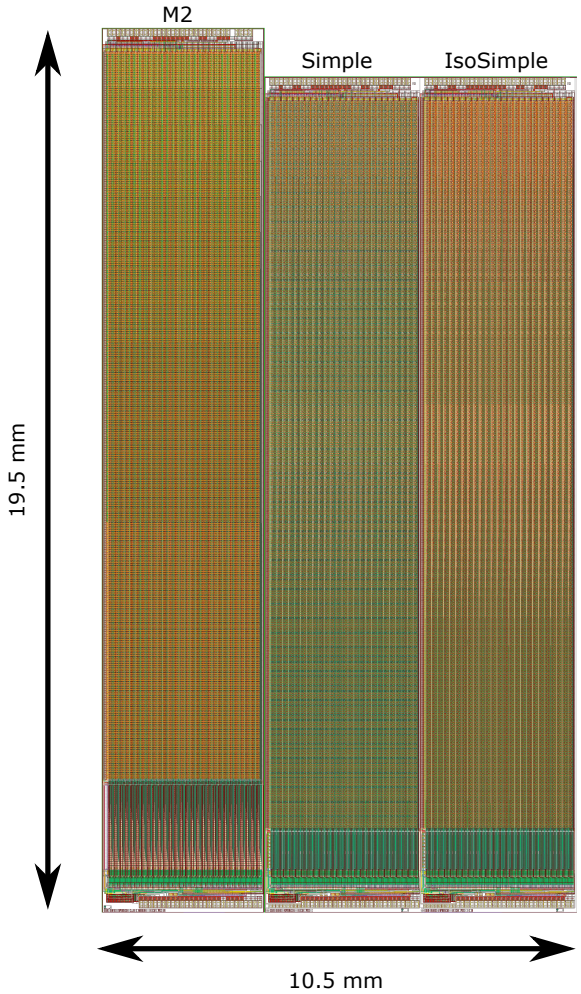


Figure 11.1: The layout of ATLASpix1. All three matrices have bond pads on the top and bottom edge. Simple and IsoSimple share most characteristics, M2 differs from them considerably.

The ATLASpix1 is not full reticle size (21.4×22.0 mm^2), but 10.5×19.5 mm^2 and covers about half the available area. It has been fabricated on substrates of different resistivity, ranging from 20 Ωcm to 1100 Ωcm. Some samples have been thinned down to a thickness

between 70 µm and 200 µm. The matrices have been separated from the MuPix and the smaller chips, but not from each other (Figure 11.3).

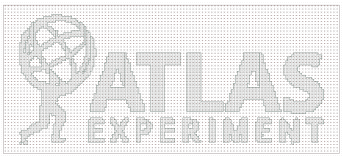


Figure 11.2: The ATLAS logo can be found on the chip.

ATLASpix1 is designed to satisfy the requirements for the upcoming upgrade of the outermost layer of the inner tracker of the ATLAS detector. The logo of ATLAS experiment has found its way onto the final layout (Figure 11.2).

Like the H35Demo, the ATLASpix1 comprises several fully independent matrices: Each is approximately 3.5 mm wide. One is 19.5 mm and two are 18.4 mm long. A fully monolithic readout has been implemented in each matrix. It is capable of higher rates, zero suppression and provides additional hit information. This is a great improvement compared to the standalone readout of the H35Demo. Capacitive signal transmission to readout chips is no longer implemented.

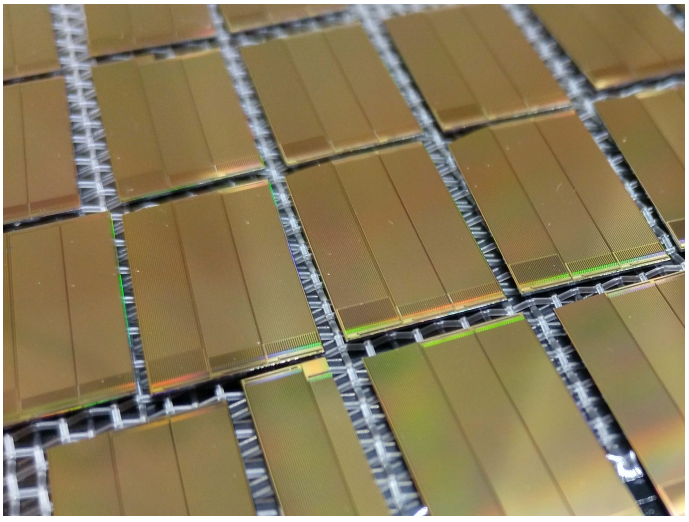


Figure 11.3: Photograph of the diced sensors. The three ATLASpix1 matrices of a die were not separated from each other.

The ATLASpix1 has been designed by the Karlsruhe Institute of Technology ASIC and Detector Laboratory, the University of Heidelberg and the University of Geneva, under the leadership of Prof. Perić (Figure 11.4).

Figure 11.4: All designers left their signature on the final layout. Each matrix has its own signature. ATLASpix1_Simple and ATLASpix1_IsoSimple were designed by Prof. Perić, M. Prathapan and H. Zhang. ATLASpix1_M2 was designed by Prof. Perić and M. Prathapan with support by F. Messaoud, E. Vilella and H. Zhang.

11.1 Three matrices – commonalities and differences

The ATLASpix1 consists of three matrices: ATLASpix1_M2, ATLASpix1_Simple and ATLASpix1_IsoSimple. All matrices share the same analog front-end. The charge sensitive amplifier is similar to the design of H35Demo. A comparator is located inside each pixel,

thus digitized signals are transmitted to the digital periphery. It comprises the End of Column blocks (EoC) and ReadOut block (RO). The address and time stamp of a hit are generated here. Additionally, the hit data package is expanded by analog information in form of a secondary time stamp for Time over Threshold (ToT) calculation.

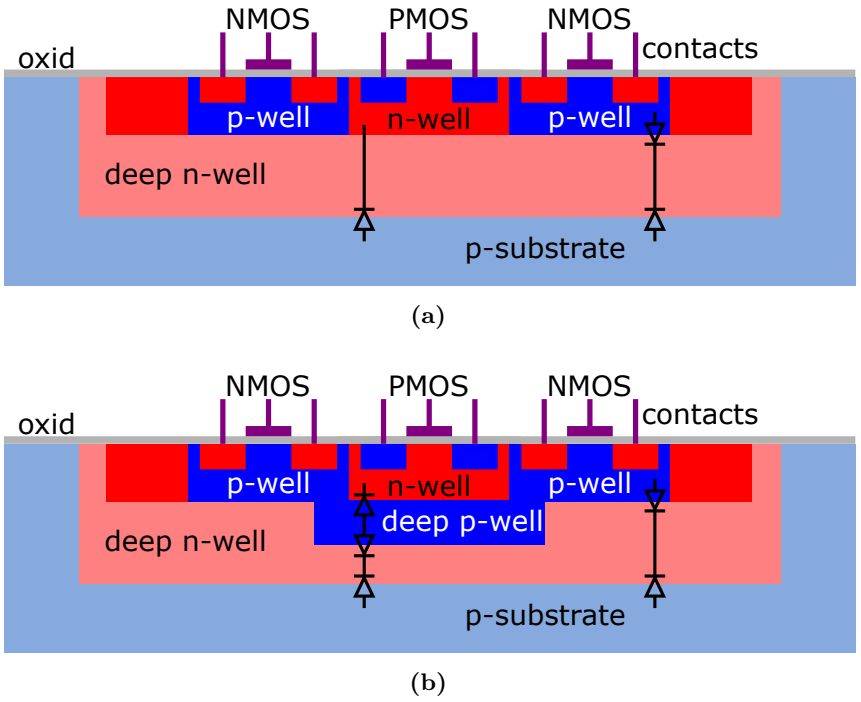


Figure 11.5: In standard HV-CMOS process (a) only NMOS transistors are isolated from the substrate, as p-well, deep n-well and p-substrate form two diodes with inverted polarity. PMOS transistors are not isolated, as only a single diode is formed. The non-standard deep p-well (b) isolates PMOS transistors by forming two additional diodes.

Both Simple and IsoSimple matrices share the same readout scheme. The difference is the additional deep p-well in the pixels of the IsoSimple matrix, which isolates PMOS transistors from the deep n-well. The process variation is shown in figure 11.5. This isolation is necessary to use an in-pixel CMOS-comparator, because the binary signals of the comparator would cause crosstalk to the pixel input. The Simple matrix features an NMOS-comparator, with similar performance, but with larger power consumption and less radiation tolerance. Its NMOS transistors are isolated from the substrate by the standard deep n-well. The deep p-implant is not part of the standard aH18 process.

The M2 matrix has a different pixel geometry and features a triggered readout. Additionally, hit information is stored in buffers which are not mapped one-to-one to pixels, but to groups of pixels.

All matrices are designed to be radiation tolerant to at least $5 \cdot 10^{15} \mathrm{n_{eq}/cm^2}$ and 10 MGy [85]. This is achieved by choice of technology, usage of radiation tolerant transistor layout in key positions and smart circuit design (see chapter 3.5.2).

The used 180 nm technology operates at a lower supply voltage than the previously used 350 nm technology. The nominal main voltage is 1.8 V and the amplifier voltage is 1.0 V. The analog power consumption has been further reduced by optimizing the design. A first measurement on power consumption delivered ≈ 100 mW/cm^2. It consists of analog and digital power consumption for ATLASpix1 (≈ 2 cm^2):

$$P = 1.8\,\mathrm{V} \cdot (\underbrace{60\ \mathrm{mA}}_{\text{analog}} + \underbrace{30\ \mathrm{mA}}_{\text{digital}}) + 1.0\,\mathrm{V} \cdot \underbrace{40\ \mathrm{mA}}_{\text{amplifier}} = 202\ \mathrm{mW} \tag{11.1}$$

Another great improvement compared to previous designs, is the bias generation. All ATLASpix1 matrices generate the required voltages and currents themselves in dedicated bias blocks, according to configuration. Each voltage is generated in a 6-bit DAC. An amplifier provides sufficient current capability to supply the entire matrix. All voltages are also routed to bond pads for probing or to override them externally (fail save). This innovation reduces the need for the external supply voltages and renders the test environment more tidy. In standard operation, only three voltages from external power supplies are needed: VDD (1.8 V), VSSA (1.0 V) and VGate (2.0 V).

11.2 ATLASpix1_Simple and ATLASpix1_IsoSimple

Both Simple matrices share the same layout. This applies for the entire digital periphery and the pixel matrix with only few exceptions.

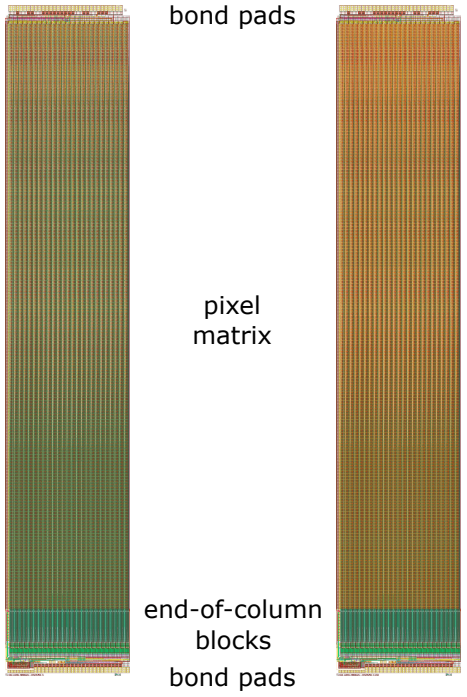


Figure 11.6: The layout of the ATLASpix1 matrices Simple (left) and IsoSimple (right).

Figure 11.6 shows the layout of the Simple (left) and IsoSimple (right) matrix. The wire bond pads are placed on the short sides of each sensor. Readout cells and periphery are located on the lower edge and occupy about 8% of the sensors total area. The remaining $> 90\%$ is occupied by the pixel matrix.

11.2.1 Pixel matrix and signal transmission

The pixel matrix comprises 25×400 pixels. Each pixel is 130 µm wide and 40 µm high. The large aspect ratio is due to the intended application in the inner parts of ATLAS ITk, which requires a high spatial resolution in only one direction. The sensor diode of a pixel is formed by the p-substrate and the deep n-well. The latter contains the front-end electronics: a charge sensitive amplifier and a comparator. The standard comparator is realized only with NMOS transistors, the deep p-well of the IsoSimple matrix allowed the implementation of a CMOS comparator, which is considered more radiation tolerant and power saving, as it allows the use of both NMOS and PMOS transistors (see chapter 3.5.2). Signal transmission to the digital periphery is carried out digitally.

The layout of a pixel with NMOS comparator is shown in figure 11.7. Equipped with an additional deep p-well and CMOS comparator, the pixel layout shown in figure 11.8,

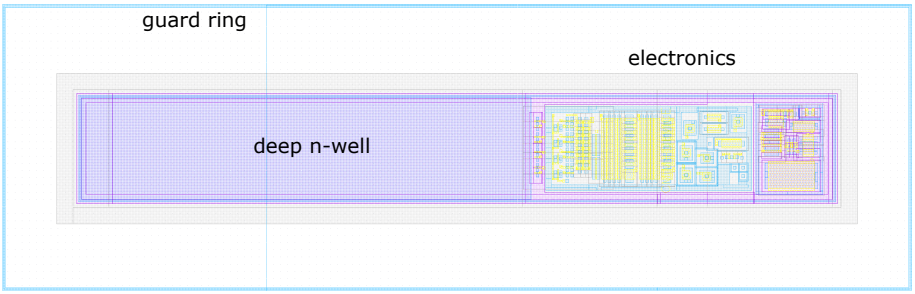


Figure 11.7: Pixel layout in the Simple matrix. The pixel outline is formed by a guard ring, indicated by the light blue line. The electronics (yellow and blue) is located in the right area of the deep n-well (gray).

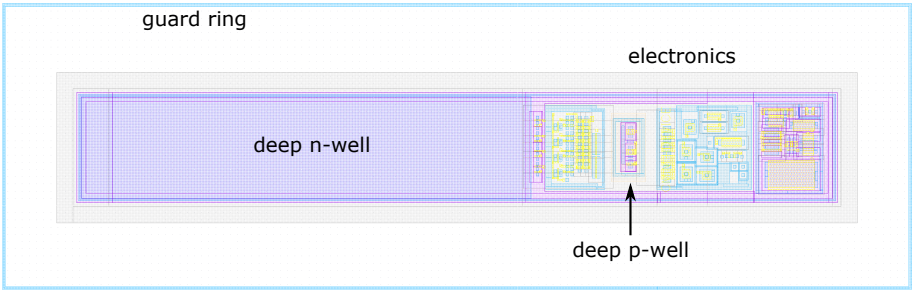


Figure 11.8: The layout of the IsoSimple pixel is very similar to the Simple pixel. The difference is a CMOS comparator instead of an NMOS comparator, which requires a deep p-well (white area) in the left third of the electronics part.

promises a lower power consumption. The pixel outline is indicated by the light blue line. The gray area indicates the deep n-well, which forms the sensor diode together with the p-substrate. Amplifier and comparator are located on the right side of the pixel. The digitized signal from the comparator is transmitted to the periphery by a voltage signal generated by a line driver. Analog information is not lost: The length of the transmitted digitized pulse scales with the size of the analog signal.

11.2.2 Readout scheme

Each pixel is connected to an individual readout cell in the periphery. In this readout cell, the signal is strengthened by a receiver for further processing. Both the rising and the falling edge are converted into pulses. These pulses trigger the generation of a time stamp each. The rising edge indicates the time of impact (time stamp 1), whilst the time stamp triggered by the falling edge is used to calculate the Time over Threshold (ToT), which is an indirect measure for the charge generated in the pixel diode. The running speed of the primary time stamp is deduced from the state machine speed. At nominal state machine speed, the minimal time between two time stamps is 12.5 ns, but can be extended to upto 200 ns by configuration. The speed of the secondary time stamp for ToT calculation can be adjusted individually. It can be set to a value suitable for the expected signal duration. The principle of adaptive sampling is illustrated in figure 11.9

The hit word generated in the readout cell is composed of the row address (8 bit), the first time stamp (10 bit) and the secondary ToT time stamp (6 bit). The hit words of a group of readout blocks are gathered in the End-of-Column block (EoC). One group consists of the upper or lower half of a pixel column. In case of several hits per half-column, a static priority logic determines the order of processing.

An EoC takes a hit word from one of the connected readout cells and adds the column address to it (8 bit). The data of the 50 EoCs are passed on to encoding and serialization

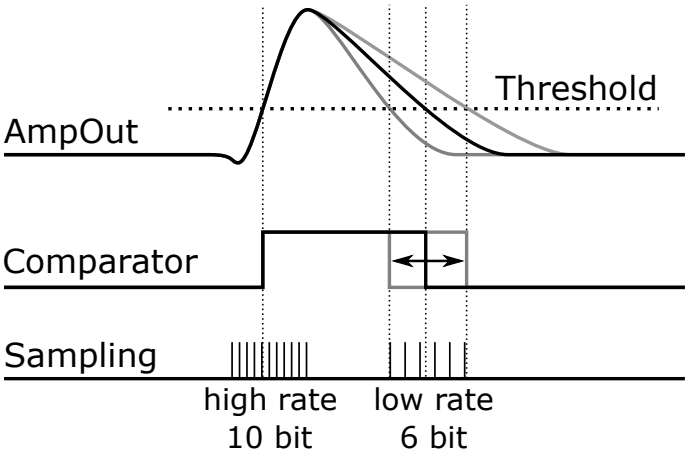


Figure 11.9: The graph shows the analog signal (AmpOut), the response of the comparator (asynchronous) and the sampling frequency of synchronization. The sampling frequency of both primary and secondary time stamp can be adapted to the requirements of an experiment. The leading edge is sampled with 10-bit precision down to a precision of 12.5 ns. The falling edge is sampled with 6-bit. 25 ns precsion is possible but not advisable as the falling edge has uncertainties typically in the order of 50 ns.

with a priority logic similar to the connection between readout cells and EoC.

ATLASpix1_Simple has one readout link. That means that the data from the entire chip have to be transmitted to the receiving FPGA via a single differential line. The transmitted words are 8 bit long. Therefore, the 32-bit hit data are chopped into single byte data words. Additionally, debug words and comma words can be sent at well-defined positions. Each byte undergoes an 8-to-10-bit encoding [86]. This encoding increases the data to transmit by 25%, but is beneficial for data integrity and DC-balance.

The transmission speed from sensor to FPGA is up to 1.6 Gbit/s. That means ever 25 ns a hit word can be send out, if comma words are neglected. The fast clock necessary for serial data transmission can be provided externally or is generated on chip by an oscillator and a Phase-Locked Loop (PLL). The latter is adjusted to a slower external reference clock. External clocking is only advisable for slow operation, as it gets more and more difficult to provide a high quality external clock with increasing speed. The PLL on the other hand only works for fast operation, as it can only tune the oscillator around its designed parameters.

11.3 ATLASpix1_M2

The matrices discussed in the previous section were, apart from the CMOS comparator and its deep p-well, identical. The ATLASpix1_M2 matrix differs substantially from them. This is already indicated by the different size of the periphery as can be seen in figure 11.1. The readout of M2 does not follow the column drain approach, but a readout scheme, called Parallel Pixel to Buffer (PPtB) readout. Furthermore, triggered readout is implemented. The matrix size remains almost the same, however the pixel size, and therefore the number of pixels differs.

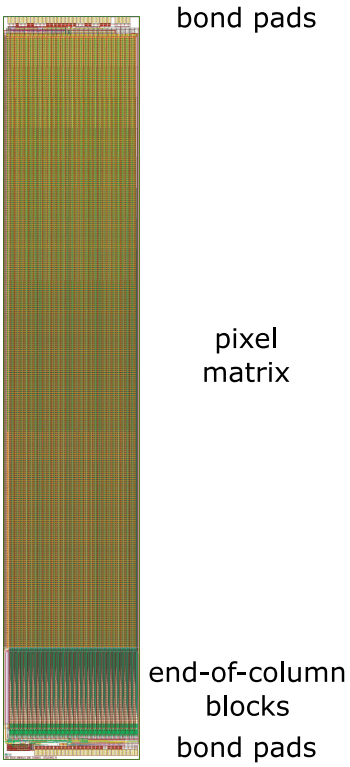


Figure 11.10: Layout of M2. The area of the pixel matrix is the same as for the other ATLASpix1 matrices. However, the pixel size is smaller and therefore the total number of pixels is larger. The increased number of pixels and the more complex readout pattern require a larger periphery.

The layout of ATLASpix1_M2 is shown in figure 11.10. Located at the upper and lower edge are the bond pads. On the lower side, the insensitive digital periphery is located. It covers approximately 15% of the chip's total area. The active pixel matrix covers over 80%.

11.3.1 Pixel matrix and Signal transmission

The pixel size of the ATLASpix1_M2 is 60×50 µm^2. That reduces the area of a pixel compared to ATLASpix1_Simple by 43%. Also the shape is not of large aspect ratio anymore, but nearly quadratic. The active area remains the same, therefore the number of pixels is increased. The pixel matrix comprises 17 920 pixels in 56 columns and 320 rows, 80% more than the Simple and IsoSimple matrices.

The layout of a pixel is shown in figure 11.11. The front-end electronics is almost the same as in the Simple version. A charge sensitive amplifier, an NMOS comparator, but two line drivers to send a digitized voltage signal from pixel to periphery.

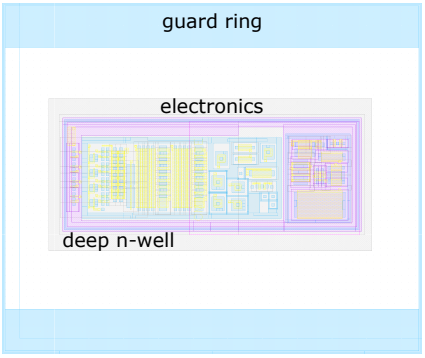


Figure 11.11: Layout of a pixel in the M2 matrix. The front-end electronics is the same as in a Simple pixel, but a second line driver is implemented. The pixel size is significantly smaller (60×50 µm^2) and nearly square.

11.3.2 Readout scheme – pixel grouping

Pixels of the ATLASpix1_M2 are not connected one-to-one to readout cells in the periphery. The increased number of pixels in the matrix and the limited space available for connections between pixels and periphery demand data reduction already on this stage. Therefore, the pixel matrix is organized in 28 double columns with 40 pixel blocks each. Each pixel block contains 16 pixels.

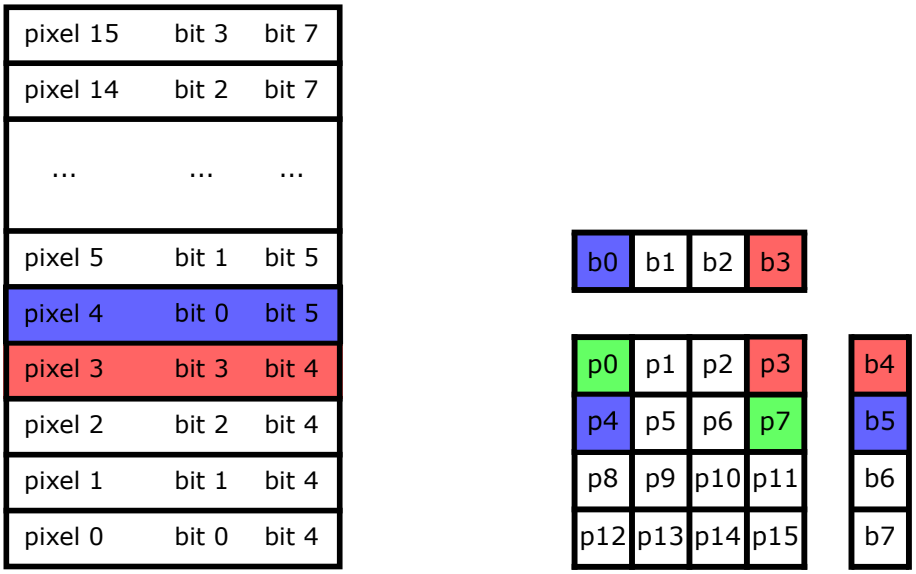


Figure 11.12: An illustration of the pixel grouping and readout implemented on the ATLASpix1_M2. The left figure shows the physical position of the pixels of a group. The right figure illustrates the encoding. The colors give an example of a problematic event. pixels in blue and red are actually hit, but the readout cell will also register the green pixels as hits.

These 16 pixels of one block share 8 transmission lines to their readout block. One pixel is connected to two lines, the transmission is encoded. In case of single pixel hits, the performance is the same as for one-to-one connections. When multiple pixels of a group are hit simultaneously, there is a possibility for ghost hits. This is assumed a minor issue, as a ghost hit only occurs when a particle leaves a cluster hit in the detector plane and the ghost hit, would only changes its size, but not its time nor its location.

The functionality of the pixel grouping is shown in figure 11.12. The color gives an example

for an event: When hit, each pixel sends a pulse on two of the eight bit lines to the periphery. If a single pixel is hit (blue), bits 0 and 5 are activated. The readout cell will register one hit in pixel 4. If, additionally, in the same time stamp, pixel 3 is hit (red), bits 3 and 4 are activated, too. This situation can not be resolved by the readout cell. It will record the two pixels actually hit, but also two ghost hits (green).

We distinguish three types of events:

- Events that are not affected by the encoding and are correctly recorded by the readout cell. This includes all single pixel hits but also most events with more than one firing pixel.
- Events that contain ghost hits, but can be reconstructed offline, like the example given above. Offline analysis shows that the recorded cluster event has two pixels next to each other and two in a certain distance. Therefore, the distant ones can be identified as ghost hits and are discarded.
- Events that can not be correctly reconstructed. This happens, when even offline analysis can not distinguish between the actual cluster and ghost hits.

In order to evaluate the impact of the encoding, a small Monte Carlo simulation has been set up. It randomly generates particle trajectories through a single pixel group. According to the pixels that have been hit, the active bit lines are determined (Figure 11.12a). From these active bit lines, it is reverse calculated, which pixels were hit (Figure 11.12b). In case of a multi-hit event, an attempt to identify and remove possible ghosts is started. Finally, the result is compared to the actual event and the result is evaluated according to above's list.

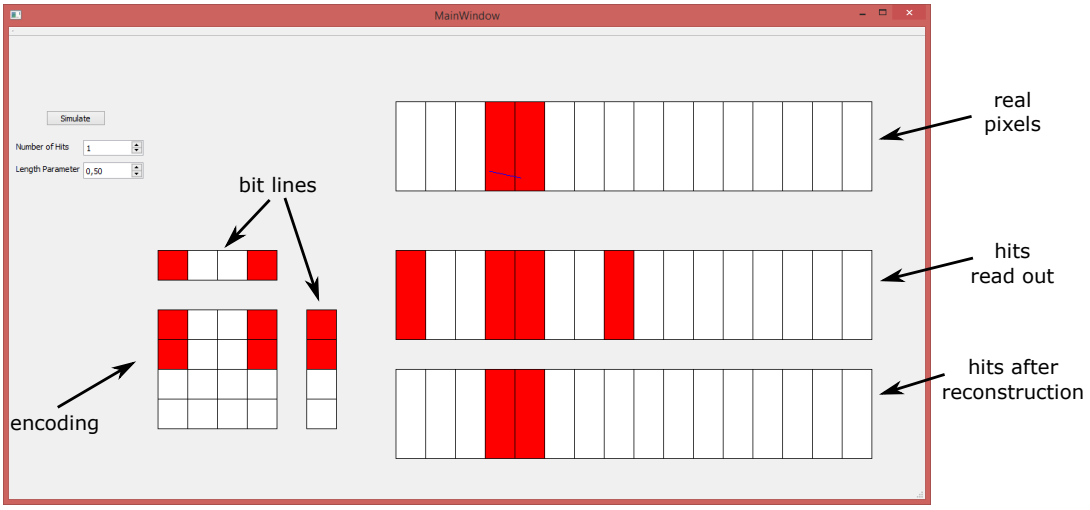


Figure 11.13: The picture shows the GUI of the Monte Carlo simulation. The top right matrix represents the physical pixels, with hit pixels in red. The resulting encoding is shown on the left. The second row on the right side displays, what pixels are registered as hit in the readout cell. The last row shows the hits after off-chip reconstruction.

Figure 11.13 shows a screen shot from the simulation. With increasing size of the simulated charge cloud, the rate of correct reproduced event drops. However, in the domain of event sizes that have been observed with HV-CMOS sensors in source- and beam-tests, the rate of correct recorded events is high. Over 80% of all events, do not need reconstruction. In another 10% of the events ghost hits can be easily identified and discarded. Only in about 6% of the events, ghost hits can not be identified. As this happens primarily on events with a large cluster and many hits, the fraction of ghost hits compared to real hits is even smaller. This means, that a detector, equipped with this kind of pixel grouping, would

deliver a small amount of ghost hits additionally to the real hits, which is not considered a major impairment.

11.3.3 Readout scheme – parallel pixel to buffer

The pixel grouping is not the only innovation of the M2 matrix compared to the Simple and IsoSimple matrices. In contrast to the other matrices, the pixels of M2 have no one-to-one mapping to buffers. The hit information of one pixel group is stored in a shared buffer with four entries. In general, a shared buffer is an advantage over one-to-one mapping. Therefore, the absolute number of buffers on a sensor with PPtB might be reduced compared to a sensor with one-to-one mapping. How many buffers can be saved depends strongly on the type and rate of the expected events. Hits distributed homogeneously over a PPtB matrix, allow for a smaller number of buffers than hits concentrated in a small area. These considerations lead to the development of the ReadOut Modelling Environment[1] (ROME). This simulation uses physics simulation data from ATLAS ITk as input to a simulation of pixel geometries, and readout architectures[2].

11.4 Readout data

The total hit data set contains 8 byte. It comprises 4 byte data from the sensor and 4 byte additional data from the FPGA. The Simple and IsoSimple matrices use the 32 bit as follows:

Name	Description	Bits
column	Address of the readout column. Is used to calculate the geometrical column.	8
row	Address of the geometrical and readout row. Is used to calculate the geometrical column.	8
event time stamp	Gray coded event time stamp.	10
ToT time stamp	Gray coded secondary time stamp. It represents the point in time, of the falling edge of the analog signal.	6

The M2 matrix uses the 32 differently, because of the pixel grouping:

Name	Description	Bits
column	Address of the readout column. Is used to calculate the geometrical column.	8
group	Address of the pixel group in the column.	6
pixel code	Address code of the pixels in a pixel group. It may contain more than one hit.	8
event time stamp	Gray coded event time stamp.	10

The time stamps are transmitted Gray-encoded. The 4 bytes from the FPGA are reserved for trigger time stamp and trigger index, generated on FPGA from an additional source, e.g. scintillator or reference detector.

[1] Development of ROME was started by the author and R. Schimassek, later development was fostered by R. Schimassek alone.

[2] ROME wiki: `https://git.scc.kit.edu/jl1038/Readout_Simulation/-/wikis/home`

12. The Mu3e experiment

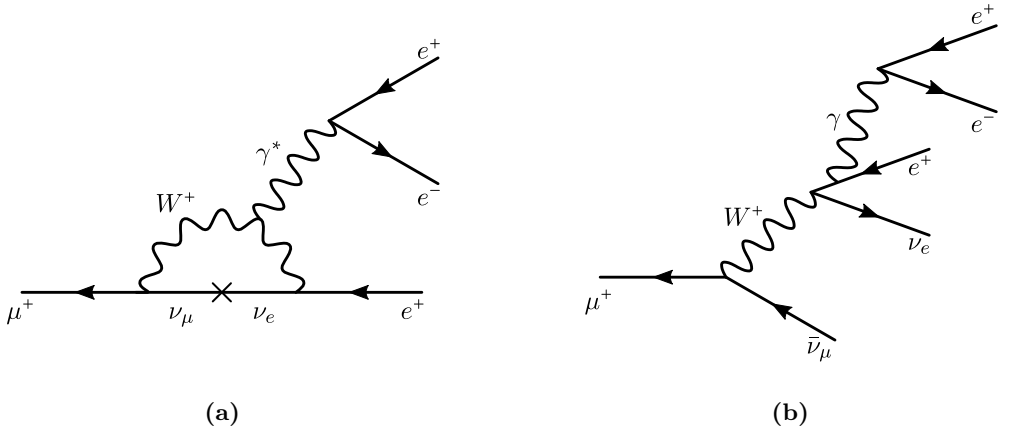
The Mu3e experiment is located at the Paul-Scherrer-Institute (PSI) in Switzerland, under development at the time of writing [87, 88]. Some components are already under construction.

Mu3e has been designed to measure the decay of a muon to three electrons $\mu \to eee$ with previously unreachable precision in a very clean environment. It follows the approach of high intensity at a comparably low energy. Thus, new particles will rather be produced as virtual particles, as their mass is expected to be larger than the provided energy.

12.1 Physics goal of Mu3e – physics beyond the Standard Model

The decay $\mu^+ \to e^+e^-e^+$ is of special interest for the search of physics beyond the standard model.

Within the Standard Model, this process is not permitted, as it changes the lepton flavor. However, numerous experiments have shown that lepton flavor is not conserved for neutrinos. In 2015, the Nobel Prize for Physics has been awarded to scientists of Super-Kamiokande Observatory [89] and Sudbury Neutrino Observatories [90] for the discovery of neutrino oscillations [91]. The existence of neutrino oscillations is not only proof of physics beyond the Standard Model, but it further indicates that neutrinos have a non-zero mass (see chapter 19).

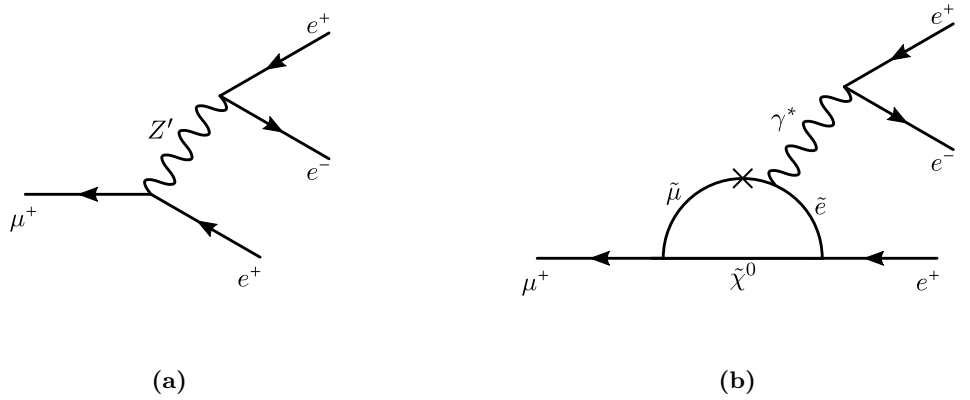


(a) **(b)**

Figure 12.1: The background processes expected at Mu3e. The extremely rare process involving neutrino mixing (a) and a process producing two additional neutrinos (b). (After [88])

Neutrino mixing allows the process $\mu^+ \to e^+e^-e^+$ as shown in figure 12.1a. This process is very rare $\mathcal{O}(10^{-50})$ [92] and is not expected to be observed by Mu3e.

The second background process is the decay $\mu^+ \to e^+e^-e^+\nu_e\bar{\nu}_\mu$ (Figure 12.1b), which has to be discriminated from the neutrinoless process Mu3e is looking for.



(a) **(b)**

Figure 12.2: Mu3e searches for processes involving particles beyond the Standard Model. One example process with a Z'-boson is shown in (a), supersymmetric particles are involved in the process shown in (b). (After [88])

Theoretical particle physicists have suggested extensions to the Standard Model that could greatly increase the branching fraction of $\mu^+ \to e^+e^-e^+$ up to a level that can be found by Mu3e. Two of them are shown in figure 12.2: Lepton flavor violation involving supersymmetric Z' or other bosons is an often postulated process [93] on tree level.

Similar to the loop level background process by neutrino mixing, a loop-level involving supersymmetric particles is another possibility increasing the branching fraction.

These or other $\mu^+ \to e^+e^-e^+$ processes have to be discriminated from $\mu^+ \to e^+e^-e^+\nu_e\bar{\nu}_\mu$ background. As neutrinos hardly interact with matter, their detection is not feasible. Therefore, neutrino involvement has to be detected indirectly as missing energy. A detector had to be designed, capable of discriminating the different processes with extreme high reliability, to identify the still rare flavor violating $\mu^+ \to e^+e^-e^+$ events.

12.2 The detector for Mu3e

The before mentioned considerations result in the following demands on a detector for Mu3e [88]:

High rate capability The goal is to find neutrinoless μ-decays as rare as 10^{-16}. Previous experiments have set upper boundaries to 10^{-12} (SINDRUM 1986) [94] or 10^{-13} (MEG 2012) [95]. To keep the measurement time feasible, an event rate of $2 \cdot 10^9$ Hz is needed.

Geometric acceptance The kinematics of the searched processes is not known, therefore as much area around the interaction point as possible should be monitored with sensors.

Momentum resolution Whether or not neutrinos were produced in an event has to be determined by reconstructing the mass of the observed particles. Sensitivity of Mu3e depends on the correct discrimination of signals from background via center of mass reconstruction. The resolution of the reconstructed mass indicates directly the sensitivity (Figure 12.3). For the projected sensitivity of 10^{-16} at 95% confidence level (blue line), the average reconstructed mass resolution has to be better than 0.5 MeV/c^2.

Vertex resolution At the requested rate, each recorded frame will consist of several events. Their superposition might be seen as accidental background (e.g. $\mu \to ee\nu$ plus

$\mu \to eee\nu\nu$). This can be prohibited by reconstruction of the vertices of every individual event with high precision.

Time resolution The combination of high event rate, small number of events per frame and very good event separation, results directly in the request for a very good time resolution.

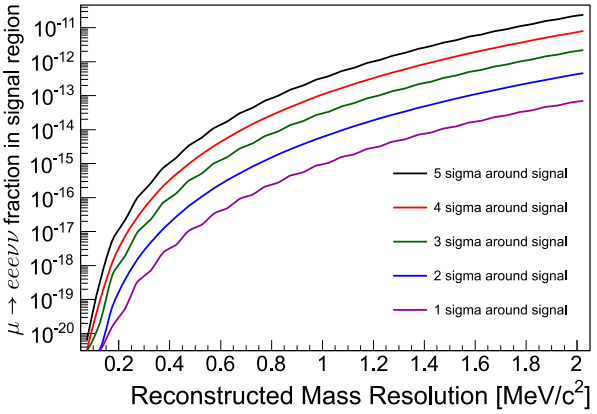


Figure 12.3: The plot shows the sensitivity of the projected Mu3e experiment as a function of the reconstructed mass resolution. Each line stands for a certain confidence level. (From [88])

These requirements have converged to the design of the Mu3e experiment. Figure 12.4 shows the most important components of Mu3e [88].

The high-intensity muon beam is projected on a Mylar target [96]. Its shape is optimized to distribute the decay vertices over a larger volume, to reduce the probability of background events by accidental superposition of vertices.

Pixel layers equipped with MuPix HV-CMOS sensors provide excellent spatial resolution and a time resolution better than 25 ns. The inner pixel layers are arranged close to the target, measuring the emerging particles, while the outer pixel layers monitor the trajectories of both outgoing (middle segment) and recurling particles (up-steam and down-stream segments).

Scintillating fibers in the middle segment provide timing information better than 1 ns with moderate spatial resolution. The scintillator tiles in the other segments promise even better timing (better than 100 ps), but with even less spatial resolution.

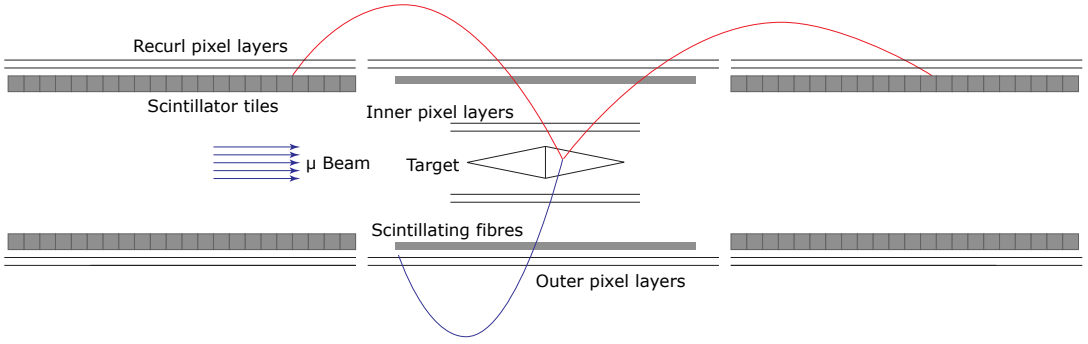


Figure 12.4: Side view of the Mu3e detector with an example event. In red positrons, in blue electron. (From [88])

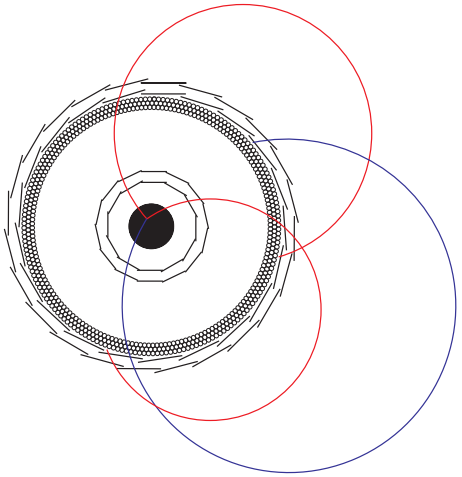


Figure 12.5: Transversal view of the detector. (From [88])

All pixel layers, the scintillating fibers and the scintillator tiles have to be more than 99% efficient.

A strong magnetic field of 1 T bends emerging particles back into the detector on a helical trajectory. Their shape depends on charge and momentum of the particle, which allows their identification. In figure 12.4 (side view) and 12.5 (transversal cut) example trajectories are shown: two positrons (red) and one electron (blue), which emerge from the same vortex.

High spatial resolution and precise timing will be achieved, while the amount of detector material as well as the material of support structures has to be as small as possible. This applies especially for the inner pixel detector, as angular variations have a larger effect on the measured momentum if they are closer to the vertex. More material means more multi-scattering (this issue is addressed in chapter 2.3.3), which has the potential of significant reduction of spatial resolution. The inner pixel layers will be assembled from chips thinned down to 50 µm, held by thin Kapton foil. Cooling is provided by pumping thinned helium gas through the otherwise evacuated detector.

13. The MuPix8 ASIC

The design of MuPix8 is closely related to that of ATLASpix1. They are not only implemented and produced next to each other on a joint submission in *ams* aH18 technology with a minimum feature size of 180 nm, but also share a large number of design features and concepts. This applies particularly to front-end electronics, bias generation, clock generation and data readout off the chip to the outside world. But there are also substantial differences. ATLASpix1 has been described in chapter 11, the design features both chips have in common will not be repeated.

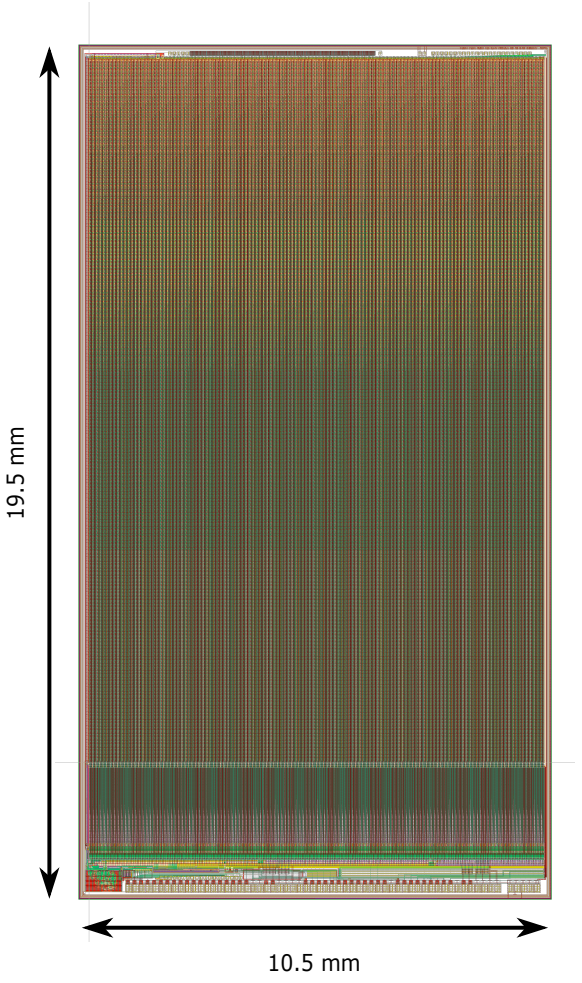


Figure 13.1: The layout of MuPix8.

The MuPix8 layout is shown in figure 13.1: It is 10.5 mm wide and 19.5 mm high which is approximately half the reticle's area. The periphery, located on the bottom edge, covers 15% of the total area. This is comparable to the ATLASpix1_M2 relations, and considerably larger than the periphery of the other ATLASpix1 matrices.

Its name indicates the purpose of the MuPix8: The Mu3e experiment. This ASIC is the 7$^{\text{th}}$ sensor in a series with the goal of equipping the projected Mu3e experiment with HV-CMOS sensors. It combines most design features relevant for a final chip on the area approximately half a reticle. In fact, it has – in certain aspects – more features than the final design will have. Namely, this is the case for the signal transmission from pixel to periphery, where after evaluation one option will be picked. The same is the case for the two different types of analog information generated in the periphery.

The designing of MuPix8 was led by Prof. Perić. The MuPix8 was developed by ASIC designers associated with Karlsruhe Institute of Technology and Heidelberg University, but other members of the Mu3e collaboration contributed, too (Figure 13.2).

Figure 13.2: All designers left their signature on the MuPix8 layout. It was designed by Prof. I. Perić, A. Weber, H. Augustin, H. Zhang, M. Prathapan, N. Berger, R. Leys and R. Blanco.

13.1 Pixel matrix

The Mu3e experiment is substantially different from the ATLAS experiment, although both track the trajectory of charged particles. From these differences, the need for a square pixel arises, unlike the rectangular pixels for ATLAS ITk. The MuPix8 pixels are 80×81 µm^2 large. Their area of 6480 µm^2 is 24% larger than the one of ATLASpix1_Simple.

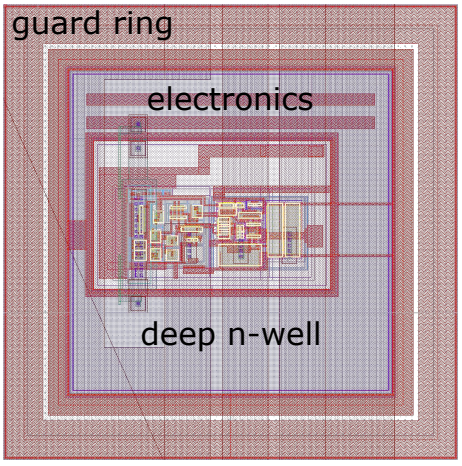


Figure 13.3: The pixels of MuPix8 have a square shape.

The pixel matrix is formed by 25 600 pixels arranged in 128 columns and 200 rows. The matrix is divided into 3 parts with one readout link each (in figure 13.1 from left to right): Part A with 48 columns, Part B with 48 columns and Part C with 32 columns.

The pixels of all parts are identical in shape (Figure 13.3) and front-end, which is the same as in ATLASpix1. In contrast to ATLASpix1, not a digital signal but the analog signal is transmitted to the periphery. The analog pulse of the amplifier is too weak to be transmitted directly over a distance of about 2 cm. Therefore, a line driver is connected to the amplifier output. Transmitting an analog signal is beneficial for noise in the matrix, as the voltage changes are not that fast and large. On the other hand, an analog signal is more sensitive to picking up noise on the way.

The line driver is implemented as source follower (Figure 13.4). The amplifier output is connected to *Vin*. The difference between the pixels in Part A and both other parts is

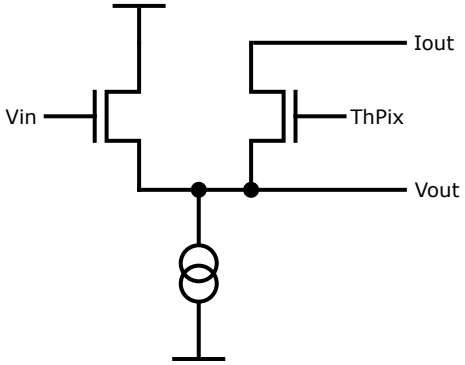


Figure 13.4: Schematic of the line driver in MuPix8. It is located inside each pixel and transmits the amplifier output as voltage (*Vout*) or current (*Iout*) signal to the periphery. (After [97])

the used output: While in Part A *Vout* is connected to the transmission line, it is *Iout* in Part B and C. Therefore, pixels in A are called voltage mode pixels and the pixels in B and C are called current mode pixels. The transmission mode is expected to have an influence on time-walk and transmission delay uncertainty.

13.2 Readout periphery

The signal line from each pixel is connected to an individual readout cell. In case of current mode transmission the signal is converted back to a voltage signal by a line receiver. The voltage signal is passed on to two comparators with independent thresholds. Each comparator stores a time stamp at the time of threshold crossing. These comparators can be operated in two different modes, the ADC (or ramp) mode and the two-threshold mode. In both modes, the voltage signal from the pixel is connected to the inverting input of both comparators. The reference voltage for comparator 1 is *ThHigh*, one of two configurable thresholds. The second threshold *ThLow* is the reference for comparator 2. Both thresholds are global voltages, but can be adjusted locally by tuning circuits controlled by tune bits. The rate at which the comparators' outputs are sampled is adjustable. Two configuration bits activate one of two operation modes:

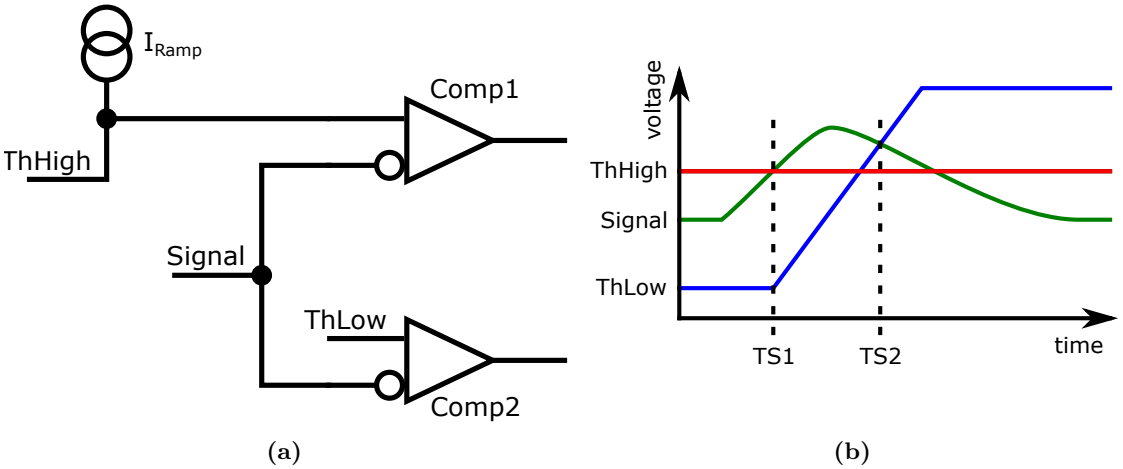
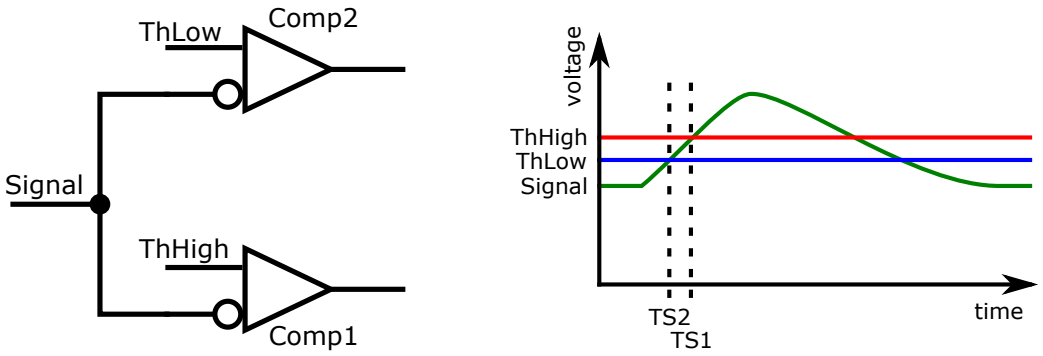


Figure 13.5: Function principle of the the ramp-mode digitizer. It uses the voltage-to-time approach to measure analog information. The simplified circuit is shown in (a), the most important signals over time are shown in (b). The analog information can be used to correct for time-walk. (After [97])

The ramp-mode is an indirect measurement of the signal height by generating two time stamps (*TS*). The maximally simplified schematic is shown in figure 13.5a. Its functionality is illustrated in by figure 13.5b, which depicts the most important voltages.

In idle state, *ThLow* is at the preset voltage level, the current source I_{Ramp} is switched off. When the incoming voltage signal (green) crosses *ThHigh* (red) at comparator 1, a control signal is generated, which triggers the storing of *TS1*. The same control signal activates the current source I_{Ramp}, rising *ThLow* (blue). At the moment it reaches the voltage level of the signal, comparator 2 triggers the generation of *TS2*. The difference of *TS1* and *TS2* is a measure for the analog pulse height of the original signal and therefore can be used to correct for the time-walk.



(a) **(b)**

Figure 13.6: Function principle of the two-threshold-mode. It does not digitize analog information but avoids time-walk. Its simplified schematic is shown in (a), the relevant signals over time are plotted in (b). (After [97])

The second mode, the two-threshold mode and its basic circuit (Figure 13.6a) is similar to the time-over-threshold measurement implemented on ATLASpix1. However, the outputs of both comparators are used differently as shown in figure 13.6b: *ThLow* is set right above the signal's baseline. In fact, it is too low for normal operation, as the slightest noise induced variation of the baseline is sufficient to trigger the comparator. Anyway, the time stamp of every threshold crossing event is preliminarily stored. There is only one cell to store the time stamp and it is overwritten at every threshold crossing event. This ongoing process is stopped when the signal crosses *ThHigh* and the time stamp stored at this point is passed on for readout. In this way, the benefits of a low threshold *ThLow* suppressing time-walk, and a high threshold *ThHigh* rejecting noise, are combined. This is an option for precise measurement of the traversing time of a particle passing through the sensor, however no analog information is recorded.

The readout state machine and serializers are the same as in the ATLASpix1_Simple matrices. The column drain readout approach is realized using priority logic. As MuPix8 is larger than the individual ATLASpix1 matrices, it has three independent readout links, one for each part. They can be multiplexed into a fourth readout link.

Four bytes carrying hit information are transmitted to the FPGA. They comprise the following data (in order of transmission):

Name	Description	Bits
column	Address of geometrical and readout column.	8
time stamp 2	Gray coded secondary time stamp. Meaning depends on selected mode	6
time stamp 1	Gray coded event time stamp.	10
row	Address of the geometrical and readout row.	8

14. The Multi-purpose Adapter Board

From the experience gained on the characterization of the H35Demo, the conclusion was drawn that an improved modular characterization setup was beneficial for many aspects of HV-CMOS characterization:

- **Development speed:** A modular setup requires less development time for a new chip.
- **Measurement speed:** Performing the same measurement on different samples is quicker and easier, as less connections to the carrier PCB are needed (*Plug and play*).
- **Costs:** Some functions, such as bias voltage generation and injection circuit, are needed in virtually every characterization setup and their components are considerably costly. Therefore, it is reasonable to put them not on every carrier PCB but to render them reusable.

Moving functionality from the chip carrier PCB to an intermediate board, is an effective way to improve the characterization process. Work on new chips does not require the development of a new setup, but only of a new carrier PCB. However, not every functionality is necessary in the characterization of every sensor. Therefore, two circuits with expensive components have been moved to daughter boards: Bias voltage generators and test signal injectors.

14.1 MAB hardware

The layout of the eponymous Multi-purpose Adapter Board (MAB) is shown in figures 14.1 and 14.2. The connection to the FPGA board is established via an LPC-FMC connector (Nexys_Con) on the bottom side. Just as in the H35Demo setup, the NexysVideo evaluation board is used. The differential LVDS signals can be terminated on the adapter board, to translate current signals into a voltage signals. If only differential signals are required, the grounds of FPGA and sensor can be separated. Single ended signals need a common ground.

The signals are routed to four box headers (CON1 - CON4). This type of connector was chosen for better mechanical stability, price and availability. All connectors are implemented twice: TAP1 - TAP4 are duplicates of CON1 - CON4 and no components are soldered to them. They allow easy probing of all lines routed to the carrier PCB, as the probes of most oscilloscope fit into the pin holes.

The power supply of the device under test is routed via those connectors, too. The power supply of the NexysVideo can provide some of these voltages. The interface to connect additionally some arbitrary external voltages are six power plugs.

Three smaller box headers are located on the lower end of the board. VoltageBoards can be plugged into two of them on the top side. The connector on the bottom side can be equipped with an InjectionBoard.

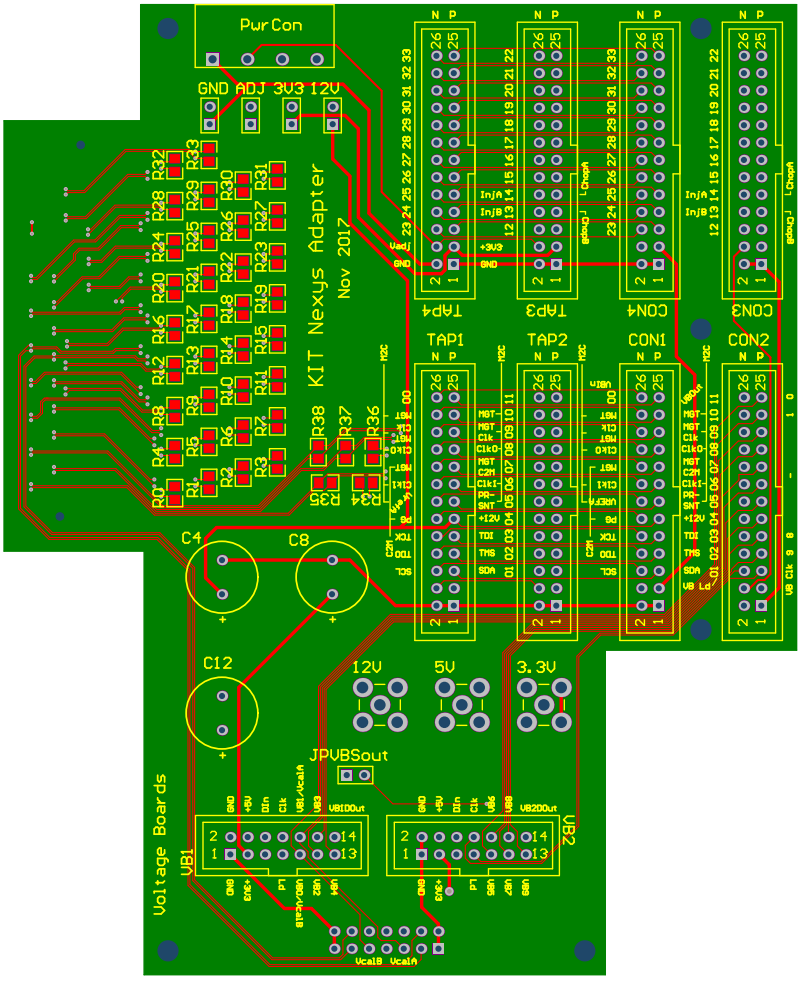


Figure 14.1: Top side layout of the Multi-purpose Adapter Board. On the right side are the box headers to connect to the carrier PCB and their duplicates to probe all signals. On the lower edge, VoltageBoards can be plugged in. External power supplies can be connected to a power connector *PwrCon*.

Figure 14.3 shows the MAB equipped with three daughter boards, plugged into the commercial NexysVideo FPGA board.

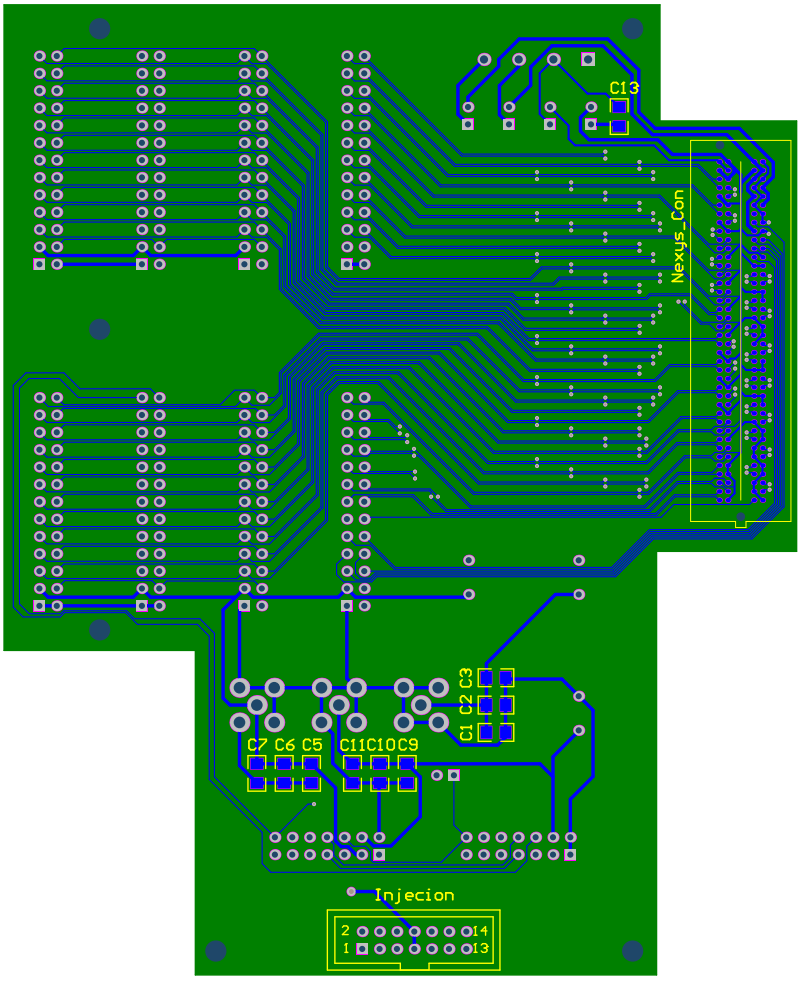


Figure 14.2: Bottom side layout of the Multi-purpose Adapter Board. It connects to the FPGA board by a LPC-FMC connector on the right side. On the lower edge, an InjectionBoard can be plugged in.

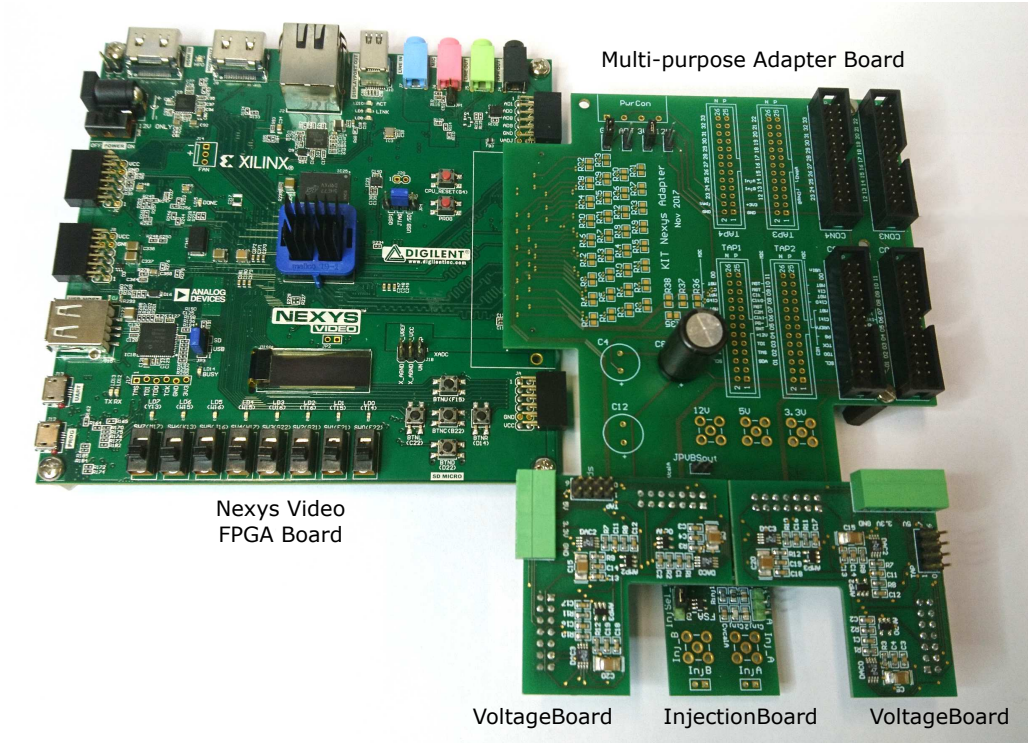


Figure 14.3: Photo of the Multi-purpose Adapter Board setup. The MAB is connected to a NexysVideo FPGA board and is equipped with three daughter boards (two VoltageBoards and one InjectionBoard). A DUT may be connected to the four box headers on the right.

14.1.1 VoltageBoard

The most expensive components of the MAB system are the high-precision LTC1658 14-bit digital to analog converters (DAC) [98]. Essentially, every characterization setup needs high-precision bias voltages generated by external DACs. The number however, varies from setup to setup and might even change during measurements on the same chip. Therefore, the bias voltage generation has been migrated to small daughter boards, the VoltageBoards[1].

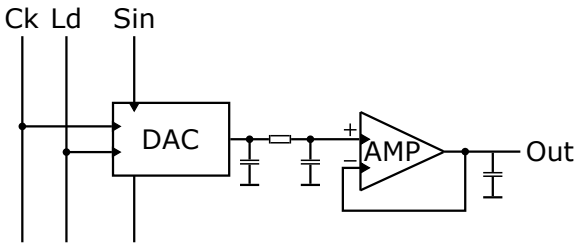
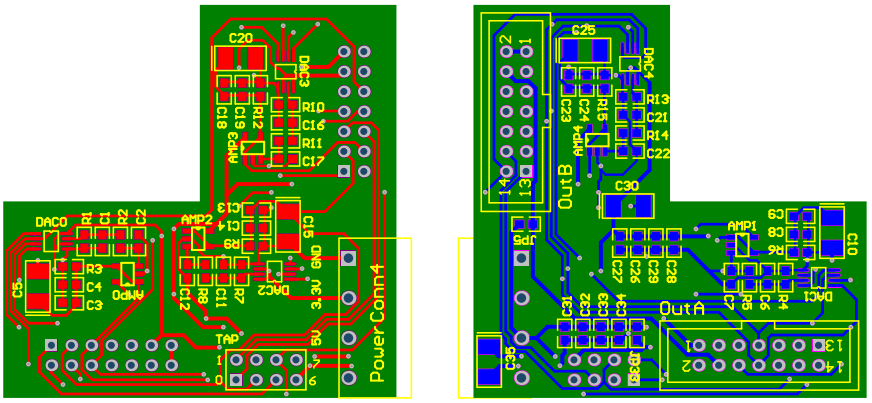


Figure 14.4: Simplified schematic of one voltage source.

The generated voltages depend on the configuration of the DAC, which is software controlled, via FPGA. The simplified schematic of one bias generation circuit is shown in figure 14.4. Each board has 5 bias voltage generation circuits. The DACs are connected in a daisy-chain pattern with common clock (Ck) and load (Ld) signals. The data bits (Sin) are shifted through the daisy chain.

The DACs provide a precise, but weak voltage. It has to be considered that the circuit supplied with this voltage draws a small current. Therefore, an LMV721 amplifier [99] is interposed between DAC and output in impedance converter configuration. From this circuit, several miliamperes can be drawn at constant voltage.



(a) Top **(b)** Bottom

Figure 14.5: Layout of the VoltageBoard. Up to two can be connected to the MAB. Each can generate five voltages with 14-bit precision.

The layout of the VoltageBoard is shown in figure 14.5. Its plug connects to the box headers on the top side of the MAB.

Two VoltageBoards can be plugged into one MAB. Therefore, the MAB system can provide a maximum of ten bias voltages.

[1]Notation without space according to previous presentations and design resources. Applies also to InjectionBoard and the components of GECCO in chapter 16.

14.1.2 InjectionBoard

A charge injection to an HV-CMOS sensor is carried out by the injection circuit. It generates a steep negative voltage pulse, by discharging a 1 nF capacitor. The amount of electrons pushed into the sensor depends on the voltage step. The injection voltage can be set between 0 V and supply voltage, but is effectively limited to 0.05 V to 2.7 V by the DAC on the VoltageBoard. It needs to be adjustable in order to study the reaction of the device under test to signals of different charge.

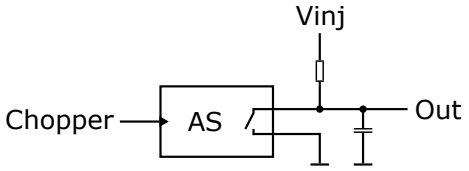
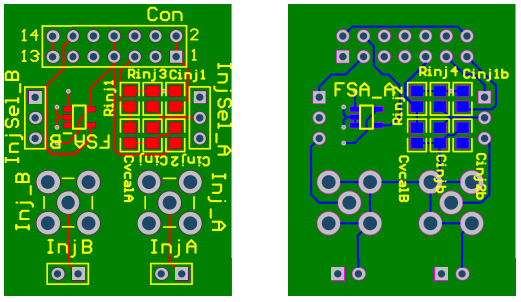


Figure 14.6: Simplified schematic of injection circuit. The capacitor is charged up to an arbitrary voltage. When the analog switch is closed, it is rapidly discharged, generating a negative voltage step.

The simplified schematic of the InjectionBoard is shown in figure 14.6. The state of the analog switch (AS) is controlled by a digital signal from the FPGA (chopper). The injection voltage is defined by a VoltageBoard. In idle state, the capacitor is charged via the resistor up to V_{inj}. When the analog switch is closed by the chopper, the capacitor is shorted to ground, thus quickly discharged and a negative voltage step is emitted. After opening the switch again, the circuit resets.



(a) Top **(b)** Bottom

Figure 14.7: Layout of the InjectionBoard. It can be plugged into the MAB. One board has two injection circuits.

Figure 14.7 shows the layout of the InjectionBoard. One board has two injection circuits. The injection pulse can be routed directly to the DUT via box headers on the MAB, or it can be directed to a low a low impedance connector (SMA or BNC).

14.2 Software improvements

Firm- and software for the MAB setup are similar to the ones of H35Demo. Those have been described in chapter 6.5.

The changes having the biggest impact on characterization, have been made in software[2]. Instead of writing the configuration of the chip into a bit-vector, the configuration is encapsulated in a dedicated class. The configuration object receives its properties on start of the software in the form of a series of DACs or bits. A name is appointed to each and the bit order is defined (MSB first, LSB first or custom bit order).

[2]The software improvements are team accomplishment of KIT ADL, mainly by the author and R. Schimassek.

During operation, the values or parameters can be adjusted by addressing the names instead of certain bits. Furthermore, the order of the bits in one DAC can easily be altered, as well as the order of the entire bitstream, which is generated on request for configuration from the stored parameters.

Encapsulation of bitstream generation code relieves the MAB user from testing generation code when applying it to a new chip, but only the bit order has to be tested.

A similar approach is chosen for the configuration of the MAB's daughter boards. The configuration of the DACs on the VoltageBoard are stored in a dedicated class. Each object of this class addresses one VoltageBoard. It stores the input comprising the voltage name and voltage value and translates it into a binary pattern to be sent to the VoltageBoard via FPGA.

The third configuration class does the same thing for the FPGA's pattern generator which generates the chopper for the InjectionBoard.

14.3 Summary of the Multi-purpose Adapter Board

The Multi-purpose Adapter Board System means a great improvement for chip characterization at KIT ASIC and Detector Laboratory. The hardware is no longer single use, but universal components have been moved to modular, intermediate PCBs, which can be compiled according to the requirements of the DUT. The only piece of hardware that has to be developed for characterization of a new chip is a simple carrier PCB. This means a speed-up in development, and at the same time a speed-up at characterization of multiple samples of the same chip.

New firm- and software functions have been developed, which are tailored to the functions of the MAB setup and to the most common properties of DUTs. These modules and classes can be easily utilized whenever a new characterization software is needed.

The success of this system is best illustrated by a list of chips that have been measured using the Multi-purpose Adapter Board system:

- ATLASpix1
- ATLASpix2
- MuPix8
- MuPix9
- LFoundry LFATLASpix1
- LFoundry LFATLASpix2
- LFoundry LFATLASpix3
- LFoundry LFATLASpix4
- CCPD 180 nm

The MAB has been distributed to collaborating institutes located in Jülich (De), Marseille (Fr), Milan (It) and Munich (De).

15. Characterization of ATLASpix1 and MuPix8

Characterization of ATLASpix1 and MuPix8 are conducted on a joint hard- firm- and software. The characterization setup is based on the Multi-purpose Adapter Board system. In principle soft- and firmware are similar to the ones for H35Demo which has been presented in chapter 6. The modifications necessary for this new chips are shown in Appendix F, together with the carrier PCB.

15.1 Analog calibration

The front-ends of ATLASpix1 and MuPix8 are identical. Therefore, the analog calibration is shown only once for MuPix8 and is valid for both. It provides information on the gain and noise of the front-end of a single pixel.

For analog calibration, the output of the charge sensitive amplifier (CSA) of one pixel is connected to an output pad of the chip. An oscilloscope is used to probe this signal. The maximum amplitude of each triggered waveform is filled into a histogram, while the chip is illuminated either with X-rays from an ^{55}Fe source or electrons from a ^{90}Sr source.

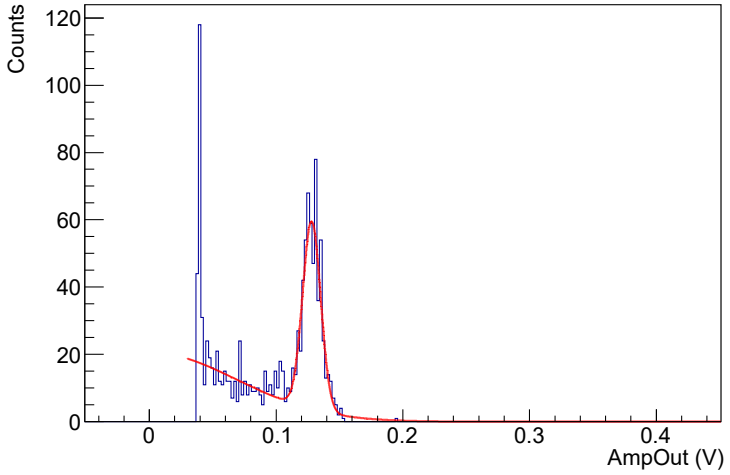


Figure 15.1: Histogram of the analog pulse height. The sensor was illuminated with 5.9 keV X-rays from an ^{55}Fe source.

Figure 15.1 shows the analog pulse height spectrum obtained from a measurement with ^{55}Fe. The peak appearing close to the low energy cut-off originates from noise signals that exceed the 50 mV threshold.

The peak towards higher energies is the ^{55}Fe-signal. A Gaussian fit was applied to both peaks. The fit of the combination of both is drawn in red. The signal of ^{55}Fe generates

1639 electron-hole pairs in silicon and is found at

$$U_{\text{AmpOut}}\left(^{55}\text{Fe}\right) = 128\ \text{mV} \pm 7\ \text{mV} = 128\ \text{mV} \pm 5.5\%. \tag{15.1}$$

Even though the voltages of both main (*VDDA*) and amplifier power supply (*VSSA*) are lower than the ones of the H35Demo, the output signal of the aH18 front-end is 60% larger (compared to equation 7.1).

For comparison, test signals have been injected to the same pixel with the same settings and the analog response has been measured. A 0.3 V injection results in a 110 mV amplifier signal. Therefore, a 0.349 V injection generates the same signal as X-rays from ^{55}Fe decays.

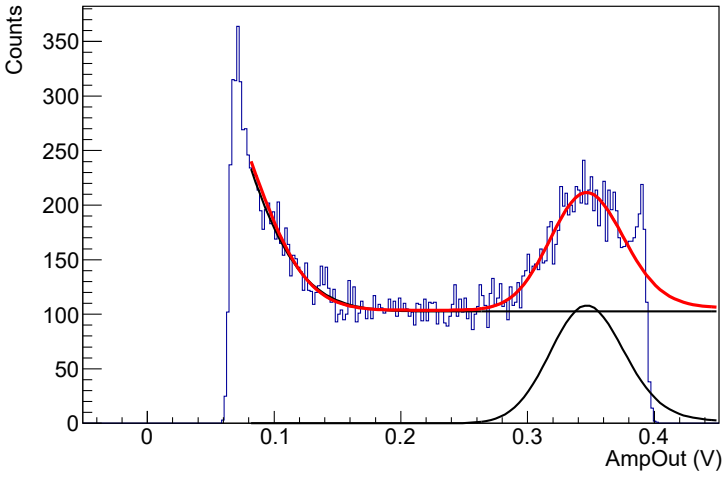


Figure 15.2: Spectrum of an analog single pixel measurement. The pixel was illuminated with ^{90}Sr-electrons at 0.546 MeV. In black, the fits for noise background and Landau-Gaussian signal are drawn, the combined fit is drawn in red. The upper end of the distribution is cut off as the amplifier reaches saturation.

For illumination with ^{90}Sr-electrons, the sensor diode is reverse biased with −10 V. The DUT has been produced on 80 Ωcm substrate. The collected data were filled into the histogram in figure 15.2. Again, the noise peak has been cut off by the oscilloscope's threshold. The histogram is limited towards higher energies as the amplifier saturates at its supply voltage.

The most probable value (MPV) of the electron signal was determined by fit. The fit function is composed of a Gaussian-shaped noise peak with offset and the Landau-Gaussian convolution for the signal. Both fit components are shown individually as black lines. The combined fit is drawn in red. From the fit, the MPV is extracted and translated from volt to electrons using equation 15.1:

$$U_{\text{AmpOut}}\left(^{90}\text{Sr}_{\text{MPV}}\right) = 347\ \text{mV} \pm 34\ \text{mV} \tag{15.2}$$

$$Q\left(^{90}\text{Sr}_{\text{MPV}}\right) = 4443\ e^- \pm 435\ e^- \pm 5.5\% \tag{15.3}$$

The increased gain, together with moderate depletion voltage and substrate resistivity, is sufficient to saturate the amplifier. This saturation on the high energetic end of the spectrum can not be correctly depicted by the fit function and is therefore excluded.

At first glance, this appears as disadvantage compared to H35Demo, which had a higher dynamic analog range due to higher supply voltage (3.3 V) and lower gain. However, both MuPix8 and ATLASpix1, except for the M2 matrix, come with an analog measurement

integrated in the fast readout. The secondary time stamp saves the point in time, when the trailing edge of the signal crosses the threshold again. Together with the event time stamp, the time over threshold (ToT) can be calculated. The dynamic range of ToT is not limited by the saturation voltage of the amplifier, but is still a good measure for the charge deposited in the pixel diode, when the analog signal has reached saturation.

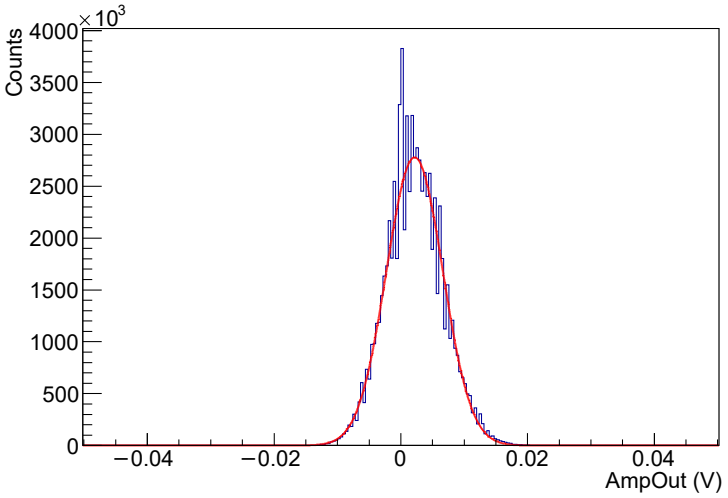


Figure 15.3: The histogrammed baseline noise of the in-pixel amplifier.

Additionally to the two previous measurements, the baseline noise of the amplifier has been recorded (Figure 15.3) The noise sigma is

$$\sigma U_{\text{noise}} = 4.44 \text{ mV} \tag{15.4}$$

$$\sigma Q_{\text{noise}} = 57\ e^- \pm 5.5\%. \tag{15.5}$$

Assuming signals from MIP-like electrons of the ^{90}Sr-measurement, the signal-to-noise ratio (SNR) is:

$$\text{SNR} = 78 \pm 7 \tag{15.6}$$

Just like the H35Demo, the aH18 sensors have an analog buffer between amplifier and output pad. Therefore, the analog noise measured is not identical with the noise at the output of the amplifier, but can still be used as relative measure.

15.2 HitBus – time over threshold measurement

Both the ATLASpix1 matrices and the MuPix8 have a HitBus output. The HitBus is a characterization and debug feature. This pad can be connected to one or more outputs of the first comparator after the in-pixel amplifier. When the amplifier crosses the comparator's threshold, the HitBus transmits a pulse, as long as the analog signal exceeds the threshold. In case of ATLASpix1, this pulse is positive, in case of the MuPix8 it is negative.

Figure 15.4 shows a screen capture from an oscilloscope probing the injection signal (green), the amplifier output (purple) and the HitBus (blue). A short time after injecting the charge, the amplifier's output voltage level rises. Once above the threshold, the HitBus is triggered until the analog signal drops below the threshold again.

The HitBus transmits a digital signal, but the pulse length is an analog measure signal pulse size. Its dynamic range is larger than the pulse height measurement of the amplifier signal. Figure 15.5 illustrates the response of the amplifier (purple) and HitBus (blue) to

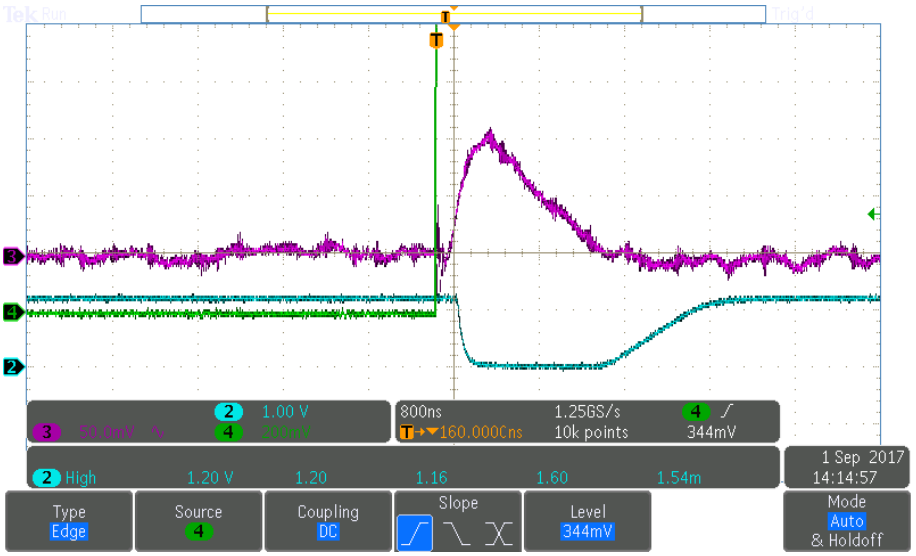


Figure 15.4: Screen capture of the Tektronix oscilloscope. The analog output of the amplifier (purple) of a pixel on MuPix8 starts to rise shortly after the test signal has been injected (green). The HitBus shows a signal, once the analog signal has crossed the comparator's threshold and decays after it has dropped below.

different charge signals. Detection threshold is depicted in green. Small signals (a) can be measured by both analog signal height and HitBus pulse length. For a certain signal, the amplifier saturates (b). A charge signal exceeding the saturation point (c) will still stretch the analog signal in length, and therefore increases the pulse length on the HitBus. However, observing only the pulse height is not sufficient to determine the original charge signal.

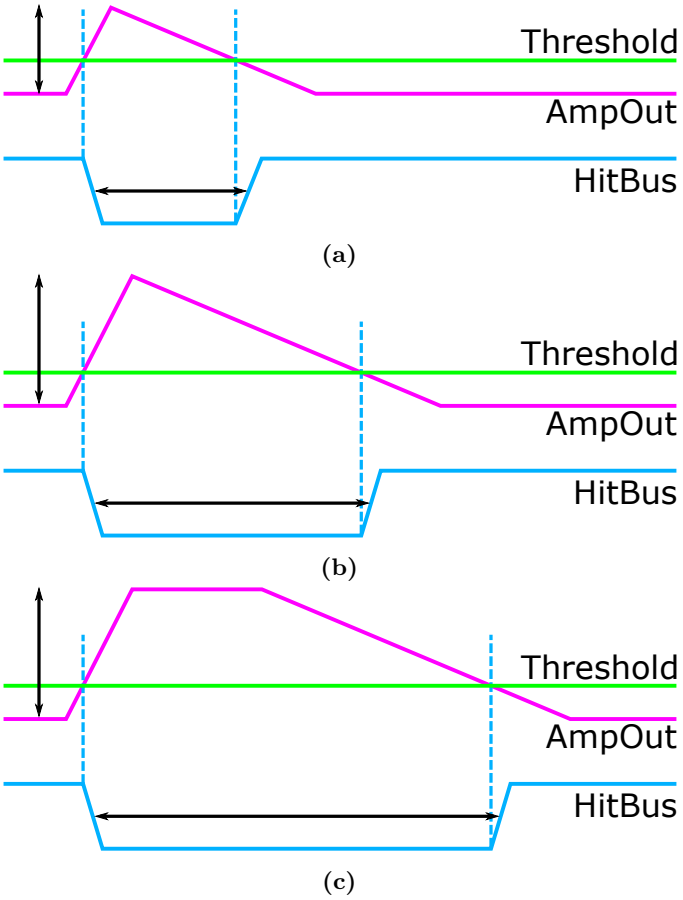


Figure 15.5: Illustration of the dynamic range of amplifier output (purple) and HitBus (blue). Three cases of signals above threshold (green) are discriminated:
(a) Signal well below saturation: The analog pulse height, indicated by the black arrow, is a measure for the amplifier's input signal.
(b) Signal at saturation: The amplifier output voltage reaches its maximum value.
(c) Signal above saturation: The the pulse height is insufficient to calculate the amplifier's input signal.
Time over threshold is independent of the amplifier's pulse height. It increases beyond the saturation point of the amplifier.

15.2.1 Calibration of HitBus

The HitBus characteristic depends strongly on the comparator's threshold setting. Test signal injections can be utilized after applying new settings to evaluate the changes. Test signal generation is linear within known limits, but has to be calibrated, too.

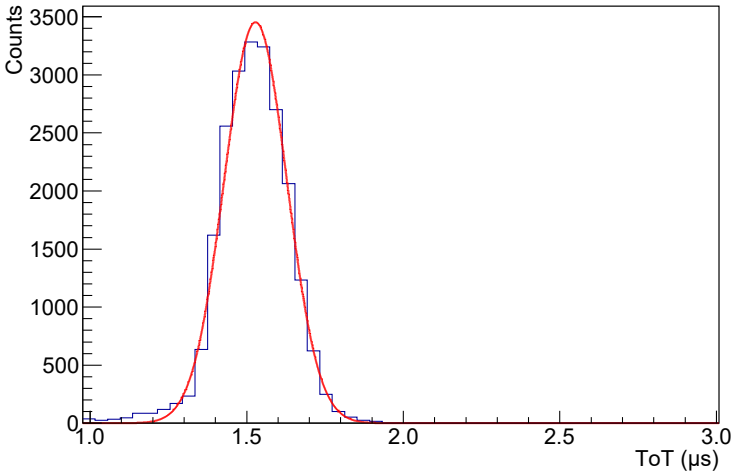


Figure 15.6: ToT histogram of signals from ^{55}Fe decays measured with an oscilloscope on HitBus.

The pulse length of the HitBus is recorded while the sensor is illuminated with X-rays from an ^{55}Fe source. The data are filled into a histogram and a Gaussian fit is applied (Figure 15.6). The mean ToT in this measurement is 1.528 µs.

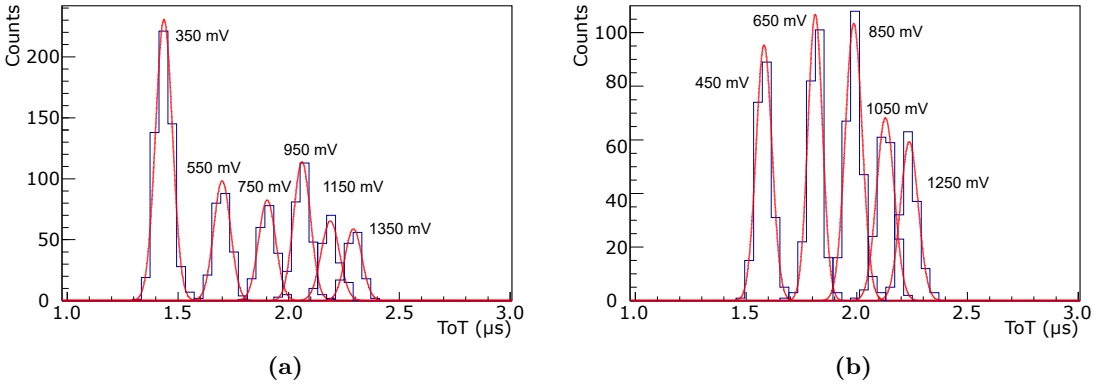


Figure 15.7: ToT histograms of test signal injections ranging from 350 mV to 1350 mV. (Plotted in two graphs.)

The pulse length of the HitBus is now recorded for test signal injections of different voltages. Figure 15.7 shows the ToT histograms of eleven voltages. For better discriminability, the eleven histograms are divided into two plots. From the Gaussian fits, the parameters in the following table are extracted. The injections have been calibrated using the previous ^{55}Fe measurement.

Injection (mV)	Injection (e^-)	ToT (µs)	σToT (µs)
350	1541	1.437	0.038
450	1981	1.581	0.036
550	2421	1.699	0.038
650	2862	1.812	0.033
750	3302	1.902	0.039
850	3742	1.986	0.038
950	4183	2.060	0.038
1050	4623	2.129	0.040
1150	5063	2.188	0.041
1250	5504	2.237	0.040
1350	5944	2.293	0.038

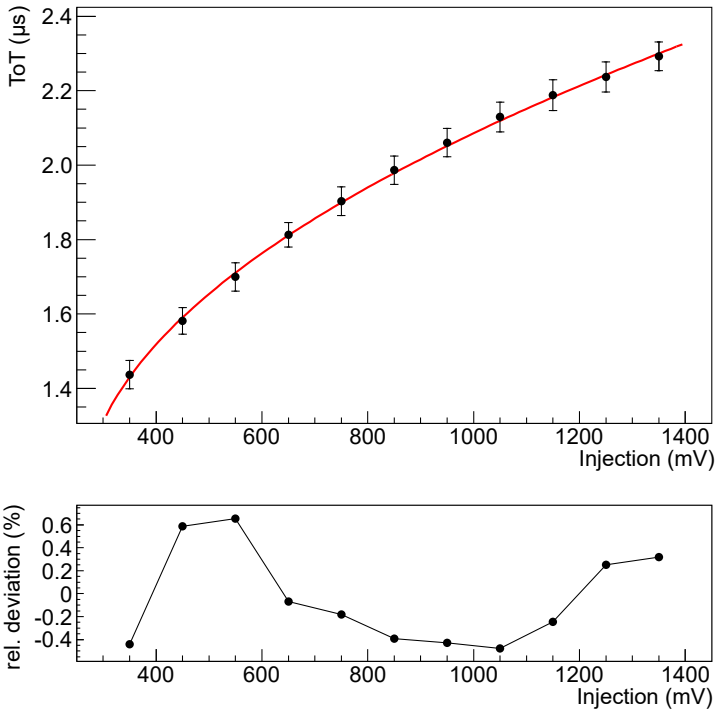


Figure 15.8: Time-over-Threshold (ToT) of HitBus as a function of injection voltage. A square root fit has been applied. The relative deviation of fit and data points is shown below.

The values from this table are visualized in figure 15.8. Unlike the amplifier output (AmpOut) measurement, the relation between signal and ToT is not linear, but of square root shape. However, the exact shape of the square root depends on the chip settings, for some settings, it can be assumed linear.

15.2.2 Charged particle detection using HitBus

Once the HitBus pulse length has been calibrated, it can be used to measure the charge deposited in the pixel diode by charged particles. Therefore, a ^{90}Sr source is placed above the sensor and a pixel is connected to the HitBus.

The pulse lengths of the incoming signals are recorded. The resulting spectrum shows two peaks. A lower energetic peak originating, from secondary processes and charge sharing and a higher energetic peak formed by signals caused by electrons traveling directly from

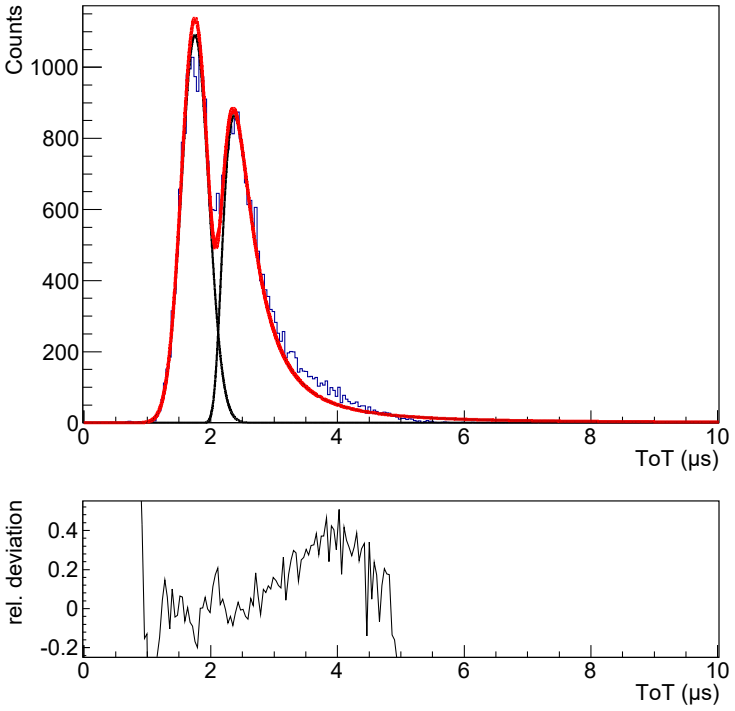


Figure 15.9: ToT histogram of ^{90}Sr. The background is fit by a Gaussian, the signal by a Landau-Gaussian (both black). The combination of both is shown in red. Other than in the analog measurement (Figure 15.2), no saturation is visible. However, the Landau tail can not be perfectly fit, due to the non-linear parametrization of ToT.

the source to a single pixel. The latter is fit by a Landau-Gaussian distribution. The low energetic peak is assumed Gaussian. The fits are shown in black, their superposition in red.

In contrast to the Gaussian part of a ^{90}Sr-histogram measured via the AmpOut, the background Gaussian is not cut off. The AmpOut cutoff originates from the threshold set on the oscilloscope to discriminate signals from background. In case of the HitBus, the comparator threshold has not only a discriminatory effect, but also an effect on the

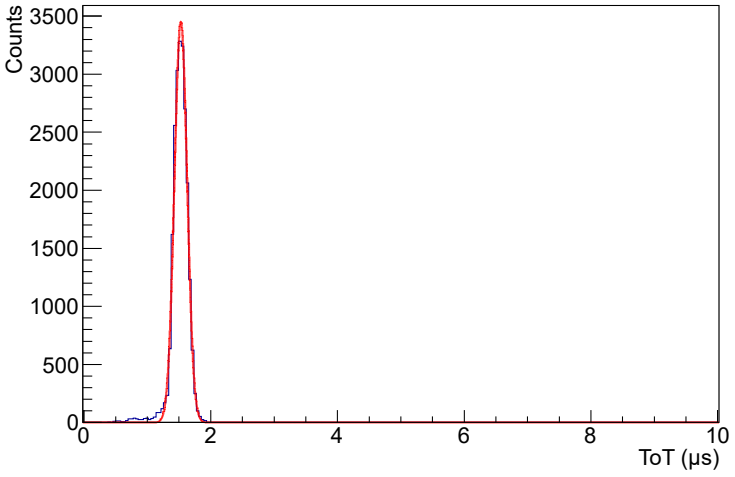


Figure 15.10: ToT histogram of ^{55}Fe with the same settings as the ^{90}Sr measurement in figure 15.9.

pulse length. Both in analog measurement and HitBus measurement, the number of background signals is reduced by a higher threshold. But a higher threshold of the HitBus comparator shifts the spectrum to lower values. The Most Probable Value (MPV) of the Landau-distribution is 2.370 µs.

The ToT duration is translated to electrons using an ^{55}Fe calibration (1639 e^-) at the same settings. The respective histogram is shown in figure 15.10 and has a mean of 1.528 µs. Assuming a linear characteristic of ToT, the MPV in this measurement would translates to 2540 e^-. The square root behavior (Figure 15.8) however leads to a MPV between 3000 e^- and 4000 e^-.

15.3 Fast readout – time over threshold

Whilst AmpOut and HitBus are debug features, streaming digital hit data via a readout link is the most potent way of reading information from ATLASpix1 and MuPix8. AmpOut can only probe a single pixel of the first column at a time. The HitBus can theoretically probe all pixels at the same time, but merges the hit information on a single channel. All measurements with more than a few pixels activated or with more than a tiny hit rate will not deliver useful data. Both AmpOut and HitBus need external measurement devices like oscilloscopes and offline analysis to access the data.

The Fast Readout directly provides digitized data, which only needs basic decoding in order to be read- or displayable. Furthermore, each readout link is capable of streaming out the data from all pixels concurrently up to the event rate it has been designed for.

The hit word of all matrices is 32 bits long and holds information on the hit location and timing. All matrices, except the ATLASpix1_M2 matrix, additionally send analog information. The analog information investigated in this section is the Time over Threshold (ToT) calculated from the event time stamp and the secondary time stamp generated from the falling edge of the amplifier's signal.

15.3.1 Calibration of fast readout

A first impression of Fast Readout is gained by externally generated test signal injections. The generated readout data are recorded and the ToT is calculated from primary and secondary time stamps. The injections have been sent to a single pixel and ToT has been calculated during decoding. For each injection voltage, a ToT-histogram has been made.

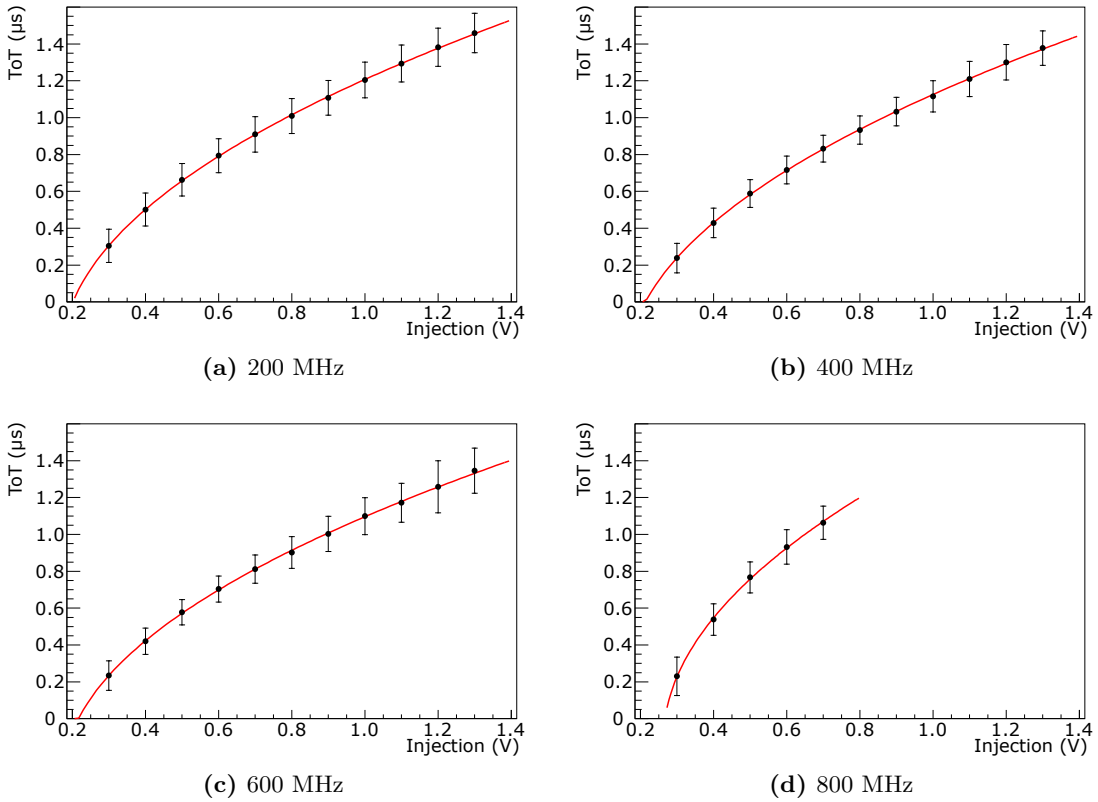
Figure 15.11 shows the mean and sigma of these histograms. As expected from the HitBus ToT measurements, the relation between injected charge and ToT is not linear but of square root shape.

The Fast Readout requires more settings than the HitBus or AmpOut: Additionally to the configuration for DACs and matrix, the readout state machine has to be configured and several clocks are to be generated accordingly.

The starting point of a configuration is the choice of readout speed. This speed determines how fast the readout link sends out bits to the FPGA. As mentioned in the sensor description, a fast readout speed is generated from a slow external reference clock by a PLL. A lower readout speed requires external clocking. All other settings have to be chosen appropriately.

The step width of both time stamps scales with the speed of the readout clock. However, their speed can be individually reduced by configuration.

For the plots in figure 15.11, it was tried to not change the configuration, except for the readout speed. This is only possible for the readout speeds 200 - 600 MHz. For 800 MHz,



(a) 200 MHz **(b)** 400 MHz

(c) 600 MHz **(d)** 800 MHz

Figure 15.11: Plots (a) to (d) show ToT as a function of signal at different clock speeds. (a) to (c) have been measured at the exact same settings with external clocking. d) The measurement at 800 MHz requires internal clocking (PLL). The dynamic range is reduced as the 6-bit secondary time stamps reaches overflow faster, here above an injection voltage of 0.8 V.

the operation of the PLL requires distinct changes in the configuration. The components' dimensions of the PLL were chosen to generate high clock speeds, it is not possible to operate it stable at low frequencies.

To each plot a fit of the form

$$\theta\left(U_{\text{Inj}}\right) = a + b\sqrt{U_{\text{Inj}} - U_0}$$

was applied (red). Where θ is the ToT, U_{Inj} the injection voltage and the free parameters are a, b and U_0. The fit parameters of plots with external readout clock, appear very similar. Direct comparison of the fit curves of each injected signal of different readout speed, however reveals that ToT measured with faster readout speed is smaller, compared to measurements with slower clock. The change per 100 MHz is $\approx 5\%$.

The dynamic power consumption of digital logic circuits scales with the operational frequency (see Appendix A.1):

$$P \propto f.$$

The supply voltage remains constant and therefore an increased current has to provide the additional power. The increased current in turn, might cause different voltage drops in some parts of the chip. The outcome is slightly shifted voltages and transistor working points, resulting in a decreasing ToT.

In figure 15.11d the same plot measured with an internally generated readout clock of

800 MHz is shown. For internal clock generation many settings had to be changed, thus it is not directly comparable to the other plots, but the square root shape remains.

The most notable change is the absence of injection signals stronger than 0.7 V. Compared to the speed of the readout state machine, the ToT is too large. The state machine passes the hit data on, while the signal is still above threshold and no secondary time stamp was generated. In consequence, the readout data contain only random values for the secondary time stamp and ToT can not be calculated.

This issue is tackled by slowing down the readout state machine providing enough time to generate the secondary time stamp, while keeping the high time stamp speed. This is to be kept in mind when preparing the detector for a measurement campaign: Not only the time stamp speed has to suit the expected signal, but also the speed of the readout state machine.

15.3.2 X-ray detection using fast readout

HV-CMOS sensors are designed to detect charged particles passing through the detector. Photons are useful in characterization of the detector. They deposit a known charge, unlike charged particles which deposit a Landau-distributed charge.

Radioactive sources are not easily available and are limited to few energies, but do not require additional equipment. X-ray tubes produce a wide spectrum of X-rays, which can be sent on targets made of a single element. They emit X-rays with their respective characteristic energies.

An ATLASpix1 sample was placed next to an X-ray tube (for the irradiation facility see Appendix B.2), shielded from the tube's exit, but facing a target-holder. The target-holder can be equipped with a target from a selection which has a single dominant characteristic energy. The used targets and their K_α energies are:

Element	Symbol	K_α-energy (eV)	e^--h^+-pairs generated in Si by K_α
Iron	Fe	6403	1779
Copper	Cu	8048	2236
Zinc	Zn	8639	2400
Molybdenum	Mo	17479	4855
Silver	Ag	22163	6156
Tin	Sn	25271	7020

While the sample was illuminated, the Fast Readout was activated. ToT was calculated from the primary and secondary time stamp offline. The result was filled into a histogram for each pixel and a Gaussian fit was applied.

Figure 15.12 shows the mean ToT of each pixel. The data were collected during a measurement with a copper target. Most ToTs are between 0.5 µs and 1 µs. Values outside this range originate typically from failed fits or noisy pixels. However, within each half-matrix a vertical drop from top to bottom is visible. This is being investigated in chapter 15.3.3.

Instead of creating histograms for each individual pixel, a histogram for the entire matrix can be made. Such histogram is shown in figure 15.13. The data are collected from a sensor illuminated via a copper target.

The mean and sigma of measurements with different target are shown in figure 15.14a as a function of the charge generated by the K_α-energy. The large uncertainty of the higher

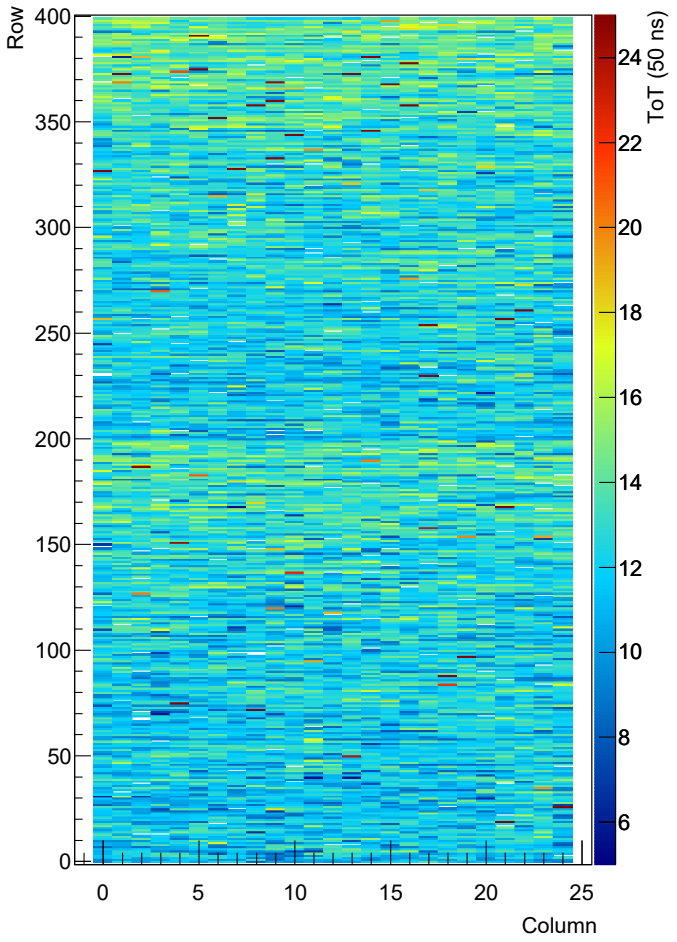


Figure 15.12: Mean ToT histogram measured on ATLASpix1_Simple illuminated with X-rays from a copper target. The matrix appears divided in a top and bottom half. In each half the higher rows have a higher ToT. The matrix has not been tuned.

energetic data points is a result of a number of noisy pixels, which add a large number of noise hits to the histogram. The resulting histogram comprises two overlapping peaks, noise and tin signal. A double Gaussian-fit delivers good mean values for each, but with poor uncertainty. Selecting a pixel without significant noise hits results in a similar diagram, but with greatly reduced error bars (Figure 15.14b).

Figure 15.15 compares the histogram of the full matrix (a) to the histogram of a single pixel of an X-ray measurement (b) with tin target. The large peak of the full matrix histogram is the noise peak. The bump on its trailing edge is the tin signal. The single pixel measurement shows no noise peak and only few background signals.

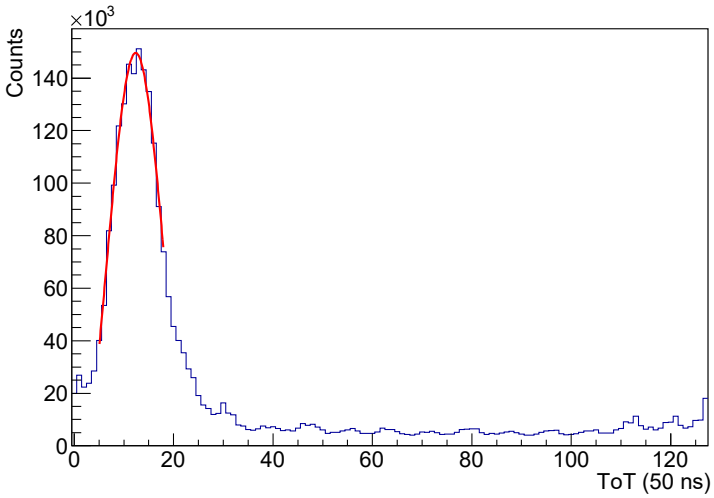
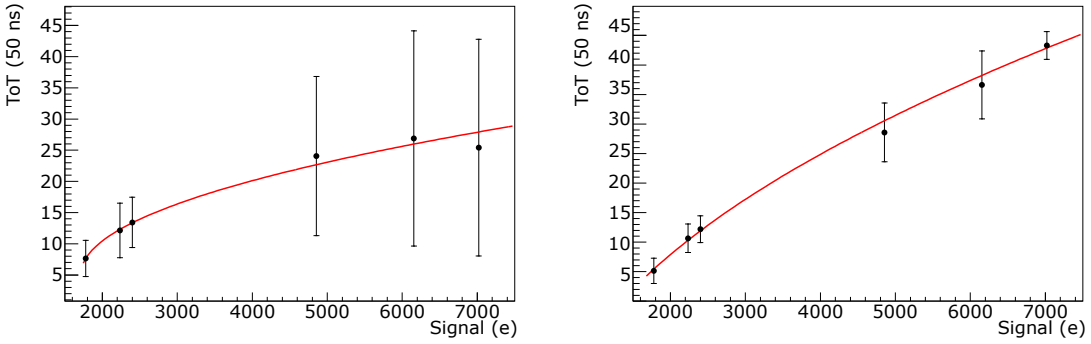
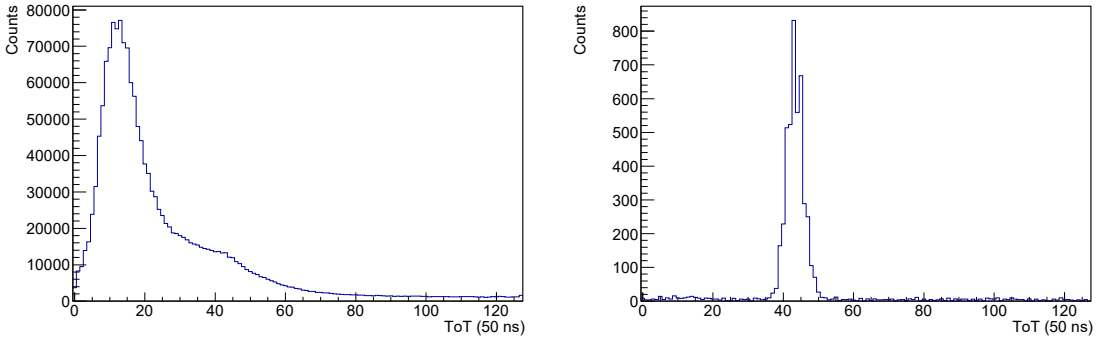


Figure 15.13: Combined spectrum of all pixels illuminated with X-rays from a copper target. Mean and standard deviation are determined by Gaussian fit.



(a) All pixels **(b)** Single pixel

Figure 15.14: ToT as a function of X-ray signal.
Plot (a) shows the result from all pixels combined. Higher energetic X-rays cause a lot of failed measurements, thus large errorbars: Some pixels suffer from a too fast state machine, others from a large number of noise hits.
Plot (b) shows the data from a single pixel measurement.



(a) All pixels **(b)** Single pixel

Figure 15.15: Tin spectra of all pixels (a) and a selected pixel (b).

15.3.3 Matrix effects in ATLASpix1_Simple

Analysis of the ToT of all pixels when illuminated with X-rays shows a distinct inhomogeneity as seen in figure 15.12. This behavior is investigated more in detail in this section.

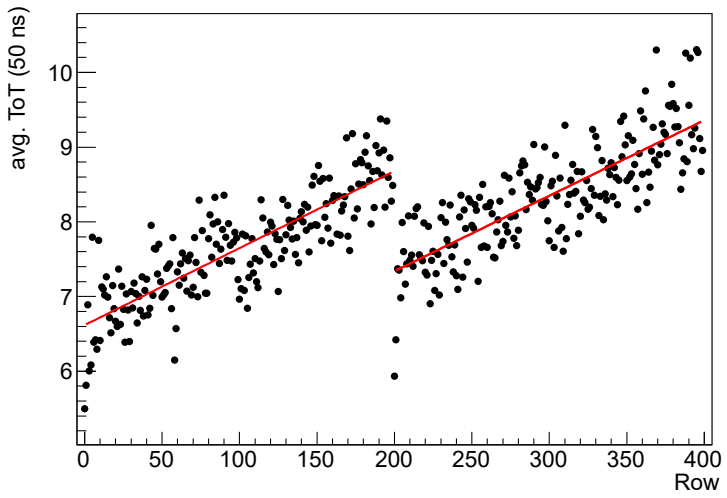


Figure 15.16: The plot shows the average ToT of each row. The data were taken on an iron target X-ray measurement. Higher rows have a higher ToT. This behavior appears twice, once for the pixels of each readout column.

The row-dependency of ToT is more distinct when the mean ToT of the pixels in a row is plotted. Figure 15.16 shows the mean ToT over the row index for an iron target measurement. There are two readout columns (ROC) per pixel column. The bisection appears between row 199 and row 200. This is the cut between the two readout columns: the first ROC is connected to pixels 0-199 of the pixel column, the second ROC is connected to pixels 200-399. Therefore, the main reason for this behavior is to be found in the periphery. Most likely a voltage drop in row direction at the line receivers or the edge detectors of the ROC cause this behavior. However, the matrix has an impact, too. Otherwise, the two fits would be at the same height, but the fit on the pixels on the top half of the matrix is shifted towards higher mean ToTs. This difference might be caused by a voltage drop of the external supply voltage, which is connected to the bond pads on the top edge of the chip.

The equivalent analysis for a column dependency is inconclusive due to the limited number of columns.Horizontally, only 25 columns cover 3 mm. Vertically, 2×200 pixels cover nearly 20 mm.

15.3.4 Matrix effects in MuPix8

Although MuPix8 shares many design features with the ATLASpix1 and has been produced in the same submission. But some design features are different, e.g. the readout structure and matrix wiring. Therefore, the signal gain in column and row direction of MuPix8 is investigated.

For this measurement, the matrix is illuminated with X-rays from an ^{55}Fe source. The analysis is the same as described before. The readout data of the first readout link, which is connected to the pixels of the first 48 columns, are recorded.

Figure 15.17 shows mean ToT over row index. In contrast to the same plot of ATLASpix1_Simple, the mean ToT is dropping rapidly with increasing row number. The mean ToT of ATLASpix1_Simple was increasing linear for higher rows.

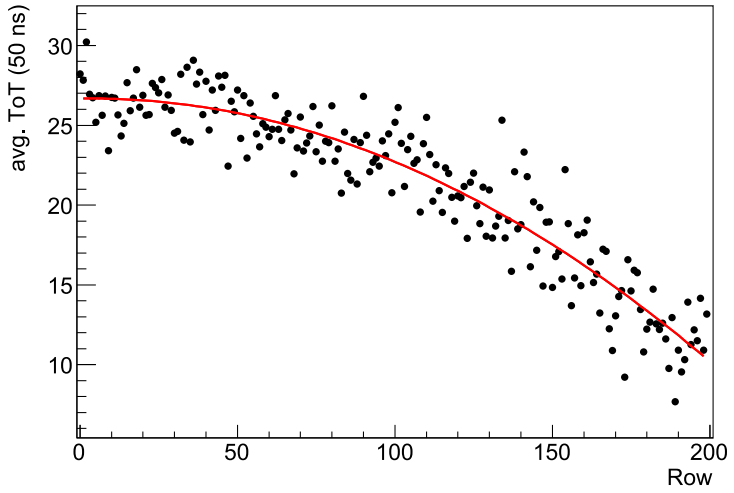


Figure 15.17: Average ToT of each row. The data were taken on an ^{55}Fe source X-ray measurement. ToT is dropping faster than linear. The applied fit is a second order polynomial.

The reason for this behavior is the connection between the pixels in the matrix to the readout cells in the periphery and the way how the signal is transmitted. Signal variation in ATLASpix1 is dominated by voltage drops in the periphery. In MuPix8 the variations are caused by different length of the metal connections from pixel to readout cell. The MuPix8 transmits analog signals from pixel to periphery, thus the signal is more sensitive to transmission distance. ATLASpix1 has the comparator located in pixel and transmits quasi digital signals which are less sensitive to line length.

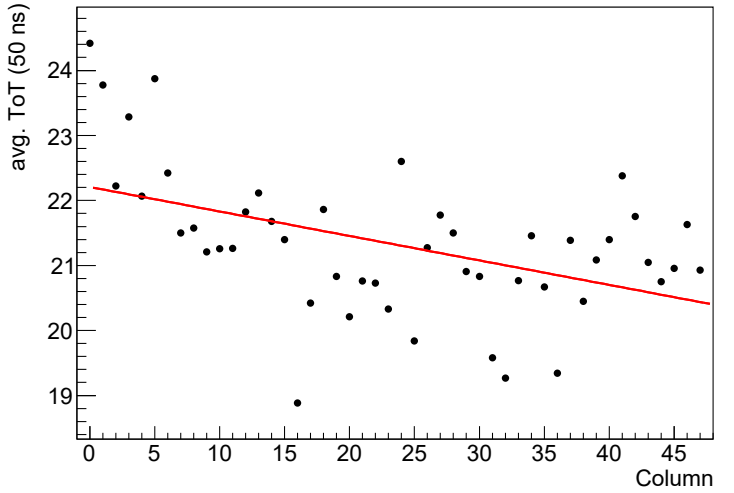


Figure 15.18: Average ToT of each column. The data were taken on an ^{55}Fe source X-ray measurement. Higher rows have a lower ToT. The left columns have longer ToT than those on the right.

The measured area of the MuPix8 is approximately 3 mm wide with 48 columns. Figure 15.18 shows the average ToT in column direction. The number of pixel columns is larger than in the previously investigated ATLASpix1 matrix. Therefore, the slightly dropping trend can be observed. The most likely explanation for this behavior is the increasing distance between both pixels and readout cells to the bias block with increasing column. A voltage

drop along the supply lines is the most probable explanation for a column dependency of the signal gain. However, the actual voltage level can only be probed at the bond pads. An example calculation for the voltage drop along a supply line if shown in Appendix A.6.

15.3.5 Charged particle detection

Instead of X-ray photons, the sensor is now illuminated with MIP-like electrons from a ^{90}Sr source. The following measurement investigates the front-end, which is similar in MuPix8 and ATLASpix1. The exit window of the source is small compared to the area of the sensor. Therefore, a hit map is made to verify alignment of source and sensor.

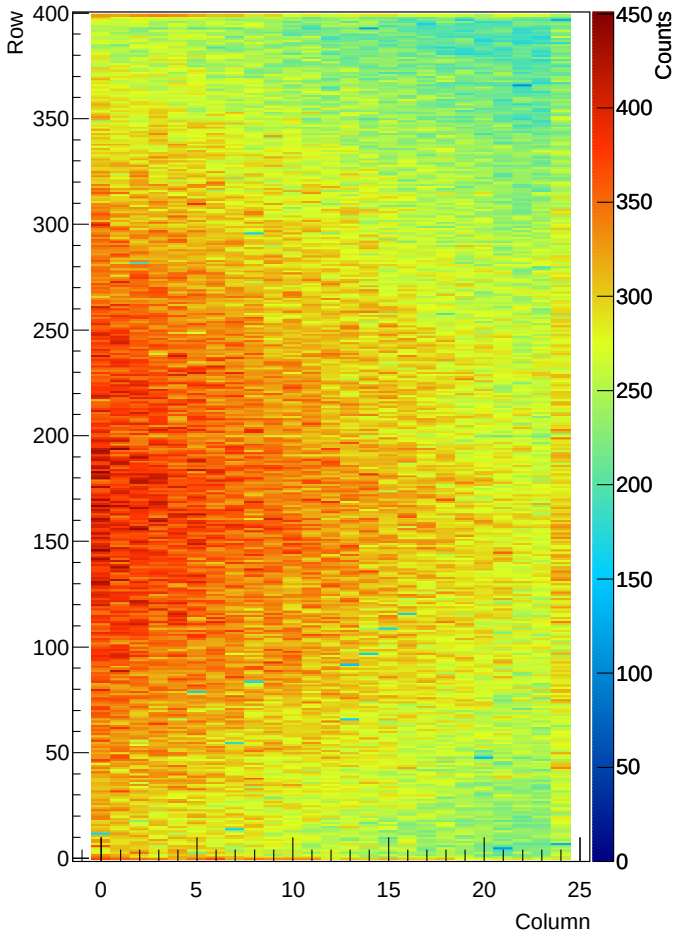


Figure 15.19: Hit map plotted from data collected during a ^{90}Sr source measurement. The source was placed above the front side of the sensor, slightly left of center.

The hit map is shown in figure 15.19. The source is slightly of the center, but covering the active area sufficiently. The first and last of both column and row show a higher activity than the rest of the matrix. This increased noise behavior has been observed for most characterized HV-CMOS sensors and originates from edge effects of the pixel outlines. These pixels are not as well protected against leakage and surface currents as pixels in the middle of the matrix, which causes additional noise. Sensitivity is not affected and therefore this is not considered a substantial issue.

Figure 15.20 shows the ToT spectrum from a measurement with 40 V depletion voltage. The data collected from the entire matrix were used. The MPV is 20 clock cycles or 1 µs.

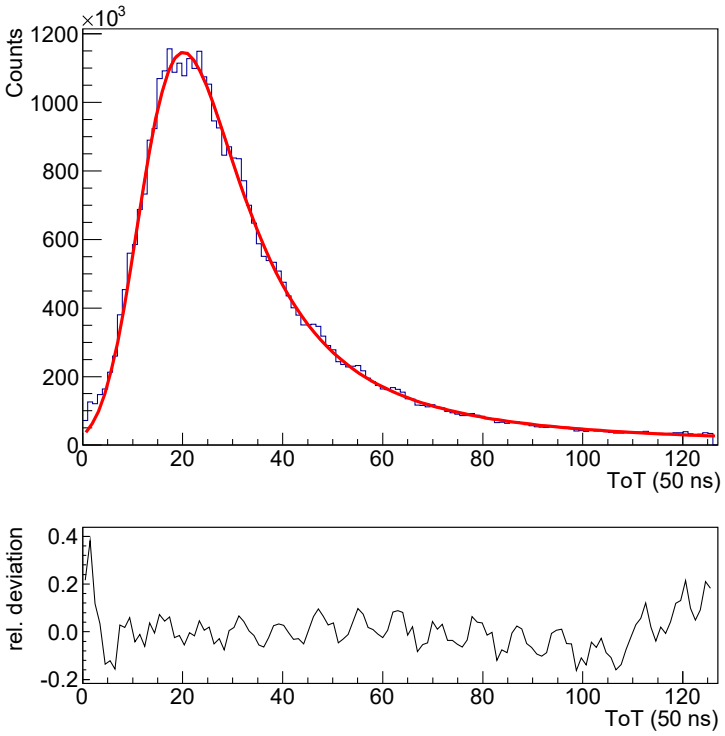


Figure 15.20: ToT spectrum measured on the full matrix illuminated with ^{90}Sr. The most probable value is determined by Landau-Gaussian fit to 1 µs at 40 V depletion voltage. Relative deviation of fit and data points is shown below.

Increasing the depletion voltage to 60 V shifts the most probable value (MPV) of ToT to 26 clock cycles or 1.3 µs. A translation to a number of electrons is not feasible at this point, as a calibration on pixel level is required, as well as cluster-summing[1].

15.3.6 Cluster size

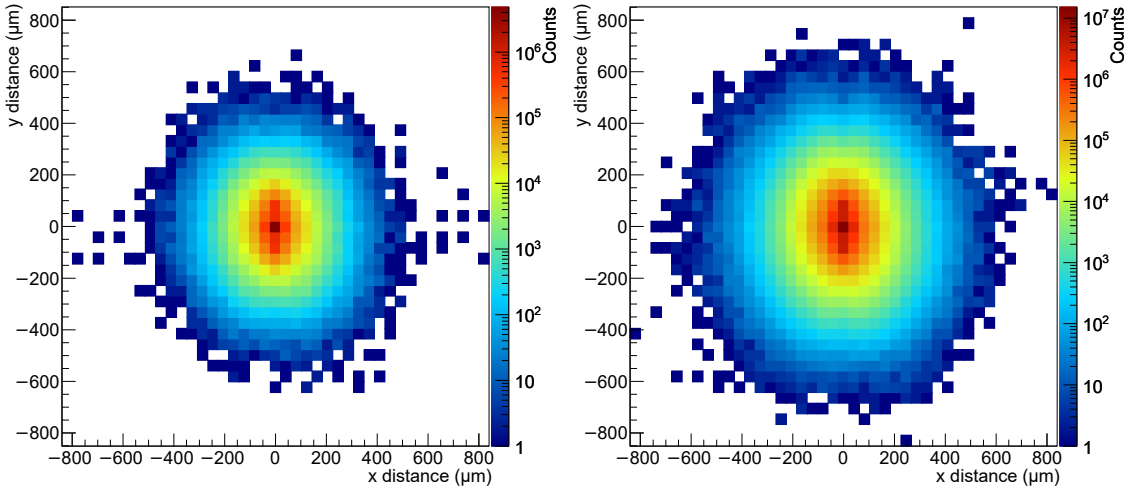
Charged particles interact with the sensor material along their path and generate electron hole pairs. The charge is not only generated exactly on the trajectory, but with decreasing probability in greater distance. The size and shape of the resulting charge cloud depends on the particle path's inclination and its energy. The size of the charge cloud can easily exceed several hundred µm. Therefore, a significant number of the events spreads over several pixels. The ensemble of all pixels responding to a particle impact is called cluster.

The data collected in a measurement do not come with cluster information. In order to reconstruct clusters, the data are analyzed with respect to event location on the sensor, event time and position in the data set as follows:

1. The first entry of a data set is chosen as cluster seed.
2. The single pixel hits are read out not in chronological order. Theoretically, they can be sorted by time stamp offline. However, the 10-bit time stamp is not sufficient for this purpose: Hits with similar time stamp might be several time stamp overflows apart. Therefore, only a limited number of hits is checked versus the seed pixel. In this example, 15 are checked after the last cluster member has been identified and the search is aborted.
3. To join a cluster, the inspected hit has to be close to the outline of the forming cluster. In this example the distance can be not larger than 300 µm.

[1]Such measurement conducted on ATLASpix3 has been presented in [133]

4. Finally, the time between the candidate and any of the cluster members has to be smaller than a selected duration, in this example 500 ns.
5. If no further member of a cluster is found, the cluster is removed from the raw data and saved to the clustered data set. Then the process starts over for the next cluster.



(a) -10 V **(b)** -70 V

Figure 15.21: The superposition of all clusters of a ^{90}Sr measurement on MuPix8, first matrix. The clusters have been calculated from plain readout data and their center was shifted to (0|0). (a) for -10 V depletion voltage, (b) for -70 V.

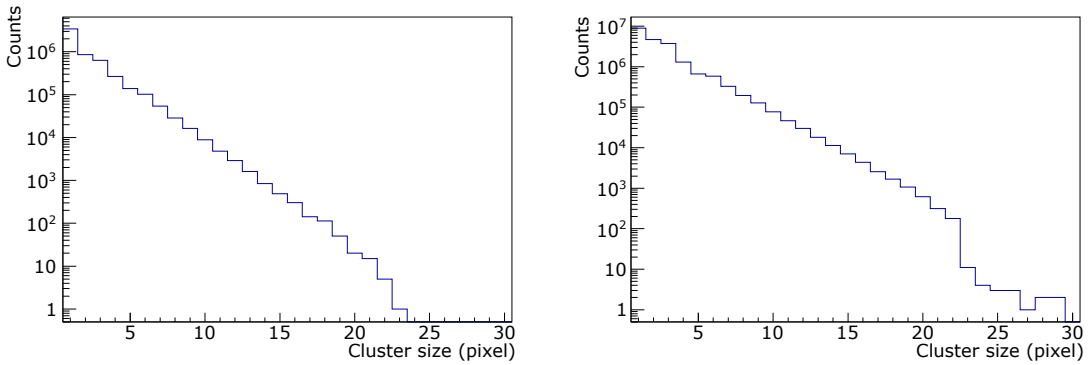
The center of each cluster is calculated as arithmetic mean. Figure 15.21 shows the superposition of all clusters of a data set. A bin is increased if a pixel with the respective relative location to the cluster center is found. In this way, sub-pixel resolution is achieved.

The cluster shape represents the shape of the underlaying charge cloud. But other properties can affect the shape, too. The shown plots have a slight asymmetry of x- and y-direction. As the inclination of the particles from the source is negligible and the pixels are nearly of quadratic shape (measurement on MuPix8), the most probable reason is crosstalk along the signal lines connecting the pixels of the matrix with the periphery. Lines of vertically neighboring pixels run next to each other. A small analog signal right below detection level on one of these lines can be lifted above detection level by a strong signal being transmitted on an adjacent line.

The average size and shape of the clusters depend on the incoming particle, its energy and the target material. The recorded size and shape however, depend on the collection efficiency of the generated charge and where this charge is collected. Only charge collected by drift can cause a signal strong enough to cross detection level. Therefore, the recorded size of the cluster depends on the used depletion voltage, as only charge in the depletion zone is collected in time. Figure 15.21a shows the cluster shape with an applied depletion voltage of -10 V. If the measurement is repeated with a higher depletion voltage of -70 V, the change in cluster size is directly visible in figure 15.21b.

For quantitative analysis the number of hit pixels per cluster is filled into histograms. In figure 15.22 two of these histograms are displayed for depletion voltages -10 V and -70 V. The used data sets are the same as in the previous cluster shape plots. The extracted key-value is the mean cluster size.

In the presented measurement, the mean cluster size N_{cluster} is 1.878 for depletion voltage -10 V and 2.357 for -70 V. The same analysis was applied to data sets with other depletion



(a) -10 V **(b)** -70 V

Figure 15.22: Histogram of cluster size for -10 V (a) and -70 V (b) depletion voltage. The cluster size is here the number of 80×81 µm^2 pixels that contribute to a cluster. The mean cluster size at -10 V is 1.878, the mean cluster size at -70 V is 2.357.

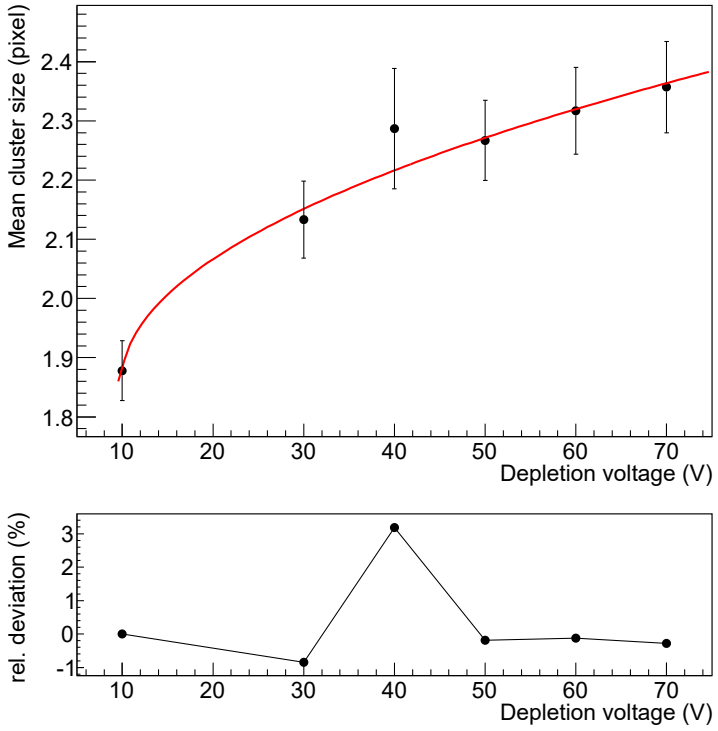


Figure 15.23: The mean cluster size as a function of depletion voltage. The square root fit function is drawn in red.

voltages and the result is shown in figure 15.23. The collected charge is proportional to the depletion depth d, which is proportional to $\sqrt{U_{\mathrm{HV}}}$. Therefore, a square root fit is applied.

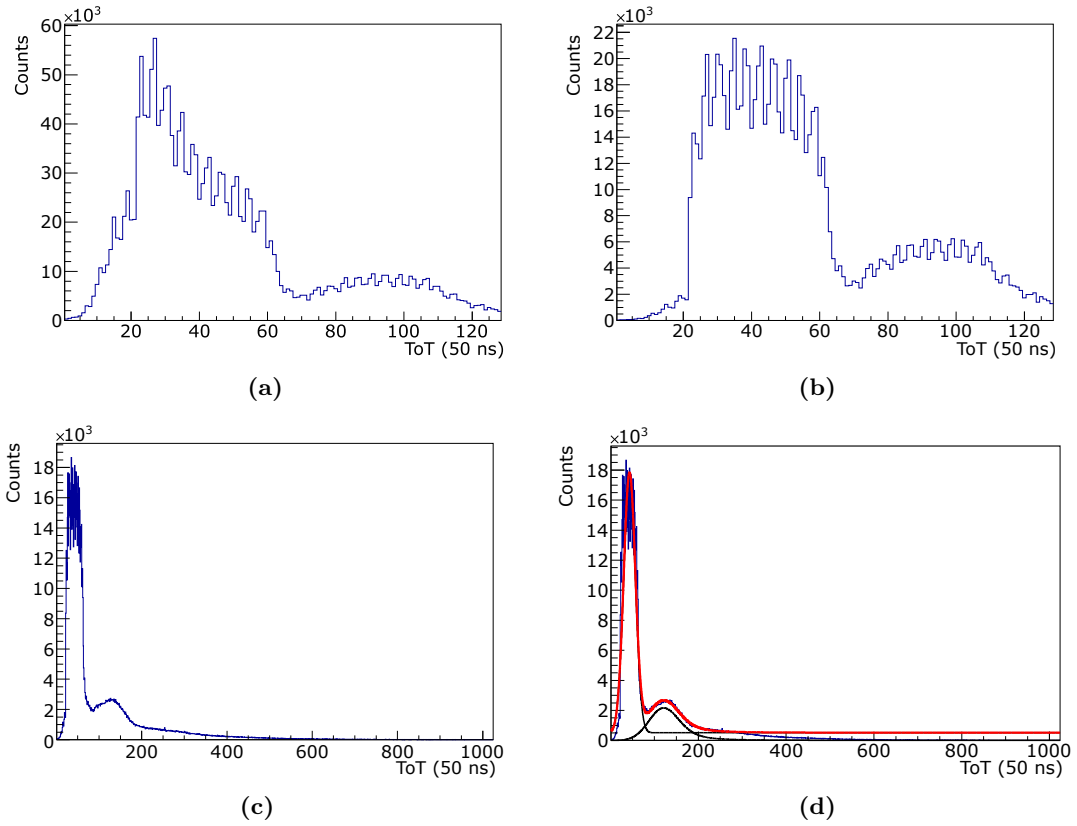


Figure 15.24: Each plot shows a ToT histogram measured on a ^{90}Sr illuminated sensor. The individual ToT values of a cluster are handled in four different ways:
(a) Single-pixel ToT values without clustering.
(b) Clustered data and only the largest ToT value per cluster are used.
(c) All ToT values of a cluster summed up.
(d) Same as (c) but with applied fits to signal and noise peak.

The fact that a significant number of events causes clusters, requires advanced analysis of the read out data to determine the most probable value of a ToT-spectrum. Figure 15.24 shows three ways, how to treat multi-pixel events:

(a) A spectrum without clustering. The ToT of every hit is filled into the histogram. The maximum ToT before overflow is 128. The periodical ripple of the data is being addressed below.

(b) Only the pixel hit with largest ToT is picked for this diagram. This approach is already an improvement. It is often used to improve time resolution as large signals suffer less from time-walk. The block-shaped background peak has its origin in hits that have been read out before a negative edge could be detected and thus have a random number for the secondary time stamp. The lower edge of this block is defined by the speed of the state machine controlling the time stamp generation. Lower values can not cause this issue. The upper edge at 64 is due to the secondary time stamps maximum value $2^6 = 64$. Slowing down the fast readout state machine of MuPix8 solves this problem, but reduces the number of hits that can be read out in a given time.

(c) The third spectrum is obtained by summation of all ToT values of a hit. In this way also large charge depositions can be displayed. The signal peak becomes more

dominant. However, the summation of ToT values is not very precise and comes with significant systematic errors:
Calibration of ToT is conducted with single pixel events. This method respects that ToT can only be measured above a certain threshold. Adding two or more ToTs does not result in the equivalent value of a larger hit in a single pixel, because the virtual sub-threshold part of the ToT is neglected.
The second systematic error is that ToT is not linear with deposited charge but of square root behavior.
For an improved measurement, every single pixel would need individual calibration with X-rays over the entire expected range. Then, ToT could be translated on pixel level to a charge which can then be summed up.

(d) The same histogram as in c), but with applied fits. The box-shaped background is fit by a Gaussian function. The signal is fit by a Landau-Gaussian convolution. The combined results suit the relevant range.

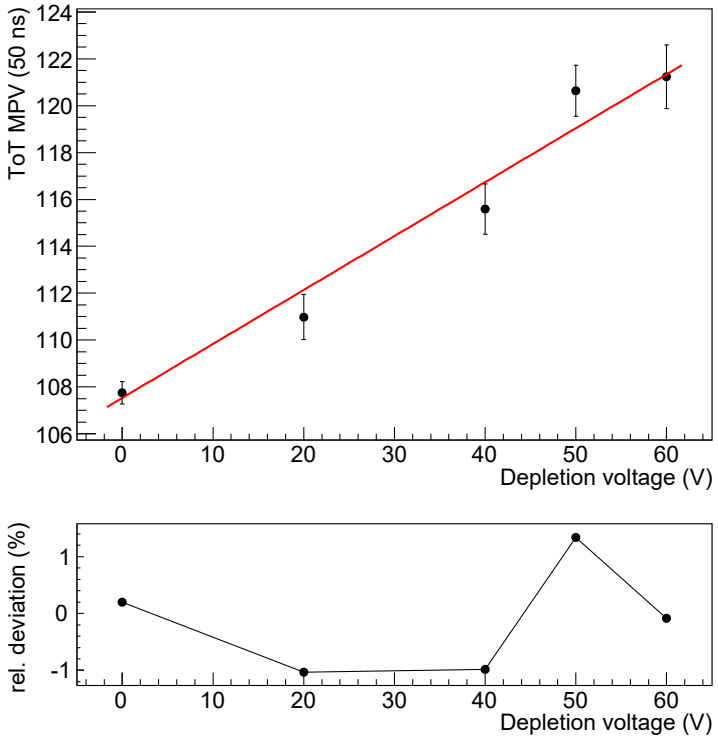


Figure 15.25: Most probable value of cluster-summed ToT as a function of depletion voltage.

From these given alternatives of MPV determination, the summation of ToT of all pixels of a cluster delivers the most sophisticated spectrum. The most probable values of the spectra measured at different depletion voltages is shown in figure 15.25.

The total ToT θ_Σ grows linear with increasing depletion voltage U_{HV}. Apparently, the

discussed non-linearities cancel each other out:

$$\theta_{\text{pixel}} \propto \sqrt{U_{\text{HV}}} \tag{15.7}$$

$$N_{\text{cluster}} \propto \sqrt{U_{\text{HV}}} \tag{15.8}$$

$$\theta_{\sum} \neq \sum_{\text{cluster}} \theta_{\text{pixel}} \tag{15.9}$$

$$\Rightarrow \theta_{\sum} \propto U_{\text{HV}} \tag{15.10}$$

A more precise calculation requires a translation of ToT to generated charge per pixel with consecutive cluster summation. As the pixel gain is not homogeneous, this procedure requires individual calibration of every pixel.

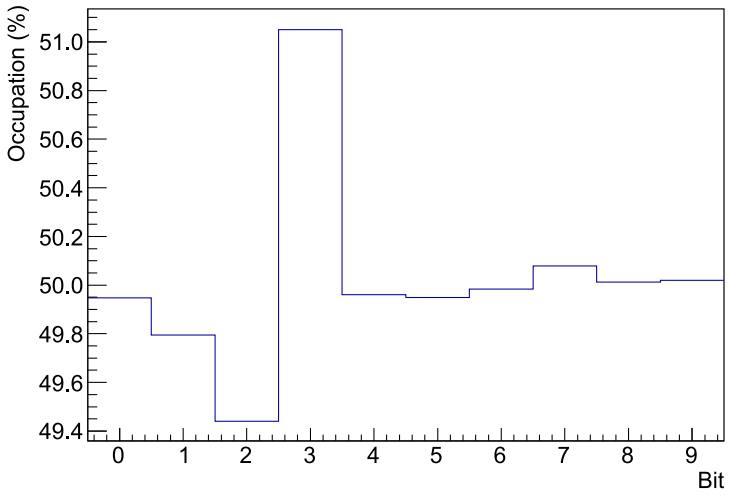


Figure 15.26: Appearance of bits 0 to 9 of the primary time stamp in a source measurement. The occupation of every bit should be 50%. Bit 3 is over- and bit 2 underrepresented.

Some ToT spectra show a distinct ripple. The ToT measurement is based on two counters, primary and secondary time stamp. The ripple appears in spectra, that have been measured on chips running with high speed > 600 MHz, but vanishes again if the time stamp speed is reduced by configuration. The explanation is that the time stamp registers are too slow for the applied clocking speed. Figure 15.26 shows a histogram for the bits of the primary time stamp. As radioactive decay is a randomized process, all bits should be set or cleared with the same probability. However, bit 2 is set less often and bit 3 is set more often than 50% of the time. This bit switching issue is the reason for the ripple in the spectra.

15.4 Tuning

Matrix measurements, such as X-ray illumination studies, have shown the typical non-uniformity due to production variations and design imperfections. This behavior is expected and therefore a tuning circuit has been implemented in MuPix8 and all ATLASpix1 matrices. The compensation mechanism is similar to the one implemented in the H35Demo trimming[2] circuit (see chapter 7.5). The globally set comparator threshold Th_{global} is adjusted locally for each pixel-comparator. The effect of the tune bits has changed with respect to the previously used circuit: All tune bits increase the local threshold Th_{local}. Therefore,

[2]The naming has been changed by the chip designers from *trimming* at the H35Demo to *tuning* on all later sensors discussed here. In order to be in line with the respective publications, this change of term is respected in this thesis, too

equation 7.5 has to be adjusted:

$$Th_{\text{local}}(N_{\text{RAM}}) = Th_{\text{global}} + (s(VPDAC) \cdot N_{\text{RAM}}) \tag{15.11}$$

Where VPDAC is the setting of a current source which defines the scaling of the tuning value N_{RAM}. Each N_{RAM} is three bits wide, thus can take the values 0 to 7.

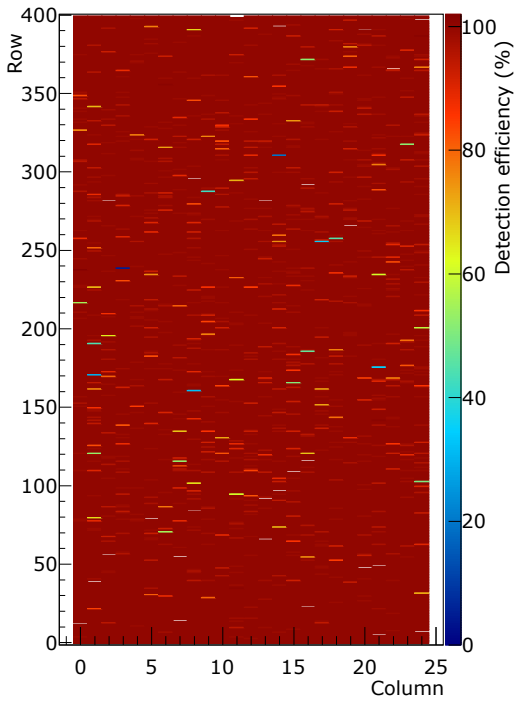
In this section, the tuning procedure and performance is shown on the example of ATLASpix1_Simple. The identical procedure, programs and tools have been used for tuning MuPix8 and ATLASpix1_IsoSimple, a variation of those has been used in the tuning of ATLASpix1_M2.

The effect of tuning is inspected by injecting test signals of constant charge into all pixels while monitoring the detection efficiency. All RAM values are set to 0, the global threshold is set well above noise level of all pixels and the injection voltage is set to a value so most pixels show 100% efficiency.

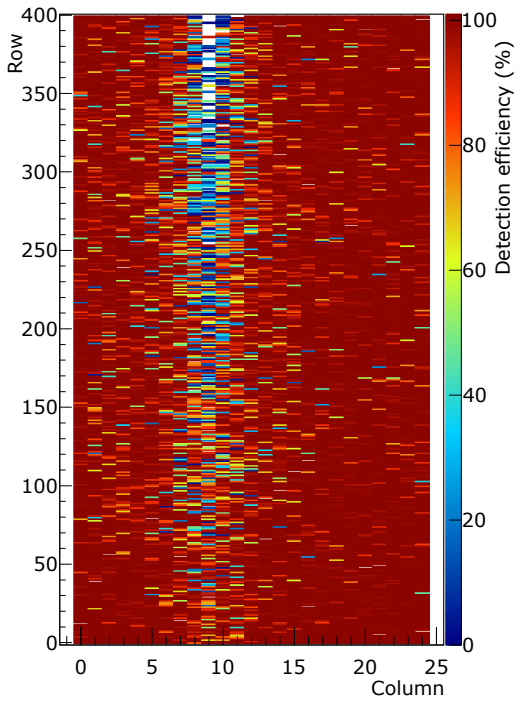
The tuning value, thus the local threshold, is then gradually increased and for each value the measurement is repeated. Once detection threshold is reached, the efficiency drops. The tuning scaling factor has been set to a large value ($VPDAC = 62$) to make sure an effect of N_{RAM} can be observed.

The four plots of figure 15.27 show the detection efficiency of the ATLASpix1_Simple matrix for every pixel with a certain RAM setting. While the efficiency is homogeneously high in the first plot with RAM setting 0 (lowest local threshold), the efficiency drops from the central region towards the left and right edges for RAM values > 0. As this behavior has not yet been observed in the previous measurements with deactivated tuning circuit (RAM = 0), it is concluded that the in-homogeneity is caused by a non-uniform impact of the RAM value on the local threshold. If the pixels gain caused this behavior, it would have been observed in the previous X-ray measurements, too. The best explanation is that *VPDAC* drops from the central region towards the edges, affecting the thresholds of the central columns more than the others.

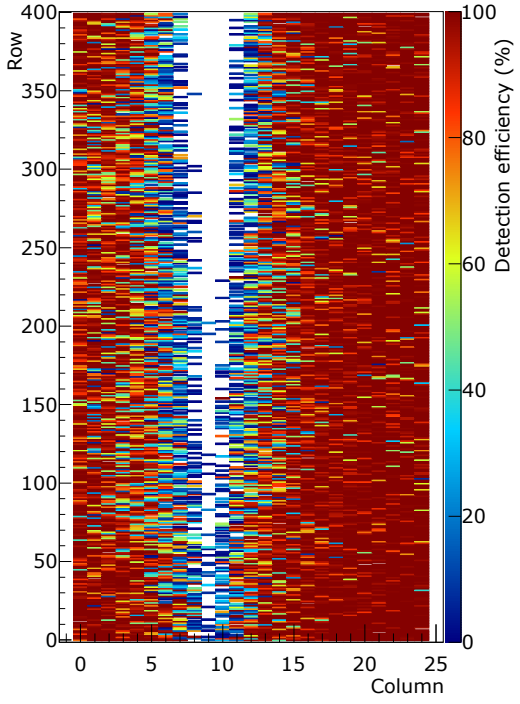
Although tuning is still possible, this issue should be addressed in the design of future chips. For successful tuning, the RAM value has to be set high enough to have a sufficient effect even in columns 0 and 24. Consequently, the step size in the central region is larger than with a uniform scaling of the RAM value.



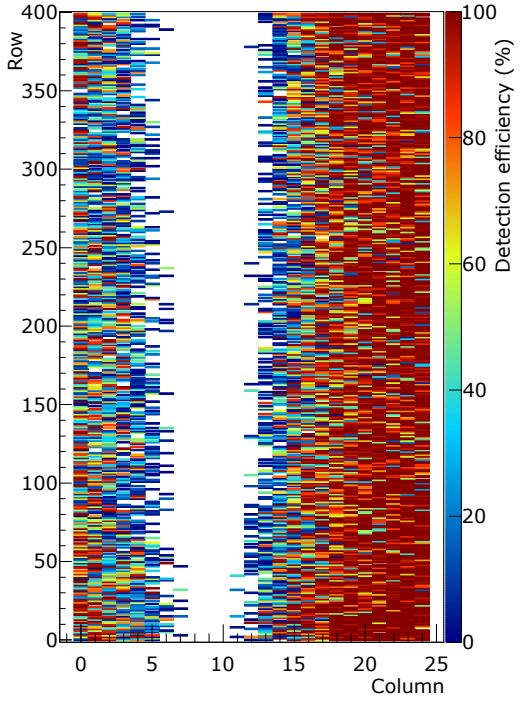
(a) All RAMs set to 0



(b) All RAMs set to 2



(c) All RAMs set to 4



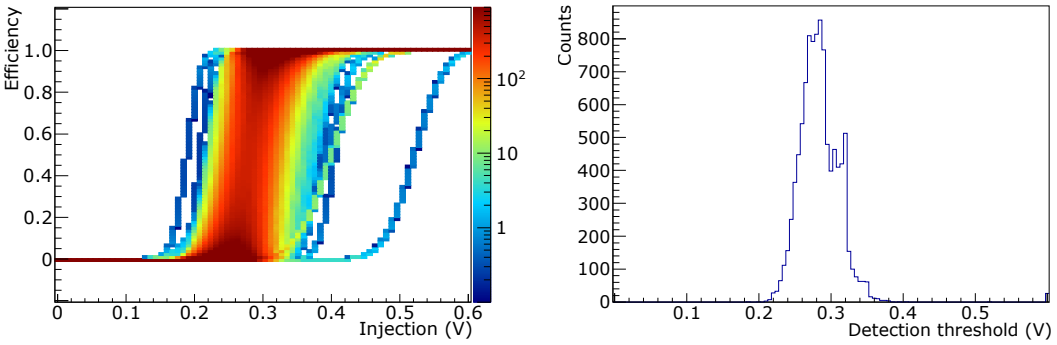
(d) All RAMs set to 7

Figure 15.27: The four plots show the detection efficiency of the same constant signal. All tuning RAMs are set to the same value. A higher value in RAM increases the comparator threshold, thus decreases the detection efficiency. The effect is stronger in the central region than on both sides.

The tuning procedure is based on S-curve measurements. It has been discussed before (see chapter 7.5), therefore only the major steps are listed here:

- S-curves for every pixel and every RAM value are measured.
- From the data set a good value for *VPRAM* and a target detection threshold is deduced.
- For every pixel the S-curve closest to the target threshold is chosen.
- The found RAM values are written back into the pixel RAMs.
- Global threshold is lowered right above noise level.
- A set of S-curves with the found RAM setting is recorded for verification of tuning.

Each pixel RAM (3 bit) has one bit more than the pixel RAMs in H35Demo (2 bit), therefore eight S-curves per pixel need to be measured. A binary search has been considered. However, the temporal benefit is minimal and is likely to introduce problems: Many S-curves can be measured simultaneously, but the detection threshold requires fitting, which increases the dead time between two measurements, which reduces the theoretical benefit of the binary search. Another problem is that in case of a failed fit or noisy data, the decision whether to shift the S-curve left or right for the next iteration might be wrong. Such mistake might be detected, but offline correction is impossible. If another bit should be added to the pixel RAM (then 4), binary search is inevitable to keep the tuning time reasonable.



(a) **(b)**

Figure 15.28: Tuning is based on S-curve measurements. (a) shows the S-curve density plot of all pixels before tuning (S-curve density plot is introduced in chapter 7.5.1). (b) shows the distribution of the detection threshold (inflection point of S-curve). The standard deviation is $\sigma = 25$ mV.

Figure 15.28 shows the S-curve distribution before tuning with all RAM values set to 0 and the extracted detection thresholds as histogram. The same S-curves have been measured for the other seven possible RAM values. The parameters and values of all eight are saved for further analysis.

The optimal target threshold is found by brute force: The data of all S-curves are virtually tuned to all[3] possible target thresholds and the target threshold with the smallest S-curve spread is chosen. The expected spread of tuning S-curves as a function of target threshold is shown in figure 15.29. The best target threshold is found at 0.311 V.

After tuning the measured S-curves moved closer together as shown in figure 15.30. Thus detection threshold distribution is much more narrow ($\sigma = 11$ mV) than before tuning ($\sigma = 25$ mV). The distribution is of near Gaussian shape but entries towards higher and lower thresholds are over represented.

[3]The step size between two investigated target thresholds is smaller than the precision of the DAC defining the global threshold voltage.

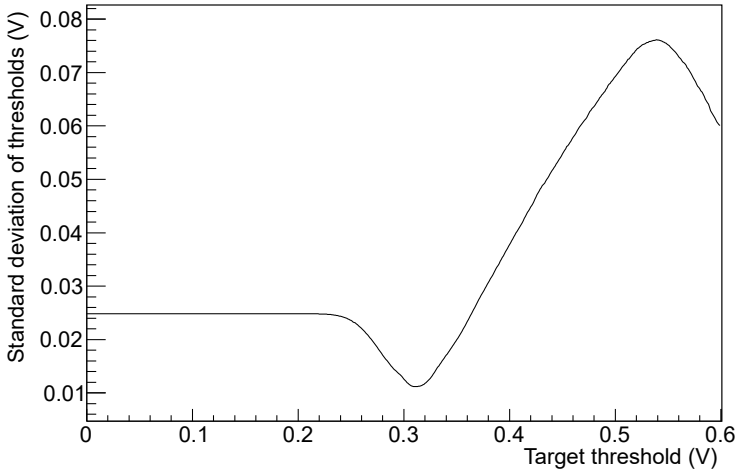
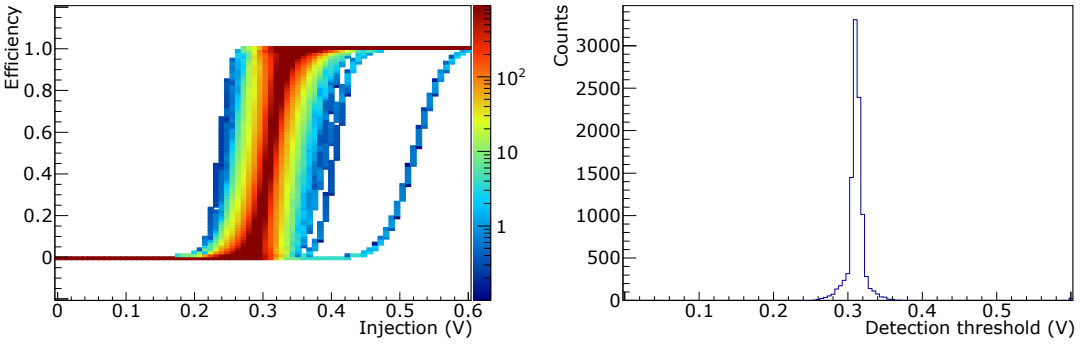


Figure 15.29: The plot shows the standard deviation of a virtual tuning as a function target threshold. S-curves have been measured for all pixels and all possible tuning values. In order to determine the best tuning target threshold, a virtual tuning is performed for every possible target threshold and the respective threshold distribution is calculated. The virtual target threshold with the smallest distribution is picked as target for the real tuning.



(a) **(b)**

Figure 15.30: S-curve density after tuning is shown in (a), the histogram with their detection threshold in b). The reduction of the spread compared to the same plots before tuning (Figure 15.28) is evident. The standard deviation is $\sigma = 11$ mV.

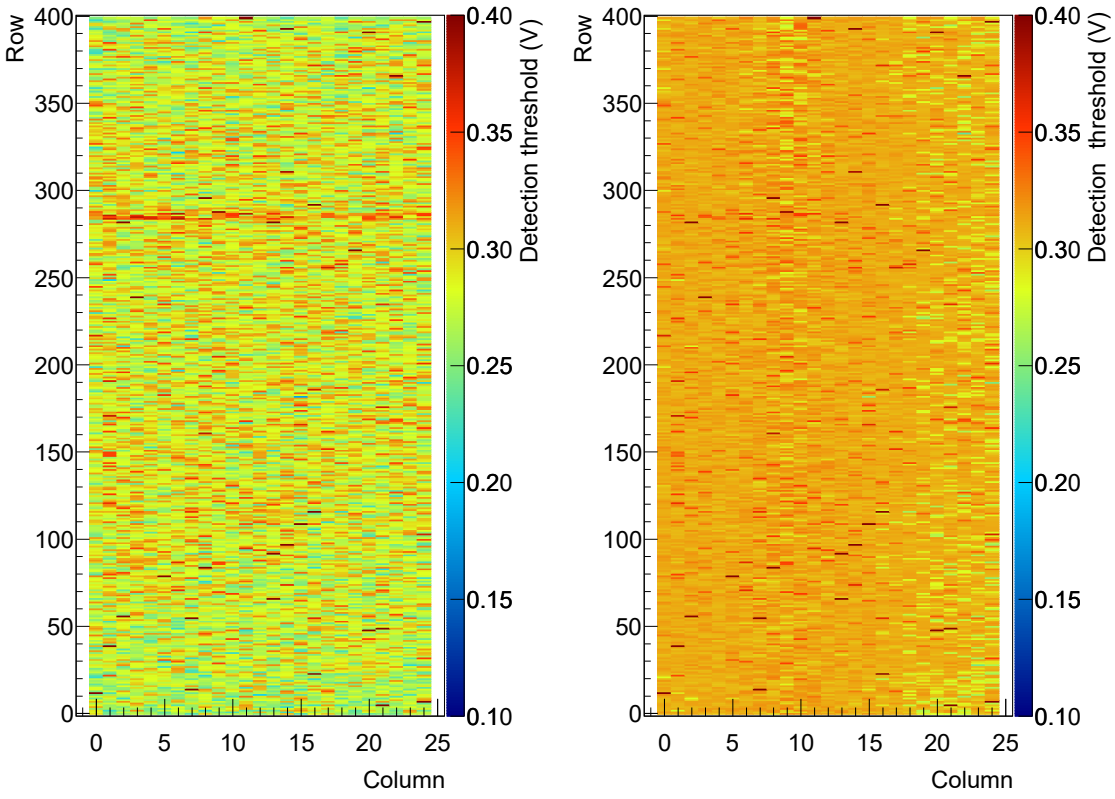


Figure 15.31: The plots show the detection threshold of each pixel before (a) and after tuning (b). In both, the special distribution is quite homogeneous. The average is higher after tuning (same scale).

The spatial distribution of the thresholds is quite homogeneous both before and after tuning (Figure 15.31). It is notable that the average detection threshold is higher after tuning than before. This is reasonable as the tuning mechanism implemented in ATLASpix1 can only increase the comparator threshold.

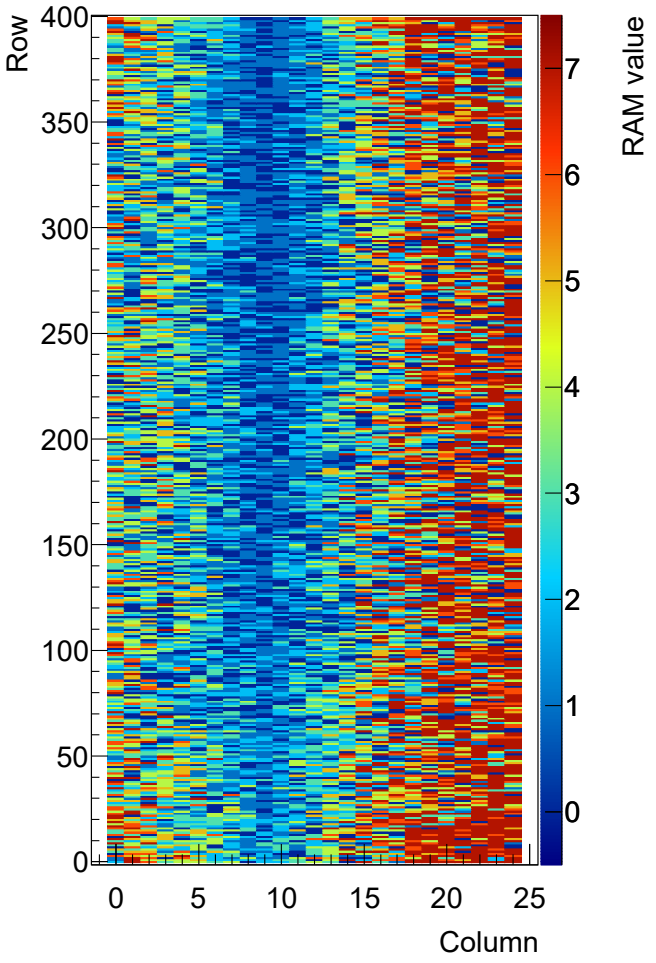
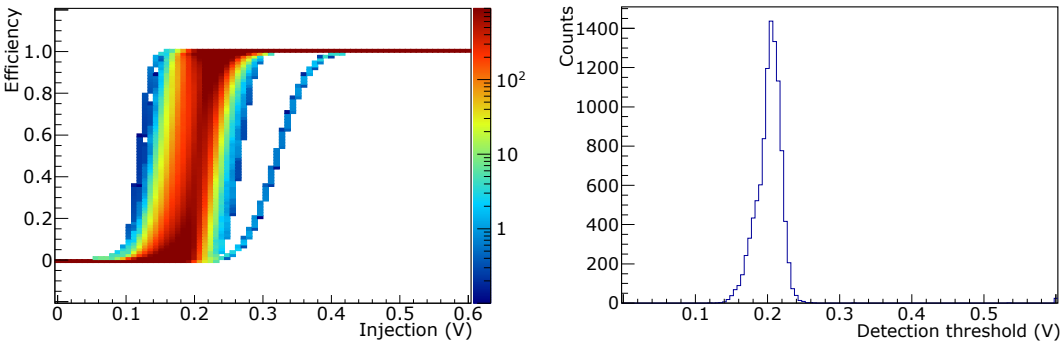


Figure 15.32: During the tuning process, the RAM value for every pixel is calculated to obtain the optimal detection threshold. The plot shows the RAM value for every pixel. Generally speaking, the sides need a higher value than the pixels around column 10.

The spatial distribution of the RAM, on the other hand, is highly in-homogeneous (Figure 15.32). The comparator threshold of pixels around column 10 are more affected by the same RAM value than pixels further to the sides of the matrix. This finding confirms the hypothesis made during testing of the tuning mechanism.



(a) (b)

Figure 15.33: After tuning, the global threshold can be set lower than before tuning. The S-curves are measured again (a). The threshold histogram (b) reveals the effect: The sensor is significantly more sensitive, however the distribution has slightly increased to $\sigma = 17$ mV.

Finally, the global comparator threshold is lowered from 0.820 V to 0.755 V. The comparator's baseline remains at 0.700 V. The effective threshold is only 55 mV, before tuning this threshold would cause heavy noise all over the matrix. After tuning, noise rate is still virtually zero and the S-curves can be measured again. They are shown in figure 15.33 together with a histogram of their detection thresholds.

The spatial distribution of the detection thresholds (Figure 15.34) shows a slight drop towards the right side. The mean detection threshold is now significantly below the value before tuning, but the spread has increased with respect to the situation after lowering the threshold. It can be deduced that the distribution of the global threshold voltage is also not perfect across the matrix. Consequently equation 15.4 has to be adapted:

$$Th_{\text{local}}(x, y, N_{\text{RAM}}) = Th_{\text{global}}(x, y) + (s(VPDAC,\text{x,y}) \cdot N_{\text{RAM}}) \tag{15.12}$$

The tuning performance can be summed up by the statistics of the detection threshold histograms as follows:

detection threshold	μ (mV)	μ (e)	δμ (%)	σ (mV)	σ (e)	δσ (%)
before tuning	284	1338	±0	25	117	±0
best RAM values	311	1466	+9.5	11	51	−56
best RAM and lowered threshold	202	952	−28.8	17	80	−31

After tuning the sensor has a nearly 10% increased mean threshold, but the spread of S-curves is reduced by 56%. Lowering the global comparator threshold after tuning lowers the mean detection threshold compared to the untuned sensor by nearly 30%. However, as the global threshold is not evenly distributed, the spread of S-curves worsens, but is still 31% better than before tuning.

Depending on the application, a user might choose a homogeneous matrix over a minimized detection threshold, or decides for maximized efficiency on the cost of a reduced homogeneity. One way or the other, tuning greatly improves the performance of the sensor.

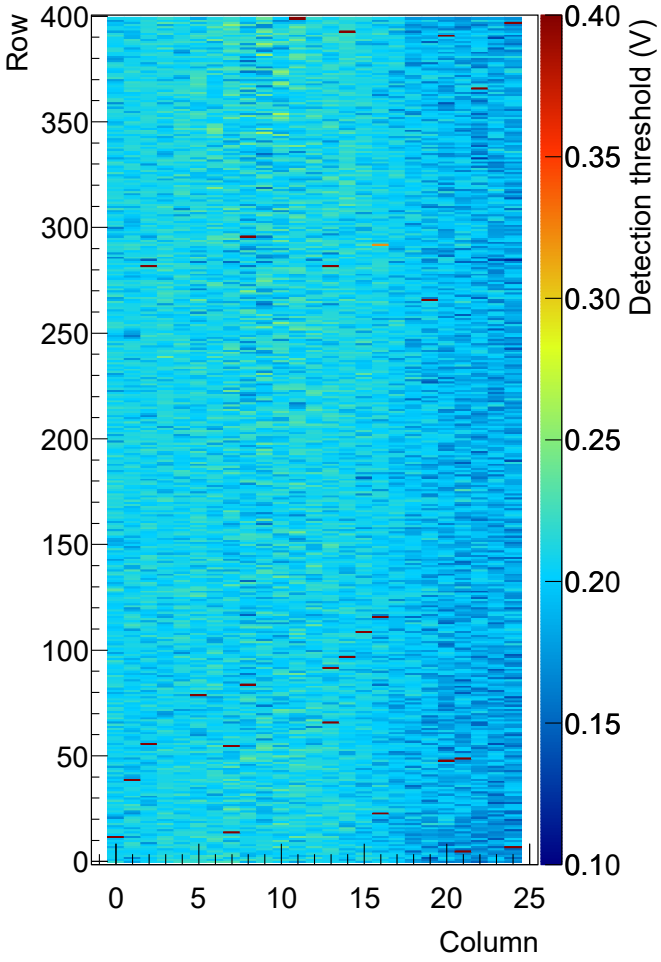


Figure 15.34: After tuning, the average detection threshold can be significantly lowered. For comparability, the scale is identical with the scale in figure 15.31.

15.5 Time resolution

The time resolution of a single pixel depends on the charge collection time. In case of HV-CMOS sensors it is neglectable compared to the time-walk effect. For the time resolution of an entire detector, the differences between pixels, such as transistor mismatch effects and delays due to different signal line lengths contribute, too.

Both time-walk and inter-pixel delay differences can be measured by S-curves. The previously used efficiency S-curve measurement function was modified in two ways:

- The test signal injections are now synchronous with the *SyncRes* signal. *SyncRes* is generated by the FPGA and resets the time stamp counters of the sensor. The injection is triggered a fixed period of time after it, thus a perfect detector would report back always the same time stamp, regardless the signal height or pixel location.
- The data from every detected hit, primary and secondary time stamp, are recorded individually.

The S-curve measurement is conducted on the whole ATLASpix1_Simple matrix, starting at 0.8 V or 3800 e^-. From the collected data, plots as shown in figure 15.35 are formed. Besides the efficiency S-curve (black), the averaged measured delay between *SyncRes* and detection time stamp is plotted in blue. In this analysis, both time-walk and inter-pixel

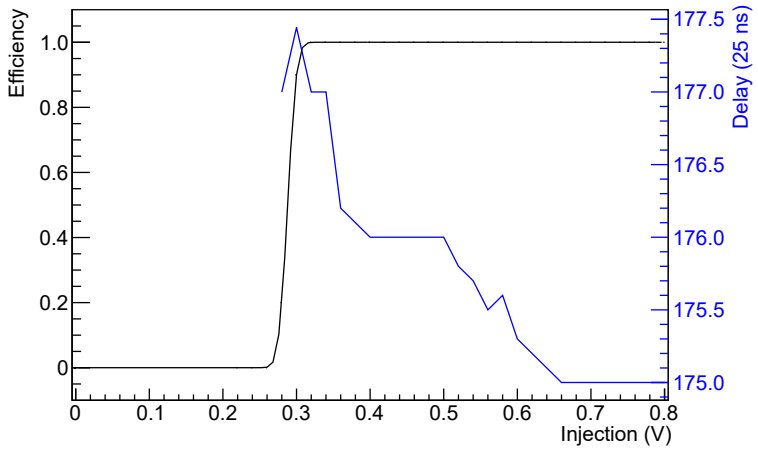


Figure 15.35: The graph shows an efficiency S-curve (black) and the time of signal detection (blue). The injections were triggered with fixed delay after time stamp 0. Due to the time-walk effect, higher signals are detected faster than lower signals.

delay differences are to be measured. The time-walk is calculated as the difference between largest and smallest averaged delay. The time stamp speed is 25 ns. In the given example the time-walk is 177.5 – 175 = 2.5 = 62.5 ns.

The performance of different pixels is determined by the averaged delay measured at 0.78 V injection voltage. A high value is chosen, as the time-walk effect is minimal for large signals, so the inter pixel effect is dominant.

15.5.1 Matrix time resolution

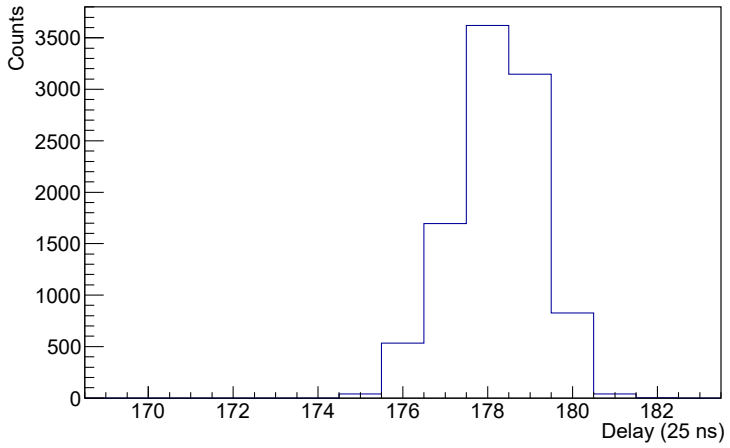


Figure 15.36: The relative delay between constant injection (0.78 V) and time stamp generation of every pixel has been measured and is displayed as histogram.

The matrix time resolution is the RMS of the delays measured on all pixels. The delay histogram in figure 15.36 has a mean delay of 178.2 and an RMS of 1.0. This means an uncertainty in timing of ±25 ns.

Signals are digitized in-pixel by a comparator. However, the transmission lines from pixel to periphery are not driven with CMOS signals[4], but with a elevated logical 0. The voltage

[4]Logical 0 is 0 V, logical 1 is 1.8 V

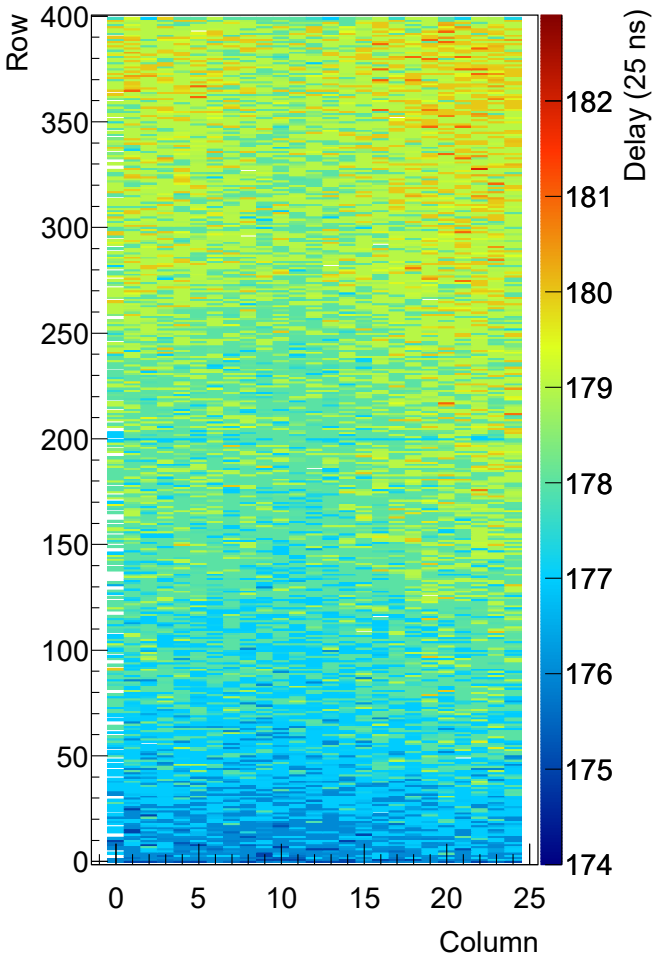
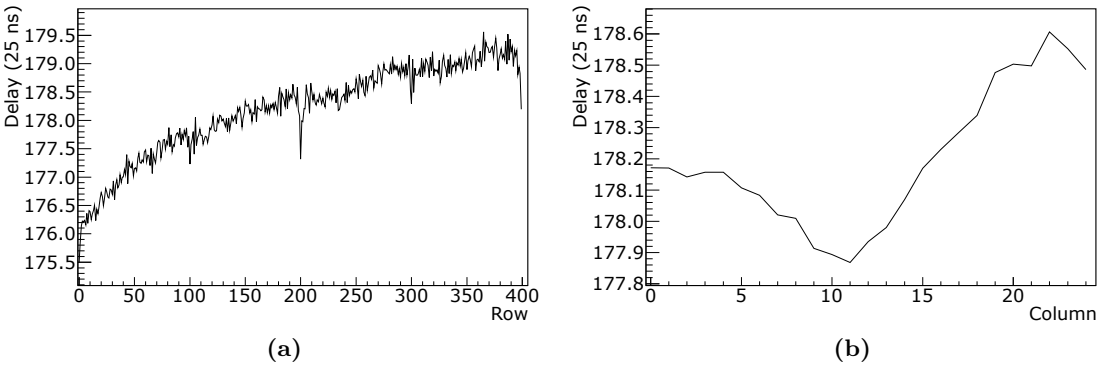


Figure 15.37: The delay of every pixel in the matrix. Pixels close to the periphery respond faster then pixels further away. Pixels around column 10 are faster than pixels on both sides.

level of the logical 0 can be defined externally and is by default 1.8 V - 0.7 V = 1.1 V. The logical 1 remains 1.8 V. These lower digital steps are chosen for the same reason for which the time stamp is generated in the periphery: noise reduction. Fast switching signals with high voltage differences cause noise which might be caught up by the input of other pixels, resulting in fake signals.

The downside of time stamp generation in periphery is that signals pick up additional delay on their way from pixel to periphery. This behavior is investigated in figure 15.37. The delay between test signal injection (plus offset) and time stamp generation of every pixel is plotted. The color code reveals two features: There is a distinct drop of delay from top to bottom. This is expectable, as the lines from the top matrix are longer than those closer to the periphery.

The second feature is a valley in the middle of the matrix. It has the same shape that has been discovered before in the tuning process. This measurement has been conducted on a tuned sensor, therefore it is reasonable to see this behavior as the shifting of the comparator threshold has a direct impact on the time of detection: The rise time of the analog signal is finite, therefore a higher threshold makes the comparator react later to a signal. The tuning effect is much smaller than the effect by line length.



(a) (b)

Figure 15.38: Graph (a) shows the average signal delay as a function of row. Graph (b) shows average delay as a function of column.

Figure 15.38 shows the average delay as a function of row, or column respectively. In column direction the maximum spread is 17 ns. The contribution to the time uncertainty is therefore 4.9 ns.

The contribution in row direction is with 87.5 ns spread or 25.3 ns uncertainty larger. The HL-LHC requires a time resolution of better than 25 ns, therefore a single source of uncertainty can not use up this entire budget, especially as time-walk and other sources contribute significantly as well.

The delay is caused by the line impedance, but the capacitance is already minimized for the used technology, further reduction is not possible. One solution is to hold a lookup table (LUT) in the FPGA to correct for this deterministic effect. Anyway a reduction on-chip is beneficial. Beside the line capacitance, the loading voltage determines time constant. The discrimination of the transmission line is dynamically set between high and low level. As mentioned before the low level can be set externally. Shifting the lower level closer to the higher level is expected to improve the absolute line delay as well as its uncertainty. The lower level is now set to 1.8 V - 0.5 V = 1.3 V and the measurement is repeated. As long as the levels can be clearly discriminated, even under the eventual influence of external or internal noise, lowering the difference between high and low level is not expected to have a negative effect.

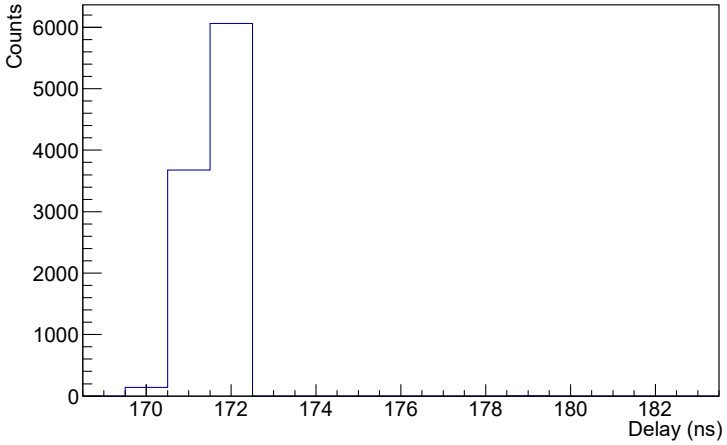


Figure 15.39: Both absolute delay and more importantly spread of delays is significantly reduced by a reduced voltage distance between logic levels. The default value is 0.7 V, for this measurement it was reduced to 0.5 V.

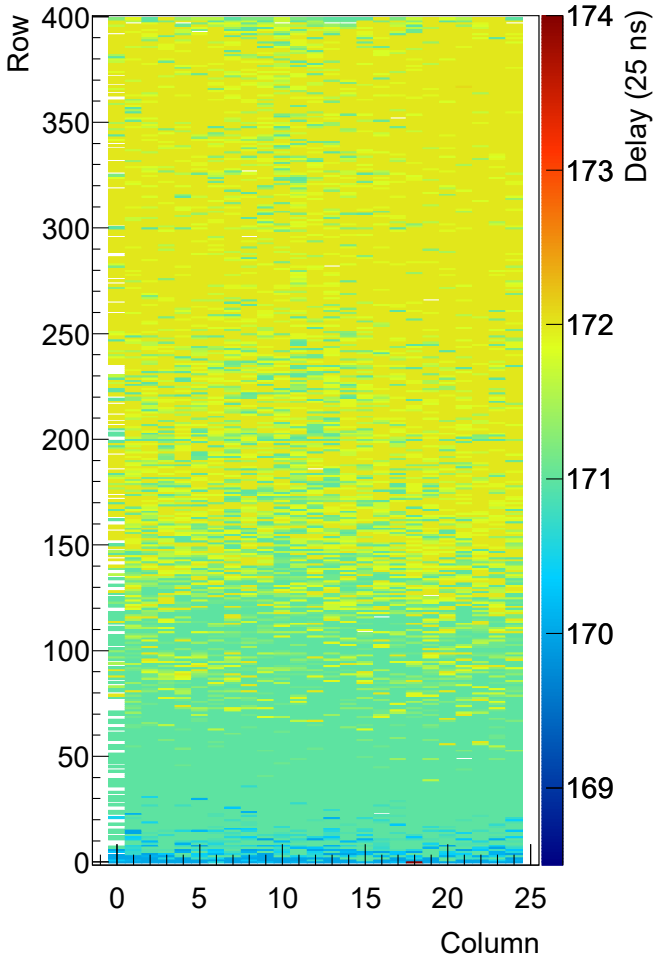
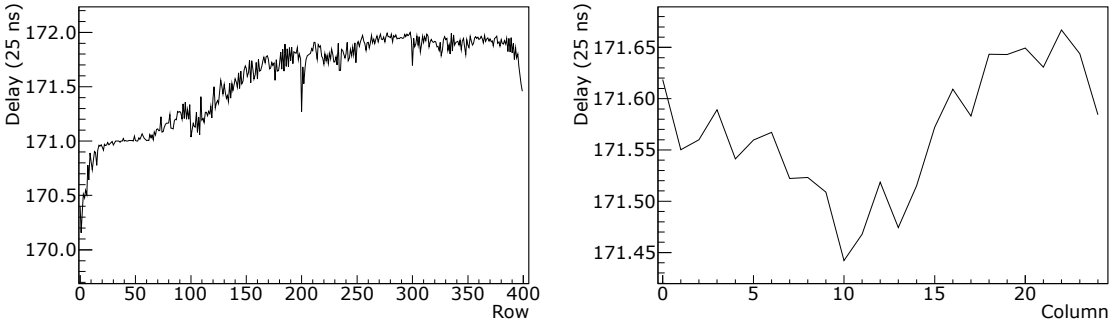


Figure 15.40: Delay of every pixel in the matrix after reduction of the voltage distance between logic levels. The drop along columns remains.



(a) **(b)**

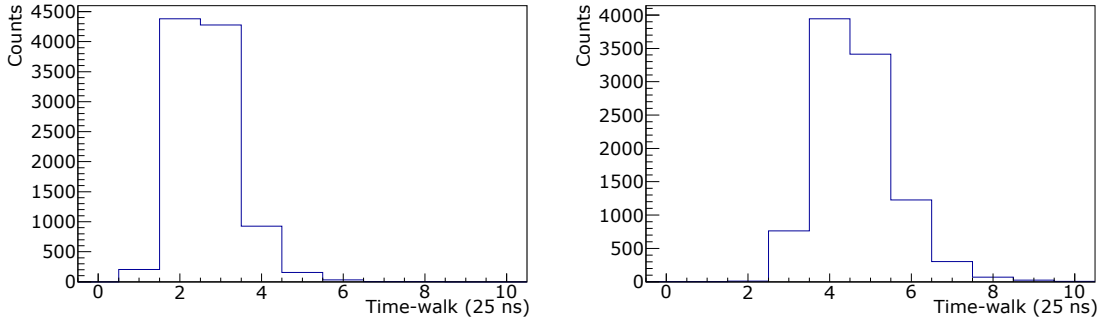
Figure 15.41: Mean delay as a function of row (a) and as a function of column (b). The general behavior remains, however the differences are significantly reduced. In row-direction (b), it is even close to significance level.

Figure 15.39 shows the histogram of all delays with increased lower logical level. The mean has been reduced by 6.6 (165 ns) from 178.2 to 171.6. Also the spread is reduced by 13.3 ns or 53% to only 12 ns.

Looking at the matrix delay plot, the finding is confirmed (Figure 15.40), but a drop from top side to periphery remains. Averaging the measured values of delay along column or row direction shows that both spreads have profited from the smaller voltage difference between logical high and low. The remaining spread by line length difference (Figure 15.41a) is ≈ 50 ns, meaning an uncertainty of 14.4 ns. The delay effect by tuning (Figure 15.41b) has been reduced to a spread of less than 6 ns or an uncertainty of 1.7 ns.

15.5.2 Time-walk

The other dominant contribution to the total time uncertainty originates from time-walk. The rise time of the amplifier is in the order of 100 ns, therefore the time-walk is of the same order, however the comparator has a finite response time, too.



(a) (b)

Figure 15.42: The histograms show the time-walk of all pixels for a logic level difference of 0.7 V (a) and 0.5 V (b). Apparently, the time-walk suffers from the reduced logic level difference.

The measured time-walk of every pixel was filled into the histogram in figure 15.42. The left histogram was created from data measured with 0.7 V between logic levels, the right one with 0.5 V. The mean time-walk with default logic levels is 2.6 time stamps or 65 ns. With reduced logic level difference, the uncertainty by time-walk is even more increased to 115 ns.

The time-walk is stronger in pixels closer to the periphery (Figure 15.43). A column dependency does not exist.

The time-walk increases with decreasing row, independent of the exact logical voltage difference. Figure 15.44 shows the row dependency of time-walk.

Time-walk appears while digitizing the analog signal of the amplifier by the comparator. Therefore, it is consequential that it is affected by changes to their supply voltages. The observed behavior is explained by the switching speed of the line driver adjacent to the comparator's output. Changing the lower potential changes the Gate-Source voltage of the drive transistor and thereby the switching speed is reduced.

The increasing time-walk from top edge to lower edge is most likely a voltage distribution problem. *VMinus*, the lower logical level is fed into the sensor on the top edge. It is connected to *VDD*, the high potential, via the line drivers. The mere number of line drivers lets both voltages drift towards each other. As *VDD*, the main supply voltage, is much stronger, better distributed and fed by more connections from off the chip, *VMinus* drifts up, with increasing distance to its bond pad on the top edge.

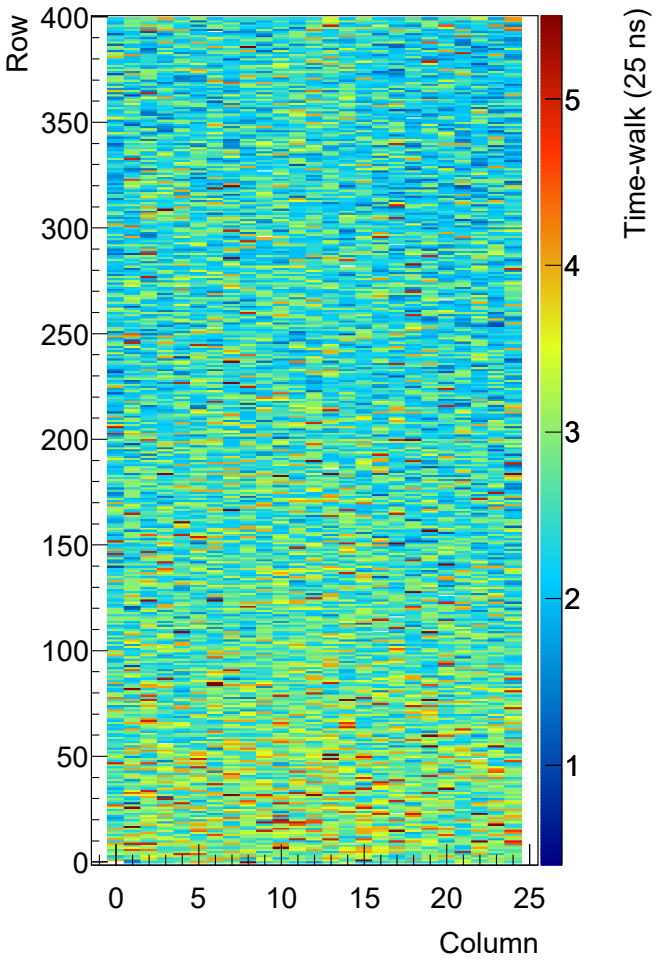
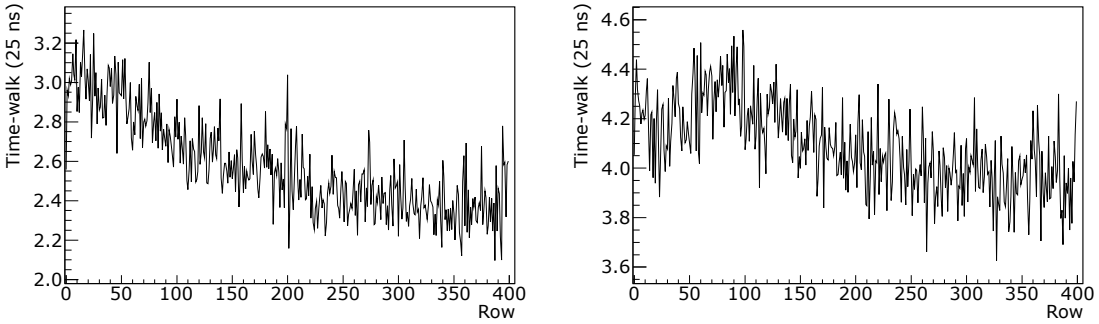


Figure 15.43: Time-walk of every pixel. The time-walk is larger in pixels close to the periphery on the lower edge.



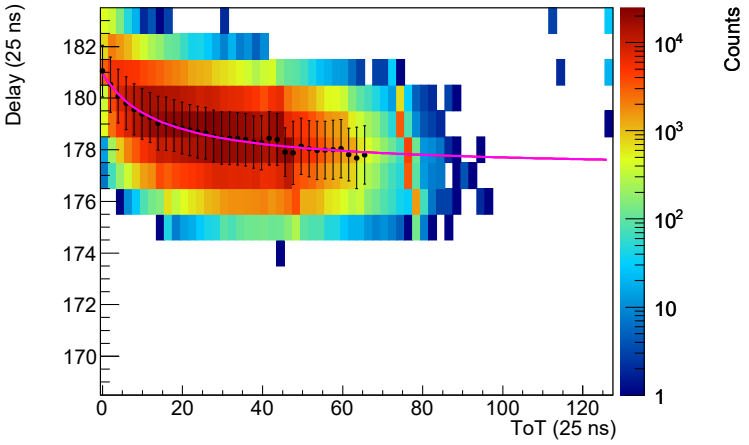
(a) **(b)**

Figure 15.44: The row dependency of time-walk at 0.7 V (a) and 0.5 V (b) logic level difference. The row dependency is stronger for the default value. Overall time-walk is stronger with reduced logic level difference.

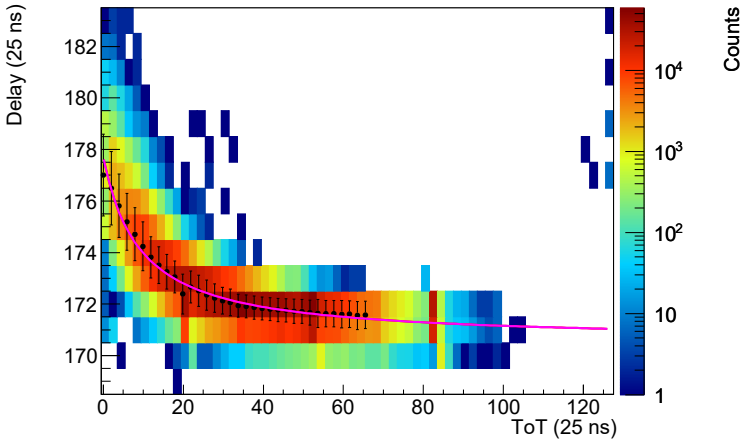
15.5.3 Timing corrections

Time resolution and the influence of sensor settings have been discussed in the previous sections. Knowing the reason for timing uncertainties and their characterization has the potential to improve the overall time resolution.

Two major sources for time uncertainty have been identified: Time-walk effect and transmission time difference introduced by line length differences. Both have been measured and modeled in order to correct for the respective effect. This procedure is being presented in this section on two data sets. Half of each data set was used to determine the correction parameters, which were then applied to correct the other half.



(a)



(b)

Figure 15.45: The plots show relative delay as a function of ToT. The time-walk effect delays hit detection of small signals (low ToT). The time-walk is stronger with lower logic level difference in (b), than with default levels in (a). A fit is applied to each. The plots are made from a subset of the collected data.

Figure 15.45 shows the measured delay over ToT, one plot per setting. The color map shows the number of events with the respective combination of delay and ToT. For every value of ToT (plot: column) with sufficient entries, the mean and standard deviation is calculated (black dot and errorbars). Then a fit is applied to those points, drawn in magenta. The fit function is of the form:

$$t(\theta) = t_{0\,\mathrm{TW}} + \frac{a}{\theta - \theta_0} \tag{15.13}$$

Where t is the relative delay and θ is ToT. The fit parameters are $t_{0\ \mathrm{TW}}$, θ_0 and a. Without time-walk this simplifies to

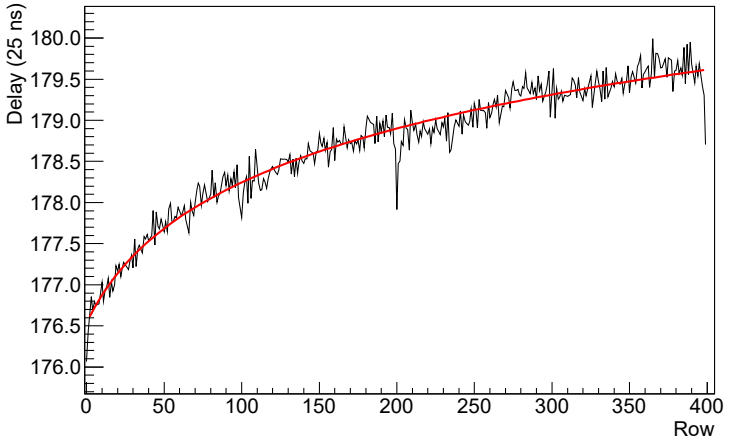
$$t_{\mathrm{no\ TW}} = t\left(\theta \to \infty\right) = t_{0\ \mathrm{TW}} \tag{15.14}$$

and the correction term is

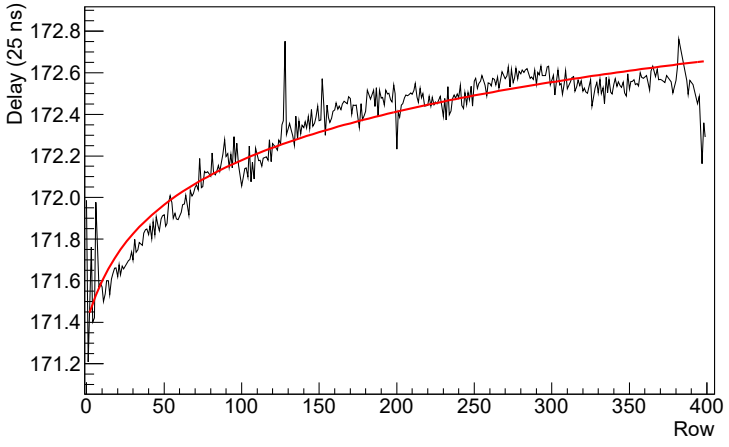
$$t_{\mathrm{corr\ TW}}\left(\theta\right) = t_{\mathrm{no\ TW}} - t\left(\theta\right) \tag{15.15}$$

$$t_{\mathrm{corr\ TW}}\left(\theta\right) = -\frac{a}{\theta - \theta_0}. \tag{15.16}$$

This term can be used to correct the event time stamp for the time-walk effect by incorporation of ToT. The free parameters are determined by fit to a data subset.



(a)



(b)

Figure 15.46: Relative delay as function of row number for default (a) and reduced logic level difference (b). The plots are made from the same subset of data as figures 15.45. A logarithmic fit is applied.

The row dependency is of logarithmic behavior. Therefore, a function of the form

$$t\left(k\right) = t_{0\ \mathrm{row}} + b \cdot \log\left(k - k_0\right) \tag{15.17}$$

is fit to the plots in figure 15.46. They show the mean delay as a function of row. Again, t is the relative delay and k is the row number. The fit parameters are $t_{0\ \mathrm{row}}$, k_0 and b. The

correction term is calculated relative to the delay of the first row:

$$t\,(0) = t_{0\ \mathrm{row}} + b \cdot \log\,(-k_0). \tag{15.18}$$

to

$$t_{\mathrm{corr\ row}} = -b \cdot \log\,(k - k_0). \tag{15.19}$$

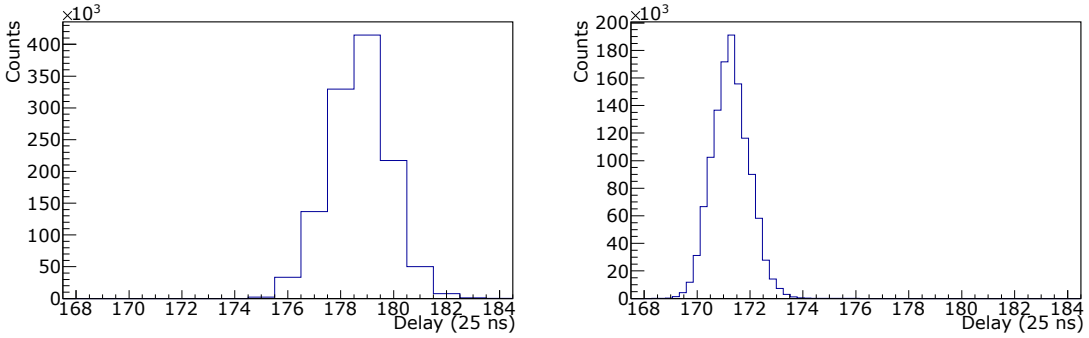
Constant contributions are neglected, as the time uncertainty does not depend on the absolute delay. The free parameters are obtained by fit.

Together the timing correction is

$$t_{\mathrm{corr}} = t_{\mathrm{corr\ TW}} + t_{\mathrm{corr\ row}} \tag{15.20}$$

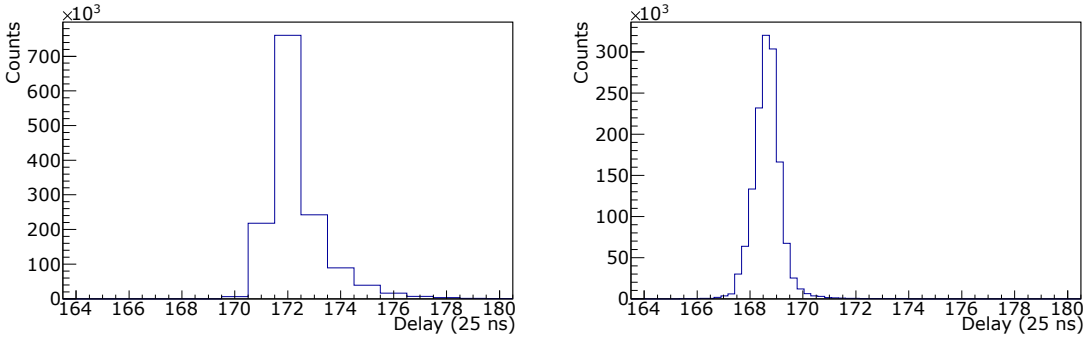
$$t_{\mathrm{corr}} = -\frac{a}{\theta - \theta_0} - b \cdot \log\,(k - k_0). \tag{15.21}$$

These corrections are now applied to data that have been collected with the same settings as the training data.



(a) no correction **(b)** corrected

Figure 15.47: Signal delay distribution before correction (a) and after correction (b). The measurement was conducted at default logic levels. The corrections were calculated from one data set and applied to another. Before corrections the RMS is 29.03 ns, after both corrections applied, it is 17.45 ns.



(a) no correction **(b)** corrected

Figure 15.48: Signal delay distribution before correction (a) and after correction (b). The measurement was conducted at reduced logic levels. The corrections were calculated from one data set and applied to another. Both corrections reduce the RMS from 27.58 ns to 13.08 ns. The procedure is the same as in figure 15.47.

The time resolution is measured with and without corrections by filling time stamps into a histogram. These histograms are shown in figures 15.47 and 15.48. Before correction, the

binning is limited to 1 as raw data were used. After correction the time stamp is a float value, that allows smaller bins.

Beside the shown histograms, the effect of only one correction term has been investigated. For comparability the measured uncertainties are compiled in a table:

correction	ΔU_{logic} = 0.7 V		ΔU_{logic} = 0.5 V	
	RMS (TS)	RMS (ns)	RMS (TS)	RMS (ns)
none	1.161	29.03	1.103	27.58
time-walk	1.027	25.63	0.623	15.58
line length	0.879	21.98	1.063	26.58
both	0.698	17.45	0.523	13.08

In every constellation each correction term has a positive influence on time resolution. In case of reduced logic level difference, the effect of the time-walk term is stronger, as the uncorrected time-walk is larger. In case of default logic levels, the line length term has a larger effect, as the time-walk is small. The best time resolution is obtained for reduced logic levels and both corrections. The remaining RMS is 13.08 ns.

15.6 Irradiated ATLASpix1 and MuPix8 sensors

During their possible service in ATLAS ITk outermost layer after high luminosity upgrade, a detector is expected to be irradiated with $10^{15}\,n_{eq}/cm^2$. The general radiation hardness of HV-CMOS sensors has already been demonstrated [19, 100, 21, 101, 102, 103, 104]. Nevertheless, the radiation hardness has to be confirmed for every new sensor and technology. Previous radiation hardness studies were not conducted on sensors produced in *ams* aH18, but for example on sensors in *ams* H35 or *ams* H18.

Two sample sets of ATLASpix1 have been irradiated with protons in a cyclotron (see Appendix B.1) to the total fluence of $6 \cdot 10^{14}\,n_{eq}/cm^2$ and $10^{15}\,n_{eq}/cm^2$. However, characterization of irradiated samples turned out to be challenging for two reasons: The sensors have been produced in three batches. A pre-production batch and two main-production batches. The samples of all were distributed within the collaboration, so that from the first batch not enough sensors were available for irradiation. The main-production had a limited yield of about 30-50%. The common point of failure was the broken shift register. Sensors for irradiation can not be electrically tested before irradiation, as no probe card is available, thus up to 2/3 of them will not work, due to production problems, not due to irradiation.

The sensors of the main-production have been thinned down to 70, 100 or 200 µm, rendering handling complicated. For irradiation, the samples are glued to Kapton-foil. But for characterization the foil needs to be removed or cut out in order to glue them onto carrier PCB. From thicker chips it can be striped easily, thinned chips would break, they have to be cut free right at the edge of the chips. Additionally, 70 µm chips have a moderate probability of breaking during wire-bonding.

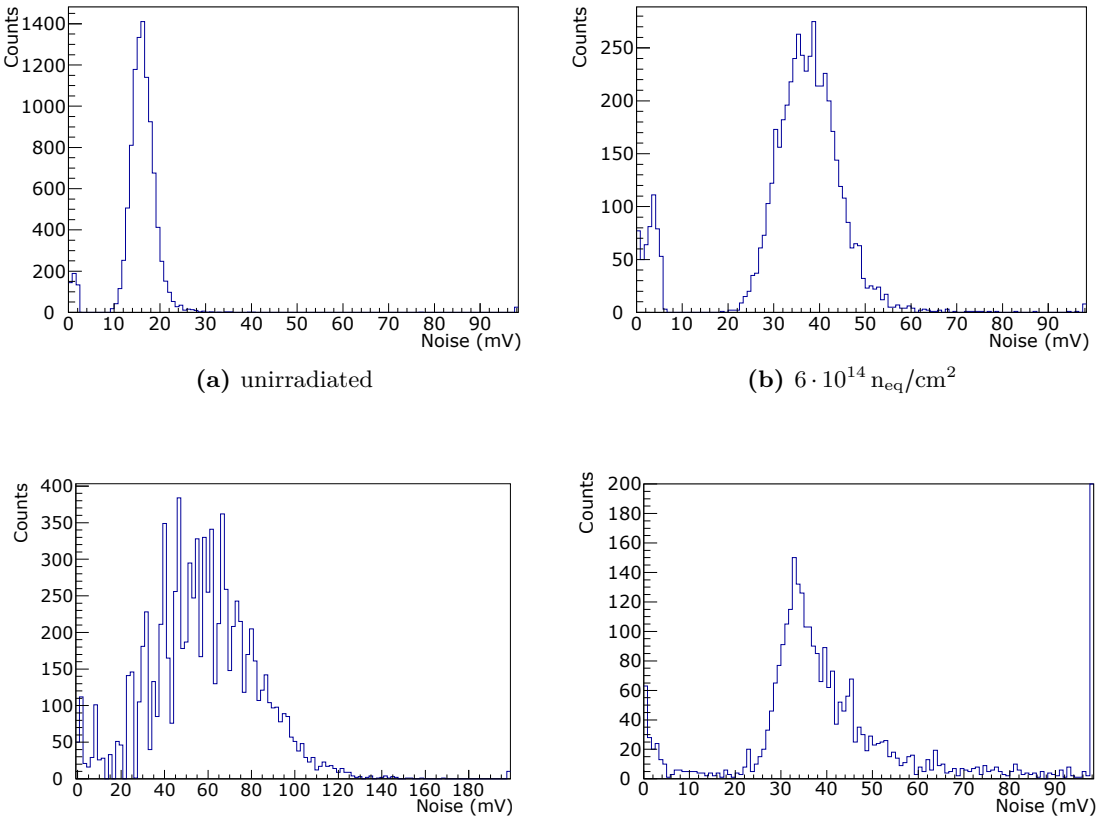
Of eight irradiated chips, six made it intact to a bonded sample. Each ATLASpix1 sample has three independent matrices, so 18 matrices were bonded without mechanical defects and 16 were tested. Only six matrices have the configuration shift register intact.

15.7 S-curve noise of irradiated sensors

Samples that have been produced correctly, survived handling during irradiation and assembly, are put to operation. Due to radiation damage the default configuration had

to be adjusted individually for every matrix. The leakage current of the sensor diode's depletion voltage suffered from NIEL damage and has significantly increased in all sensors. The current drawn by the amplifier and main supply is larger than in unirradiated samples, but the change is not homogeneous. Some samples draw more than others at same fluence and some parts of the same chip show different changes. However, these variations might be caused by production issues as well.

A good way to evaluate the impact of irradiation (after functionality has been verified), is to measure efficiency S-curves and compare them to those of other samples. Fitting error-functions to the measured data points delivers a value for noise, which sums up imperfections and disturbances picked up on the way from signal generation in the sensor diode to readout in the digital periphery.



(a) unirradiated

(b) $6 \cdot 10^{14}\,\mathrm{n_{eq}/cm^2}$

(c) $10^{15}\,\mathrm{n_{eq}/cm^2}$, 9 days annealing

(d) $10^{15}\,\mathrm{n_{eq}/cm^2}$, 56 days annealing

Figure 15.49: Noise histograms of sensors before irradiation (a) and after a fluence of $6 \cdot 10^{14}\,\mathrm{n_{eq}/cm^2}$ (b). Histograms (c) and (d) have been measured on a sensor irradiated with $10^{15}\,\mathrm{n_{eq}/cm^2}$ after short annealing (c) and long annealing (d).

The histograms in figure 15.49 show the noise, measured on individual pixels as transition width of efficiency S-curves. The mean and standard deviation of each histogram have been calculated:

histogram	fluence ($\mathrm{n_{eq}/cm^2}$)	annealing time (days)	noise (mV)		noise ($\mathrm{e^-}$)	
			μ	σ	μ	σ
(a)	0	-	15.8	4.1	74	19
(b)	$6 \cdot 10^{14}$	27	34.6	12.2	163	58
(c)	$1 \cdot 10^{15}$	9	58.7	23.1	277	109
(d)	$1 \cdot 10^{15}$	56	36.9	15.7	174	74

Irradiation to $6 \cdot 10^{14}\,\mathrm{n_{eq}/cm^2}$ doubles the noise compared to an unirradiated sensor. Further irradiation to the expected life-time dose of $10^{15}\,\mathrm{n_{eq}/cm^2}$ increases the noise to approximately 3.7 times the original noise. Annealing over several weeks however, reduces the noise again to a value around 200% the original noise floor.

The measured samples have been irradiated in a very short time (minutes) compared to their life time in particle physics experiments (years). Therefore, it is reasonable to anneal the samples for some time. In experiments, the annealing process can happens during the slow irradiation, if the operational temperature is not too low, or in maintenance cycles. The measured noise floor after irradiation is strong, but manageable. The expected signals are much higher ($\gg 2000\ \mathrm{e^-}$) than the noise ($< 300\ \mathrm{e^-}$) measured right after a life-time dose of $10^{15}\,\mathrm{n_{eq}/cm^2}$ has been applied. Even the noisiest pixel that has been observed ($\approx 700\ \mathrm{e^-}$) can discriminate signal from background.

15.8 Sub-pixel efficiency measurement by fine focus laser scan

The charge collection depends on the depletion depth (total charge) and the electrical field (speed and efficiency). Both are affected by radiation damage and substrate resistivity. Furthermore, they are not homogeneous. A pixel can have areas of high and low efficiency at the same settings. The differences in sensitivity of a single pixel is investigated in this section.

The following measurements have been conducted on ATLASpix1_M2 matrices using the pulsed laser setup developed at KIT-ADL. The laser setup is described in the Appendix B.3.

15.8.1 Measurement principle

The output of the amplifier of one pixel is connected to an oscilloscope, set to automatic triggering. If a signal crosses the threshold, which is set right above baseline, the event is recorded. In case of no threshold crossing a random trigger is prompted and the baseline is recorded. The recorded data are transmitted to the computer where the signal height is determined as difference between the averaged baseline and the maximum point. The laser is moved parallel to the sensor plane and after each step 20 waveforms are read out from the oscilloscope.

Figure 15.50 shows the averaged signal for every measurement point in the *xy*-plane. The step size is 1 µm. Measuring the 37 000 points took about 36 h. Therefore, this resolution is not feasible for a measurement series. The upcoming analyses were carried out on scans with 5 µm or 10 µm resolution.

The initial expectation of the resulting picture was a smeared out square in the size close to the pixel size of $60 \times 50\ \mu\mathrm{m}^2$, a convolution of the sensitive volume with the double-cone-shaped laser beam. By design, the axis of chip and stages are quite well aligned. However, the measured picture is different: The assumed outline of the measured pixel is indicated by the magenta line. The lower two third (1) of the pixel show the expected shape. The upper third (2) shows insensitivity on the right side. All measured samples and pixels show this insensitive corner. The best explanation found is the used 904 nm laser. Its penetration depth in silicon is 15 µm [105]. However, the surface of the sensor is covered with several metal layers, which damp the laser beam. Illumination from the backside of the sensor would solve this problem, but this is not an option here: The purpose of this measurement is to compare the charge collection of different depletion voltages, substrate resistivities and fluences. The used sensors however, are of different thickness (70 - 700 nm) and the irradiated sensors have leftovers of glue and Kapton foil on the backside, which held the sensor in position during irradiation. As the shape of the insensitive area is the same on all measured pixels, the relative change can be examined anyway.

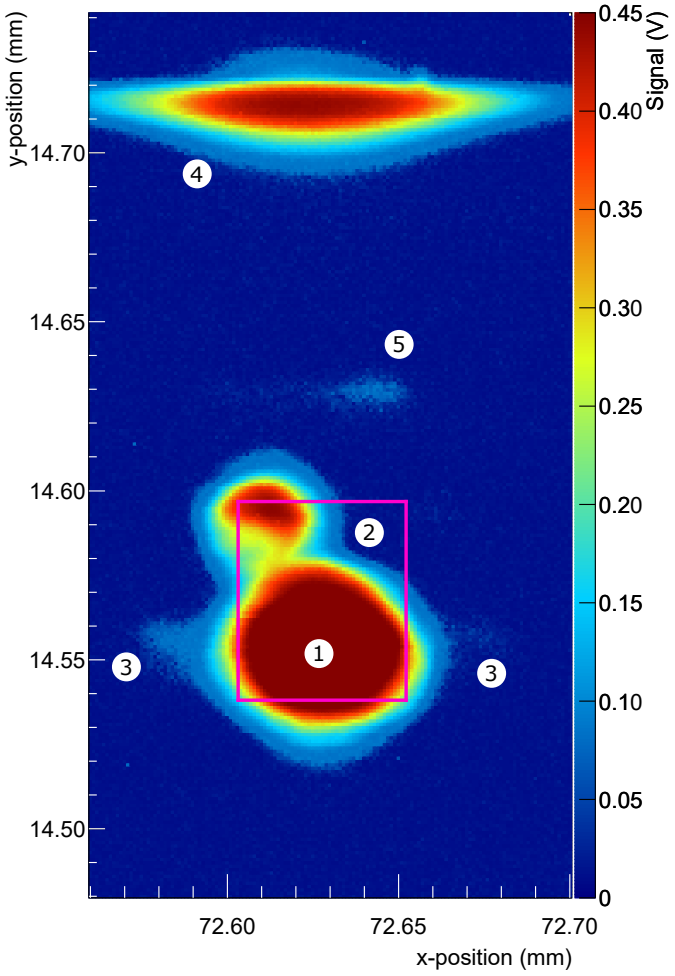


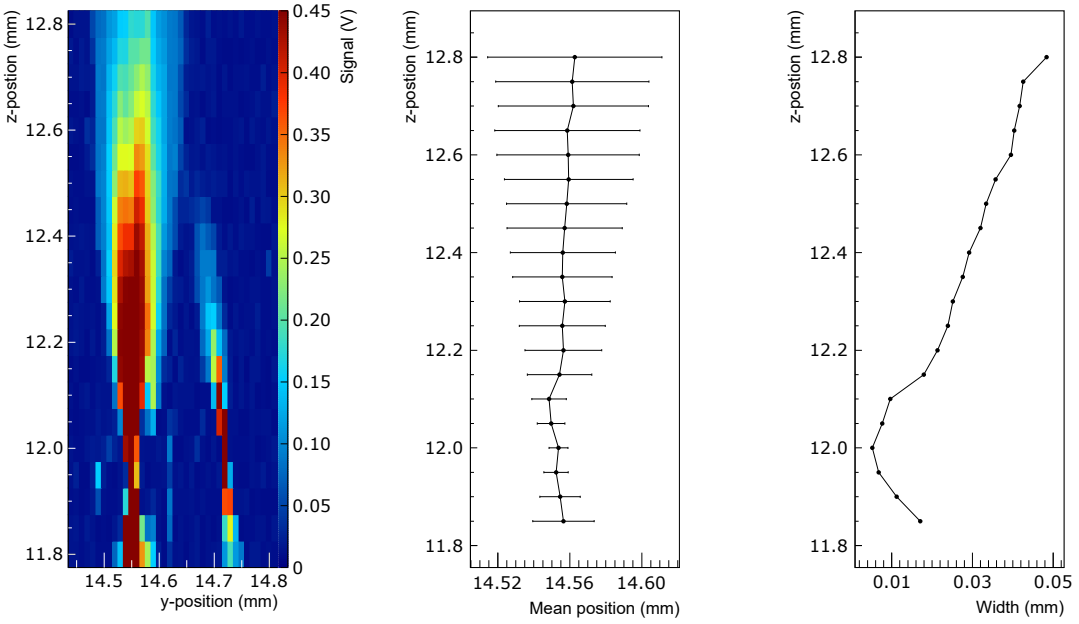
Figure 15.50: Mean analog signal as function of the laser's position. It is moved over the surface of the chip and the analog response of a single pixel is monitored. The pixel outline is indicated by a magenta line. High sensitivity is found at the pixel's location (1), but also area outside its outline appear sensitive: cross-talk from neighbor pixels (3), reflection from chip edge (4) and reflection of the laser on the chip's backside (5). Due to front-side illumination, some area inside the pixel is insensitive to the laser, as it is shielded by top metal layers (2).

The readout periphery is located to the right (positive x-direction). The pixels of one column are more prone to crosstalk than row-neighbors for two reasons: The transmission lines from pixel to periphery run in column direction, therefore the lines of column neighbors run parallel over a long distance. The second reason is the pixel grouping of ATLASpix1_M2. A group consists of 16 pixels, which are all located in the same column next to each other (Figure 11.12). The grouping happens in the pixel matrix, which means that these pixels share the same readout lines, which directly couples them to each other. This effect is visible as small sensitive areas (3). Depending on the depletion voltage, these areas are more or less distinct, but their center has exactly the distance 50 µm to (1), the x-direction pixel pitch.

Only the amplifiers of column 1 pixels can be connected to the analog out pad. The edge of the silicon chip is visible in (4). When the laser beam hits the edge, it is reflected into

the chip.

The slightly sensitive area (5) could not be traced to a feature of the chip. It is possibly a backside reflection or noise introduced by hitting the AmpOut line.



(a) **(b)** **(c)**

Figure 15.51: Before the measurement, the focus (optimal z-position) has to be found. This is achieved by scanning the xz- or yz-plane. Plot (a) shows the signal as a function of position in the yz-plane. The width of the active area as a function of depth and location is plotted in figure (b) and width as function of depth is shown in (c). The z-value that results in the smallest lateral distributed active area is used as focus (here: 12.0 mm).

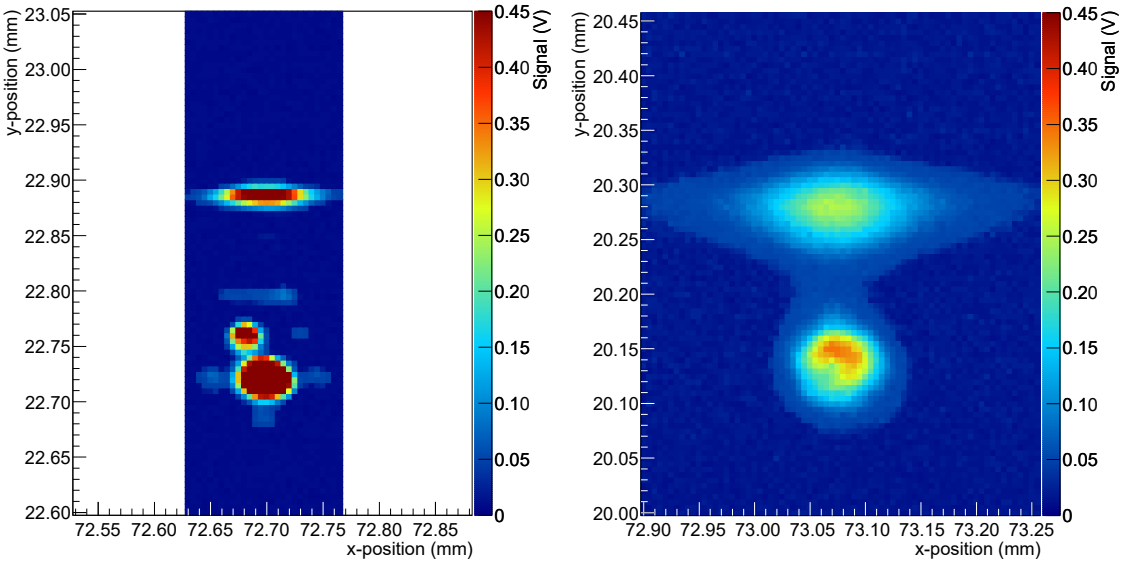
Before a picture like the previous one can be measured, the optimal focus distance between laser and chip has to be found. If the laser is too far out of focus ($> 1.5\,\mathrm{mm}$), the pixel does not pick up a signal even from direct hit. First, pixel ($50 \times 60\ \mu\mathrm{m}^2$) and laser beam ($\mathcal{O}(1\,\mu\mathrm{m})$) have to be brought into alignment. This is a semi-automated process. It has to be made sure, that the laser points at the pixel itself and not at an edge which reflects the laser.

Once the location of the selected pixel has been narrowed down to a few hundred µm, a scan in z-axis and another axis is started. Figure 15.51a shows the plotted result. The pixel is located at $y = 14.55$ mm. The responsive area to the right is the edge of the silicon and can be ignored. The shape can be compared to an hour glass: The lateral extension is minimal when the focal distance is reached. It is in the order of the pixel size. Moving the laser up or down widens the laser spot seen by the sensor and causes the double cone shape until the signal gets too weak to be detected.

Offline, the area of interest is identified, here it is $y = 14.45 \ldots 14.65$ mm. For every z-value the mean and standard deviation of the signal is determined by fit. The result is shown in figure 15.51b. The standard deviation is the parameter to be minimized, it is plotted in figure 15.51c as a function of z-setting. In this example the best focus is at a z-axis setting of 12 mm.

15.8.2 Sensitive area as a function of depletion voltage

Depletion voltage does not only change the thickness of the depletion zone, but also its lateral extension. The sensitive area of the same pixel has been laser-scanned four times for different depletion voltages. Unfortunately, the responding area shows a strong dependence on the focus of the laser. The focus determination is not as precise as necessary for this particular measurement ($< 5\,\mu m$). It can not be reproduced with sufficient precision after exchanging the DUT.



(a) **(b)**

Figure 15.52: Analog signal as a function of laser position. With good focus, all features are clearly delimited (a), but already slightly off focus they become blurry (b).

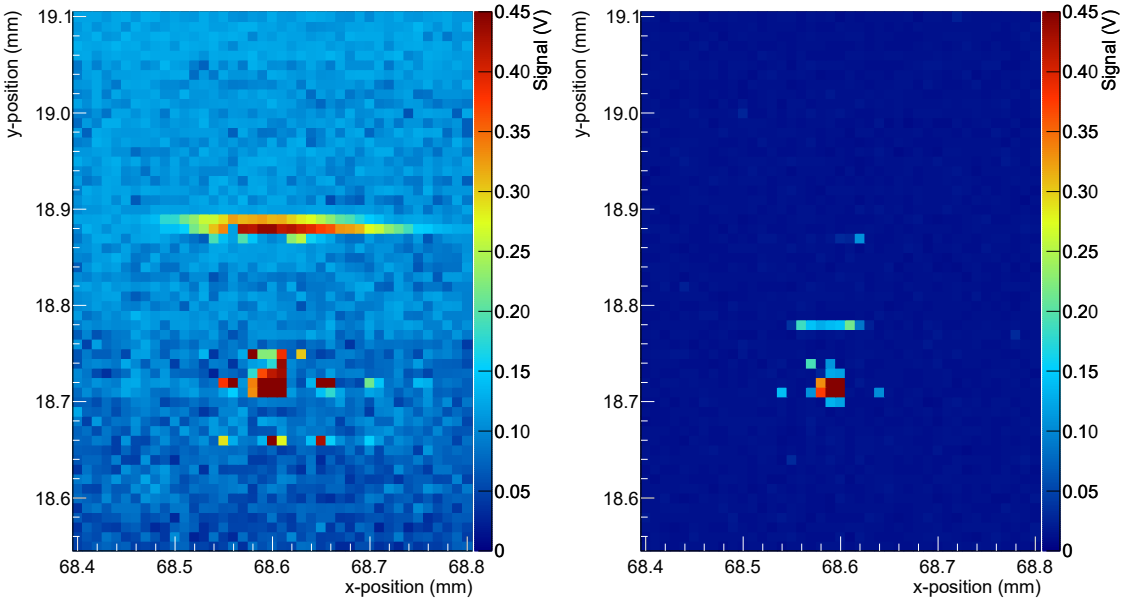
Figure 15.52 compares a scan of good focus (a) to a scan slightly off focus (b). The x- and y-scalings are the same. To avoid the systematic error of focus variations, the relative change of responsive area as a function of depletion voltage is calculated for different samples.

Additionally, the measurement resolution had to be reduced to 10 µm in favor of a feasible measurement time.

Figure 15.53 shows the scans of an irradiated sensor at -40 V (a) and at -1 V (b). At higher depletion voltage, the sensitive area is larger and the reflection of the die's edge is visible, which nearly vanishes for minimal depletion voltage. The baseline noise rises 5-fold for -40 V to $\sigma = 0.065$ V. Depletion voltage-dependent noise originates from leakage current (shot noise). The noise rises for all measured samples, the increase is in the order of a few percent for unirradiated samples and multifold for irradiated ones:

sample designator	fluence (n_{eq}/cm^2)	noise (mV) at HV = -1 V	noise (mV) at HV = -40 V	change (%)
AnIrr1	$6 \cdot 10^{14}$	13.0	65.4	502
AnIrr4	$1 \cdot 10^{15}$	11.6	64.0	553
An1	-	11.7	13.1	-1
An3	-	8.8	8.7	11

The noise of all investigated samples has been determined and was used to set a detection threshold for each sample individually. Pulses above threshold are treated as signals for



(a) **(b)**

Figure 15.53: Analog signal as a function of laser position, measured on the pixel of an irradiated sensor. Irradiated sensors have an increased leakage current, the caused shot noise increases the baseline noise. This effect is stronger for higher depletion voltage (a), than for lower voltage (b). Additionally the plot for higher depletion voltage reveals the temperature dependency of noise: The measurement setup is locked in a small box, which heats up during operation. The measurement points with small y have been measured first, the chip had a lower temperature, thus the noise is smaller.

calculation of the sensitive area. It is not possible to apply the same threshold to all samples as the gain of the amplifier is not uniform:

sample designator	fluence (n_{eq}/cm^2)	depletion voltage (-V)	sensitive area (×100 µm^2)	change (%)	mean signal (mV)	change (%)
AnIrr1	$6 \cdot 10^{14}$	40	28	100	279	100
AnIrr1	$6 \cdot 10^{14}$	20	21	75	260	93
AnIrr1	$6 \cdot 10^{14}$	10	17	61	266	95
AnIrr1	$6 \cdot 10^{14}$	1	13	46	275	99
AnIrr4	$1 \cdot 10^{15}$	40	28	100	251	100
AnIrr4	$1 \cdot 10^{15}$	20	19	68	319	127
AnIrr4	$1 \cdot 10^{15}$	10	18	64	339	135
AnIrr4	$1 \cdot 10^{15}$	1	16	57	333	133
An1	-	40	57	100	201	100
An1	-	20	72	126	177	88
An1	-	10	71	125	185	92
An1	-	1	55	96	217	108
An3	-	40	55	100	82	100
An3	-	20	62	113	74	90
An3	-	10	52	95	83	101
An3	-	1	36	65	150	183

Some findings of the above table were expected, others came rather as a surprise:

- Higher depletion voltage increases not only the depletion depth but also the lateral depleted area, the sensitive area. The difference between largest and smallest measured area is between 50% and 100%.
- The relative size of the active area at −1 V compared to −40 V is smaller for irradiated sensors than for unirradiated ones.
- The absolute size of the sensitive area of irradiated samples appears smaller than the area of unirradiated samples, but the main reason is the higher threshold of irradiated samples to effectively exclude noise. It might be interpreted as a reduction of the signal-to-noise ratio.
- The average signal is only of limited significance. A larger averaged signal might indicate sharper borders of the sensitive area, which might be caused by actually sharper limits, or simply by a slightly less focused laser beam. Both might be the reason for lost charge in the measured pixel. The missing charge is not necessary lost, but can be collected in other pixels (charge sharing).
- The unirradiated samples show an increase in active area, if the depletion voltage is reduced from −40 V to −20 V, whereas a decrease was expected and is observed at the irradiated samples and at other voltage changes. The explanation is that these samples are at −40 V already in full lateral depletion. The electric field of neighboring pixels collects all charge beyond the pixel border. At lower depletion voltages, the field is not as strong and not as deep. Therefore, charge sharing is less likely, as all charge generated by the laser beam is dragged by drift to the closest electrode. This hypothesis is supported by the drop in average signal when reducing depletion voltage from −40 V to −20 V. The charge of a focused laser beam pulse is collected by a single pixel and thus the average measured charge is larger for −40 V, than it is for −20 V depletion voltage.

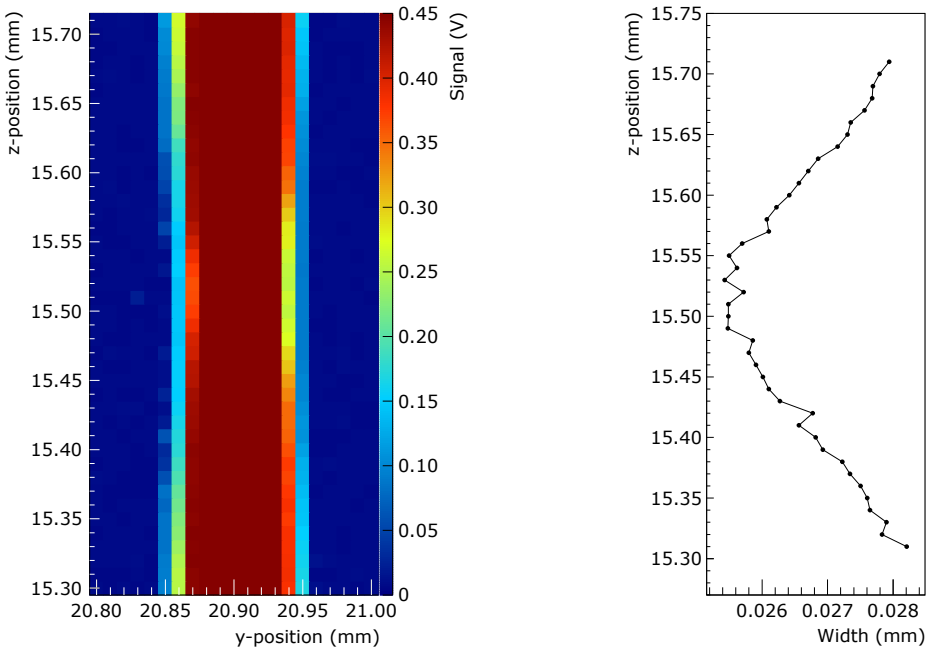
  This effect is not visible on irradiated samples. The additional bulk defects are being ionized by the applied depletion voltage. This is an intrinsic protection against radiation damage of HV-CMOS sensors and is an advantage over other CMOS sensor technologies, but reduces the depletion depth. For full lateral depletion, a higher depletion voltage than −40 V has to be applied.

  The effect is stronger in sensors produced on high resistive substrate (sample An1, 200 Ωcm), than on sensors with lower resistivity (sample An3, 80 Ωcm). Higher resistivity means larger depletion and therefore requires less additional depletion by reverse biasing.

15.8.3 Backside illumination

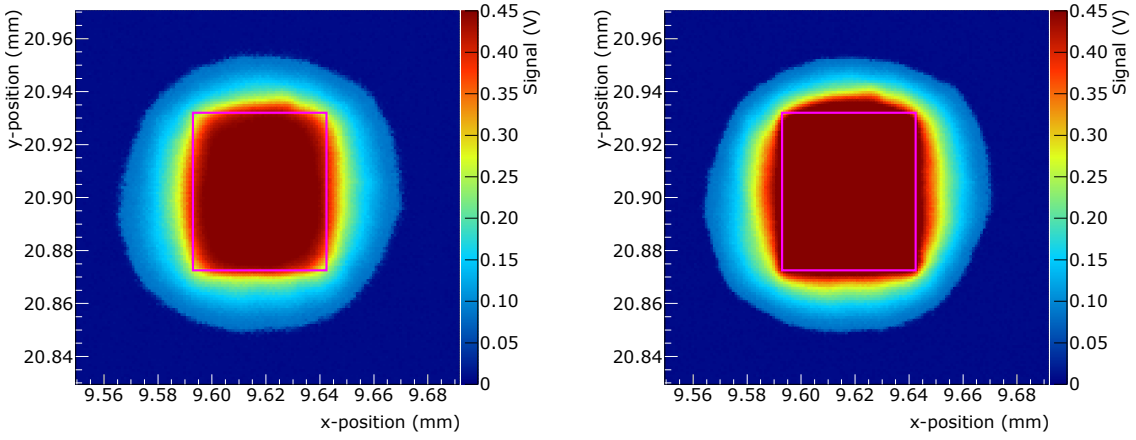
Front side illumination allows comparison of the active area of different sensors: thickness, irradiation, resistivity. Backside illumination can only compare thinned sensors, but has the advantage over front side illumination that no metal layers absorb the beam. Therefore, backside illumination can depict the actual extension of the active area. The absence of absorbing material in the way of the laser beam allows the reduction of laser intensity by approximately ten orders of magnitude. With this setting the amplifier still reaches saturation, but only for a short time (< µs). The intensity used for front side illumination is more than sufficient to permanently saturate the amplifier at a laser pulse repetition rate of 100 Hz.

The point of best focus has been determined in the same way as for front side illumination. Figure 15.54 shows the used yz-scan. The distribution for fixed y-value is not of Gaussian shape. Since the shape is boxlike, a double error-function is fit and the distance of both



(a) **(b)**

Figure 15.54: The 2D-histogram (a) shows the yz-scan of a backside laser measurement. Backside illuminated sensors are less sensitive to z-variations. The z-value with best focus is found by fitting the width of the sensitive area for each distance setting (b), it is 15.52 mm.



(a) **(b)**

Figure 15.55: Laser illumination from the backside is not blocked by the top metal layers, therefore the full area can be sensitive. The plots are made from data collected on a unirradiated sensor at −1 V (a) and −40 V (b). The magenta line indicates the pixel location.

determines the z-setting for further measurements. Notable is that the focus is less sensitive to z-variations (compare to figure 15.51). It is found for the z-stage set to 15.52 mm.

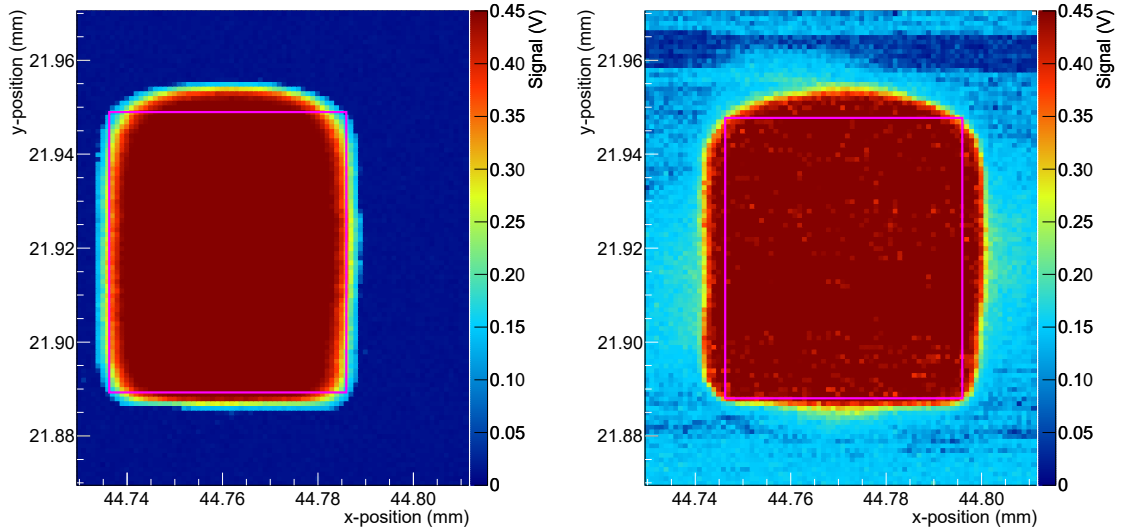
Figure 15.55 shows the two dimensional histograms of the average analog signal measured on a single pixel as a function of laser position. The data have been taken on an unirradiated sensor of 80 Ωcm resistivity and 70 µm thickness with 1 µm resolution at a depletion voltage of −1 V (a) and −40 V (b). The magenta line indicates the most probable location of the pixel (pixel size 50×60 µm^2). Already at minimal depletion voltage, the entire pixel area is sensitive, however the charge collection efficiency is slightly reduced in the very tips of each corner. At −40 V the entire area is fully sensitive. Even area below adjacent pixels is sensitive, probably by generating charge below the depleted volume, from where charge is collected by diffusion and not by drift with reduced efficiency and speed. The sensitive area outside the pixel outline in positive y-direction appears more bulgy than the other three boarders as here is no pixel located to collect charge itself.

The influence of the applied depletion voltage on the sensitive area is evaluated by considering the area with averaged signal above 400 mV as fully sensitive:

depletion voltage (V)	sensitive area (µm^2)	baseline noise (mV))
−1	2737	8.0
−10	3133	8.0
−20	3157	8.0
−40	3603	7.6
−60	3644	7.7

The area of a pixel is 3000 µm^2. At −1 V area with reduced sensitivity can be found, but already at −10 V the entire area reacts with nominal strength to laser pulses. With increasing depletion voltage the area grows larger.

The same measurement has been conducted on a sensor, which has been exposed to a fluence of 10^{15} n_{eq}/cm^2 with the same resistivity and thickness cut from the same wafer.



(a) **(b)**

Figure 15.56: Laser illumination of an unirradiated sensor at a depletion voltage of −1 V (a) and −40 V (b). Beginning breakthrough in (b) causes shot noise independent of the laser's position.

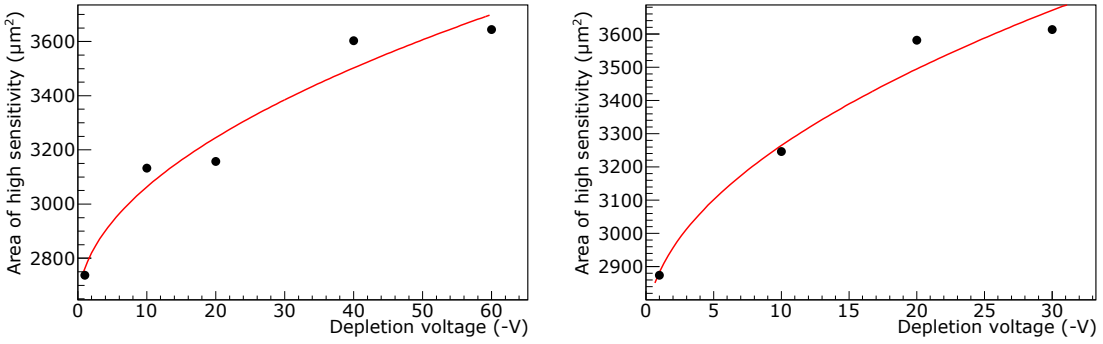
Figure 15.56 shows the averaged signal for −1 V and −40 V depletion voltage. Most notable is that the transition from sensitive area to insensitive area is much sharper than on the unirradiated chip. This indicates a reduced travel length of charge carriers in the non-depleted volume below the depleted volume. The irradiation induced defects can trap charge, act as recombination center for complementary charges and reduce the average free path length by scattering, especially if the generated charge carriers are not divided and not collected quickly by drift in the depleted volume.

The crisper image displays the bulge, due to no adjacent pixels, on the edge pixel towards positive y-values more clearly.

At −40 V, the irradiated sensor is already very close to breakthrough voltage and has a significantly increased leakage current generating a lot of noise. The threshold level of the sensing oscilloscope had to be raised for this measurement. Anyway, a significant number of the comparably small noise signals made it above threshold, reducing the average signal in the otherwise homogeneous area in the middle of the pixel. The baseline noise in the insensitive area easily exceeds 150 mV. In an experiment, this kind of noise would be unacceptable and counter measures would be indicated.

The sensitive area and noise have been determined in the same way as before for the unirradiated sensor:

depletion voltage (V)	sensitive area (µm^2)	baseline noise (mV))
−1	2874	12.2
−10	3246	14.2
−20	3581	18.2
−30	3613	8.6
−40	3462[5]	144



(a) **(b)**

Figure 15.57: Area of high sensitivity as a function of depletion voltage of an unirradiated sensor (a) and a sensor exposed to $10^{15}\,n_{eq}/cm^2$ (b). High sensitivity is defined as the area with average analog signals greater than 400 mV, approximately 80% of a saturated signal.

The area of high sensitivity as a function of depletion voltage is shown in figure 15.57 for an unirradiated and an irradiated sensor. The findings confirm the square root behavior seen in the front side illumination measurements.

The noise of the unirradiated sensor is practically independent of the set depletion voltage, where the already 50% elevated noise of the irradiated sensor reaches extreme amplitude when a value close to the breakthrough voltage is set.

[5]The increased threshold renders this measurement incomparable to the other measurements.

15.9 Effect of temperature on HV-CMOS sensors

The impact of leakage current on the noise floor has been impressively illustrated in figure 15.53a and 15.56b. Irradiation induced bulk damage boosts the leakage current and therefore noise (see chapter 3.4). An effective countermeasure is cooling of the sensor in order to reduce charge generation in the depleted sensor volume.

Cooling was not foreseen in the design of the chip carrier. The option of thermally coupling a peltier-element to the backside of the chip was discarded in favor of a window in the PCB for undisturbed access to the backside of the sensors.

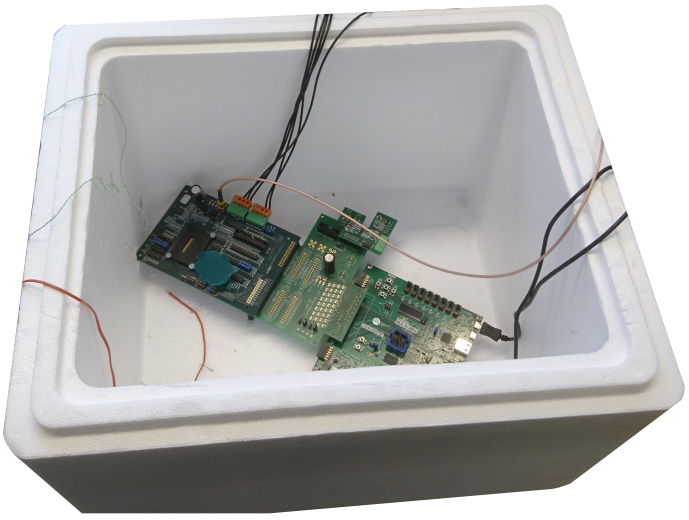


Figure 15.58: Picture of the characterization setup. Solid carbon-dioxid is placed on the floor of the box and in the free corners for cooling. During measurement the box is closed by a fitting top cover.

Therefore, air-cooling is the only possibility to reduce the temperature of the sensor. For this purpose, FPGA-board, MAB and carrier PCB with sensor are placed in an insulated box (Figure 15.58). The floor of the box was covered with dry ice. The temperature is measured by three precision temperature probes connected to a logger (see Appendix C). One is buried in frozen CO^2, one is fixed at the wall of the insulation box at half its height and the third measures the chip's temperature, located next to the chip on the carrier PCB.

The temperature can be manipulated by changing the distance of the dry ice to the chip. Below $-50\,°C$ the amplifier output starts uncontrollably oscillating, measurement is possible above only.

15.9.1 Leakage current of sensor diode

The measurement of leakage current is conducted at room temperature ($\approx 23\,°C$) and at $(-22 \pm 2)\,°C$, as this is a temperature regime widely used in detector systems for high-energy physics.

IV-curves have been measured at both temperatures on nine sensors: four unirradiated sensors, three sensors irradiated to a fluence of $6 \cdot 10^{14}\,n_{eq}/cm^2$ and two irradiated to $10^{15}\,n_{eq}/cm^2$. The voltage is set to values between $0\,V$ and a voltage slightly below breakthrough voltage.

Figure 15.59 shows the leakage current as a function of depletion voltage at room temperature (a) and -22°C (b). IV-curves measured on unirradiated sensors are depicted in black, sensors irradiated to $6 \cdot 10^{14}\,n_{eq}/cm^2$ are colored red and green lines were exposed to $10^{15}\,n_{eq}/cm^2$. As the leakage current of the unirradiated sensors is three orders of

magnitude smaller, than those of irradiated sensors, the same data are shown with reduced range on the y-axis in (c) and (d).

The breakthrough voltage is the location of the knee in the IV-curve. For every curve, the value is taken from the plot, as well as the leakage current at -40 V:

sample designator	resistivity (Ωcm)	fluence (n_{eq}/cm^2)	breakthrough voltage (V)		leakage current at -40 V (μA)	
			23 °C	-22 °C	23 °C	-22 °C
A200	200	-	-65	-75	-0.295	-0.089
A200A	200	-	-71	-76	-0.128	-0.057
An1	200	-	-70	-75	-0.364	-0.101
An2	200	-	-62	-65	-0.346	-0.107
An3	80	-	-64	-73	-0.386	-0.328
AnIrr1	200	$6 \cdot 10^{14}$	-48	-50	-284	-0.638
AnIrr5	200	$6 \cdot 10^{14}$	-52	-54	-342	-0.419
AnIrr3	200	$1 \cdot 10^{15}$	-53	-56	-215	-0.082
AnIrr4	80	$1 \cdot 10^{15}$	-57	-59	-248	-2.262
AnIrr6	200	$1 \cdot 10^{15}$	-53	-54	-190	-1.155

The comparison of the measured IV-curves and extracted key values delivers some characteristics:

- The leakage current, at applied voltage between 0 V and breakdown, is 1000 times smaller for unirradiated samples compared to irradiated ones. Cooling reduces the current to values comparable to unirradiated sensors.
- The breakdown voltage depends on the fluence. However, there is no significant change between $6 \cdot 10^{14}\,n_{eq}/cm^2$ and $10^{15}\,n_{eq}/cm^2$.
- Comparison of 80 Ωcm and 200 Ωcm samples suggests that the leakage current of the cooled sensor is about two to three times as high in low resistivity substrate samples, as in high resistivity substrate.
- The breakdown voltage is only little affected by temperature.

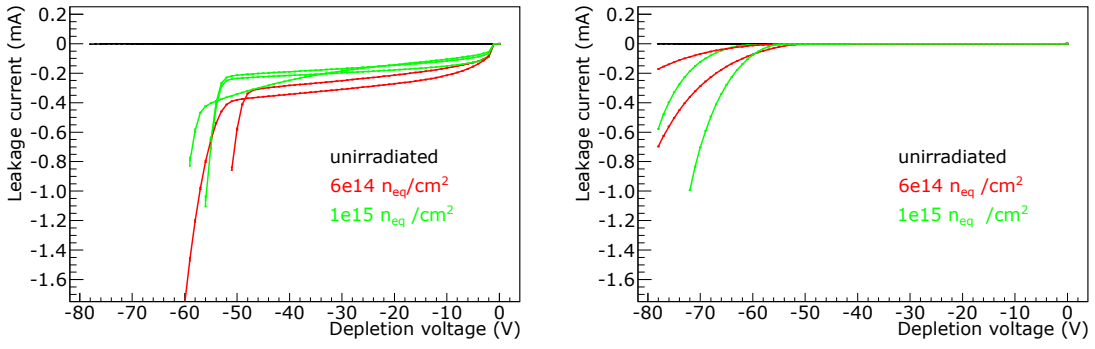


Figure 15.59: Leakage current as function of depletion voltage of unirradiated sensors (black), sensors irradiated to $6 \cdot 10^{14}\,n_{eq}/cm^2$ (red) and $10^{15}\,n_{eq}/cm^2$ (green) at room temperature (a) and -22°C (b).

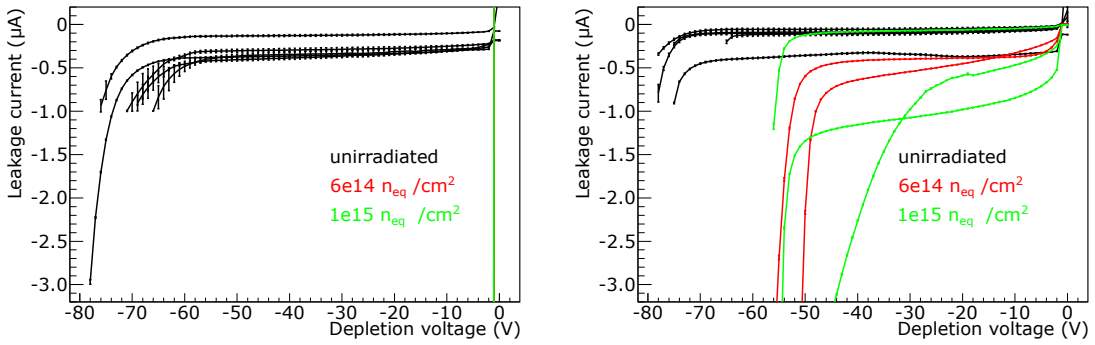


Figure 15.59: Leakage current as function of depletion voltage with reduced y-range to depict the three orders of magnitude smaller leakage current of cooled or unirradiated sensors. Plot (c) shows the same data as (a), plot (d) shows the same data as (b).

15.9.2 Signal and noise as a function of temperature

The effect of temperature on the sensor's leakage current has been investigated in the previous section. Now, the signal amplification and noise as a function of temperature is to be investigated. In order to exclude effects by leakage current, an unirradiated sensor has been used with the depletion voltage set to -40 V, well above breakdown voltage. The measurement setup stays the same (Figure 15.58), but an Iron-55 source is placed above the sensor. The analog output is monitored by an oscilloscope and every waveform is read out and stored. Temperature is permanently logged with 30 measurements per minute.

The insulation box is filled with dry ice as high as possible, without the ice touching the setup. The initial measured temperature at the sensor after closing the box is approximately -40 °C. Due to imperfect insulation and heating of the setup, the dry ice starts to sublimate. The temperature rises as the amount of dry ice decays. Once all dry ice is consumed, driven by waste heat, the temperature rises above room temperature to approximately 30 °C. This process takes approximately 12 hours. It has been repeated four times on four consecutive days. During each cycle, the Iron-55 source illuminates the sensor and the waveforms are tagged with generation time stamp and stored for offline analysis.

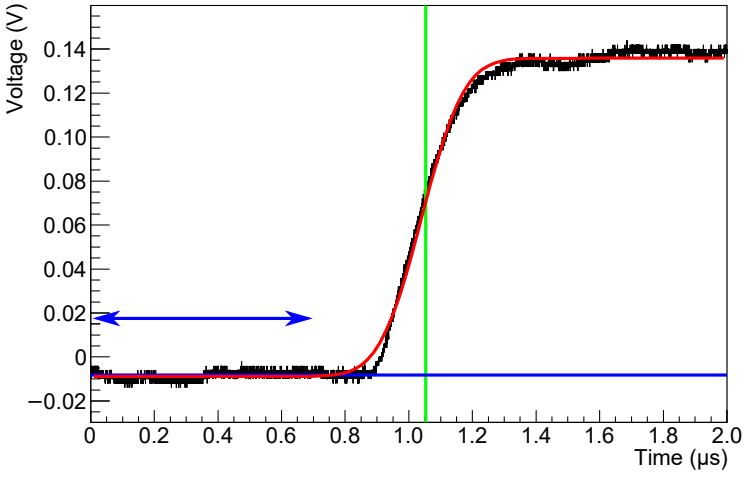


Figure 15.60: The amplifier output is recorded by an oscilloscope and stored on the PC. Only the rising edge is recorded. The key values are extracted by fit and calculation.

The generation time of a waveform file is used to obtain the temperature at the time of measurement from the data of the temperature logger. The relative time uncertainty has been determined to be less than two seconds. This is insignificant for temperature tagging, as the temperature changes over hours, not seconds.

From these waveforms the key properties are extracted. This is illustrated in figure 15.60. The recorded time period is limited to 2 µs, but with 10 000 points. With this setting, only the rising edge of the signal pulse is recorded and is fit by an error function (red), from which the middle point is extracted (green). Two thirds of the points to the left of it, are used to determine the baseline noise, indicated by the blue arrow. Also from the error function extracted are maximum and minimum, their difference is the signal height.

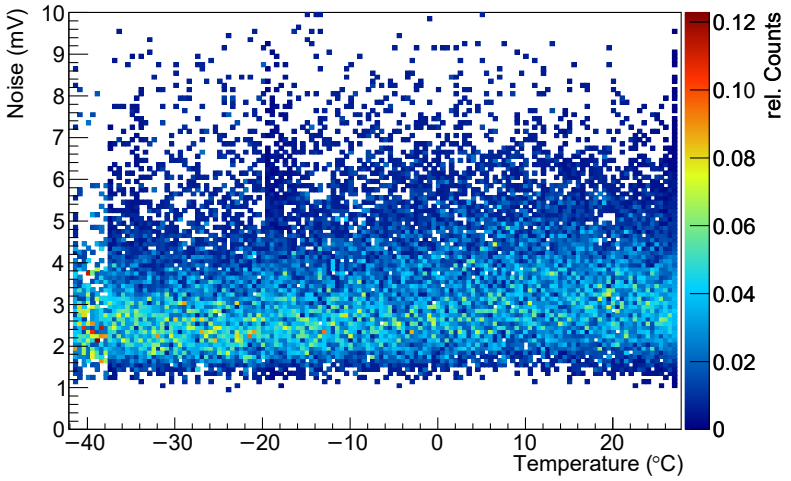


Figure 15.61: The histogram shows the correlation of baseline noise and temperature. The noise is extracted from every recorded waveform and tagged with temperature information. The entries of each temperature (column) is normalized.

The combination of baseline noise and temperature of every waveform is shown in figure 15.61 as 2D histogram. As the number of entries per temperature strongly varies, the histogram is normalized for identical temperatures. A definite trend can not be observed.

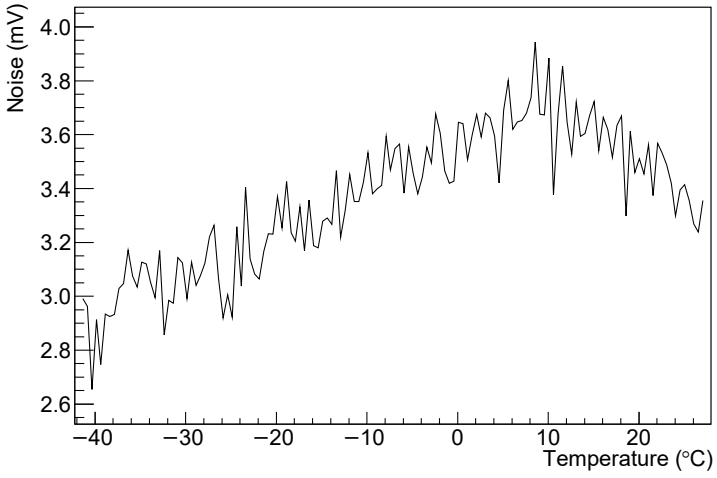


Figure 15.62: Mean noise as a function of temperature. A distinct maximum is reached at 10 °C. The measured values cover the range from 2.8 mV to 3.8 mV.

For easier analysis, the mean noise of every temperature is calculated and plotted as a

function of temperature in figure 15.62. This visualization reveals that, in contrast to the expectation, a noise maximum is located around 10 °C. The mean noise drops for smaller and larger temperatures linearly. Low temperature is expected to reduce leakage current and thereby noise in the measured temperature range. The drop of noise for high temperatures might be due to a change of conductivity in the bias DACs. Already a small change of the feedback setting can affect the baseline noise. At higher temperatures, the respective DAC might vary its output slightly causing the observed decrease of noise.

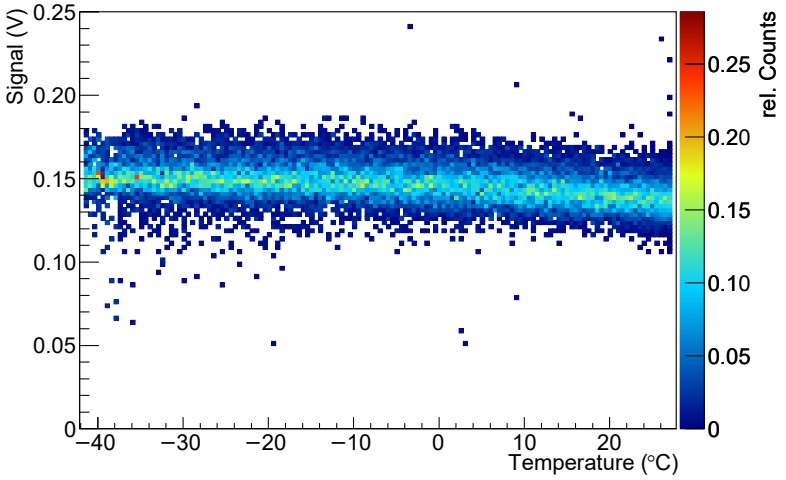


Figure 15.63: Correlation plot of signal and temperature. The signal has been determined by fitting an error function to the waveform, every value is filled into the 2D histogram in the respective temperature column. Each column is normalized. The signal is slightly larger at smaller temperatures.

The same 2D-plot has been made for the signal height. Figure 15.63 shows the relative distribution of signal heights for temperatures between -40 °C and 30 °C. Already in this representation of data, a trend can be observed: lower temperature leads to higher signal.

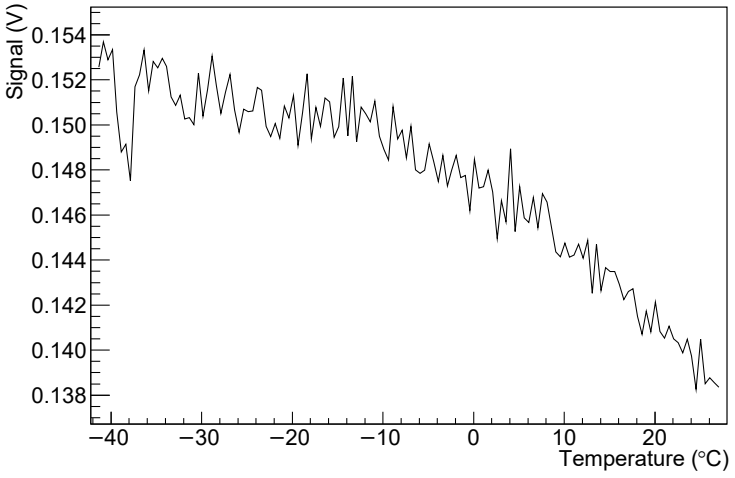


Figure 15.64: The mean signal as a function of temperature. It rises logarithmically with decreasing temperature from 138 mV to 153 mV (+11%).

To quantify this finding, the average signal height is plotted as a function of temperature in figure 15.64. The impression of an increasing signal for decreasing temperature is confirmed. The increase between room temperature and -40 °C is about 10%. In agreement with

the noise characteristic, an influence of temperature-dependent DAC performance can be deduced.

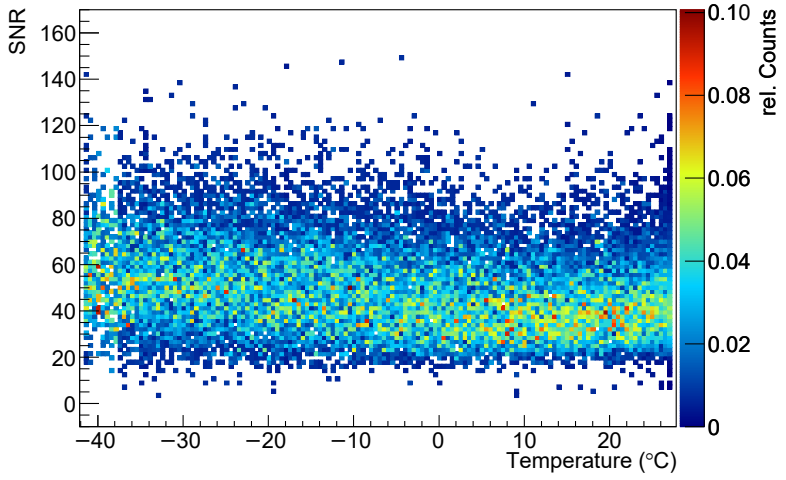


Figure 15.65: The combination of signal and noise information is used to calculated the signal-to-noise ratio.

The most interesting temperature-dependent value is the combination of signal and noise, the signal-to-noise ratio (SNR). Relative occurrence of the combination of SNR and temperature is shown in figure 15.65. A clear trend is not visible, however the spread of entries seems broader below 0 °C, than above.

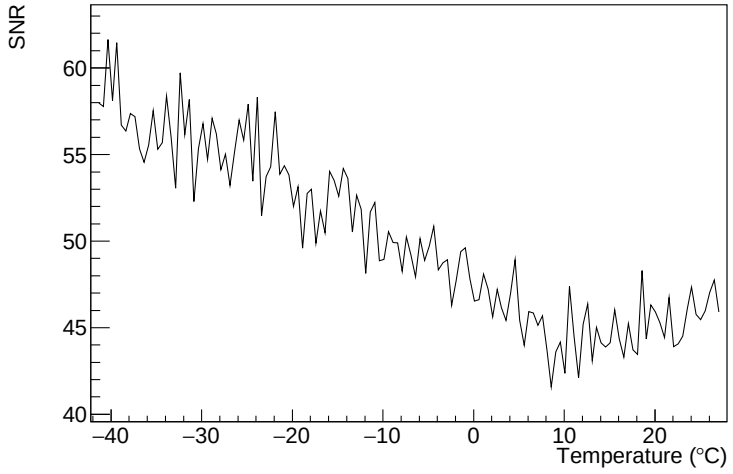


Figure 15.66: Correlation plot of the signal-to-noise ratio and temperature. The averaged SNR per temperature reveals a minimum at 10 °C, below and above of which, the SNR is rising. In the measured temperature range, mean SNR varies from 42 to 61.

The mean SNR for given temperature is depicted in figure 15.66. The SNR has a distinct minimum at 10 °C. Below that and in the measured temperature range, cooler means better SNR. Both signal and noise contribute to it. The increase at -40 °C compared to room temperature is approximately 30%. Even the more feasible operational temperature of -20 °C, grants an improvement of approximately 20%.

16. GECCO – the GEneric Configuration and COntrol Setup

A second version of the Multi-purpose Adapter Board (MAB) has been developed to satisfy the need for large-scale sensor setups and multi-sensor assemblies for a fast development speed, high modularity and plug-and-play characterization. The size has been reduced, ground separation between sensor and FPGA is now standard and a versatile hardware interface allows future upgrades. The software and firmware packages have been updated to the new hardware and new universal functions have been implemented. In order to increase its recognition value, a logo has been designed for the GECCO system (Figure 16.1).

Figure 16.1: Logo of the GECCO project.

16.1 Hardware

In comparison to the MAB system, the main board has changed the most. With the exception of a few signals needed to control the functions of GECCO, all lines from the FPGA are routed 1:1 to the output connector, a 16x PCI express card slot with 164 pins, seen in figure 16.2 and 16.3 (from FMC left to PCIe right).

This card slot brings several advantages over the previously used box headers: Higher integration density allows more connections without the need for compromises. The only needed component is a card slot, which is widely available in several variants due to its extensive use in computers. The boards that carry the device under test do not need a connector, but are directly plugged into the PCI express slot. The power supply for the DUT is no longer to be connected to the carrier board, but can be routed via the PCI express connector. Therefore, the number of components on the carrier board in general can be reduced to few decoupling capacitors and no additional connections are required, which significantly speeds up serial testing of samples.

The large number of free pins on the PCI express connector allows still a large number of additional connections, after the numerous power lines with two connections each are routed. These remaining free pins are used by the function cards. The GECCO has eight 1x PCI express card slots. The different function cards are introduced below. Any type can be plugged into any slot. Each slot has input lines for configuration, power lines and eight output lines, routed to the carrier PCB.

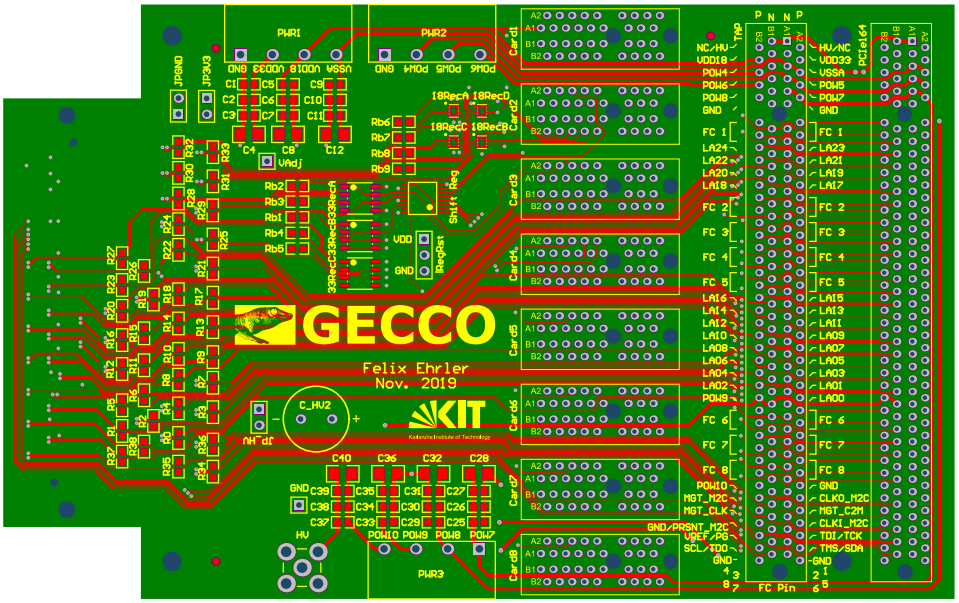


Figure 16.2: The top layout of the GECCO main board. The main components are from left to right: FMC connector to FPGA, optional line termination resistors, power and high voltage plugs (top and bottom), differential line receivers and a shift register for configuration of the DUT and function cards, eight card slots (1x PCIe), two PCIe connectors to device under test and for signal probing.

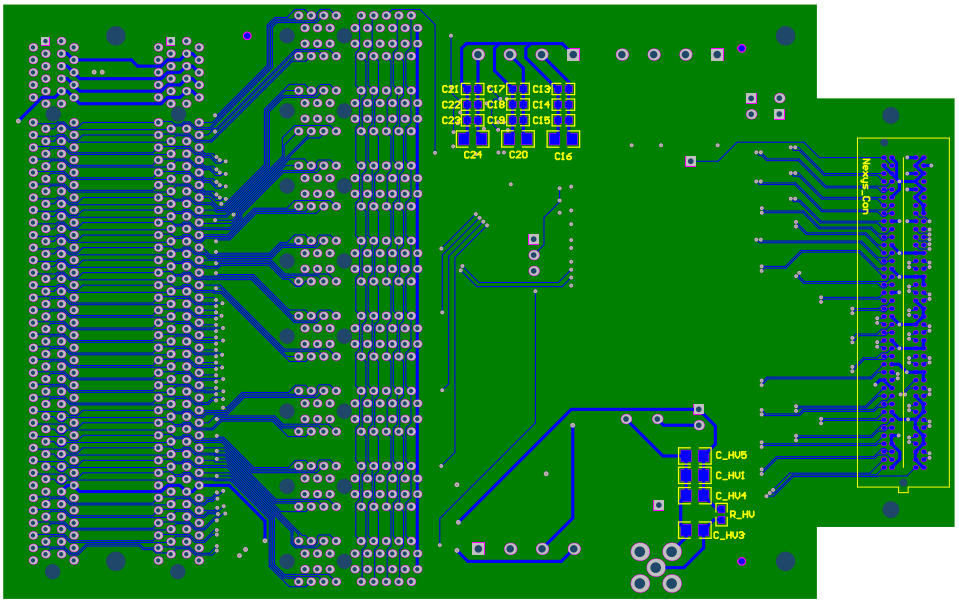


Figure 16.3: The bottom layout of the GECCO main board. This side is mainly used for signal routing. The internal layers are used for ground and power planes.

VoltageCard

The VoltageCard is basically a minor revision of the previously used VoltageBoard. Instead of five output voltages, the combination of DAC and amplifier is now implemented eight times per card. This card can be used to bias transistors and define baselines, even small currents can be drawn. Power supply by VoltageCards is not possible. Its layout is shown in figure 16.4.

InjectionCard

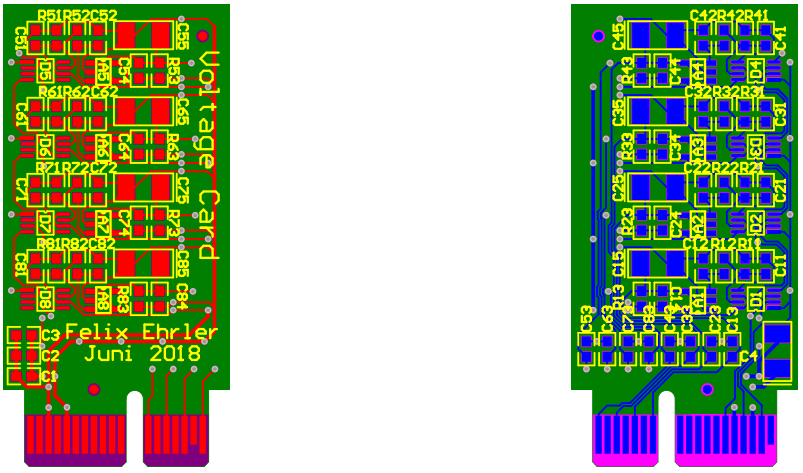
The new InjectionCard is actually a combination of the previously used InjectionBoard with two voltage generation circuits as used on the VoltageCard. Therefore, the new InjectionCard is standalone and does not require the input of voltage levels. One InjectionCard can generate two independent injection signals. The injection is triggered by single ended chopper signals. In order to separate grounds of FPGA and GECCO, the differential signals from FPGA are translated to single ended signals in differential line receivers on the GECCO main board. Its layout is shown in figure 16.5.

ConfigCard

The ConfigCard is the simplest card available. It connects the single ended configuration lines on the GECCO main board to pins of the PCI express connector of the DUT. Thus, it is possible to route the single ended configuration lines on GECCO to a certain, convenient area of the carrier board. Its layout is shown in figure 16.6.

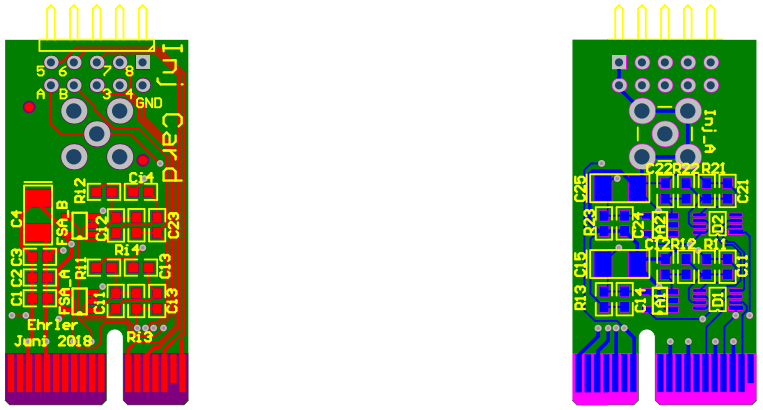
PowerCard

The PowerCard brings an additional plug to connect arbitrary signal, voltage or power lines from external devices to the DUT. It could be connected directly to the DUT, but that would contradict the paradigm of a single connection – the PCI express connector. Its layout is shown in figure 16.7.



(a) (b)

Figure 16.4: The layout of the VoltageCard. It's up to eight voltages are generated according to the configuration received from the FPGA between 0 and 3.3 V.



(a) (b)

Figure 16.5: Layout of the InjectionCard. It houses two test signal injection circuits. The switching is triggered by an FPGA signal (chopper). Both chopper and injection signal can be probed on an additional connector. Two voltage generation circuits have been added to the card to render it independent of a VoltageCard.

(a) (b)

Figure 16.6: Layout of the ConfigurationCard. This card bridges the four configuration signals, data (Sin), clock 1 (Ck1), clock 2 (Ck2) and load (Ld) from GECCO main board to the respective card pins of the PCI express card and to the DUT. At this point the differential signals from the FPGA are already translated to single ended CMOS signals.

(a) (b)

Figure 16.7: Layout of the PowerCard. It features the possibility of connecting four additional power or signal lines to a setup. Only in case of power lines, the decoupling capacitors should be assembled.

The active cards are powered by 3.3 V. The maximum output levels have to be adjusted in the software, according to the DUT. The power supply for GECCO can be provided by the NexysVideo FPGA board, but it is not recommended as this requires a common ground and the levels are not noise free.

The configuration of the cards is done by three single ended CMOS signals, which are generated on the GECCO main board by differential line receivers. The differential signals are driven by the FPGA. Differential signal transmission from FPGA to GECCO is used as it does not require a common ground. Configuration of up to eight cards by only three signals requires a special scheme. Daisy chaining of the cards is not possible, as not every slot has to be equipped.

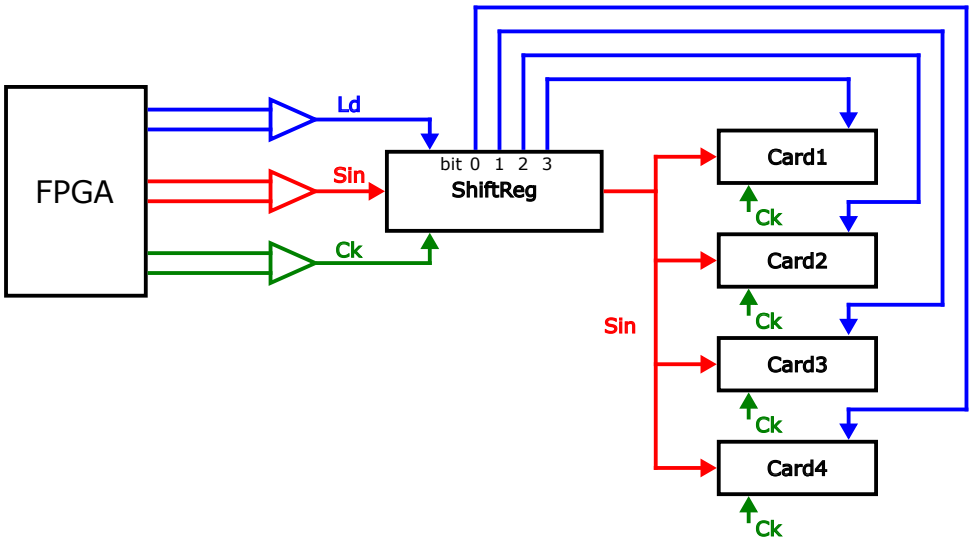


Figure 16.8: The configuration of the cards is based on a fan of shift registers. The card registers are in parallel, after the shift register generating the load pulse. All cards receive the configuration, but only one receives a load pulse.

Figure 16.8 shows how configuration of the cards is realized. The schematic shows only four cards, while the real board has eight. The configuration is shifted through a shift register and then parallel into all cards. After the card configuration, the configuration of the register is sent. It consists of seven '0's and one '1'. The load pulse is sent to the register only. The single 1 sets one of the outputs high, sending a load pulse to the respective card. Therefore, even though the configuration has been sent to all cards, only the selected card will store it. A series of '0's and a load pulse for the shift register resets all load signals. The clock signal is shared by the shift register and all cards.

The card slots of GECCO have been designed so that allows the design of new cards with new features, which were not part of the original GECCO kit:

RegisterCard
: The RegisterCard holds a shift register with eight output signals. An application for this card is setups with multiple sensors. The configuration of the sensors is then done in a similar scheme as the configuration of the function cards.

TriggerCard (projected)
: The TriggerCard is an extension to the GECCO system for charged particle measurements with scintillator and photo multiplier tube (PMT). It is going to house a comparator, which will compare the output signal of the PMT to an arbitrary threshold. The comparator output can be fed to the FPGA for automated counting (compare chapter 7.7.2).

As already mentioned is the supply voltage for all functions of GECCO 3.3 V. The DUT might require a different supply voltage, and therefore the single ended configuration signals have to be translated to this voltage level and the VoltageCards should not be allowed to generate output voltages exceeding the supply voltage of the DUT. The VoltageCard's upper output limit is set by the control software, but the voltage level of the configuration line is defined by the supply voltage of the DUT, which at the same time is connected to special differential line receivers. These special line receivers translate the differential signals from the FPGA to a single ended signal that can be understood by the DUT. Their input and output stages have separated power supplies. The input is powered by the GECCO main supply, while the output stage is powered by the DUT-specific power supply.

The list below sums up all connections provided by GECCO for a device under test on the 16x PCI express connector:

- 9 power lines with two pins each. One is reserved for the GECCO supply voltage of 3.3 V.
- 1 additional power line with two pins for high voltage. It is decoupled with high voltage capacitors, SMA connector is used and the pins on PCIe are separated from others.
- Several ground pins.
- 23 differential signal pairs. Can be used single ended in case of common ground.
- 1 gigabit link, forwarded from FPGA.
- 8×8 lines are connected to the eight card slots that can be equipped with function cards.

16.2 Software and firmware

The firmware has been taken over from the Multi-purpose Adapter Board. Only minor changes and updates were required.

The software package required some updates to comply with the card configuration. For rapid development, container classes for VoltageCard and InjectionCard have been written. To use a card, only the used card slot is to be given.

Additionally, bug fixes and general improvements to the software package have been implemented over the development time.

16.3 ATLASpix1 based beam telescope

The GECCO system has been designed to support more than one DUT (Figures 16.9 and 16.10). In order to track particles, a telescope based on ATLASpix1 sensors was developed for the GECCO system. It consists of a mother board, to be plugged into the GECCO main board, which routes signals to the telescope layers. Additionally, some line receivers are used to generate load signals for each sensor, as the RegisterCard has not been available at the time of design. Up to five layers can be supported by the mother board.

The layers are simplified versions of the characterization board for ATLASpix1 (see Appendix F), without the MuPix8 part and with reduced debugging options. Only one ATLASpix1 matrix can be operated at a time (Figure 16.11). The fully assembled telescope with five tracking layers is shown in figure 16.12.

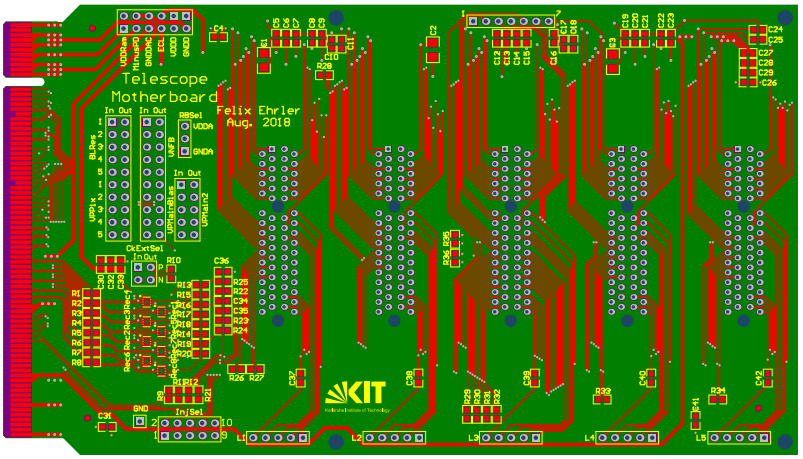


Figure 16.9: Top layout of the Beam Telescope mother board. It is to be plugged into the GECCO board.

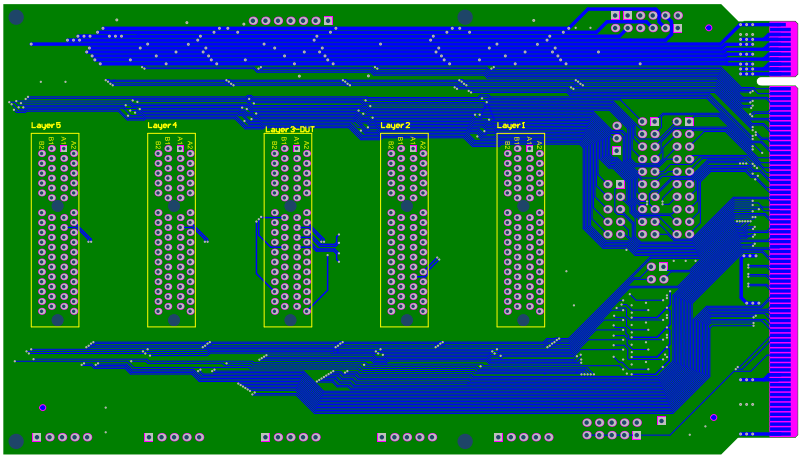
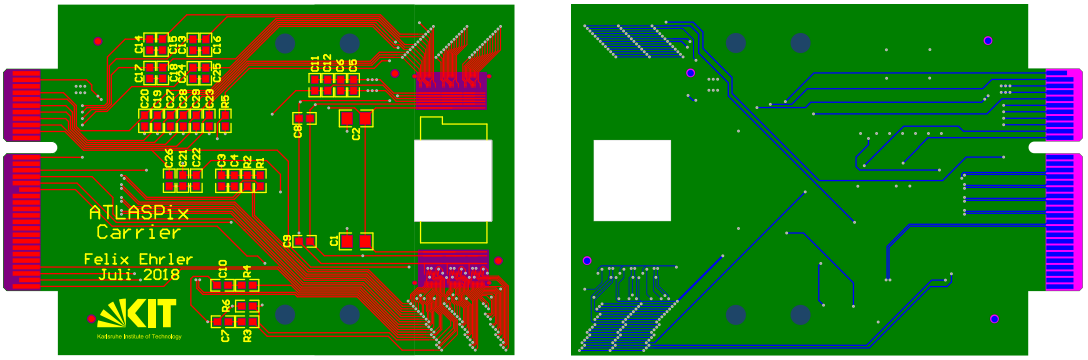


Figure 16.10: Bottom layout of the Beam Telescope mother board. Up to five layers can be connected to it.



(a) **(b)**

Figure 16.11: Layout of an ATLASpix1 layer for the beam telescope. Only one matrix at a time can be connected to the system.

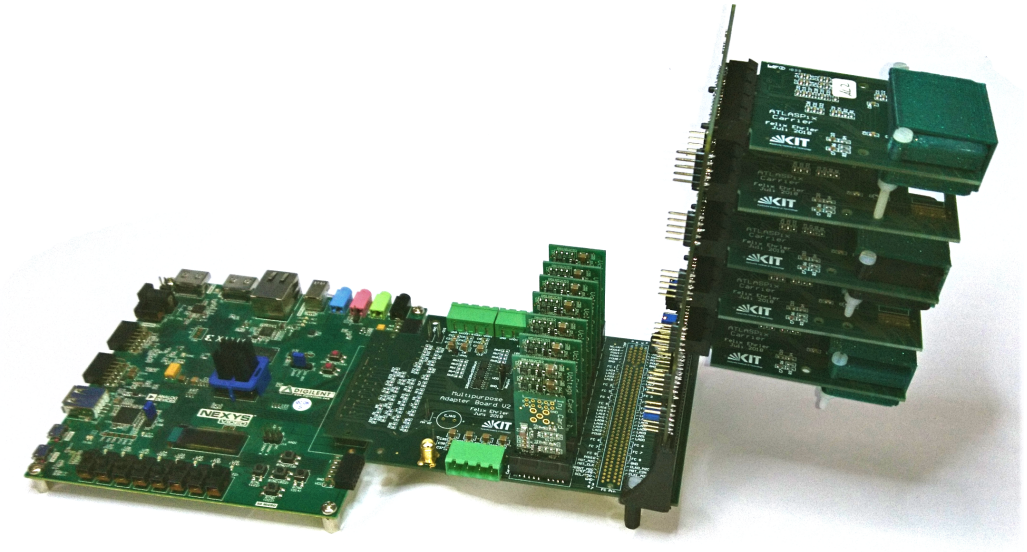


Figure 16.12: Photo of the GECCO Beam Telescope: NexysVideo FPGA Board, GECCO main board, telescope mother board and five layers.

16.4 Summary of GECCO

The GECCO system is a leap forward in sensor characterization at KIT ASIC and Detector Laboratory. It follows the path pursued by its precursor, the Multi-purpose Adapter Board System, to a new level of modularity, signal quality and rapid development, even for the largest HV-CMOS sensors or sensor assemblies.

The main board is designed for a large number of connections between FPGA and device under test, but maintains separated grounds. The functionality of the system is expandable by functions cards, which can be plugged in eight card slots. The universal interface allows not only for the connection of cards foreseen during development, but also for new ones. Each card can be plugged into any slot. This detail suits the system for a long run duty.

The software- and firmware-package has been extended and tailored for reconfiguration to new sensors. Therefore, GECCO is even quicker adaptable to new sensors than MAB.

All signal, power and voltage lines to the carrier PCB, are routed via the GECCO main board, thus for serial characterization of sensors, only the carrier PCB has to be exchanged.

So far the system has been used in the characterization of the following chips:

- ATLASpix1
- ATLASpix3
- MuPix8
- H35Demo
- 28 nm ADC
- HITPIX (counting)
- HITPIX (integrating)
- ASTROPIX

Approximately 50 units of GECCO have been produced. The minority of about 10 is in use by KIT-ADL, the others have been shipped to collaborating institutes:

- University of Edinburgh
- University of Bristol
- Rutherford Appleton Laboratory (RAL)
- University of Liverpool
- Lancaster University
- INFN Milano
- Numerous Chinese universities of CEPC collaboration, e.g. IHEP Beijing

17. HV-CMOS sensors for beam monitoring

17.1 Tumor therapy

Approximately 500.000 people are diagnosed with cancer every year in Germany [106]. Cancer and tumor treatment is challenging as the tumor cells have developed from the patients tissue. These cells can be located next and in vital organs and do not have to be clearly segregated from healthy tissue. Treatment is based on three main principles:

Resection The tumor is surgically removed. This is only possible if the tumor is accessible. Success depends on the type of tumor, especially if it segregated from the healthy tissue and if it is localized. Some missed cells are enough for a tumor to regrow.

Chemotherapy Chemicals are used against the tumor. Typically they are used to inhibit growth of tumor cells. This does not only freeze the disease, but, due to the special metabolism of some tumors, can lead to cure. Chemotherapy is not very selective and therefore the side effects can be significant, the spectrum ranges from barely notable to much worse than the primary disease and even lethal.

Irradiation The tumor, especially when being delimited from healthy tissue can be destroyed by ionizing irradiation. The tumor needs to be localized and accessible by ionizing radiation. However, radiation does not distinct between tumor and other tissue, therefore precise aiming in direction and depth is crucial. The dose necessary to destroy a tumor can not be applied in a single irradiation but needs several sessions.

Most tumor therapies include one to three of the above mentioned methods of treatment. Persistent learning, new medicine and improved technology increase the chances of successful therapy. Usage of HV-CMOS sensors in monitoring of therapeutic beams, could lead to an improved tumor therapy, with direct impact on health and well-being of the patient, by higher spatial and time resolution, compared to the wire chambers in use. In this chapter, the applicability of HV-CMOS sensors for therapeutic heavy-ion beam monitoring will be developed.

17.1.1 Irradiation therapy

Tumors can be irradiated with different types of particles, from photons, over electrons, to protons and heavy-ions. For maximum impact on the tumor at minimal impairment of the surrounding tissue, the particle beam has to be well aimed and the particle's energy has to be defined with maximum precision as it defines the penetration depth.

The penetration depth can be estimated by integration of the Bethe formula (see chapter 2.3.2). It is only valid for large particle energies and needs corrections after slow down. The resulting Bragg-curve with its distinct peak right before the particle is stopped is the prerequisite for irradiation therapy (Figure 17.1). The Bragg-peak is shifted onto the tumor

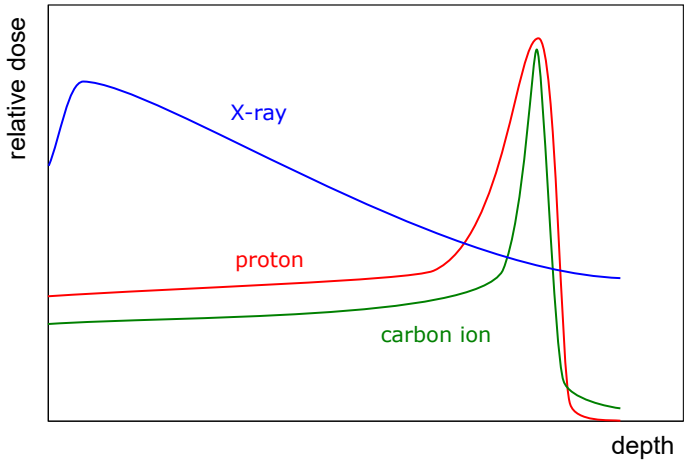


Figure 17.1: Bragg-curve of different particle types used in tumor treatment. The heavier the particle is, the more precisely the tumor can be targeted. (After [107])

volume, the tissue before is only mildly affected and right after the peak, no more energy is deposited. In practice, the particles are produced with such a good energy resolution, that it has to be smeared out again, to irradiate the entire tumor and not only a slice of it. However, the tissue before the tumor receives a certain dose, too. The amount of lost energy before the target, depends mainly on the used particle type. While light particles, such as photons and electrons, with none or small mass, deposit a significant fraction of their energy before the target, heavier particles, such as protons and ions, do less harm to the tissue before the target. Consequently, photon and electron irradiation is an option in the treatment of tumors close to the skin and tumors deeper inside the body without sensitive organs in the irradiation path. Tumors behind or even within sensitive organs, e.g. the brain, are better irradiated with heavier particles to minimize the side effects [107]. The downside of heavy-ion irradiation is the need for large-scale accelerators.

17.1.2 Beam monitoring of heavy-ion beam

Either type of irradiation requires precise planning and monitoring. The tumor has to be mapped and the respective body part has to be immobilized during irradiation. As important is the precision of the beam. This includes dose, energy and direction. While energy is controlled by the particle source, dose and direction of the beam have to be controlled by additional instrumentation.

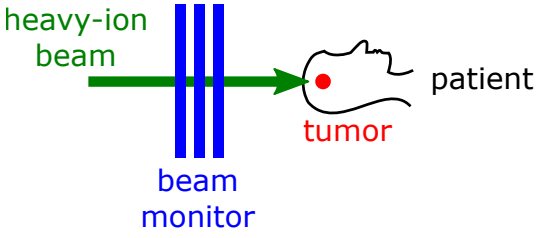


Figure 17.2: Sketch of tumor irradiation. The particle beam is monitored through out the session.

To date, therapeutic (heavy-) ion beams are monitored by a stack of wire chambers [108] (see chapter 2.1). The sketch of the principle is shown in figure 17.2. At least two are needed to determine the position in x- and y-direction, the beam shape can not be controlled and has to be calibrated in additional measurements. The independent dose control requires at least one additional layer. Wire chambers have a limited spatial resolution given by the wire distance, the alternating pattern of wire and no-wire is unfavorable for the homogeneity of the irradiation and the use of parallel patient monitoring by computed tomography scan is not possible as its strong magnetic field affects charge movement in the wire chambers.

HV-CMOS detectors are expected to improve all these short comings of wire chambers [109]: The resolution is determined by the pixel size and can be improved to sub-pixel resolution by fitting of the beam spot. The high rate of the beam does not allow readout of single hits, but in-pixel particle counting is necessary. The read out frames are then used to calculate beam position and size in real time (projection). In addition, the shape can be read out periodically. The dose is calculated from the total number of hits and the known particle energy.

At the same time, possible disadvantages of HV-CMOS sensors have to be investigated: The increased material compared to wire chambers has to be below the given limit of 2 mm water equivalent (wire chamber: 0.35 mm water equivalent). It has to be radiation tolerant enough to survive five to ten years of operation without the need for direct cooling.

A single layer of HV-CMOS sensors can fulfill the duty of at least three wire chambers: x-direction, y-direction and dose chamber. A secondary layer might be necessary for fail-save operation.

Measurements with a high rate of X-rays presented in this thesis suggest, that the leakage current could serve as additional observable for the total dose. This measurement is independent of the pixel measurement and therefore can serve as fail-save parameter, too, rendering a second layer expandable. This novel approach is to be tested.

An influence of a magnetic field on HV-CMOS sensors is not known. Therefore, CT-usage is not expected to affect beam monitoring by a HV-CMOS detector.

17.1.3 Secondary particle monitoring

Irradiation of a body with heavy particles causes a shower of secondary particles. This shower can be used to validate the location of interaction between particle beam and tissue, and therefore the precision of the treatment.

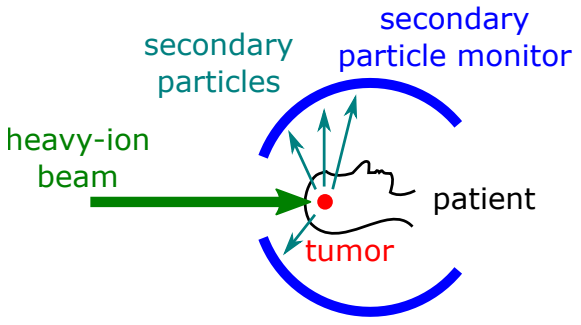


Figure 17.3: Additionally to the monitoring of the therapeutic beam, monitoring of the secondary particles generated by interaction of the primary beam with the patience's tissue is a valuable source of information for further improvement of treatment.

For a good measurement result, the sensors have to be placed around the irradiated volume (figure 17.3), high time resolution is necessary, tracking is favorable.

17.2 Monitoring an heavy-ion beam by HV-CMOS sensors

The monitoring of a beam is in two ways different from monitoring secondary particles of high-energy physics experiments:
A focused beam brings a much larger hit rate (up to 20 GHz in an area of ≈ 0.5 cm^2) than secondary particles in HEP experiments (≈ 50 MHz per 0.5 cm^2). The available sensors were not designed for this rate. Therefore, they reach overload, as soon as the particle beam hits the sensor. In this state a priority logic determines which pixel-hits are transmitted and which have to be discarded: Each readout cycle transmits one hit per column. In each column, the lowest row with a hit is selected.

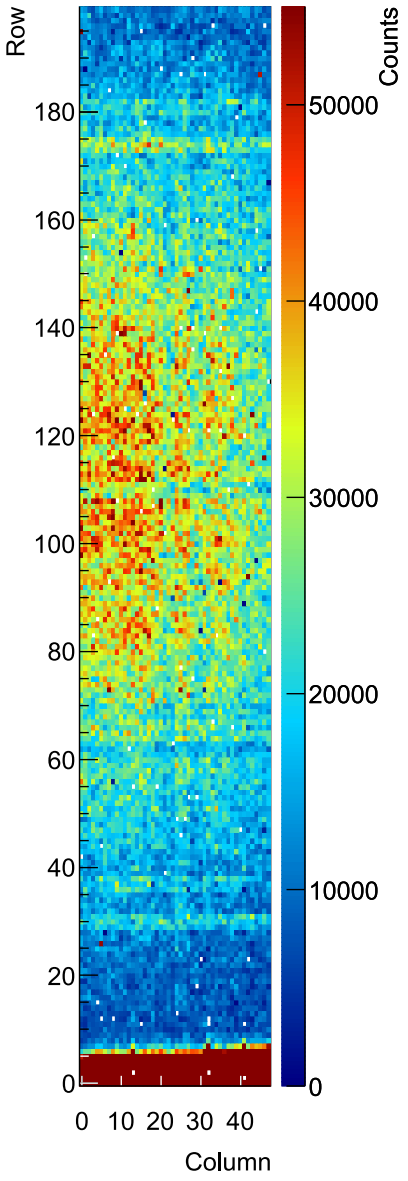


Figure 17.4: Heavy-ion beam spot recorded by MuPix8. Due to overload, only the lowest rows are read out with high efficiency (dark red area).

The signal of protons in the therapeutic energy range (48 - 220 MeV) is comparable to those of particles in HEP experiments. However, heavier particles, such as carbon ions, might act differently. This is to be investigated before HV-CMOS sensor can be used as an heavy-ion beam monitor.

On the first beam test session a MuPix8 single-chip setup (see Appendix F) and the beam telescope (see chapter 16.3) equipped with three ATLASpix1 layers have been used. Both setups were operated at the same time behind each other. The beam energy is large enough to be not overly affected by several sub-millimeter silicon sensors in its path.

The beam can be user configured to send either protons or carbon ions in spills of 10 seconds length with a 50% duty cycle. The energy of the particles, their intensity and their focus is determined by integer numbers, which refer to a particle energy per nucleon, a particle rate per second and a setting for the focal magnets. In expectation of an extremely high rate, only the lowest intensity setting was chosen.

The beam spot recorded by MuPix8 is shown in figure 17.4. It is the picture of the carbon ion beam with a focus setting of 3, which means a full width half maximum (FWHM) diameter of 11.1 mm.

Even though the intensity was set to minimum, the sensor does not have the rate capability necessary for this event rate: The color scale was cut to be able to see the beam spot. The vast majority of the read out hits originate from the lower most 5 rows. This is the area read out with highest priority in the column drain readout approach implemented on MuPix8. Hits in other rows are only read out in the short period of time at the beginning of a spill before overload is reached. It can be deduced that the actual rate is at least forty fold the capability of the used readout architecture.

Cluster analysis was inconclusive as most clusters were not read out entirely due to sensor overload. Most complete clusters were found close to the lower edge, far away of the beam center.

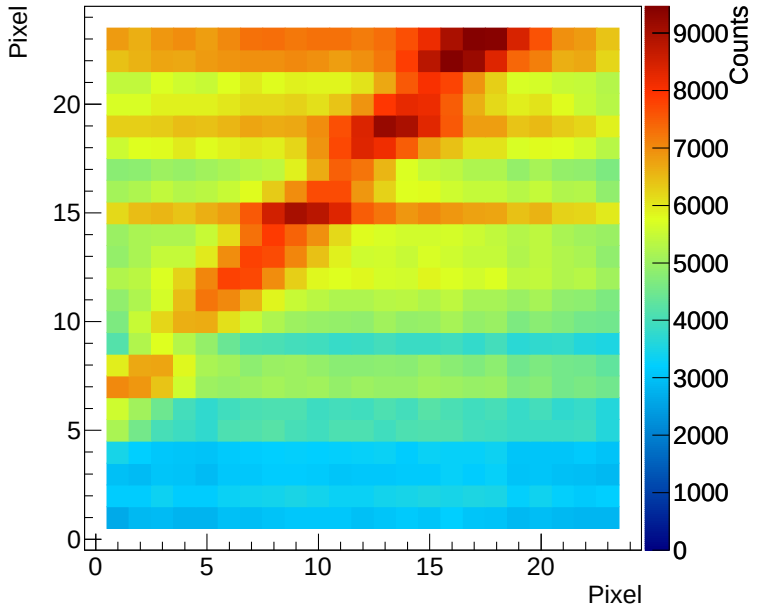


Figure 17.5: Correlation plot of two layers of the ATLASpix1 telescope: X-position of layer A over X-position of layer B. The small size is a result of the limited number of columns.

The data of the ATLASpix1 telescope show the same signs of overload, just as MuPix8, as they share the readout strategy. However, a basic correlation of the telescope layers

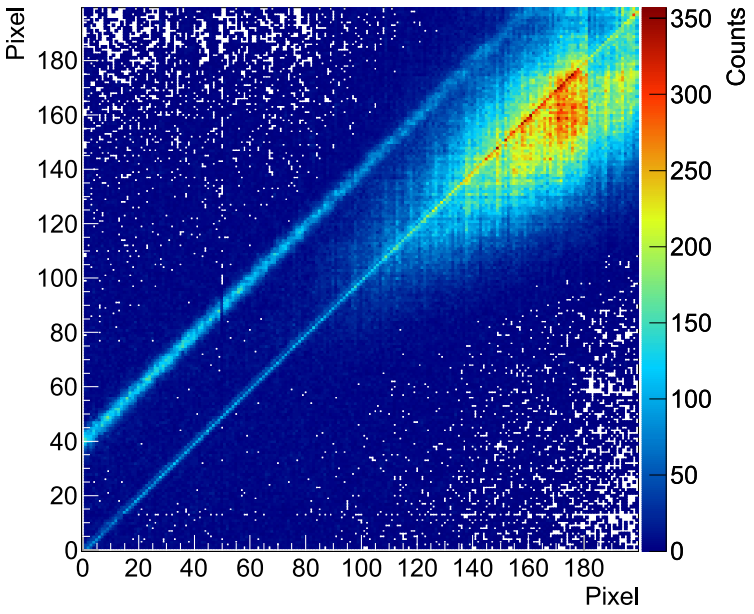


Figure 17.6: Y-correlation of two ATLASpix1 telescope layers. The correlation line (left) is disturbed by overload. The second line and the spot are results of the extreme overload.

was possible, which proves the conceptual integrity of the telescope design. Figures 17.5 and 17.6 show x- and y-correlations of the recorded data of two layers. Besides the typical correlation line starting at (0|40), two other features are visible: an angle bisecting line and an area with many entries. Both are an effect of overload. The layers run synchronously, therefore in overload, the pixels of every layer are read out in the identical order, thus the correlation line appears. The spot originates from starting and ending overload at the begin and end of a spill.

17.2.1 Beam monitoring at reduced particle rate

In order to investigate the interaction of a therapeutic beam with HV-CMOS sensors, the rate has to be reduced to a level that can be handled by the available architecture.

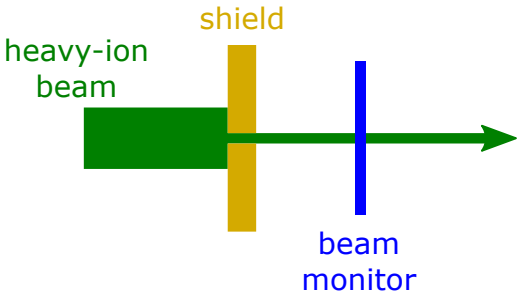


Figure 17.7: In order to reduce the particle rate to a level that can be handled by the available sensors, the beam is widened and a shield with a small hole blocks most of the beam. The shield can only block particles up to a certain energy.

Therefore, an at least 1 cm thick brass shield with adjustable opening has been designed (Figure 17.7). It is capable of stopping protons of up to 80 MeV and carbon ions up to 150 MeV/u [1].

The beam was operated at maximum focus setting, meaning largest possible beam spot, thus distributing the rate over a large area. The opening is only about 1 mm^2 large, allowing only a small fraction of the widened beam to pass. With this modification single particle properties can be measured. The following measurements with shield have been done on MuPix8.

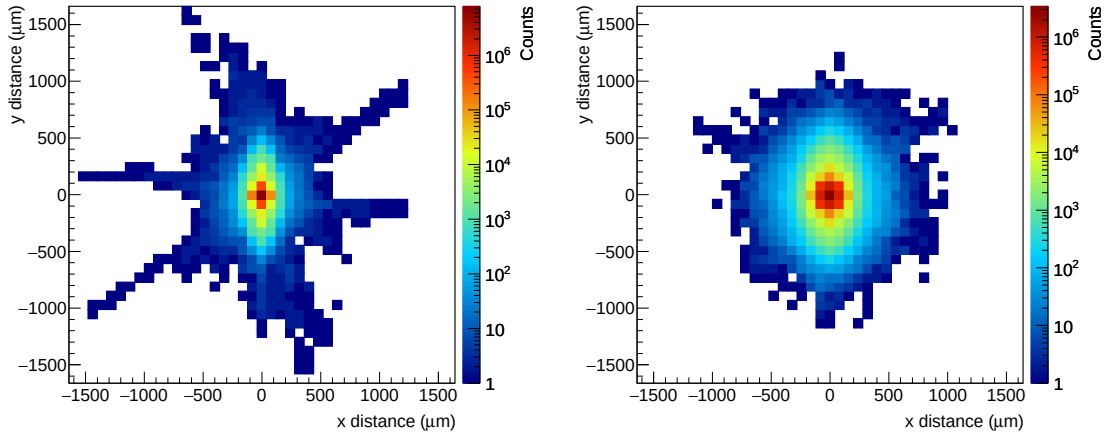


Figure 17.8: Cluster shape of minimum energy protons (a) and carbon ions (b).

The collected data have been clustered (For the used algorithm see chapter 15.3.6.) and figure 17.8 shows the shape of clusters as superposition of the proton beam (a) and the carbon ion beam (b) with lowest energy setting. The overall round shape indicates, what has been already observed during the measurement: The rate has been reduced to a level, MuPix8 can handle. Beginning rate overload is seen in cluster shape plots as half-moon with opening facing top-wards. As described before, the readout logic prioritizes pixels in lower rows but treats all columns equal. A sensor in overload will therefore transmit a half-moon-shaped subset of a large round cluster.

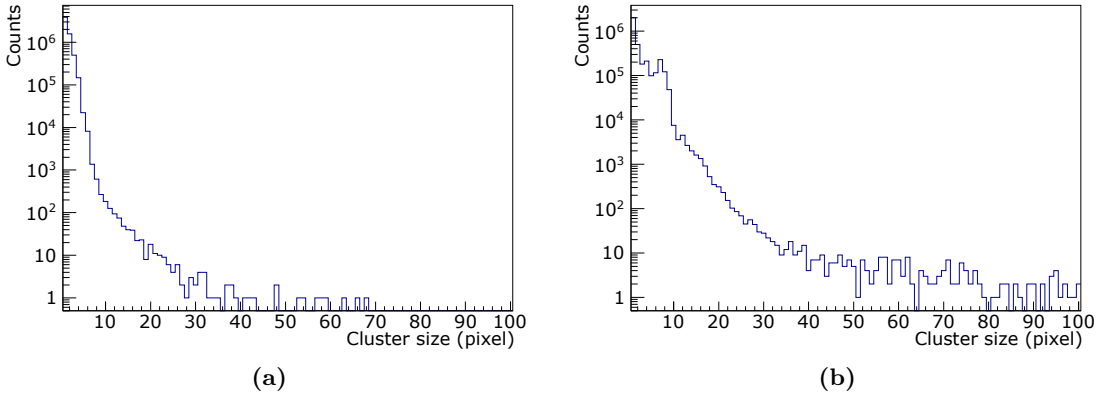


Figure 17.9: Cluster size histogram for minimum energy protons (a) and carbon ions (b).

[1] Geant simulated values by A. Nürnberg.

The size distribution of reconstructed clusters is shown in figure 17.9. What could be guessed from the cluster shape, appears clearly in the cluster size histograms: the clusters of the carbon ions are significantly larger than clusters by protons. Some clusters have around 100 member pixels, which means a cluster area of 0.6 mm^2. Proton clusters appear hardly above 40 pixels.

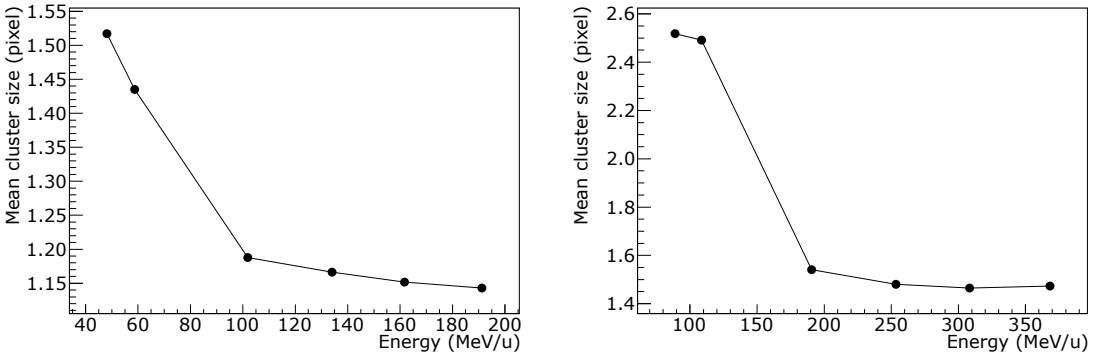


Figure 17.10: The mean cluster size as a function of particle energy for (a) proton beam and (b) carbon ion beam. At high particle energies, the shielding fails.

With increasing particle energy, the mean cluster size shrinks slightly from 1.52 to 1.43 (protons) or 2.52 to 2.49 (carbon ions), as depicted in figure 17.10. Above 80 MeV/u (protons) or 150 MeV/u (carbon ions), the shielding fails and the exploding rate allows no more read out of entire clusters. Consequently the mean cluster size drops.

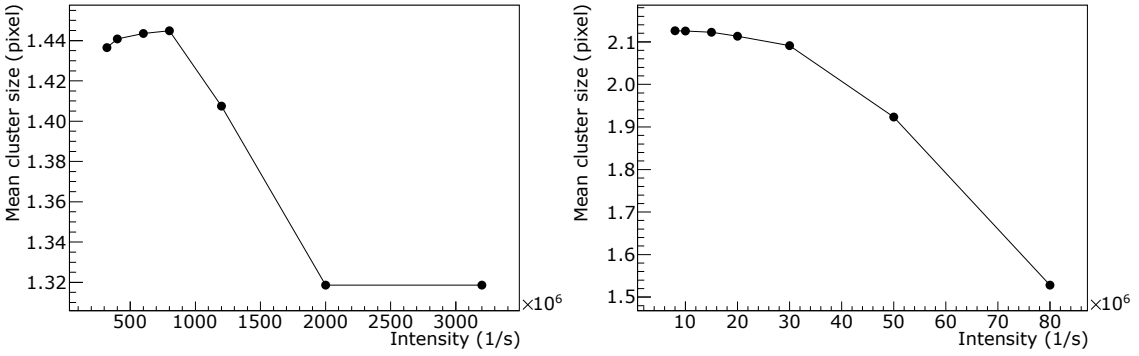


Figure 17.11: Mean cluster size over beam intensity for proton beam (a) and carbon ion beam (b).

Besides particle energy, the beam intensity is another configurable parameter. It modifies the particle rate in the range relevant for therapy. A priori, the intensity has no influence on cluster formation. But since time resolution is limited, increasing rate increases at the same time the probability of overlapping clusters in the same time frame, thus faking larger clusters. This effect can be observed in figure 17.11a (protons) for the lowest four data points. Once the intensity gets too large, the cluster size drops rapidly, as clusters are not read out totally anymore.

This effect appears not in the plot for carbon ions in figure 17.11b (carbon ions). Even with the larger clusters by carbon ions, the particle rate is too low for a significant chance for merging clusters. The recorded mean cluster size is first constant and then drops with increasing beam intensity as more and more clusters can not be totally read out anymore.

17.2.2 Depletion voltage of HV-CMOS sensor for beam monitoring

Beam monitoring of a heavy-ion beam is a special use case for a HV-CMOS detector. While in high-energy physics experiments, the particle energy is not precisely known and their distribution ranges from minimum ionizing particles to extreme signals, the energy of particles in a beam is known and precisely filtered before the particles reach the beam monitor. In HEP, high voltage is required to enlarge the depletion zone of the sensing diode to increase the signal of MIPs to a significant level with good signal-to-noise ratio. But MIPs do not occur in beam monitoring and a thick depletion zone loses importance.

A moderate voltage is expected to be sufficient for high efficiency and SNR. At the same time is the leakage current of high irradiated sensors an issue still under investigation. At the time of writing, sensors are being irradiated by the actual beam. Already after 25% of the mean annual fluence, the leakage current is increased by two orders of magnitude (extrapolated from beam spot to sensor area). The sensitivity can be maintained by adjusting the settings, but a high depletion voltage on a fully irradiated sensor might cause a large leakage current with additional noise and even thermal runaway. Therefore, the depletion voltage has to be limited and sensor cooling by forced air flow should be considered.

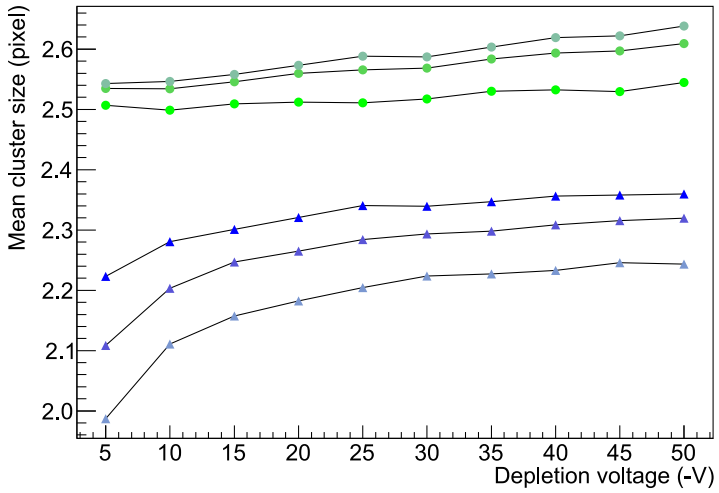


Figure 17.12: Mean cluster size as a function of depletion voltage of proton beam (blue) and carbon ion beam (green). Darker color means higher particle energy.

Figure 17.12 shows mean cluster size as a function of depletion voltage for proton beam (blue) and carbon ion beam (green). The marker color indicates the beam energy: darker means higher energy. The cluster size of protons grows about 10% from lowest to highest depletion voltage, while the cluster size of carbon ions changes only about 3%, but on a higher level. The largest change in cluster size happens between 0 V and 15 V, therefore 15 V is a good compromise between charge collection efficiency and leakage current.

The mean cluster size from proton beam shrinks with increasing energy. This is in agreement with energy deposition in the sensor material after Bethe formula. The carbon ion beam shows a rather constant cluster size over energy, but with slightly larger clusters for larger energies. This is explained by the focal behavior of the beam: If the beam energy is increased from 108 MeV/u over 128 MeV/u to 145 MeV/u, the beam spot area shrinks from 57 mm^2 over 44 mm^2 to 35 mm^2. Keeping in mind that the opening size of the brass shielding remains constant, the particle rate seen by the sensor increases with increasing particle energy. As seen before, higher rate leads to a larger mean cluster size due to increased chance of merging clusters.

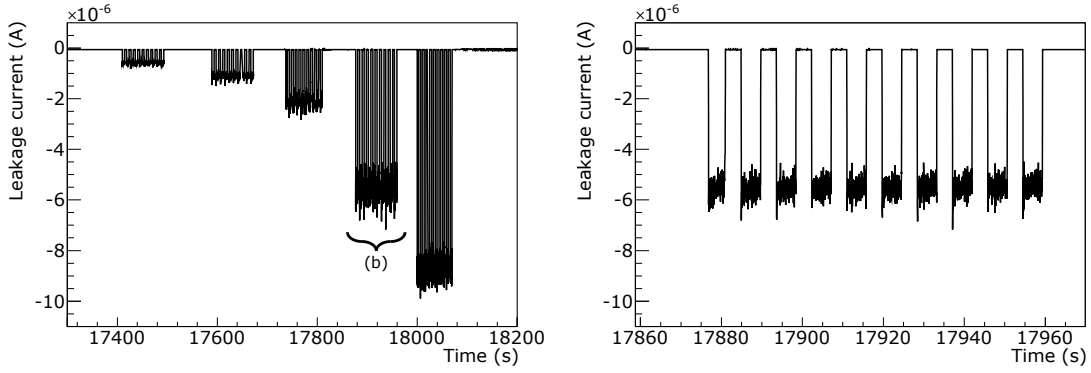
This effect does not occur on proton beam, as the particle rate per area is smaller and the shield opening stays the same, thus the proton rate in the sensor is smaller than the carbon ion rate with similar settings.

17.2.3 The sensor diode current observable

The leakage current of the sensor diode is periodically measured by a smart measurement unit (SMU). In equilibrium, it is dominated by thermal excitation.

Traversing particles leave a trace of free electrons and holes, which are dragged by the electric field to the respective electrode. This process is equivalent to a short and small current, which arises additionally to the leakage current. This signal current is in the order of 1000 e for a duration of a few nanoseconds, which is hardly measurable. In case of beam monitoring however, not only a few particles hit the sensor diodes, but more than 10^6/s in case of carbon ions and 10^8/s in case of protons, which sums up to a significant additional signal current. Its value depends on the particle type, energy and number. All these parameters can be varied in a therapeutic beam, but are well defined during an irradiation session. Consequently, the diode current can be used as safety or backup measure for the beam. A backup system has to be fully independent from the other measurement, this is the case for diode current.

Additionally, it can be used to identify the detector's area of interest. A beam monitor has to cover an area of approximately 25×25 cm^2 with about 150 individual dies. The combined data rate of those sensors is significant and central handling might turn out to be challenging. The sensor current can be used to narrow down the area hit by the beam and trigger the readout.



(a) **(b)**

Figure 17.13: Sensor diode current measured while five irradiation steps are applied (a). In this case each irradiation step consists of ten spills (b).

Figure 17.13 shows the current of a MuPix8 sensor illuminated by the carbon ion beam at 368 MeV/u. The shielding has been removed. The upper line shows the leakage current without beam, each pillar in figure 17.13a was measured with a different intensity setting, ranging from $5 \cdot 10^6$/s to $8 \cdot 10^7$/s. Figure 17.13b shows a closeup of the fourth irradiation step. The individual spills of the step are clearly visible as well as the variation during one spill.

Times with and without beam hitting the sensor can be clearly distinguished with an aspect ratio better than the measurement repetition ratio of the used SMU (< 25 ms). Fast response time is the most important requirement for exploiting the sensor current observable, to be implemented as backup system, detecting errors and triggering safety interlock.

Even in the smallest recorded current change measured on the particle/energy/intensity combination, $2{\cdot}10^8$ protons per second at 58 MeV, has a four times higher sensor current with beam compared to idle state. This value might be increased for a reticle size sensor with latest substrate and guard-ring design[2].

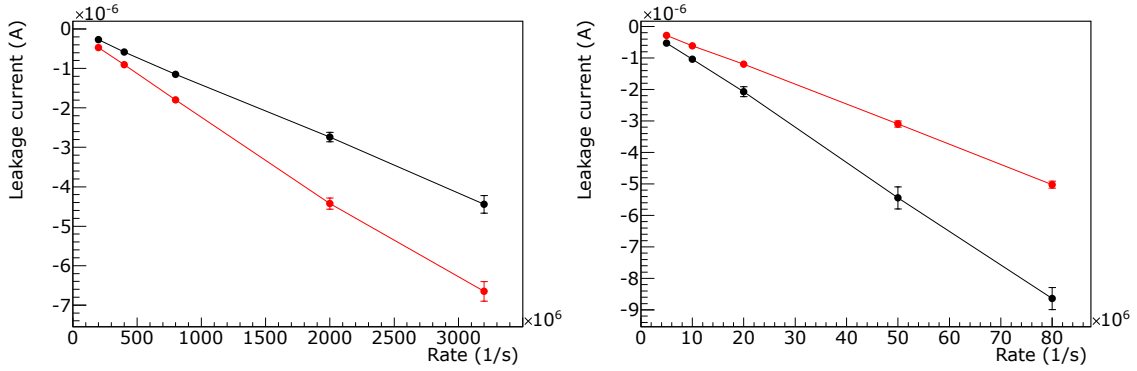


Figure 17.14: Sensor current over beam intensity for protons (a) and carbon ions (b). Red refers to high energetic particles, black to low energy setting.

The leakage current has not only been measured with one setting, but data have been collected during proton and carbon ion beams with two energy settings each, chosen at the lower and upper edge of the therapeutic range and five intensity settings. The result is shown in figure 17.14, proton beam left and carbon ion beam right. The black line represents the low energy setting (p: 59 MeV, C: 108 MeV/u) and the red line the high-energy setting (p: 191 MeV, C: 368 MeV/u).

Even though the signal current of both particle types is of the same order, the energy behavior is inverted: Higher proton beam energy leads to a higher leakage current, while a higher carbon ion beam energy leads to a lower current. This is not expected as in the used energy range higher particle energy should lead to less interaction with the sensor material (see chapter 2.3.2). The most probable explanation is the energy-dependent beam spot size: at higher energies, the beam spot shrinks at constant particle rate. As the width of the sensor is smaller than the beam, imperfect alignment will increase the number of hits on the sensor.

17.3 Beam monitoring summary

Thinned HV-CMOS sensors can be used for heavy-ion beam monitoring. The clusters of protons and carbon ions are of significant size, up to 100 pixels or 0.6 mm^2, but the mean cluster size is not larger than 3 pixels or 0.02 mm^2. However, the particle rate is too high for current designs. For the required beam position resolution of better than 100 µm, a pixel size of 200×200 µm^2 is sufficient (better than 60 µm), as the comparably large beam hits many pixels and fitting is possible. These larger pixels significantly reduce the load on the readout link and provide more space for electronics inside each pixel. This additional space has to be used for hit counting or charge integration.

The highest available particle rate setting is $2{\cdot}10^{10}$/s. Each particle causes an average of 2 hits in the beam monitor, therefore the readout has to cope with a rate of $4{\cdot}10^{10}$/s on a single detector chip.

The beam's center hits at least 8000 pixels (FWHM). Given the request for a 100 µs repetition rate, one frame has up to $4{\cdot}10^6$ hits. This means, all 8000+ pixels need to be

[2] ATLASpix3 has in idle an at least 100 times smaller leakage current at full reticle size.

read out every frame, once in 100 µs. This is hardly possible with a reasonable number of readout links (max. 4000 hits/100 µs per link possible), as there are around 150 sensors necessary to form a beam monitor. Consequently, even counting or integrating brings not the required data reduction for therapeutic real time beam monitoring. As mentioned before, projection is an option. Reading out only the sum of events in each column and each row (100 per direction and die) will reduce the rate from >8000 per frame to 200 per frame. This is a number easy to handle. For scientific beam studies and maintenance, the frame time can be relaxed and pixel-wise readout is possible to study the beam shape.

A therapeutic beam monitor has to be fail safe. This can be ensured by a backup system, which detects deviations and errors of the beam, triggering an interlock. This backup system has to be independent from the primary beam monitor. HV-CMOS brings this options for free. The principle of beam diagnostic by sensor diode current tracking has been proven possible. However, in-depth studies are necessary before it can be used for medical application.

Towards a medical applicable beam monitor, radiation hardness of HV-CMOS sensors irradiated by a focused heavy-ion beam has to be proven (under investigation during the writing of this thesis with promising results), test chips with hit counting or charge integration have to be deployed (already designed and in commissioning) and last but not least, a detector system of several sensors has to demonstrate real time beam monitoring before the system is scaled to the final 25×25 cm^2 heavy-ion beam monitor.

18. ATLASpix1 and MuPix8 summary

ATLASpix1 and MuPix8 are the first large-scale HV-CMOS sensors designed at KIT ADL in *ams* aH18 technology. Their advanced, fully monolithic readout, in combination with a proven front-end, makes them a more complete detector than earlier designs. Even though each design aims for a different experiment, they share many design features and can be seen as technology platforms for charged particle detection in general.

- The sensors were produced on diverse substrates of different thickness (100 to 700 µm) and resistivity (80 to 200 Ωcm).
- The front-end is a large-diode design with charge sensitive amplifier implemented in the collection electrode. The PMOS transistors of ATLASpix1_Iso matrix are protected by an addition deep p-well against cross-talk intermediated by the substrate.
- The analog signal is either digitized in pixel (ATLASpix1) or transmitted to the comparator in the periphery by voltage or current line drivers (MuPix8).
- The 180 nm technology allows for smaller pixels compared to H35Demo. The minimum pixel size is limited by the density of connections from pixel to periphery, transmitting hit signals. ATLASpix1_M2's pixel-grouping is a possibility to reduce the necessary number of connections.
- Hit data are aggregated in the periphery (end-of-column blocks and readout blocks). They contain information on hit location, hit time and analog information (time-over-threshold or secondary time stamp). The readout is event driven with zero suppression for maximal performance.

During characterization, both novelties and approved features were investigated. Characterization was conducted on the novel modular Multi-purpose Adapter Board System.

The most important characterization results are:

- The front-end shows excellent performance. Noise floor is less than $60\,e^-$ at the amplifier and less than $80\,e^-$ after digitization. The signal-to-noise ratio is – depending on settings – up to 78.
- Analog information can be maintained after digitization. Adaptive sampling is an effective tool to measure event time stamp and secondary time stamp at individual frequencies, perfectly adapted to the energy range of a specific experiment or measurement.
- The large signal line length differences of a 2 cm long pixel matrix introduce timing and signal variations. Pixel sensitivity can be homogenized by adjusting the global threshold locally by tune bits. After tuning remains a variation of detection threshold of $51\,e^-$. The time uncertainty is in the order of 90 ns and can be compensated offline by a look up table.

- The time-walk effect by digitization is inevitable, it is in the order of 60 ns. Calibrated analog information can be exploited to calculate corrections for the time-walk.
  After time-walk correction and matrix-effect correction applied, remains a time uncertainty of 13 ns for signals close to minimum ionizing particle energy. This is the most pessimistic view, as most particles in high-energy physics experiments cause much larger signals with virtually no time-walk.
- Samples have been irradiated by proton cyclotron to $6 \cdot 10^{14}\,\mathrm{n_{eq}/cm^2}$ and $10^{15}\,\mathrm{n_{eq}/cm^2}$. Even after the highest fluence and dose, the sensors are functional, but with increased noise level. The tuned chips have a noise of $109\,\mathrm{e^-}$, while the signal of a minimum ionizing particle in the pixel with weakest response is still $> 2000\,\mathrm{e^-}$.
- Pixels have no area of reduced detection efficiency, independent of substrate and radiation exposure (up to $10^{15}\,\mathrm{n_{eq}/cm^2}$). Prior to irradiation, a very small depletion voltage is sufficient to reach full sensitivity. After a sensor has been irradiated, a moderate depletion voltage is necessary to reach full sensitivity.
- Sensor cooling has an effect, not only on noise, but also on the signal of a sensor. Especially irradiated sensors benefit in terms of shot noise from the reduced leakage current of a cooled bulk. The temperature dependence of the amplifier response can be explained by temperature-dependent conductivity of transistors. The temperature for optimal signal-to-noise ratio is $10\,^\circ\mathrm{C}$.
- HV-CMOS sensors are suitable for beam monitoring, both in usage as beam telescope for moderate rates (single particle tracking) and in beam monitoring of high rate heavy-ion beams (beam tracking). A beam telescope based on the plug-and-play characterization system GECCO, developed during this thesis, has been proven fully functional in beam test.
  A high rate heavy-ion beam was monitored by a MuPix8 sensor. The rate had to be reduced by selective shielding. The mean cluster size of high energetic ions was – depending on depletion voltage and particle type – determined to be less than 2.5 ($\hat{=}\,16\,000\ \mu\mathrm{m^2}$). This information is required to calculate the occupancy of a detector from the particle rate. Future designs, aiming for application in therapeutic beam monitoring, have to either integrate the deposited charge or count the signals in large pixels, to comply with the high rate.
  For fail-safe operation, independent measurements are necessary, e.g. duplication of the beam monitor. However, two redundant detectors double the material budget and are therefore not favorable. The author proposes to use the leakage current of the sensor diode as observable, independent of the rest of the sensor system, instead. Besides cross-checking the primary dose measurement, this allows for coarse grain beam monitoring, to identify regions of interest. This offers the possibility to only switch on the readout of a few sensors, instead of all approximately 150 sensors of the entire beam monitor, to reduce data load.

ATLASpix1 and MuPix8 are very successful monolithic active pixel sensors, close to a possible final design for application in high-energy physics. The characterization results indicate that HV-CMOS sensors can fulfill the requirements of upcoming particle trackers. The next generation of large-scale, versatile HV-CMOS sensors, ATLASpix3 and MuPix10, developed at KIT ADL, will built on the basis laid out by ATLASpix1, MuPix8 and their characterization. Furthermore, the development of hit counting and charge integrating sensors was initiated by the success of high-rate heavy-ion beam monitoring tests.

Part IV

KIT TRISTAN integrated circuit for readout of silicon drift diodes

The KArlsruhe TRItium Neutrino experiment (KATRIN) is an experiment located at Campus North of Karlsruhe Institute of Technology. It has been designed to determine the mass of neutrinos by precision measurement of the products of tritium decay. Data taking has started and is still ongoing during writing of this thesis. The TRitium Investigation on STerile (A) Neutrinos (TRISTAN) upgrade will enable the KATRIN experiment to search not only for the mass of the electron neutrino, but also for the mass of a yet undiscovered sterile neutrino. KIT ADL has designed the KIT TRISTAN Integrated Circuit (KIT TRISTAN IC), an ASIC to amplify and read out charge signals from Silicon Drift Diodes (SDDs), which are going to be the main sensing element of TRISTAN. For this thesis, the combination of an SDD (Figure 18.1) with the readout ASIC has been investigated.

Chapter 19 introduces the KATRIN experiment in its current version and the changes required for TRISTAN upgrade.

The used SDD from Fondazione Bruno Kessler (FBK) and the KIT TRISTAN IC are discussed in chapter 20. For characterization they have been combined in a measurement setup and were connected to an FPGA.

The characterization results are presented in chapter 21, comprising comparative ^{55}Fe spectra, temperature dependency and X-ray tube measurements.

Conclusively, the results are summed up in chapter 22.

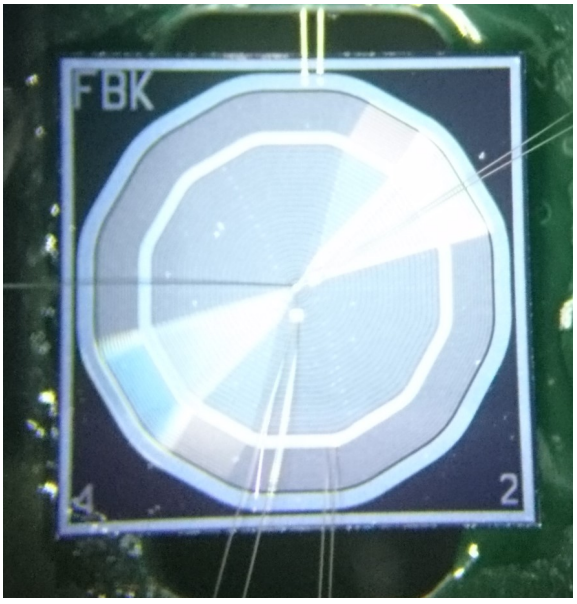


Figure 18.1: Photograph of the silicon drift diode.

19. The KATRIN experiment

The KArlsruhe TRItium Neutrino experiment (KATRIN) has been designed to determine the neutrino mass. Previous experiments were able to prove that neutrinos actually do have a mass by measuring neutrino oscillations [89, 90, 91], while the Standard Model of Particle Physics had originally assumed the neutrinos massless. The exact value of the neutrino's mass however is still unknown. An upper limit has been set to the mass of the electron antineutrino by previous experiments in Troitsk ($m_\nu < 2.5\ eV/c^2$ at 95% C.L.) [110] and Mainz ($m_\nu < 2.8\ eV/c^2$ at 95% C.L.) [111]. The KATRIN collaboration has designed an experiment to improve this limit by over an order of magnitude to $m_\nu < 0.2\ eV/c^2$ at 90% C.L. or to find an actual value for the neutrino mass [112].

19.1 Physics goal of KATRIN – weighing neutrinos

KATRIN investigates the β-decay of tritium to helium, emitting an electron and electron antineutrino, to measure the rest mass of the neutrino.

$$^3\mathrm{H} \rightarrow\ ^3\mathrm{He}^+ + e^- + \bar{\nu_e} \tag{19.1}$$

The decay energy E_0 = 18.6 keV is statistically distributed to the decay products. Tritium is chosen as it and the decay produced Helium-ion are fairly simple elements, which makes calculation of corrections to the decay energy simple. Inelastic scattering of the β-rays with nuclei in the source are rare. The decay itself is a clean, well understood process. Finally, the decay energy is low compared to other β-sources, which is beneficial for the design of a electron high-pass filter.

19.1.1 Working principle

The neutrino's mass is determined by measuring the energy of the electron E_e. The endpoint energy E_0 is known, therefore the missing energy is appointed to the neutrino.

$$E_\nu = E_0 - E_e \tag{19.2}$$

The neutrino's energy E_ν is composed of rest mass and kinetic energy.

Due to the statistical distribution of the initial energy to the outgoing particles, an energy spectrum of the electrons is expected as shown in figure 19.1a [113]. Measuring electrons close to the end point energy of 18.6 keV means that merely the energy for the rest mass has been given to the neutrino. Therefore the very end of the spectrum is of interest (figure 19.1b). The difference between the end point energy and the spectrum's cut-off is the minimum energy in the decay, that has to be appointed to the resting neutrino: its mass.

It is quite challenging to measure only the most high energetic end of an electron spectrum, which consists only of a very small fraction of the full spectrum. Therefore, KATRIN

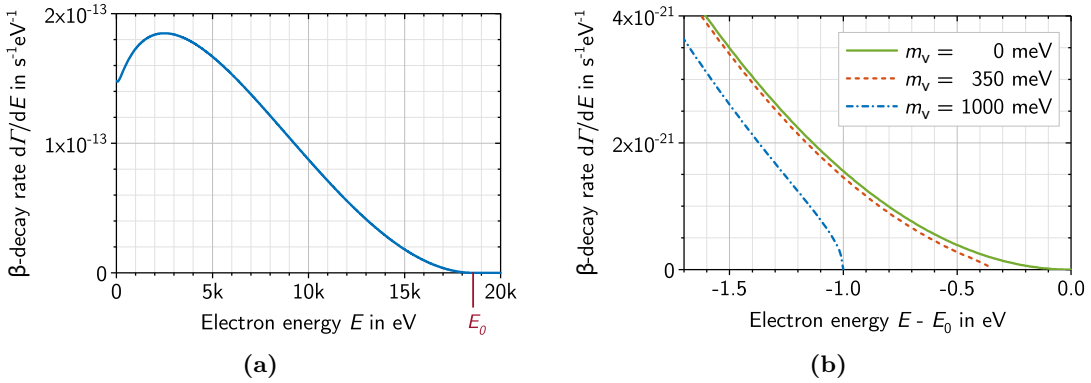


Figure 19.1: Energy spectrum of electrons emitted from tritium decay. Depending on the actual mass of the neutrino, a cut-off lower than the endpoint energy is expected. (From [113])

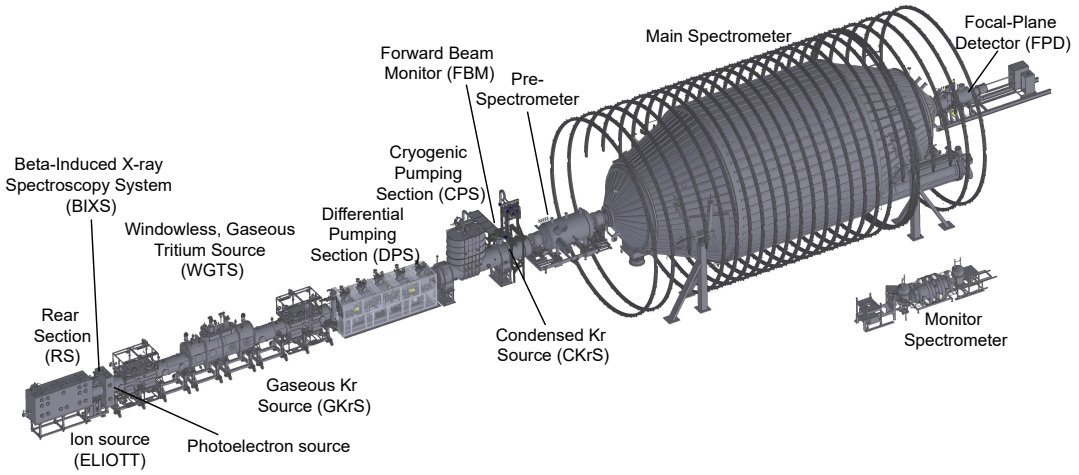


Figure 19.2: Technical drawing of the main components of KATRIN: Tritium source, pumping sections, main spectrometer and detector. (From [114], modified)

is designed as a precision high-pass filter for electrons (Figure 19.2). The tritium gas decays in the windowless, gaseous tritium source (left) and the electrons are detected in the focal-plane (right). The beam line in between is vacuum. Tritium gas escaping the source is being removed from the pipe in the pumping section before it can reach the electron spectrometer. Several secondary sources (ion, electron, krypton) are used to calibrate KATRIN.

The main spectrometer is a MAC-E filter (Magnetic Adiabatic Collimation combined with an Electrostatic Filter). Its working principle is shown in figure 19.3. Electrons (red) coming from the source enter the strong magnetic field (left) and start to circulate around the field lines (blue). Towards the middle of the spectrometer, the magnetic field strength is reduced and the circulating motion is translated in a mostly forward motion (conservation of cyclotron energy). The momentum vector p_e in the magnetic field of an electron is shown in red below the sketch. Now, that all electrons move virtually parallel, an electric field (green) is applied in the same direction, slowing down the electrons. Electrons are running against the potential of the electric field and only those with sufficient energy will make it past the potential barrier. Electrons with insufficient energy are rejected. Towards the detector, the opposite electric field accelerates the electrons to their original speed and the magnetic field is squeezed to the original size.

The energy spectrum of the electrons emitted by tritium is scanned by change of the electric field. All electrons above the set energy will reach the detector, as there is no additional

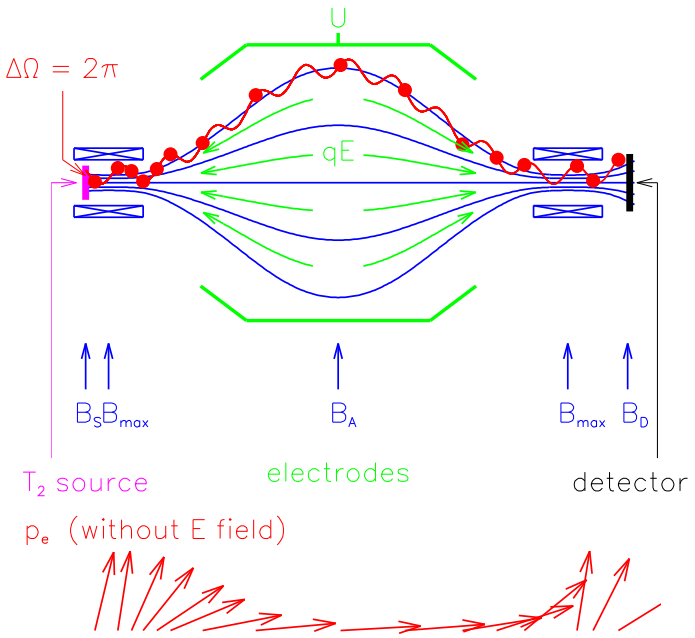


Figure 19.3: The principle of a MAC-E filter is to bend the momentum of electrons into a known plane without changing their energy, by a slowly changing magnetic field and then discriminating the electrons by an electrical field. (From [113])

low-pass filter cutting off higher energies. Consequently, only the highest part of the energy spectrum can actually be measured, due to the exponentially rising rate towards lower energies. One way to reduce the count rate for lower energies is to use additionally the Time Of Flight (TOF) information. From all electrons making it past the potential barrier, the ones arriving last are the ones closest to the set energy. Electrons arriving faster had a higher energy. However, a MAC-E-TOF for KATRIN would have required a fast alternating electrical field 'switching' the tritium source on and off to obtain a time reference.

The sharpness of the high-pass MAC-E filter depends on the difference of the maximum (4.5 T) and the minimum ($3 \cdot 10^{-4}$ T) of the magnetic field [113]:

$$\frac{\Delta E}{E} = \frac{B_A}{B_{max}} \tag{19.3}$$

This follows from the conservation of the magnetic moment μ in the spectrometer:

$$\mu = \frac{E_\perp}{B} = \text{const.} \tag{19.4}$$

19.2 The TRISTAN upgrade

The TRitium Investigation on STerile (A) Neutrinos (TRISTAN) is a proposed upgrade for the KATRIN experiment [115]. As the name indicates, this experiment is looking for sterile neutrinos by expanding the spectrum measurement of tritium decay electrons from the high energetic tail (KATRIN) towards lower energies. Sterile neutrinos are not weakly interacting and have not yet been observed. Their existence is suggested by super-symmetric extensions to the Standard Model. They are very similar to known neutrinos but with right-handed chirality, whilst standard model neutrinos are left-handed. It has further been suggested that sterile neutrinos could be Majorana-fermions, which means they are their own anti-particle. They are assumed to have a mass of a few keV, lighter sterile neutrinos have been excluded by previous experiments.

The existence of a heavier neutrino would modify the tritium decay electron spectrum

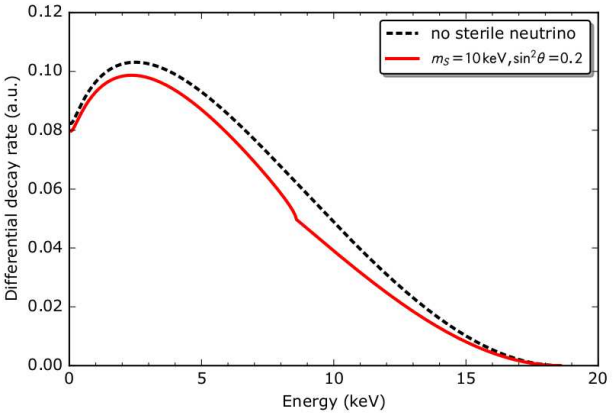


Figure 19.4: Spectrum of tritium decay electrons without sterile neutrinos (black dashed) and with a 10 keV sterile neutrino and an expected strong mixing angle. TRISTAN searches for this dip in the spectrum. (From [116])

similar to the energy cut-off described for electron neutrinos, but in the middle of the spectrum (Figure 19.4).

The search for sterile neutrinos is not only theoretically motivated, but also from experiments. Three neutrinos are known, but neutrino measurements in the vicinity of fission reactors have shown a significant deviation of the neutrinos actually observed from the expected value [117]. This lack of neutrinos indicates the existence of at least a fourth or more yet unknown neutrinos.

19.2.1 Modification of KATRIN

Scanning the low energy regime of the electron spectrum from tritium decay greatly increases the event rate. Therefore, the PIN diode pixel detector of KATRIN has to be replaced by a new detector capable of the expected high event rate of up to 10^{10} per second, with an excellent energy resolution. At the same time it has to maintains a very high signal-to-noise ratio. A good SNR minimizes noise counts, important especially for measuring the high energetic end of the tritium decay spectrum for KATRIN (Figure 19.1). The energy resolution is crucial to scan the full spectrum for TRISTAN in order to find a small step in the spectrum indicating a sterile neutrino (Figure 19.4).

Silicon Drift Detectors (SDD) are suitable for this job, as they do not only allow high rates but also provide good energy information. These silicon detectors have much larger pixels than the HV-CMOS sensors previously discussed.

The principle of an SDD is shown in figure 19.5. Like HV-CMOS, charge is collected by drift. The collection anode is located in the middle of a pixel. The backside contact is used to deplete the bulk and serves as entrance window. This reduces the interaction of incoming particles with insensitive material to a minimum, when illuminated from the backside, as the particles do not have to pass through insensitive volume. The internal field of the sensor is engineered by concentric field rings around the collection anode. This is necessary to reach maximum depletion not only in the center but also on the outskirts of the pixel. Typically, the largest ring is set to high negative potential similar to the potential of the back side contact. The anode is kept at ground potential. The rings in between receive a voltage level between those values by a series of voltage dividers. Like HV-CMOS sensors, the signal has to be amplified, but in this case by an external amplifier.

Full depletion promises high-energy resolution and charge collection by drift brings high

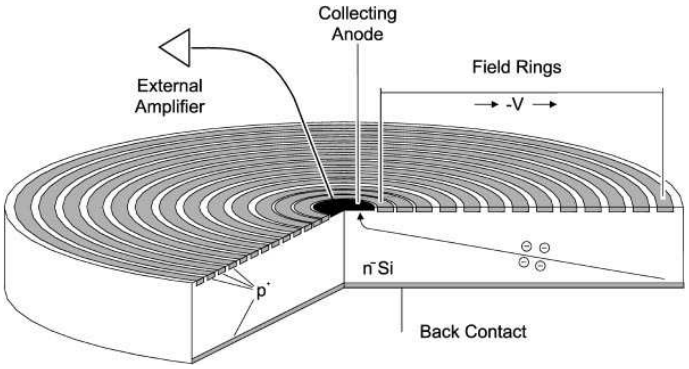


Figure 19.5: Schematic of a Silicon Drift Diode (SDD). The electric field drags the signal charge from the backside to the collection anode. The field is engineered by field rings to collect even electrons from the far edges of the sensor. (From [118])

time resolution. The small collection diode brings a small detector capacitance, which is beneficial for the charge to voltage conversion (see equation 3.2 in chapter 3).

However, SDDs have large pixels (several mm in diameter) limiting the spatial resolution, need external signal amplification and are produced in a non-standard process increasing the costs. For TRISTAN they are an excellent choice as the detector area is small compared to e.g. ATLAS ITk and particles should be stopped completely to measure their energy. Practically arbitrary cooling can be provided from the front side, which further increases energy resolution by limiting thermal noise to a minimum.

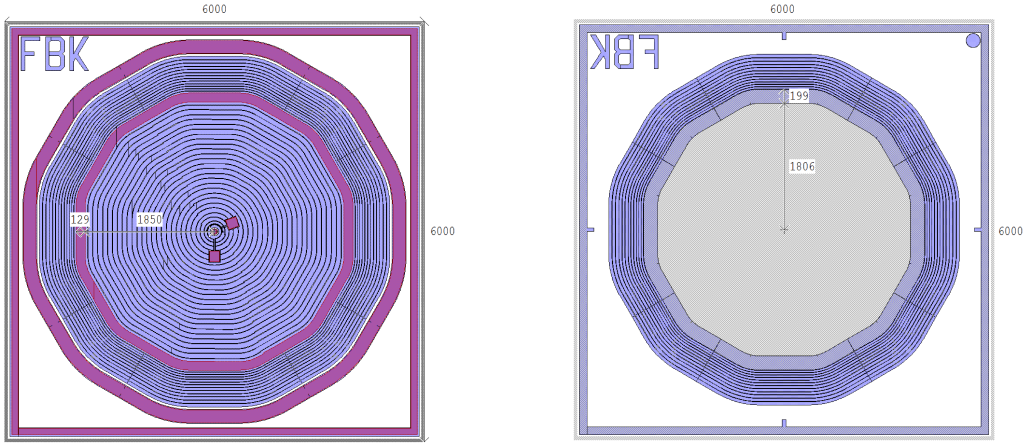
To cover a significant area (several cm), several of such pixels can be combined. Their shape is modified from perfectly round to hexagonal, so no insensitive area is created between the pixels.

20. The KIT TRISTAN integrated circuit assembly

For the TRISTAN upgrade of the KATRIN experiment, Silicon Drift Diodes have been chosen as sensor, for their excellent energy resolution, rate capability, efficiency and signal-to-noise ratio. Their charge signal has to be amplified and processed externally. For this purpose an application-specific circuit has been designed by KIT ADL, the KIT TRISTAN IC.

20.1 FBK silicon drift diode

The used sensor by Fondazione Bruno Kessler (FBK) has been designed with dodecagonal geometry as n-in-n SDD [119]. One diode is 6×6 mm^2 large.



(a) **(b)**

Figure 20.1: The used Silicon Drift Diode from the IRIS group at FBK. On the front side (a) the drift rings are visible between the inner circle and the center. The purple area can be connected by wire bonds. The entrance window is located on the backside (b), depicted in gray. (From [119])

Figure 20.1a shows the front side of the SDD. The blue color indicates the metal layers of the drift rings with passivation. The area with without passivation is drawn in purple, here connection of wirebonds is possible. The anode connection in the middle is to be connected to the readout chip. The two pads next to the center are the connection for the first drift ring and the sink pad, which is kept floating. The inner circle-shaped pad is used to provide the high negative depletion voltage to the external ring. Voltage dividers define the voltage for all drift rings between the first and the external ring. The outer circle-shaped pad connects the substrate. The area between the two rings is used to relax the field.

The backside (Figure 20.1b) of the SDD should face the incoming particles, as the gray depicted entrance window has been designed to be very thin for minimum stopping capability

and maximum sensitivity for very low-energy particles by being the backside electrode. The blue circles are metal guard rings to relax the field. The fragile entrance window can be connected to the depletion voltage in the area, where metal layers and entrance window meet. The area of the entrance window is over 10 mm^2 large.

Typically, the inner most drift ring is biased with -10 V, the outermost drift ring with -60 V and the backside with -80 V. Substrate and anode are at ground level.

20.2 The KIT TRISTAN readout chip

The KIT TRISTAN Integrated Circuit chip has been designed by KIT ADL under the guidance of Prof. Perić and was produced by *ams* AG in the *ams* H35 technology. It is a multi-channel test chip to read out different kinds of sensors. The following readout channels are implemented [120]:

- One channel for passive SDD with positive signal (hole collection)
- Two channels for passive SDD with negative signal (electron collection)
- Two channels for SDD with integrated JFET
- One 9-bit pipeline ADC

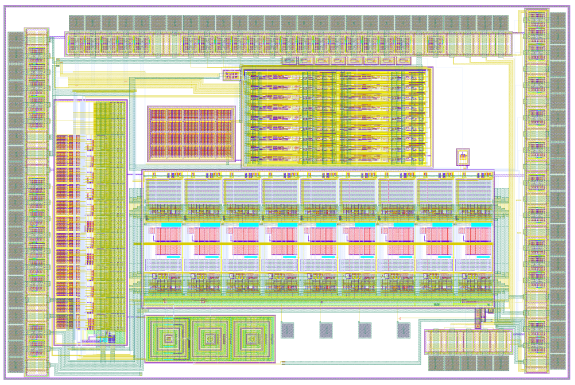


Figure 20.2: Layout of the KIT TRISTAN IC. It comprises readout channels for divers (drift) sensor types.

The layout of the KIT TRISTAN IC is shown in figure 20.2. The SDD used in the following measurement is a passive SDD with negative signal.

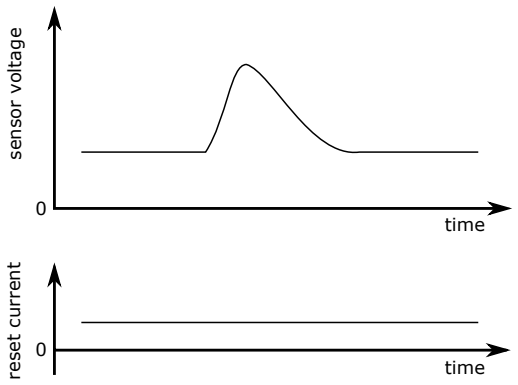


Figure 20.3: In continuous-reset mode, the sensors output voltage is kept at a constant level, the baseline voltage. The continuous-reset current compensates for the sensors leakage current. A signal is a voltage deviation of the baseline.

Sensor reset has to be provided by the readout circuit. Either the sensor can be reset continuously, this approach is simpler and has been used in the previously discussed HV-CMOS sensors (Figure 20.3). In this mode, the sensor voltage is kept at a well-defined baseline voltage, the leakage current is compensated by the reset current. A deviation of the baseline voltage is identified as signal. The height of the voltage pulse is the measure of deposited charge.

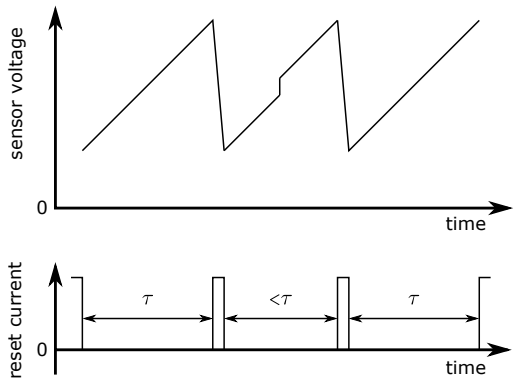


Figure 20.4: In pulsed-reset mode, the leakage current is not compensated. When the sensor output reaches the threshold voltage, a short reset current pulse is triggered to reset the sensor. A signal appears as voltage step on the otherwise steady voltage ramp. The additional charge of a signal decreases the time between two reset pulses.

In case of pulsed-reset, the baseline is no longer constant. The leakage current of the SDD constantly increases the voltage seen by the readout chips input in a slow, but steady slope (Figure 20.4). Once a preset voltage level is reached, a reset pulse is triggered and subtracts a given charge from the input, dropping the voltage level. A signal appears as a step in the slope. A simple comparator is no longer sufficient to identify a signal. Computer aided analysis has to identify the step on the slope. The step height is a measure for the deposited energy.

Alternatively, the time between two resets can be measured, it is a measure for the deposited energy during a cycle. Two hits during one cycle can not be distinguished.

If the signals are large compared to the difference between baseline and reset voltage, more than one reset pulse is needed to reset the voltage level. The number of resets right after each other can be used as a measure for the deposited energy.

The additional analysis effort by using the pulsed-reset option is rewarded by a reduced noise floor. Continuous-reset creates a persistent current flow, which introduces shot noise during the full measurement time. Short reset pulses cause only noise during reset. Charge subtraction time is much smaller (10 ns) than the time the leakage current takes to rise the baseline again.

The readout ASIC has two consecutive amplifiers. For measurements, either the first or the second is monitored by an oscilloscope.

In case of a pulsed-reset measurement, the oscilloscope, a custom comparator board or the internal comparator is used to determine the reset threshold and trigger the reset pulse.

20.3 Characterization setup

The characterization setup for the SDD and KIT TRISTAN IC comprises three parts. The commercial NexysVideo FPGA board from Digilent, an intermediate board for routing and powering, and the carrier PCB, on which next to each other the protagonists of this measurement are located.

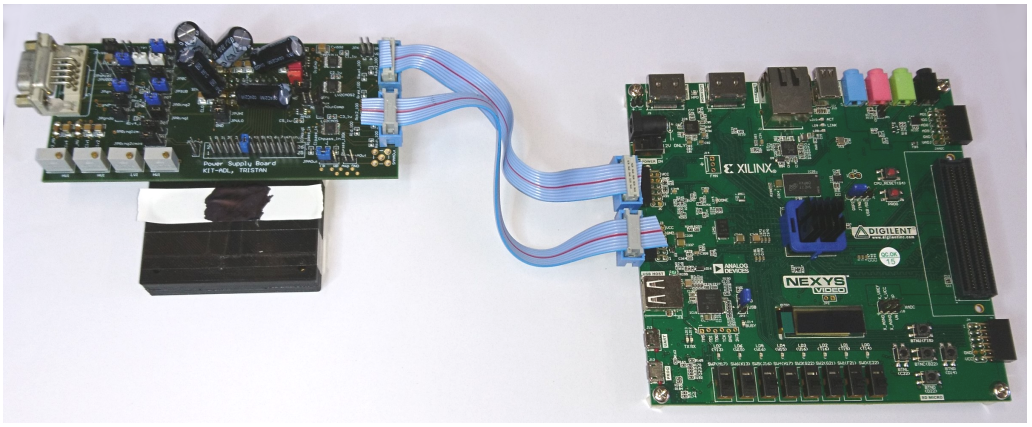


Figure 20.5: Characterization setup for the assembly of SDD and KIT TRISTAN IC. The chips are hidden in the black box.

As shown in figure 20.5, the SDD requires darkness and is therefore hidden in a black plastic box. The SDD is sensitive to visible light due to the very thin entrance window. Figure 20.6 shows the sensor carrier PCB without cover, the sensor in the middle and the KIT TRISTAN IC right below. The pictures shown in figure 20.7 were made through a microscope. The left one shows the top side of the sensor and its bond connections. The supply lines were bonded double for better connectivity. The right picture shows the backside. Through a hole in the PCB the backside entrance window contact was established. For electromagnetic shielding, the setup was placed additionally in a metal box during measurement.

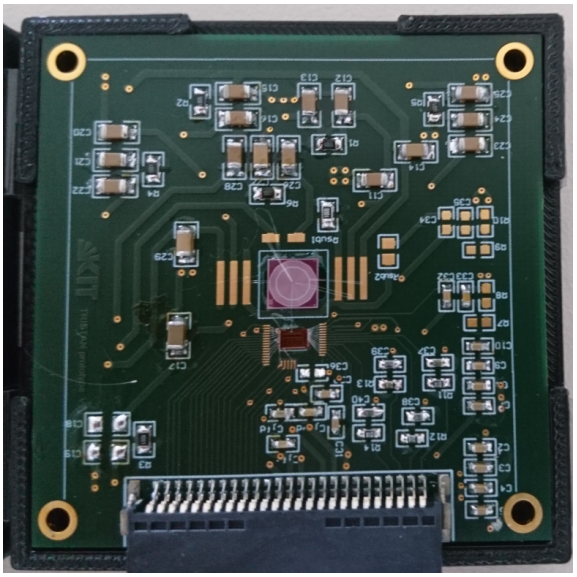
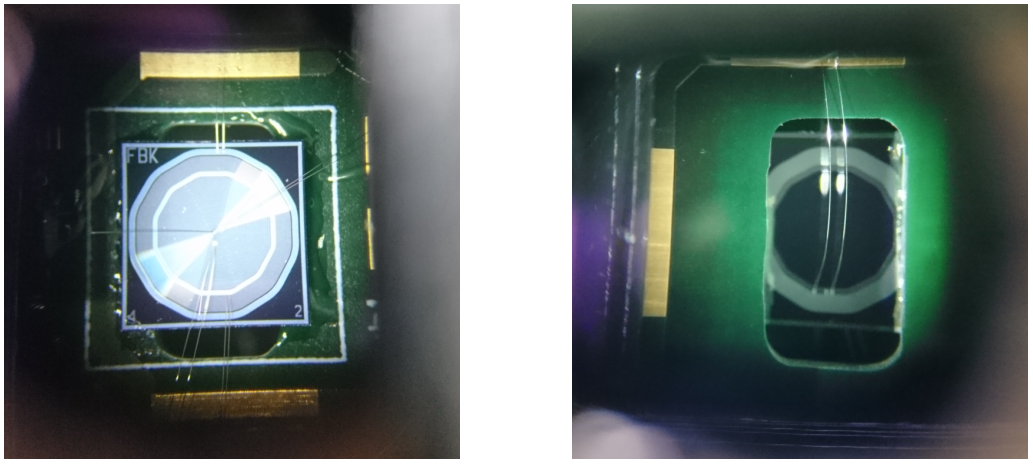


Figure 20.6: Carrier PCB without top cover. The SDD is located in the very middle, the KIT TRISTAN IC right below. The sensor's output is directly connected to the readout ASIC by a wire bond.



(a) **(b)**

Figure 20.7: Picture of the SDD through a microscope, from the front side (a) and backside through a hole in the PCB (b).

21. Characterization of KIT TRISTAN integrated circuit

The KIT TRISTAN IC has been assembled with the FBK SDD and additional hardware to a one channel test detector. It can be operated with continuous-reset or pulsed-reset. Data are taken by an oscilloscope, which is connected to the output of the first or second amplifier stage. Either waveforms or histograms are transferred from oscilloscope to computer and are stored for further analysis.

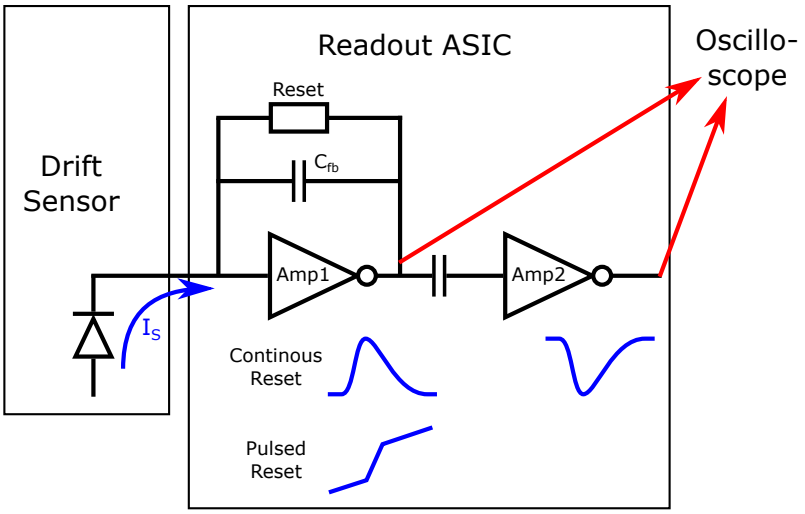


Figure 21.1: Simplified schematic of the setup. The charge signal of the drift diode is converted to a voltage signal in the first amplifier and further magnified by the second one. The reset is either continuously ('resistor') or pulsed ('switch'). An oscilloscope records the output voltage at first or second amplifier output.

The simplified schematic of the setup is shown in figure 21.1. The charge signal is generated in the drift diode and transferred to the readout ASIC. The sensor reset is either pulsed or continuous, the effect on the output of the first amplifier is shown in blue. The second amplifier has a negative pulse, but with constant baseline irrespective of the reset mode, due to capacitive coupling between the amplifier stages.

Two reset modes and two amplifier options result in four operation modes:

- Continuous-reset, signal at first amplifier: positive voltage pulse, constant baseline.
- Continuous-reset, signal at second amplifier: negative voltage pulse, constant baseline.
- Pulsed-reset, signal at first amplifier: slope with step.
- Pulsed-reset, signal at second amplifier: negative pulse (due to capacitive coupling between first and second amplifier).

21.1 Iron-55 measurement

Optimal settings for the measurement setup have been simulated by the chip designers. We know however that these simulations are not perfect. Production variations, the influence of the periphery and the setup as total, can not be simulated. Best settings are determined by altering them, starting from the best simulated configuration.

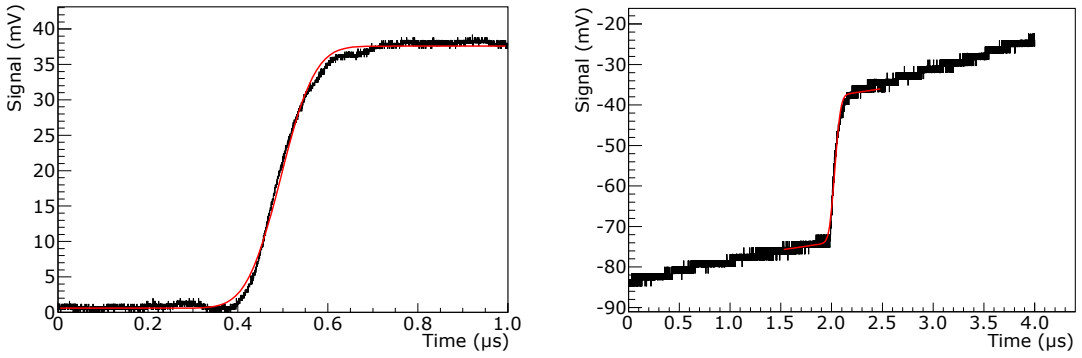
Signal source is a weak ^{55}Fe-radiation source. It is placed right next to the sensor's entrance window.

For each measured signal, the signal strength has to be determined. In case of continuous-reset, five ways of doing that have been explored. Two are based on integrated oscilloscope functions (see Appendix C), three use the stored waveform offline:

- Oscilloscope based: Maximum value minus minimum value of a triggered waveform
- Oscilloscope based: High plateau minus lower plateau
- Offline: Maximum value minus minimum value of a waveform
- Offline: Average of all points after maximum minus average baseline of a waveform
- Offline: Fit of an error-function, difference of the evaluated fit function at plus and minus infinite

The resulting values are filled into histograms. From illumination with ^{55}Fe X-rays, a spectrum is expected with two peaks from the K_α and the K_β line. A double Gaussian is fit to each and the fit parameters are extracted: Mean μ_1 and standard deviation σ_1 of K_α and Mean μ_2 and standard deviation σ_2 of K_β. These parameters are then used to evaluate the chosen settings and analysis method by calculating the zero-point error (see Appendix A.3) and the energy resolution (see Appendix A.4). The signal-to-noise ratio was calculated as ratio of K_α signal and baseline noise.

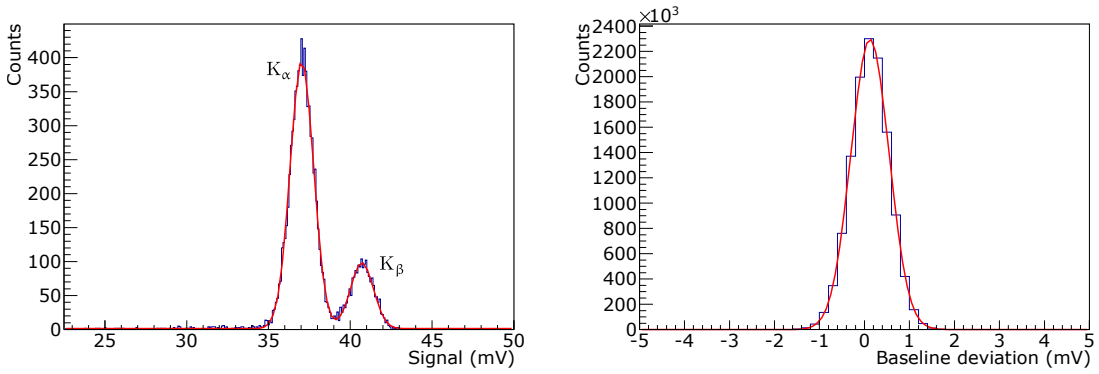
Comparison of the mentioned analysis methods show, that an error function applied to the rising edge of the signal delivers best results, therefore only numbers and plots are shown that were obtained in this way.



(a) **(b)**

Figure 21.2: Example waveforms recorded by an oscilloscope connected to the output of the first amplifier of the readout ASIC in continuous-reset mode (a) and in pulsed-reset mode (b).

The oscilloscope's time axis is narrowed so that only the rising edge of a signal is visible (Figure 21.2). In this way, all four measurement modes can be realized. For a first amplifier measurement with pulsed-reset, the rising baseline level is compensated by the oscilloscope's AC-coupling. An error-function is applied to the signal to determine its amplitude (Figure 21.2a). The rising baseline of pulsed mode is respected by an additional slope fit (Figure 21.2b). Baseline noise is determined by histogramming the deviation of the first quarter of data points from the fit function (Figure 21.3b).



(a) **(b)**

Figure 21.3: Spectrum of ^{55}Fe measured in continuous-reset mode at room temperature (a) and its baseline noise (b).

Figure 21.3a shows the ^{55}Fe spectrum measured at first amplifier with standard configuration (see Appendix G) of sensor and readout ASIC. The K_α (5.899 keV) and K_β (6.490 keV) lines are clearly separated.

This spectrum has been measured at room temperature with standard settings. In order to optimize these settings, the measurement was repeated with slightly varied settings on first and second amplifier. The results are listed in table 21.1 (first amplifier) and table 21.2 (second amplifier).

Config	$\mu_1 \pm \sigma_1$ (mV)	$\mu_2 \pm \sigma_2$ (mV)	σ_{noise} (mV)	ZPE (%)	ERes	SNR
default	36.95±0.72	40.62±0.74	0.424	-0.06	5.03	87
slow feedback	36.73±0.70	40.41±0.73	0.413	-0.01	5.13	89
VSSA = 2.5 V	33.21±0.69	36.44±0.71	0.386	-0.28	4.61	86
backside bias voltage = 90 V	36.89±0.75	40.61±0.75	0.427	0.07	4.97	86
amplifier DAC = 60	37.81±0.67	41.63±0.71	0.420	0.09	5.57	90
metal box	37.24±0.67	41.01±0.68	0.421	0.09	5.60	89

Table 21.1: Mean and standard deviation of measured ^{55}Fe spectra on the first amplifier, baseline noise and the derived values zero point error, energy resolution and signal-to-noise ratio.

Config	$\mu_1 \pm \sigma_1$ (mV)	$\mu_2 \pm \sigma_2$ (mV)	σ_{noise} (mV)	ZPE (%)	ERes	SNR
default	-193.65±4.17	-213.20±4.67	0.351	0.07	-4.42	551
slow feedback	-192.00±4.15	-211.55±4.45	0.351	0.15	-4.55	547
VSSA = 2.5 V	-171.06±4.47	-188.12±4.72	0.351	-0.04	-3.72	488
backside bias voltage = 90 V	-191.49±5.23	-211.36±5.16	0.366	0.32	-3.83	523
amplifier DAC = 60	-195.97±4.65	-216.14±4.42	0.355	0.25	-4.45	552
metal box	-193.35±4.20	-212.94±4.31	1.39	0.11	-4.61	139[1]

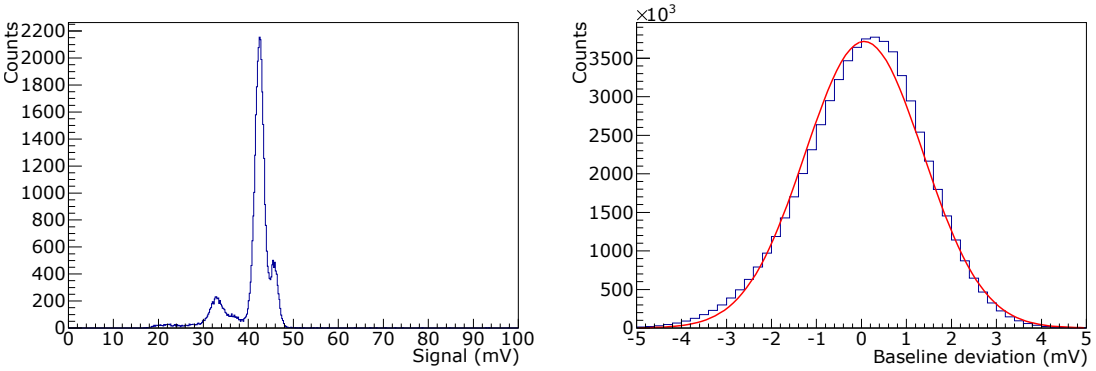
Table 21.2: Mean and standard deviation of measured ^{55}Fe spectra on the second amplifier, baseline noise and the derived values zero point error, energy resolution and signal-to-noise ratio.

The critical properties for TRISTAN are the signal-to-noise ratio and the energy resolution.

The SNR measured on the second amplifier is approximately 6 times as high (551) as SNR of the first amplifier (87). However, a price has to be paid: The energy resolution of the second amplifier is about 15% lower. In other words, the first amplifier rejects noise at lower gain and the second amplifier multiplies the signal, but adds a small uncertainty. Compared to the monolithic sensors discussed in the previous parts, all these values are much better. This is no surprise, as a special drift sensor and a readout ASIC without power limitations have been combined to a high-precision measurement tool.

The effect of variations to the configuration is small. Slower feedback, higher current for the amplifiers and the noise reduction by electromagnetic shielding have an overall positive impact on energy resolution and SNR, while higher amplifier voltage and higher depletion voltage influence the measurement in a negative way.

These findings are used in the next sections, measuring the ^{55}Fe spectrum in a chilled environment and in the measurement of X-rays of higher energy from an X-ray tube.



(a) **(b)**

Figure 21.4: Spectrum (a) and baseline noise histogram (b) recorded in pulsed-reset mode. Imperfect settings render this measurement less precise than the continuous-reset measurement.

The data taken on measurements with pulsed-reset were of insufficient quality in the first campaign. This is evident by comparing peak separation and noise floor of figure 21.3

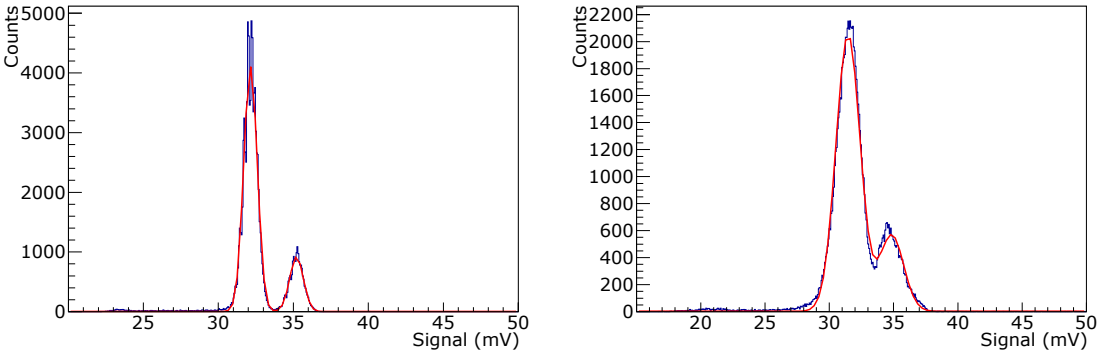
[1] After the measurement, a faulty ground connection of the metal box was discovered.

(continuous-reset) with figure 21.4 (pulsed-reset). The latter has both signal peaks nearly merged and shows a notable noise peak.

Consequently, the extracted characteristic values are an order of magnitude worse, than the values presented in tables 21.1 and 21.2. This issues leading to the poor performance of pulsed feedback measurements has been traced back to the lack of light and electromagnetic shielding of the sensor, which seems to have a larger impact in pulsed mode, than it has in continuous mode, and imperfect oscilloscope settings. Both issues have been addressed, so that the following measurements have led to comparable results for continuous and pulsed-reset mode.

21.2 Effect of cooling on the measured spectrum

Cooling reduces leakage currents both in the sensor and the readout ASIC and hence the baseline noise and signal variation. In order to quantify this effect, the entire setup is placed inside a climate chamber. The measured signals are X-rays from a ^{55}Fe radiation source. Data are taken at -5°C and -20°C at both amplifiers with pulsed and continuous-reset.



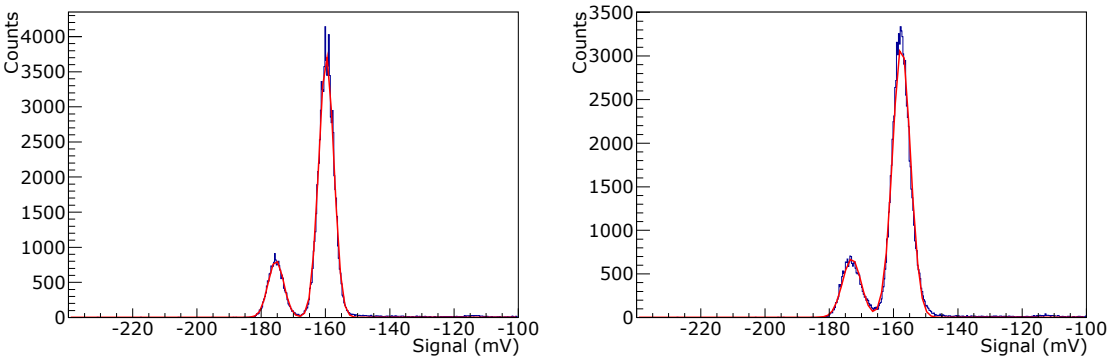
(a) **(b)**

Figure 21.5: ^{55}Fe spectra measured with continuous-reset (a) and pulsed-reset (b) on the first amplifier at −5°C.

Figure 21.5 shows the ^{55}Fe spectra measured at -5°C with the first amplifier in continuous- (a) and pulsed-reset mode (b). With continuous-reset, the peaks are more narrow and more clearly separated, compared to room temperature (Figure 21.3a). With improved settings, the pulsed-reset mode delivers a double peak spectrum, too. However, the peaks are partially merged and each peak is not of Gaussian shape. This is an effect of the data analysis. Fitting an error function to the rising edge of a signal is a good approximation for data from continuous-reset measurement, but appears insufficient for the pulsed mode, even when the rising baseline is respected. The rising edge of the signal has a more step-like shape. Continuous-reset curves starting- and end-point and thus makes the error-function more suitable.

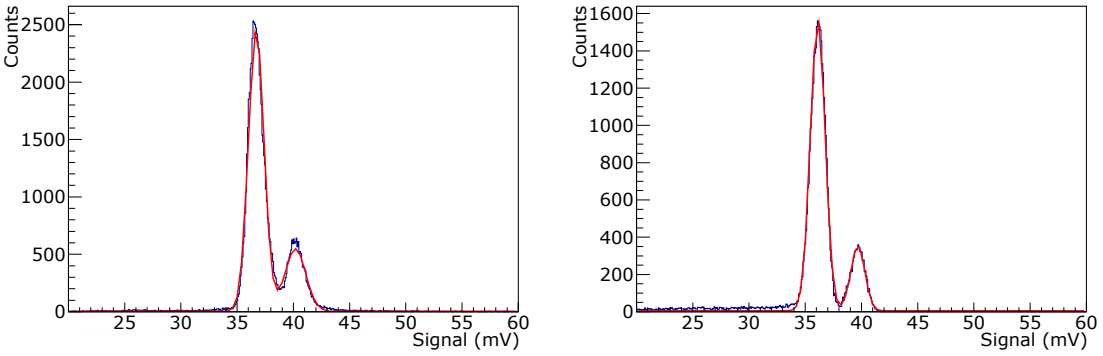
Tracing the odd shape of the pulsed-reset spectrum back to fitting problems is confirmed when looking at the spectra recorded on the second amplifier. Spectra with continuous (a) and pulsed-reset (b) are shown in figure 21.6. Both are similar with a nice peak separation. Fitting of the waveforms is identical as the AC-coupling between first and second amplifier filters the rising baseline and curves the starting- and end-point of the signal in pulsed-reset mode.

The same four spectra have been measured at -20°C and are shown in figures 21.7 (first amplifier) and 21.8 (second amplifier). Most notable is that the spectra at the second



(a) **(b)**

Figure 21.6: ^{55}Fe spectra measured with continuous-reset (a) and pulsed-reset (b) on the second amplifier at −5°C.



(a) **(b)**

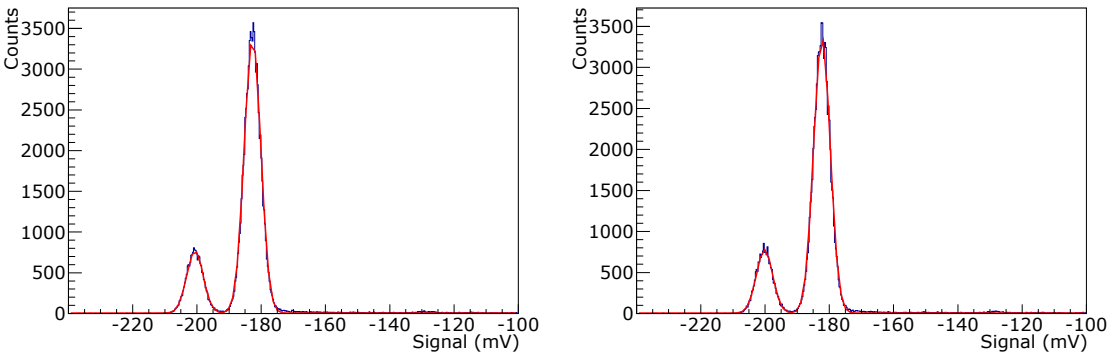
Figure 21.7: ^{55}Fe spectra measured with continuous-reset (a) and pulsed-reset (b) on the first amplifier at −20°C.

amplifier, measured with either reset mode, are indistinguishable. In comparison to the warmer spectra, the characteristic lines are more separated.

In making of the shown spectra, the baseline noise for each data set has been determined. Together with all other key values, they are comprised in tables 21.3 (-5°C) and 21.4 (-20°C).

Amp.	Reset	$\mu_1 \pm \sigma_1$ (mV)	$\mu_2 \pm \sigma_2$ (mV)	σ_{noise} (mV)	ZPE (%)	ERes	SNR
1	cont.	36.17±0.58	39.56±0.55	0.533	-0.20	6.01	68
1	pulsed	34.05±1.06	37.93±0.91	0.721	-0.23	3.94	47
2	cont.	-180.85±2.58	-198.66±2.55	1.038	-0.21	-6.94	-174
2	pulsed	-180.86±2.66	-198.65±2.64	0.777	-0.21	-6.72	-233

Table 21.3: Mean and standard deviation of measured ^{55}Fe spectra at −5°C, baseline noise and the derived values zero point error, energy resolution and signal-to-noise ratio.



(a) **(b)**

Figure 21.8: ^{55}Fe spectra measured with continuous-reset (a) and pulsed-reset (b) on the second amplifier at −20°C.

Amp.	Reset	$\mu_1 \pm \sigma_1$ (mV)	$\mu_2 \pm \sigma_2$ (mV)	σ_{noise} (mV)	ZPE (%)	ERes	SNR
1	cont.	36.70±0.69	40.18±0.93	0.382	-0.20	4.30	96
1	pulsed	36.11±0.66	39.67±0.62	1.063	-0.21	5.57	34
2	cont.	-182.71±2.56	-200.62±2.55	0.725	-0.21	-7.01	-252
2	pulsed	-182.30±2.57	-200.17±2.54	0.721	-0.21	-7.00	-253

Table 21.4: Mean and standard deviation of measured ^{55}Fe spectra at −20°C, baseline noise and the derived values zero point error, energy resolution and signal-to-noise ratio.

Comparison of the values extracted from the ^{55}Fe spectra measured at low temperatures with the values at room temperature show some characteristics:

- Signal μ_i and noise σ_i decrease, but μ_i/σ_i increases, which is beneficial for energy measurements.
- Baseline noise increases, which contradicts the expectation. However, this is not a temperature effect, but is due to longer cables to the setup in the climate chamber and the noise induced by the climate chamber itself. This has been verified by a baseline measurement in the climate chamber at 20°C.
- Reset mode has no significant effect on the measured spectrum, but on the baseline noise of the first amplifier.
- Energy resolution of the second amplifier is better than energy resolution of the first stage. At room temperature this was inverted.

Consequently, in an experiment like TRISTAN with well controllable environmental parameters (e.g. temperature and absence of external noise sources), the second stage amplifier delivers better results, both in counting efficiency and energy resolution.

21.3 X-Ray spectra

So far only the energy range of ^{55}Fe X-rays has been investigated. For this section X-ray spectra of several elements have been obtained and analyzed. The targets were illuminated by an X-ray tube, emitting a wide spectrum of X-rays up to 60 keV (see Appendix B.2).

The measurement setup can not be cooled inside the X-ray tube's housing. The cabling length had to be extended in order to connect the setup inside with power supplies outside. Finally, the X-ray tube and its cooling spreads additional noise. Consequently, no

competitive noise measurement is expected, but the spectra of the elements can be used to evaluate the linearity of the setup in a wide range.

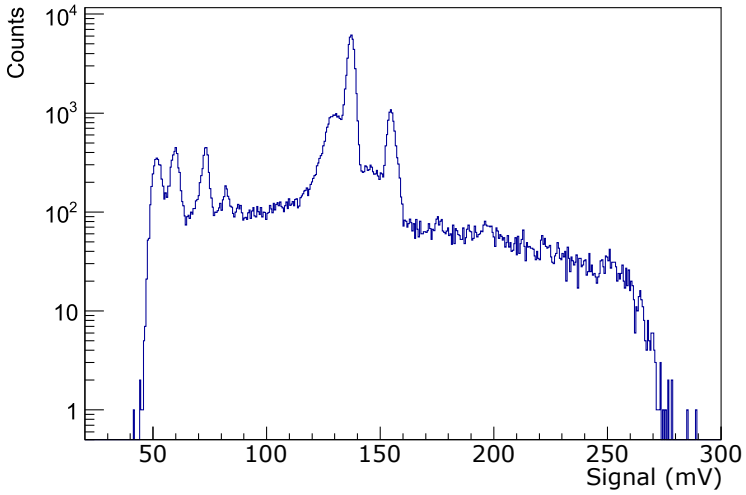


Figure 21.9: Spectrum of a silver target illuminated by the X-ray tube, measured with first amplifier at room temperature in noisy environment.

Figure 21.9 shows the spectrum measured on the first amplifier of a silver target. Each spectrum has been measured on the first and on the second amplifier. The lower cut-off is defined by the threshold set on the oscilloscope recording the waveforms. The upper cut-off refers to the maximum energy of the X-ray tube.

The two largest peaks are the characteristic lines of the used target elements (here: silver), but more peaks can be found. In this example, two peaks have been identified as the characteristic lines of zinc. These additional zinc peaks appear, as the X-ray tube does not illuminate the target exclusively, but also the area around it, which is the steel casing of the X-ray setup. Apparently zinc is part of the alloy. They are found in the spectrum of every target element. The iron lines expected from the steel casing are not found, as the threshold has been set above their energy, to reduce the signal rate to a frequency the oscilloscope can handle.

Another set of three peaks is found in most measured spectra, which could not be traced back to an element typically used in steel or other materials found in the proximity. The explanation is the way this specific X-ray tube is built. The generated X-rays are filtered by a vanadium window, which modulates the emitted spectrum [121]. The spectrum of this X-ray tube with the used vanadium window has been investigated at the time of construction and is shown in Appendix B.2. The found peaks are caused by the vanadium window.

Spectra of the following targets have been measured:

Element		Amp 1	Amp 2
Iron	Fe	yes	yes
Zinc	Zn	yes	yes
Molybdenum	Mo	yes	yes
Silver	Ag	yes	yes
Indium	In	yes	no
Tin	Sn	yes	yes
Neodymium	Nd	yes	yes

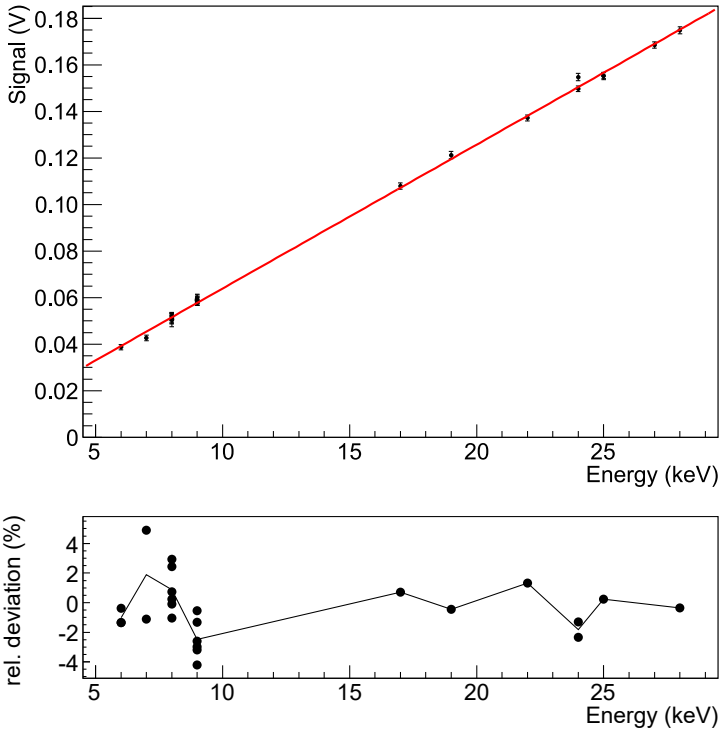


Figure 21.10: The mean signal of all peaks of all first amplifier spectra plotted over the respective X-ray energy.

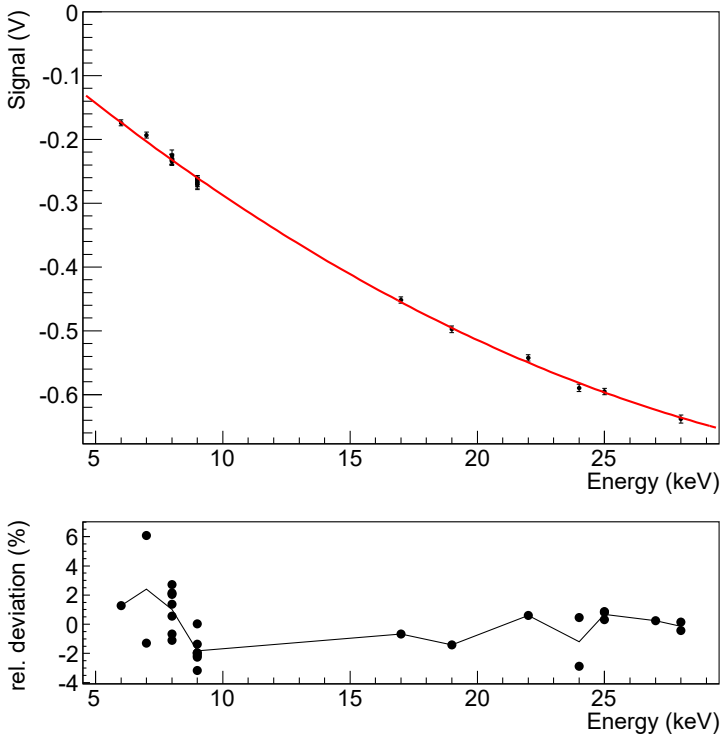


Figure 21.11: The mean signal of all peaks of all second amplifier spectra plotted over the respective X-ray energy.

Each measured spectrum has been plotted and fits were applied. Each peak was fit by a Gaussian. The mean μ and the standard deviation σ of all peaks of all spectra are plotted over the respective X-ray energy in figure 21.10. The response function of the first amplifier is linear in the measured range.

Analysis of the spectra measured on the second amplifier results in figure 21.11. The response function is not linear. The red fit is a second order function.

21.4 Energy measurement by period length

In the introduction an alternative measuring method was explained (Figure 20.4), the indirect measurement of deposited energy by measuring the time between two reset pulses in the pulsed-reset mode. The principle is simple: The time between two reset pulses is defined by the constant leakage current of the sensor. A signal adds a short current pulse, shortening the time between two reset pulses. The shorter the period between two pulses is, the larger was the accumulated charge between them. The relation between energy and period reduction is expected to be linear. If more than one particle is measured between two reset pulses, their energy can not be distinguished. The period can not be reduced to less than 0. An event with more energy necessary to reduce the period to 0, will cause several reset pulses quickly after each other. Theoretically, there is no upper limit to the energy scale.

The following section is intended as prove of principle for this measurement method on the setup of FBK's silicon drift diode and the KIT TRISTAN IC.

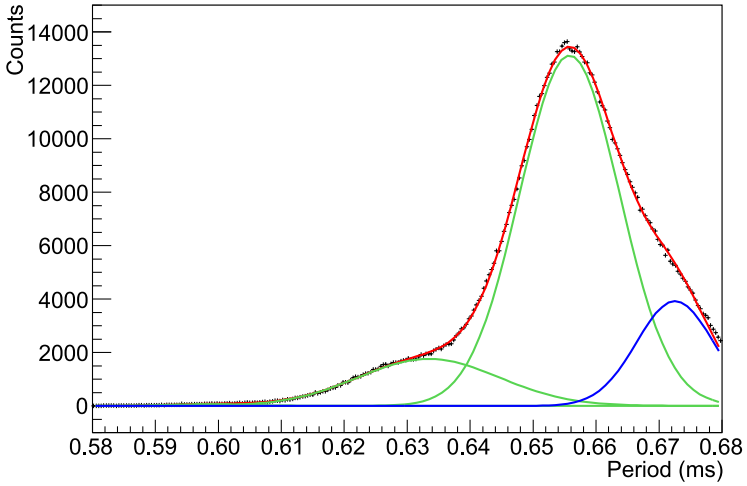


Figure 21.12: Spectrum of ^{55}Fe. The spectrum is made by measuring the time between two reset pulses in pulsed-reset mode. Depending on the number of hits within a period, the period is shortened (green). Baseline is shown in blue, the sum of all in red.

In the first test, the time between two reset pulses is measured by an oscilloscope, while the sensor is illuminated by an ^{55}Fe X-ray source. The recorded period spectrum is shown in figure 21.12. With the used settings, the mean period length without signal is ≈0.672 ms.

The spectrum was fit by a series of Gaussian functions, better seen in a logarithmic plot (Figure 21.13). The baseline is shown in blue, the signal peaks in green and the sum of all in red. The first peak is formed by a single hit during the measured period, the second peak by two independent hits during a period and the third peak by three hits (Figure 21.14).

Even though the appearance of several peaks, which refer to a different number of hits, in the spectrum is promising, the energy resolution is much smaller than in the measurement methods discussed before. However, this measurement method requires much less

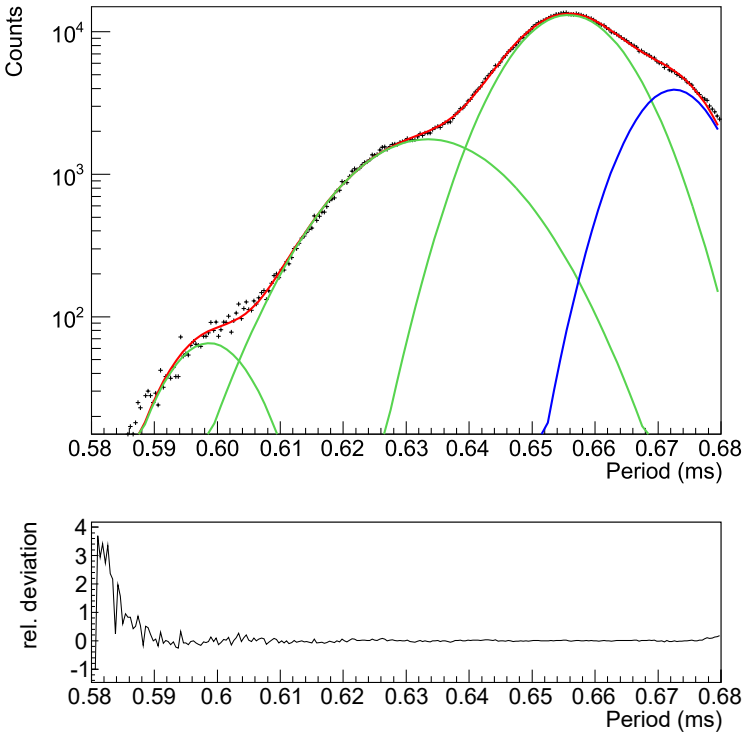


Figure 21.13: Logarithmic ^{55}Fe spectrum. Periods with one, two and three hits have been recorded (green fits from right to left). The deviation of the fit from the measured points suggests that four hits within a period have occurred a few times, too.

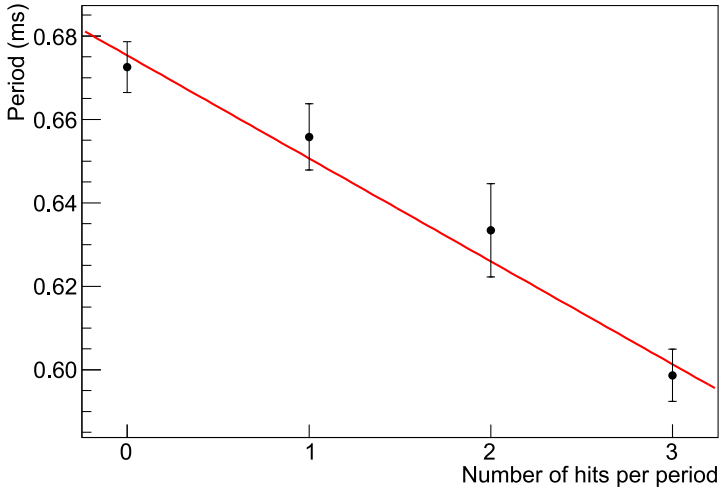


Figure 21.14: The expected linearity of this measurement is confirmed by fitting the means of the ^{55}Fe spectrum over the number of hits. The standard deviation however, is large.

computational resources. The time measurement can actually be done online in an FPGA, without the need for an oscilloscope or further data processing. This allows higher rates and the absence of an energy limit might be interesting for some applications, too.

The period spectra in the X-ray tube have been recorded for several target elements. The reset threshold has been lowered, which shortens the baseline period, to avoid multiple hits per cycle.

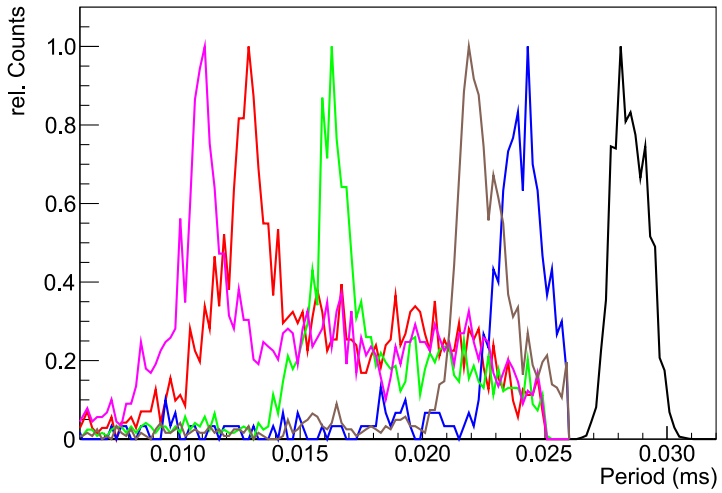


Figure 21.15: Spectra of several targets illuminated with X-rays from an X-ray tube (colored lines) and the baseline (black).

The spectra of the target elements iron, zinc , molybdenum, silver and tin are shown in figure 21.15. The baseline period is drawn in black. In agreement with the ^{55}Fe measurement, the energy resolution is reduced, compared to the waveform-based measurement. Only a single peak per target can be identified.

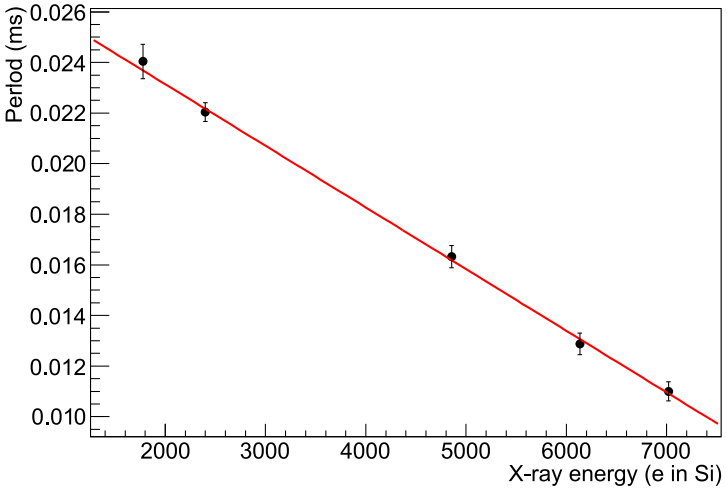


Figure 21.16: Mean of the peaks from the X-ray measurement over the respective energy of the target element. Linearity is confirmed in the investigated energy range.

The mean values of the peak of each spectrum are shown in figure 21.16 over the characteristic energy. In the investigated range, the expected linear behavior is confirmed.

22. KIT TRISTAN integrated circuit summary

A silicon drift detector has been combined with the KIT TRISTAN readout ASIC to a characterization setup, in order to evaluate this approach for the TRISTAN upgrade of the KATRIN experiment. The sensor is operated either in pulsed or continuous-reset mode. The signal is amplified and shaped by the readout ASIC, it can be probed after the first and second amplification stage.

At room temperature, the assembly shows an excellent signal-to-noise ratio of up to 552 in measurements with an ^{55}Fe-source. All measured spectra are of excellent quality compared to others measured for this thesis. This applies for the signal uncertainties and for the baseline noise.

The spectra measured in a climate chamber at temperatures down to −20°C had SNRs between 100 and 250. Lower temperatures were beneficial for the energy resolution, but the necessary variations of the setup to fit in the climate chamber had a negative effect on the baseline noise. At room temperature, measurements at the first amplifier showed the better energy resolution, but with lower SNR, compared to the second amplifier. At −20°C, the second amplifier was superior in both aspects.

The X-ray spectra of various elements, illuminated by an X-ray tube showed, beside the expected characteristic lines of the respective target elements, lines of zinc (ingredient of the steel of the setup casing) and lines of the vanadium exit window of the X-ray source. The response functions of the first and second amplifier were calculated, using the data from the X-ray spectra. In the X-ray energy range from 5 keV to 25 keV was the response function of the first amplifier linear and the response function of the second amplifier stage could be approximated by a polynomial of second order.

In a proof-of-principle study it has been shown that in pulsed-reset mode, the frequency of the reset pulses is a measure for the deposited energy and hit rate.

The characterization was so successful that an assembly of this type can be recommended for application in TRISTAN.

Part V

Conclusion

Summary

Silicon sensors are a proven tool to track particles in high-energy physics experiments. The tracking layers of the next generation of detectors will probably be equipped with High Voltage CMOS sensors: Mu3e, an experiment measuring the branching ratio of muons with unprecedented precision, will be equipped with several HV-CMOS sensor layers [87] and the experiments ATLAS [122] and LHCb [123] consider HV-CMOS sensors for future upgrades. HV-CMOS technology has been proposed for TRItium STerile Neutrino upgrade of the KArlsruhe TRItium Neutrino experiment and is under investigation for the projected collider experiments CLIC [124] and COMPASS++/AMBER [125] at CERN (Switzerland) and CEPC at IHEP (China) [126]. But also monitoring of commercial and medical beams is a possible application for HV-CMOS sensors.

The large interest in HV-CMOS technology is based on five aspects:

- Compared to hybrid sensors, the material budget can be significantly reduced. Instead of a sensing layer and a readout layer, the HV-CMOS approach combines both components on a single layer of silicon.
- There is no need for bump bonds in HV-CMOS sensors, therefore the minimum pixel size is no longer limited by the minimum bump bond pitch. Consequently, the spatial resolution can be improved.
- Time resolution is limited by the rise time of the analog front-end. Constraints on power consumption limit the rise time to moderate 20 to 100 ns. The vicinity of sensor and readout electronics allows for great improvement of the time uncertainties by maintaining the analog information for offline correction. Carrying the analog information throughout the readout chain provides information on the deposited energy in addition.
- Compared to standard CMOS technology and competitive derivatives (HR-CMOS, small diode design), HV-CMOS technology comes with intrinsic radiation hardness, granted by the high depletion voltage. Yet, the radiation hardness of specialized technologies (e.g. hybrid detectors) is unmatched. Radiation damage affects the sensing diode, but it remains operational for all in this thesis investigated fluences. Leakage current can be effectively prohibited by moderate cooling.
  Standard CMOS electronics suffers from ionizing radiation. Replacement of crucial NMOS transistors by annular transistors, adjustable biasing and other countermeasures are a powerful tool box that keeps HV-CMOS fully functional, even after having been exposed to high doses.
- The unique selling point of HV-CMOS sensors is their price tag. The costs per square meter are in the order of 85 000 €, including blank wafers, production and post-processing. The costs of hybrid detectors are in the order of 2 000 000 € per square meter. (For the detailed calculation see Appendix H.)

Possible disadvantages are power consumption and the lack of experience with these sensors in long-term experiments.

For investigation of Monolithic Active Pixel Sensors – such as HV-CMOS devices – reusable characterization setups have been developed during this thesis: The Multi-purpose Adapter Board and its successor the GEneric Configuration and COntrol systems have replaced the development- and cost-intensive single chip setups. Their modular approach in hard-, firm- and software empowers them to be adapted to new sensors and challenges, consuming minimal resources. Their success becomes apparent looking at the number of divers sensors that have been measured with them at KIT, in Germany, Europe and the world: Over sixteen sensors in at least ten institutes around the globe.

The characterization of four HV-CMOS devices has been presented in this thesis: H35Demo, ATLASpix1, MuPix8 and KIT TRISTAN Integrated Circuit. The detailed results are summarized in the respective chapters. These results have been consecutively published and presented in journals and (HV-)CMOS collaborations and influence the course of the (HV-)CMOS development.

Although, being designed for different purposes and ignoring all fundamental differences, some aspects of these four chips can be compared:

	H35Demo	ATLASpix1		MuPix8	TRISTAN IC
		Simple	M2		
spatial res. x	14.4 µm	11.5 µm	14.4 µm	23.1 µm	1732 µm
spatial res. y	72.2 µm	37.5 µm	17.3 µm	23.4 µm	1732 µm
time res.	30 ns[1]	67 ns			-
corrected time res.	4 ns[1]	13 ns			-
SNR (up to)	50	78			550
MIP signal[2]	3600 e	4400 e			-
detection threshold:					
σ before tuning	350 e	120 e	380 e[3]	230 e[3]	-
σ after tuning	230 e	52 e	95 e[3]	42 e[3]	-
max. event rate	<26.7 MHz	40 MHz	40 MHz	3 × 40 MHz	-

TRISTAN IC is different from the other investigated chips, as it is an assembly of a silicon drift diode with an HV-CMOS readout chip. Its brilliance is unmatched by other CMOS developments, as it does not compromise power consumption and signal quality. The single channel device reached an impressive signal-to-noise ratio (SNR) of 550.

Being at least one development cycle younger than H35Demo, ATLASpix1 and MuPix8 are not only fabricated in a smaller technology, but also the design is much more sophisticated: The monolithic readout has only a slightly increased nominal performance, but features zero suppression, extremely beneficial for clustered events. Furthermore is information, additionally to pure hit time and location, provided. These analog data can be used to correct for time uncertainties and reach a time resolution on matrix level of 13 ns for signals close to the signal of minimum ionizing particles with maximal time-walk. On pixel level, the time resolution can be reduced even further to less than 10 ns.

H35Demo has been produced on standard substrate (resistivity about 80 Ωcm), therefore its response to charged particles is smaller (thin depletion zone) than the response of ATLASpix1 or MuPix8 which were implemented on high resistive substrate (resistivity about 200 Ωcm, thicker depletion zone).

[1] Value of a single pixel.

[2] Signal from ^{90}Sr decay electrons were used as Minimum-Ionizing-Particles (MIPs), measured on a single pixel at first analog output.

[3] Measurement conducted by M. Prathapan (ATLASpix_M2) and A. Weber (MuPix8) under the author's supervision.

The effect of mismatch on the spread of detection thresholds is much smaller in the younger designs. The 2-bit local threshold adjustment mechanism in H35Demo was sufficient to reduce the threshold spread by 1/3. The improved design implemented in the later sensors with three bits, reduces the distribution to values between 42 and 95 electrons.

The spatial resolution depends on pixel size and is an upper limit for single pixel events. Involvement of several pixels in events – so-called cluster events – increase spatial resolution by interpolation, especially if analog information is available as in ATLASpix1 and MuPix8.

None of these sensors were designed for direct implementation in a particle physics experiment, but they powerfully demonstrated the capabilities of HV-CMOS technology and its significance for detector concepts in upcoming detector developments and upgrades.

Outlook

The next generation of full reticle size HV-CMOS sensors has been fabricated and is being characterized at KIT ADL. ATLASpix3 and MuPix10 have been developed under the influence of this thesis. Besides their size, they depict probably the last intermediate step on the path to large-scale usage of HV-CMOS sensors in high-energy physics experiments. Future experiments in high-energy physics will struggle to ignore this development.

Part VI

Appendix

A Calculations

A.1 Power over frequency

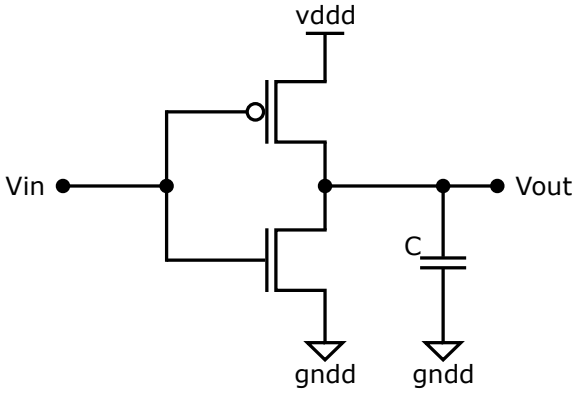


Figure A.1: *NOT* gate as an example for digital circuits.

The power consumption of digital circuits is composed of two parts, static power P_S and dynamic power P_D:

$$P = P_D + P_S \tag{A.1}$$

Digital electronics consists of gates, which change their output depending on the input. The simplest is a *NOT*-gate as shown in figure A.1. Power is consumed, as a gates output at least has to provide current to charge the line capacitance C. Switching from logic '0' to logic '1', the capacitance is charged by *vddd*. The opposite operation discharges the capacitance into ground. The stored energy E_C depends on the capacitance and voltage difference U between logic '0' and '1':

$$E_C = \frac{Q \cdot U}{2} = \frac{C \cdot U^2}{2} \tag{A.2}$$

How often this energy dissipates depends on the operational frequency f:

$$P_D = E_C \cdot f \tag{A.3}$$

Static power P_S is caused by leakage current, dominated by a constant current from *vddd* to *gndd*:

$$P_S = const. \tag{A.4}$$

Therefore, the total power consumption is given by

$$P = \frac{C \cdot U^2}{2} \cdot f + P_S \tag{A.5}$$

$$P \propto f. \tag{A.6}$$

A.2 Calculate mean and standard deviation

Calculation of mean $\bar{x}$ and standard deviation σ of a data set with N members x_i is a basic procedure:

$$\bar{x} = \frac{\sum x_i}{N} \tag{A.7}$$

$$\sigma = \sqrt{\left(\bar{x_i^2}\right) - \left(\bar{x_i}\right)^2} \tag{A.8}$$

For large data sets it is not convenient to calculate mean first and then determine the standard deviation as the formula suggests. This would require to parse the data twice. Modifying equation A.8 delivers

$$\sigma = \sqrt{\frac{\sum x^2}{N} - \left(\frac{\sum x}{N}\right)^2}. \tag{A.9}$$

From this we see which values have to be calculated while parsing the data:

Sum of values:	$\sum x$
Sum of squares:	$\sum x^2$
Number of values:	N

These can be used to calculate mean and sigma during parsing or after parsing the data for the first time.

A.3 Zero point error

Energy spectra have typically several distinct peaks. In the case of this thesis, the energy of each peak is a priori known, as not the spectrum is investigated, but the sensor measuring the spectrum.

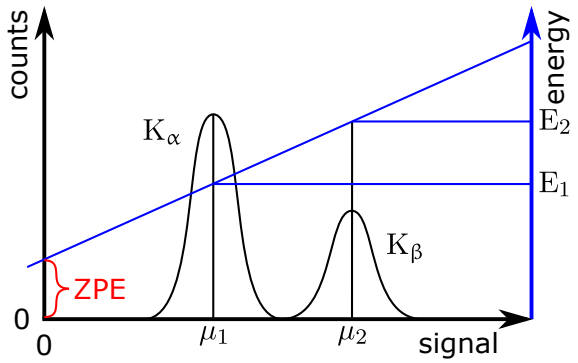


Figure A.2: A spectrum with two peaks of known energy can be used to calculated the zero point error.

The tuples of energy and peak position can be fit by an appropriate function, the simplest is the linear function. A perfect sensor would find a zero energy signal at zero point of the spectrum, if the fit line is extended. A real sensor does not necessary comply with this, therefore the zero point error is calculated.

Most often used as calibration source is ^{55}Fe, which has two dominant energies (Figure A.2). In this case the zero point error can be directly calculated as:

$$ZPE = 1 - \frac{E_1/\mu_1}{E_2/\mu_2} \tag{A.10}$$

With the peaks energies E_i and the signals' mean μ_i. In case of a more complex spectrum or several spectra, a linear function has to be applied.

A.4 Energy resolution measured by peak separation

Energy resolution is best explained looking at an energy spectrum with two peaks. The better the two peaks are distinguishable, the better is the energy resolution. If they merge, energy resolution becomes insufficient.

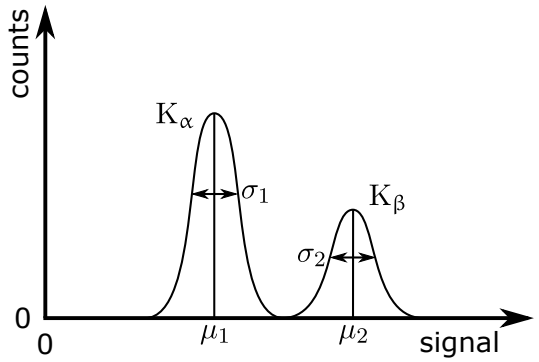


Figure A.3: A spectrum with two peaks of known energy can be used to calculated the peak separation or energy resolution.

A good energy resolution or peak separation $\Delta\Pi$ can be achieved either by a large gain, shifting the peaks further from each other, or by a low uncertainty, rendering them smaller in width (Figure A.3).

This can be cast into a formula:

$$\Delta\Pi = \frac{\mu_2 - \mu_1}{(\sigma_1 + \sigma_2)/2} \tag{A.11}$$

A.5 Spatial resolution of a pixel

The spatial resolution of discrete pixels is a key value in the design of tracking detectors, it directly determines the quality of track reconstruction (see chapter 2).

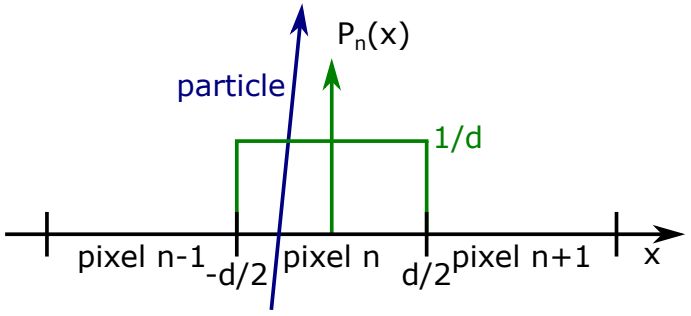


Figure A.4: The probability of a particle hitting a pixel n in a certain location is uniformly distributed $1/d$.

The simplest assumption is that a particle generates only charge in one pixel n (Figure A.4). The probability $P_n(x)$ of the particle hitting a specific spot of the pixel is therefore

uniformly $1/d$, where d is the width of the pixel. Cast into formulae this is:

$$P_n(x) = \frac{1}{d} \tag{A.12}$$

$$\int_{-d/2}^{d/2} P_n(x)\,dx = 1 \tag{A.13}$$

The information recorded by the sensor is that a particle has hit pixel n in the middle:

$$\langle x \rangle = \int_{-d/2}^{d/2} x P_n(x)\,dx = 0 \tag{A.14}$$

The uncertainty along the respected axis x is:

$$\sigma_x^2 = \langle (x - \langle x \rangle)^2 \rangle = \int_{-d/2}^{d/2} x^2 P_n(x)\,dx = \frac{d^2}{12} \tag{A.15}$$

Therefore the spatial resolution of a pixel sensor with pitch d is

$$\sigma_x = \frac{d}{\sqrt{12}} \approx 0.29\,d \tag{A.16}$$

Spatial resolution can be improved, if more then one pixel collects charge from a passing particle.

A.6 Voltage drop on a typical supply line

ASIC supply lines provide bias currents to pixels and other components of a sensor. Usually they are designed as single aluminum wires routed in column direction. Experiments show row-dependent characteristics, therefore it is reasonable to look for an explanation in the distance to the bias generation. This distance is bridged by supply lines. The main power is routed over a power grid covering the major part of a metal layer, thus suffers not really from conduction loss.

In 180 nm processes, supply line width is in the order of 280 nm and its thickness is in the order of 515 nm. Therefore, the line cross section is

$$A = 144 \cdot 10^{-9}\ \mathrm{mm}^2 \tag{A.17}$$

and the specific resistance of aluminum is

$$\rho_{\mathrm{Al}} = 2.65 \cdot 10^{-2}\ \frac{\Omega\,\mathrm{mm}^2}{\mathrm{m}}. \tag{A.18}$$

Consequently, the resistance of this supply line is

$$\frac{\mathrm{R}}{\mathrm{m}} = \frac{\rho_{\mathrm{Al}}}{A} = 184\ \frac{\Omega}{\mathrm{m}} \tag{A.19}$$

The current in such line is typically small compared to the main power supplies. For this exemplary calculation values from ATLASpix1 are used. One bias current on the whole matrix I_{total} is assumed 1 mA, which has to be divided by the number of columns to get the current in a single supply line $I_{\mathrm{col}} = 0.04$ mA. The current is not consumed at the end of the line but by all pixels along the full distance of $d = 2$ cm:

$$I(x) = I_{\mathrm{col}} - \frac{x}{d} I_{\mathrm{col}} = \left(1 - \frac{x}{d}\right) I_{\mathrm{col}} \tag{A.20}$$

The voltage drop is:

$$\Delta U_{\text{drop}}(x) = \rho_{\text{Al}} \cdot \Delta x \cdot I(x) \tag{A.21}$$

$$U_{\text{drop}}(x) = \int_0^x \rho_{\text{Al}} \cdot I\left(x'\right)\, dx' \tag{A.22}$$

$$U_{\text{drop}}(x) = \frac{1}{2}\rho_{\text{Al}} \cdot I_{\text{col}}\left(2x - \frac{x^2}{d}\right) \tag{A.23}$$

For a 2 cm long matrix, the voltage drop is 0.15 V. This is 10% of a typical voltage (1.5 V) at a typical current (1 mA/matrix).

B Irradiation facilities

B.1 Proton irradiation at ZAG

In this thesis sensors have been irradiated with protons. The irradiations have been conducted by Institute of Experimental Particle Physics (ETP)[4] at the commercial cyclotron of Zyklotron AG (ZAG)[5], located on KIT Campus North. A photo of the sample holder is shown in figure B.5.

First, negative charged Hydrogen (H^-) is accelerated in the cyclotron, the electrons are striped by foil and have an energy of 25 MeV at extraction and 23 MeV when hitting their target. The target, in this case sensors, is stored in a sample box. Due to the risk of activation of elements in the setup, operation or even assembly prior to irradiation is not possible.

The potential irradiation effect of protons in this thesis is given as neutron equivalent. That means the damage by the protons is compared to the damage 1 MeV neutrons would induce. The used unit is neutron of 1 MeV equivalent per cm^2 or $n_{eq\,1\,MeV}/cm^2$ or shorter n_{eq}/cm^2.

As protons are charged, they do not only cause NIEL damage, but also ionizing dose. This is not reflected in the unit n_{eq}/cm^2. Per $10^{15} n_{eq}/cm^2$ approximately 750 kGy ionizing dose is applied.

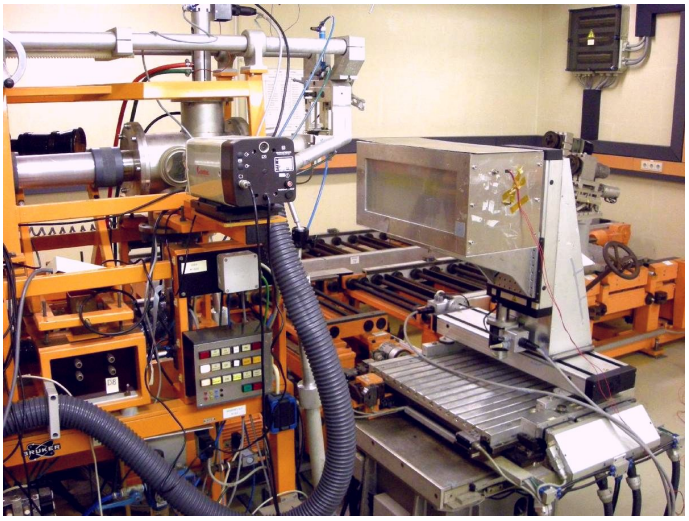


Figure B.5: Picture of the irradiation sample box. (Photo: ETP)

B.2 X-ray irradiation at ETP

In characterization of sensors X-rays are used as signal source. A wide spectrum is produced by an X-ray tube and then converted to monochromatic X-rays by sending them onto a target made from a single element, which in turn emits characteristic X-rays radiation. The irradiation facility is located and maintained by the Institute of Experimental Particle Physics (ETP)[6]. The used tungsten tube uses an acceleration voltage of typically 60 kV. The current can be varied between 2 and 60 mA. The energy spectrum (Figure B.7 [121]) of the emitted X-rays is modulated by the Vanadium window.

[4] https://www.etp.kit.edu/english/irradiation_center.php
[5] www.zyklotron-ag.de/en/
[6] https://www.etp.kit.edu/english/irradiation_center.php

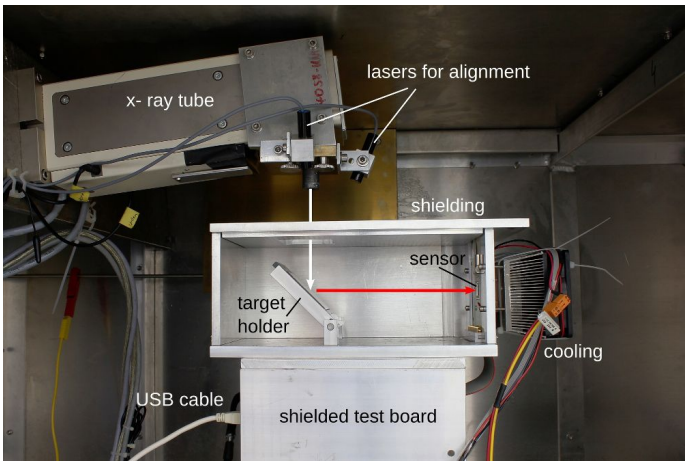


Figure B.6: Picture of the X-ray irradiation setup. (Photo: ETP)

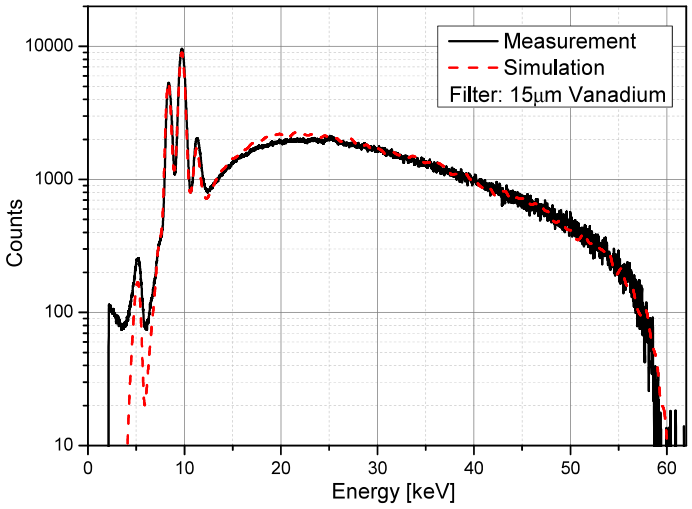


Figure B.7: Measured and simulated spectrum after a 15 µm thick V filter. (From [121])

B.3 Laser setup of KIT ADL

There are three common ways to generate signals in HV-CMOS sensors.

High energetic charged particles from sources or accelerators interact with the sensor material when passing through it. This is the mechanism, for which most HV-CMOS sensors are developed for. The time of impact can be referenced by scintillators or reference detectors. The latter can also provide spatial information.

Photons with energies that, when being absorbed by silicon, cause signals in the relevant range, can be produced by X-rays tubes or sources. Both options can neither provide timing nor spatial information.

Electrical test signal injections in turn provide a localized signal with defined amplitude and good timing information, but it is substantially different from the actual use case.

Another option is a pulsed laser with near optical frequency. The laser can be activated on request for a very short time or run freely with fix frequency. The behavior is similar to the test signal injections. The beam can be focused to few micrometers beam diameter.

In order to achieve measurements with micrometer resolution the entire setup has to be controlled with this or better precision. At the same time safety has to be maintained as the used lasers have the potential to do damage to the human eye. Figure B.8 shows a

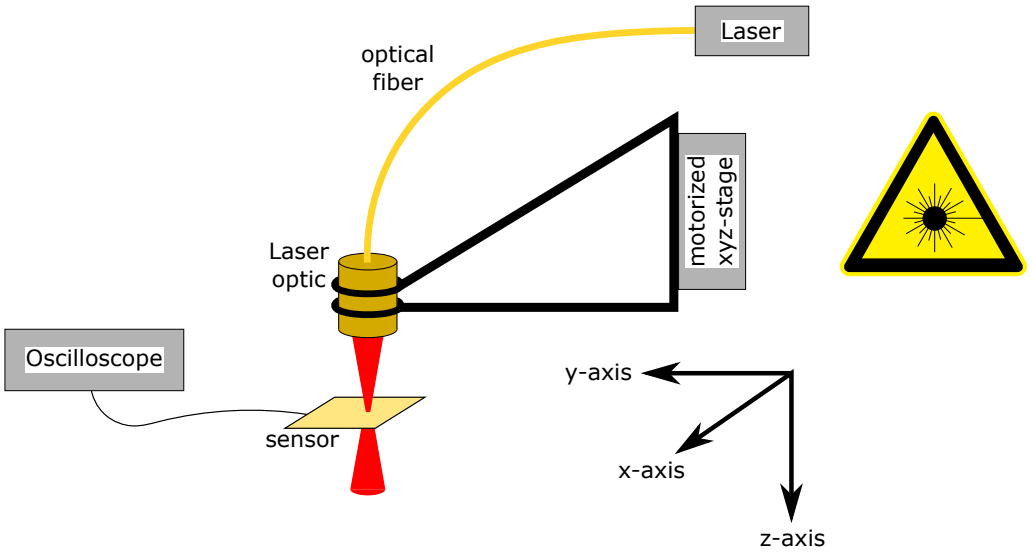


Figure B.8: The laser setup developed at KIT-ADL. The axis shown are the logical axis of the stages.

sketch with the main components. All are located in a portable aluminum box with lock to restrict access. The laser is fed through an optical fiber into the collimator, with the focal point 11 mm below. The collimator is mounted on an *xyz*-stage by a 3D-printed arm, which also holds an optical cross-laser for aiming (not shown). The precision stages move the laser along all three axis with micrometer-precision, relative to the device under test, which is fixed with screws to the same aluminum profiles as the stages.

A software package has been developed by R. Schimassek to control the motorized stages and the laser. It can be easily included in the sensor control software, in order to synchronize measurement and laser movement.

C Used laboratory tools

The list of used laboratory devices is not closed, but contains the most important tools.

Power and voltage supply The analog Toellner sources are the standard supply for all components of the system. For IV-curves or depletion voltages >50 V, the Keysight smart measurement unit (SMU) is used.

Manufacturer	Type	Outputs	Max. Voltage	Comment
Toellner	TOE8733	3	16 - 48 V	low fluctuations
Toellner	TOE8735	5	16 - 32 V	low fluctuations
Keysight	B2901A	1	210 V	SMU

Oscilloscopes Used for waveform acquisition and triggering. Built in high level functions allow fast data analysis with limited resolution. For precision measurement, the triggered waveforms are transferred to a computer. In combination with a scintillator, oscilloscopes can be used to generate reference time stamps. The compared to a pixel large area scintillator provides too many signals for Tektronix oscilloscopes, as their repetition rate is low. For this purpose, an ASIC based Keysight oscilloscope was used, which is capable of handling 1 million waveforms per second.

Producer	Type	Bandwidth	Sampling	Comment
Tektronix	DPO5054	500 MHz	5 Gs/s	built in high level functions, e.g. histograms
Tektronix	MSO5204B	2 GHz	10 Gs/s	built in high level functions, e.g. histograms
Tektronix	MDO3024	200 MHz	2.5 Gs/s	
Keysight	MSOX3054T	500 MHz	5 Gs/s	up to 10^6 wfm/s possible, required for triggering.

Thermo-logger The testo 176T4 can log the temperature measured by up to four wire-connected, remote probes. The data are stored in a log file. This allows for simultaneous temperature measurement at several points of a setup and precise offline matching to temperature-dependent measurement data.

D Other experiments at LHC

D.1 The Compact Muon Solenoid (CMS)

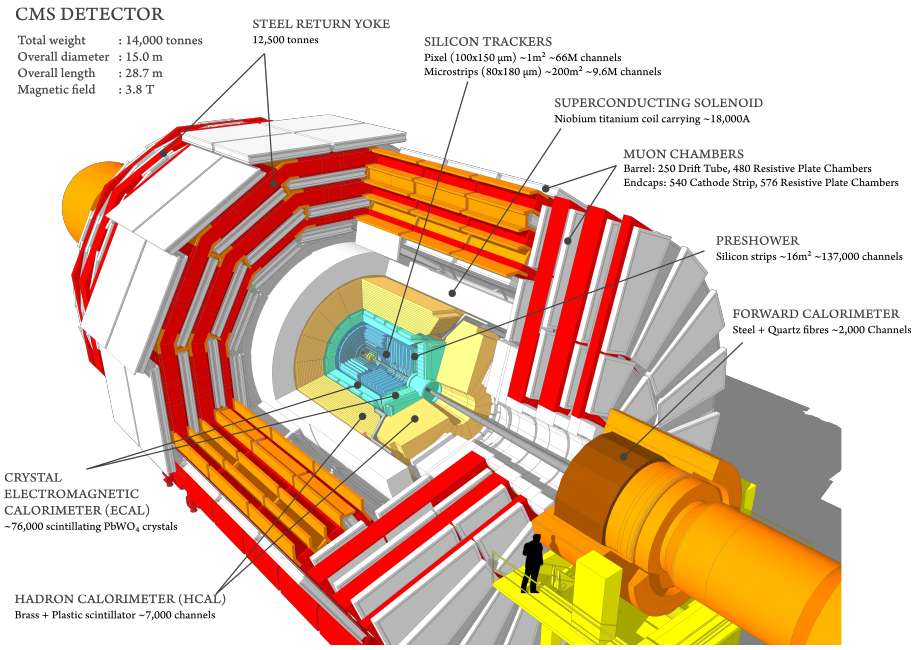


Figure D.9: Schematic drawing of the CMS detector and its main components. (From [127])

The Compact Muon Solenoid (CMS) is the second general purpose detector at LHC (Figure D.9) [43]. The main components are the same as in ATLAS: inner tracker, calorimeters and muon detector. With its 21 m length and 15 m diameter it is significantly smaller than ATLAS but has about twice its weight. The major conceptual difference is the magnetic field of CMS. With its almost 4 T, it is significantly stronger than the one of ATLAS, increasing the curvature of charged particle tracks by Lorentz force.

D.2 A Large Ion Collider Experiment (ALICE)

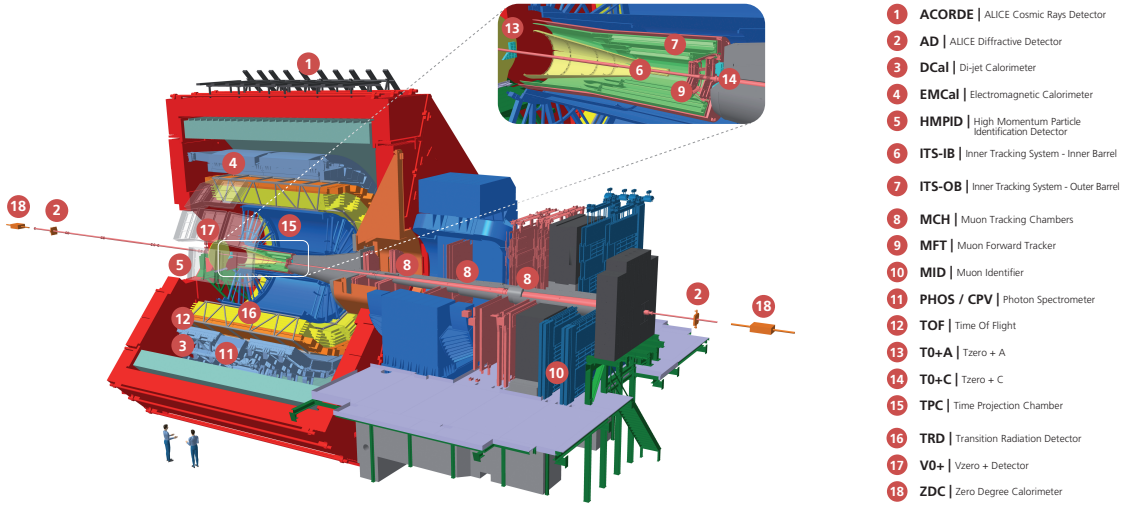


Figure D.10: Schematic drawing of ALICE. (From [128])

A Large Ion Collider Experiment (ALICE) has been designed to study mainly the collisions of high energetic lead ion beams (Figure D.10) [44]. It consists of an barrel tracker, a time

of flight detector, Cherenkov detectors and calorimeters. ALICE is not built symmetrically around the interaction point, some sub-detectors are only build in one direction around the beam pipe.

D.3 Large Hadron Collider beauty (LHCb)

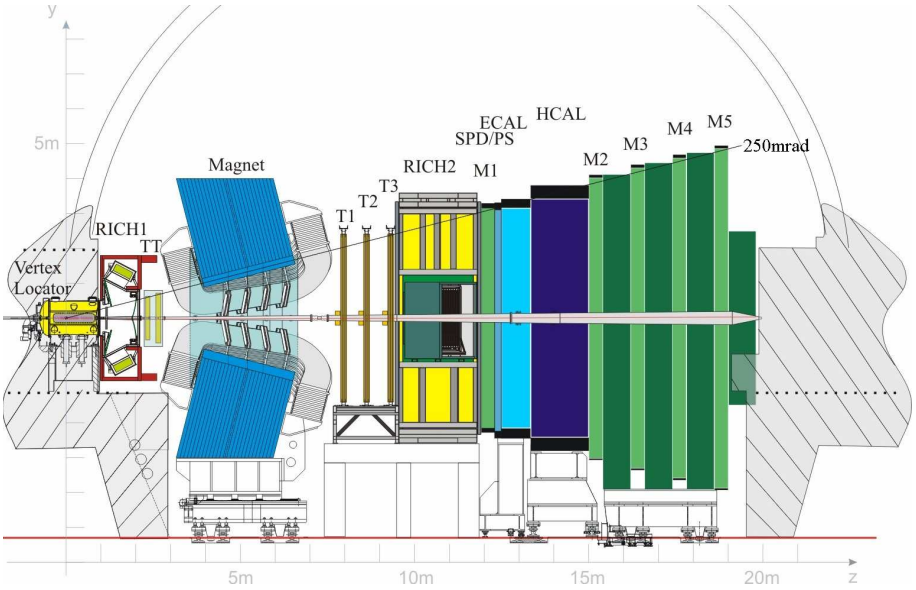


Figure D.11: Schematic drawing of the LHCb detector. (From [129])

Large Hadron Collider beauty (LHCb) is another special purpose experiment at LHC (Figure D.11) [45]. As its name indicates, is LHCb designed to investigate beauty hadron physics. The investigated events happen mainly in the forward region, this is reflected in the detectors design, which is built around the beam pipe in one direction. Around the interaction point, the vertex detector is located. On one side are the other components of the detector's sub-system: Cherenkov detectors, wire-chamber based tracker, calorimeters and muon chambers.

E Additional H35Demo measurements

This section shows some additional graphs and plots of H35Demo matrix measurements. They are explained and interpreted in chapter 6.

E.1 Detailed circuits

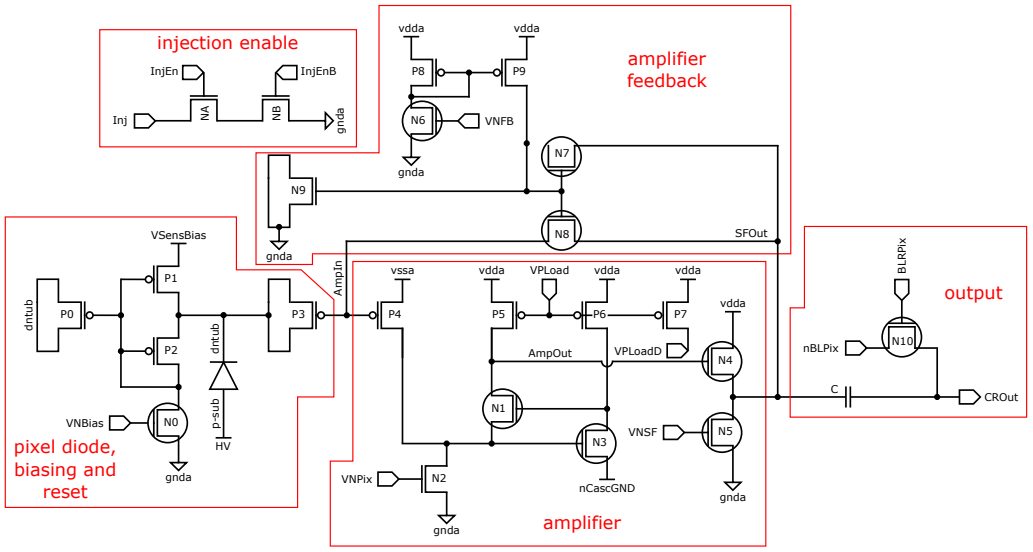


Figure E.12: The front-end of all pixels on H35Demo is the same. The signal is generated by a reverse biased diode formed between p-substrate and n-well. The charge sensitive amplifier converts the charge into a voltage signal. Its feedback can be implemented with linear or enclosed transistors. The output is capacitively coupled to the adjacent electronics line. Instead of a particle, the charge signal can also be generated by a voltage pulse coming from the injection line. (Schematic from [70], modified)

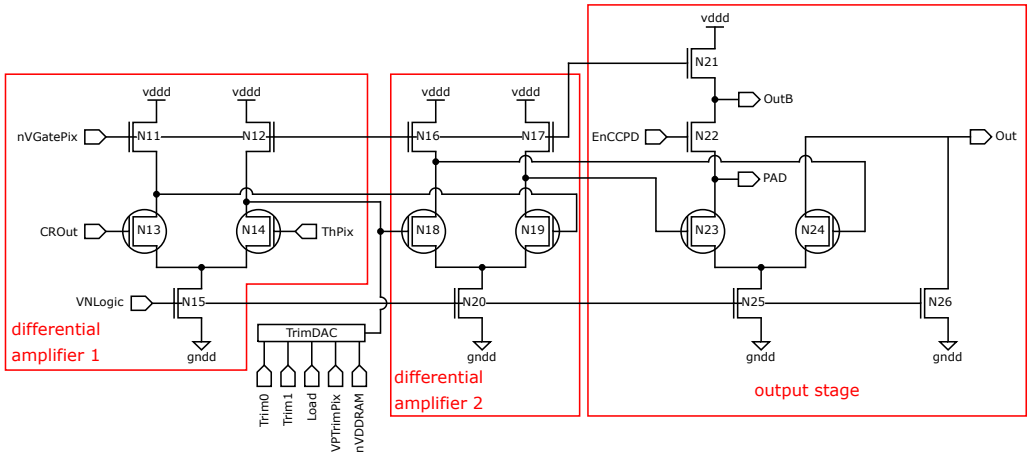


Figure E.13: Detailed schematic of the normal comparator of the NMOS matrix. It consists of two consecutive differential amplifiers and an output stage. The local effective threshold can be altered by a trimming circuit to compensate for behavioral variations. (Schematic from [70], modified)

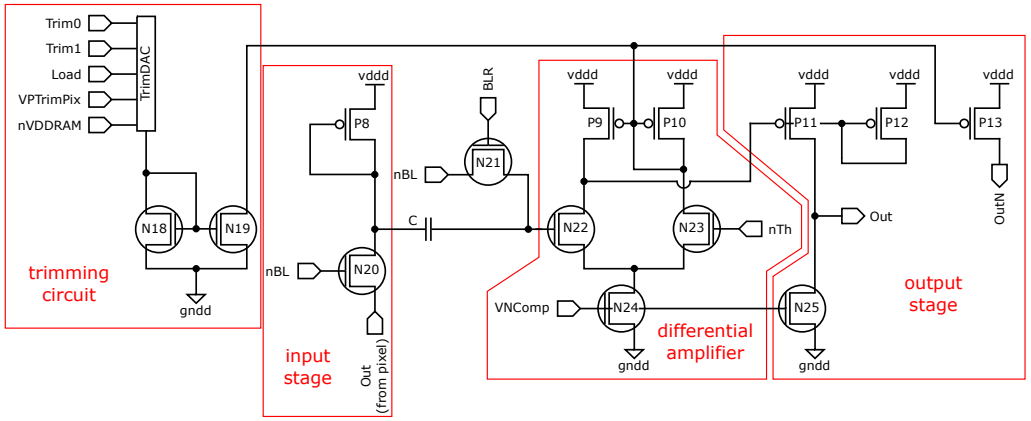
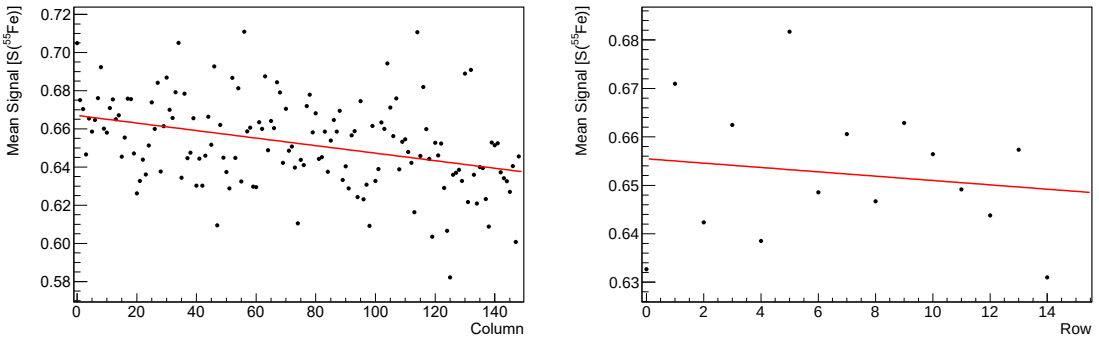


Figure E.14: The input from the pixels to the CMOS comparator in the periphery of the CMOS Matrix is AC-coupled to a CMOS comparator. The effective threshold can be adjusted by the trimming circuit. (Schematic from [70], modified)

E.2 Matrix effects in NMOS Matrix



(a) Mean signal in each column

(b) Mean signal in each row

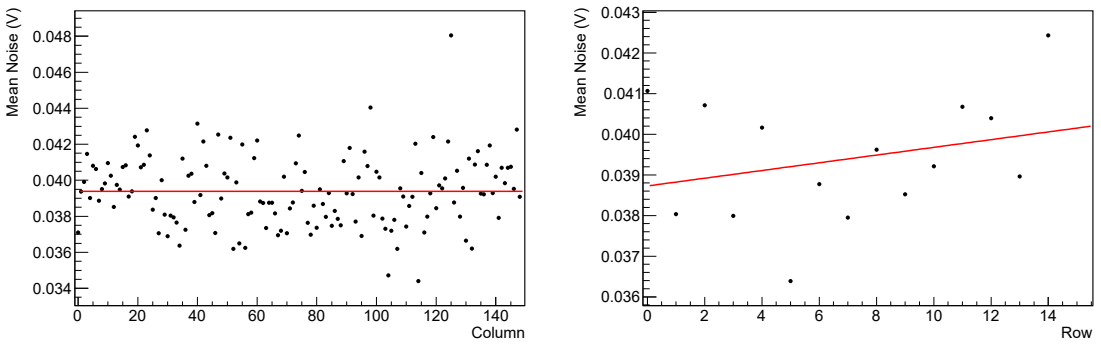
Figure E.15: The mean signal of a column in the NMOS Matrix drops with increasing distance to the bias block, which is located close to column 0. A row dependency can not be identified, probably due to the limited number of rows.

Column and row effects are displayed in figure E.15. With increasing distance to the bias block, meaning increasing column number, the detection threshold increases. This is in agreement with the findings on the Analog Matrices. The increased detection threshold suggests a decreased signal.

The rows are not long enough to identify a trend of the detection threshold, because the uncertainty of each data point is large. Looking at the data available, no row dependency can be identified.

In figure E.16a the average noise in a column is shown. Their distribution is completely flat. The row dependency in figure E.16b is without significant slope, either.

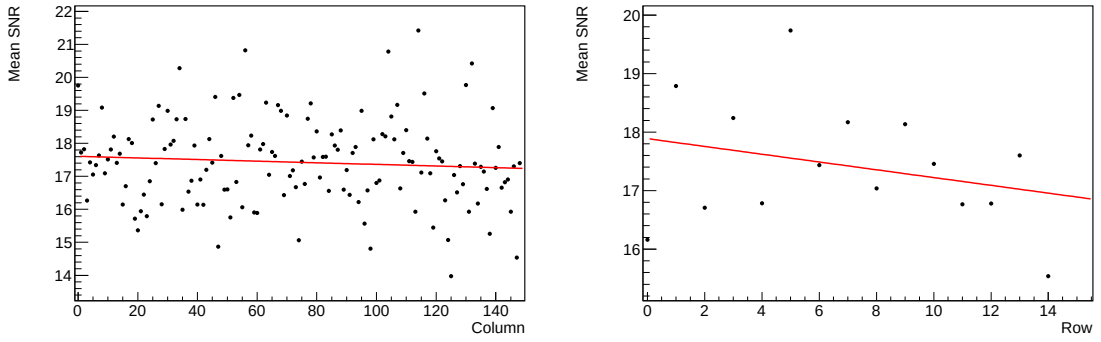
In contrast to the signal, but in line with the noise, the SNR shows no trend in column or row direction (Figure E.17).



(a) Mean noise in each column

(b) Mean noise in each row

Figure E.16: The noise in the NMOS Matrix is constant across the columns. A row dependency can not be identified with the necessary certainty due to the limited number of rows.



(a) Mean SNR in each column

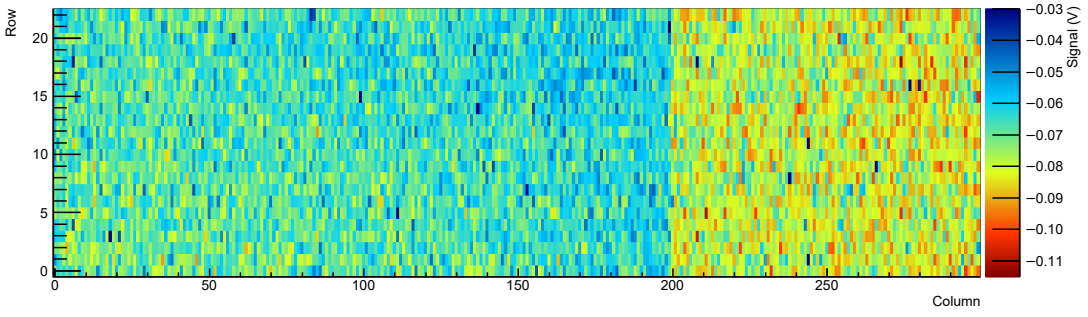
(b) Mean SNR in each row

Figure E.17: A change of SNR in column or row direction can not be identified. Both directions show no trend.

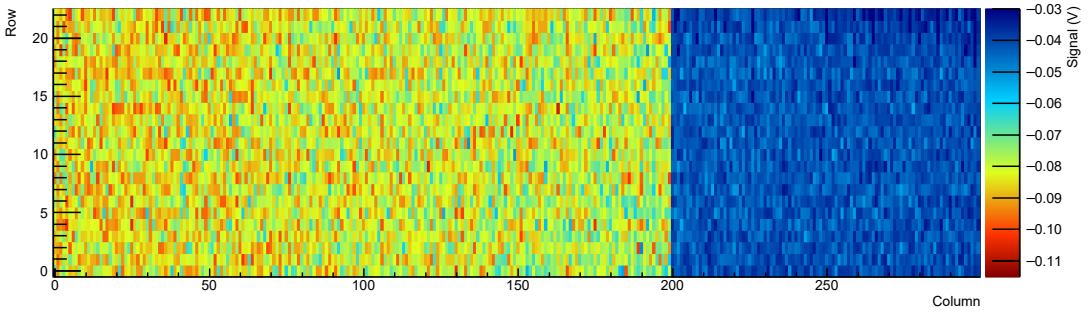
E.3 Mismatch and matrix effects of Analog Matrices A and B

The Analog Matrices do not have a standalone readout, but the second stage amplifier's output of each pixel can be connected to a monitor line. The production variations are measured by illumination of the H35Demo with X-rays from ^{55}Fe decays. An oscilloscope is connected to the monitor lines and creates a histogram for each pixel. This measurement has to be done fully automated, because the oscilloscope can only create one histogram at a time and activity of the used source is small. The time to create a single histogram with a reasonable number of entries is significant. Consequently, the total measurement time is very long. Three channels of the oscilloscope are connected to the three monitor lines of an Analog Matrix. The control software configures both the H35Demo and the oscilloscope that output of the chip and input of the oscilloscope match. Once sufficient signals have been recorded, the histogram is read out and saved. Then the measurement starts over with the next pixel.

Once the measurement is completed, ^{55}Fe-histograms of all pixels in both Analog Matrix A and B have been created. A simple Gaussian fit is applied to each.



(a) MatrixA



(b) MatrixB

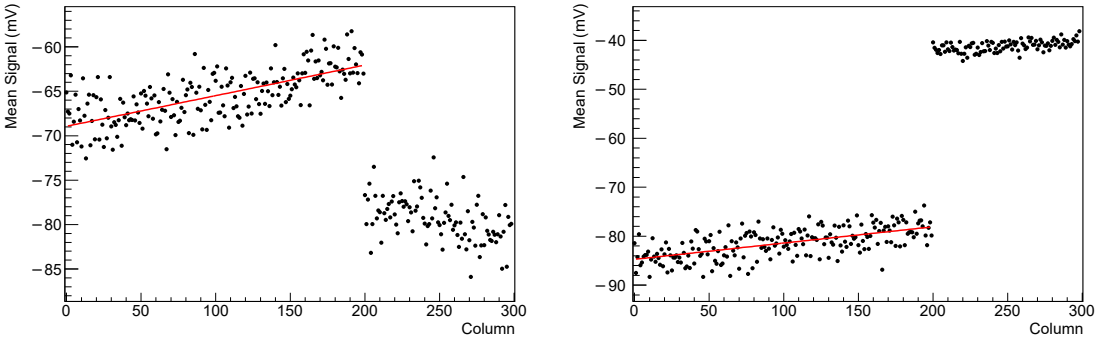
Figure E.18: Both Analog Matrices were illuminated with ^{55}Fe-X-rays and the analog output of the second stage amplifier was histogrammed for each pixel individually. The mean of each histogram is displayed on the z-axis. Note that second amplifier emits negative signals, accordingly the color coding has been inverted.
Each matrix comprises three submatrices. The largest effect of the design variations is visible in the third submatrix of both matrices. Mismatch can only be determined by comparison of pixels within one submatrix.

The mean of the Gaussian fit of each pixel in Analog Matrix A and B is shown in figure E.18. The color-coding is the same to illustrate the difference of the matrices. The used settings are the same.

At first sight, in both Matrix A and Matrix B the behavior of the third submatrix differs strongly from the rest of the matrix.

The second submatrix of each Analog Matrix is realized in baseline design. However, their behavior does not appear the same. It seems plausible that this is because each has its own bias block. Small differences in these blocks would not only explain the difference in performance, but also the difference in power consumption of about 10%. Therefore, it is reasonable to compare the submatrices one, two and three of each Analog Matrix only to each other.

In Matrix A the first submatrix with the additional deep p-tub has a 6% increased signal compared to submatrix two with the baseline pixel layout. This is not siginificant and as the difference between the two first submatrices concerns only the pixel diode, also not expected. The third submatrix, on the other hand, has signals which are on average 25% larger. As these pixels feature linear NMOS transistors, with a higher gain compared to the enclosed ones of the baseline design, this observation was expected.



(a) Matrix A **(b)** Matrix B

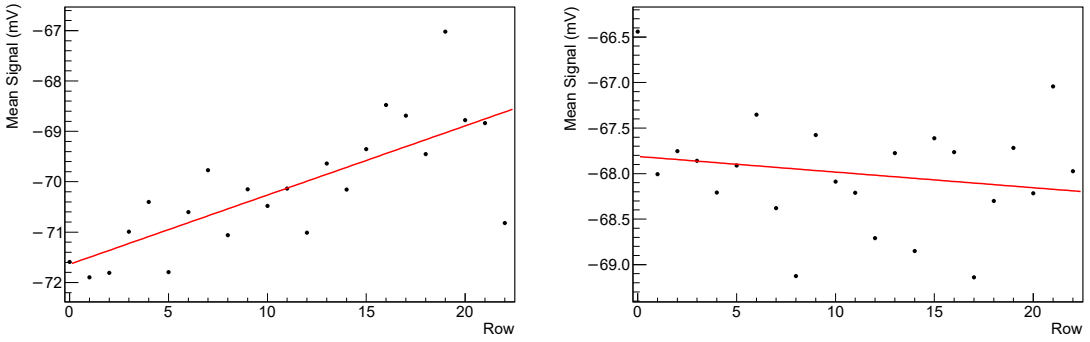
Figure E.19: A 2D-histogram is not sufficient to identify small changes in a given direction. Therefore the mean signal of a column was calculated. This data visualization reveals a dropping signal with increasing column. The reason might be the increasing distance to the bias block.

Matrix B features a deep p-tub in the first submatrix, just as in Matrix A. Again the signals of this matrix are about 5% larger than the ones of the baseline pixel design of submatrix two. As the measured matrices are totally independent of each other, this is unlikely to be a coincidence. The most probable explanation is a voltage drop of the bias voltages. The bias block is closest to the first submatrix, left of the matrices. Therefore, the voltage drop by line loss is smallest for submatrices one and largest for submatrices three. If this hypothesis is true, the signal has to drop continuously with increasing column number. Constant values within each submatrix, but a step between them would contradict this assumption.

This is investigated in figure E.19 for both matrices. The plots show the mean signal of each column comprising 23 pixels. In the first two submatrices (column 0 - 199) the signal is dropping with constant rate. The mean signal of the third submatrix of Analog Matrix A appears constant but rather wide spread, while the mean signal of the third submatrix of Analog Matrix B shows also a drop with increasing column, but on a very low level with reduced dropping rate. This submatrix has pixels with reduced gain in favor of a increased speed. The price for this increased speed is, in this measurement, a by 50% reduced signal.

The identification of a constant signal drop within the submatrices is a clear indication for a bias voltage drop in column direction. This problem could be solved by improved power distribution, e.g. broader and more supply lines. An independent supply of each submatrix might reduce the drop as well.

In row direction, Analog Matrix A shows additionally a continuously decreasing mean



(a) Matrix A **(b)** Matrix B

Figure E.20: The graphs show the row dependency of the mean signal. While Matrix A shows a reduced signal at higher row numbers, the signal of Matrix B can be considered constant.

signal, especially in the first two submatrices (see figure E.32). The signal of the third submatrix however, shows no row dependency. Figure E.20a shows the average signal height as a function of the row for all 300 columns in Matrix A. In figure E.20b the same plot is shown for Matrix B. No strong dependency can be identified here, not even for single submatrices.

The second parameter of the Gaussian fit to the histogram of each pixel is the standard deviation. As already mentioned, this parameter is an indicator of noise. Figure E.21 shows the noise of each pixel in both Analog Matrices. The overall appearance is the same as in the previous signal height 2D-plot. Submatrices one and two can not be distinguished, pixels with linear transistors show increased noise (Analog Matrix A, third submatrix), fast pixels have lower noise (Analog Matrix B, third submatrix). This is the same relative behavior of the noise as for the signal.

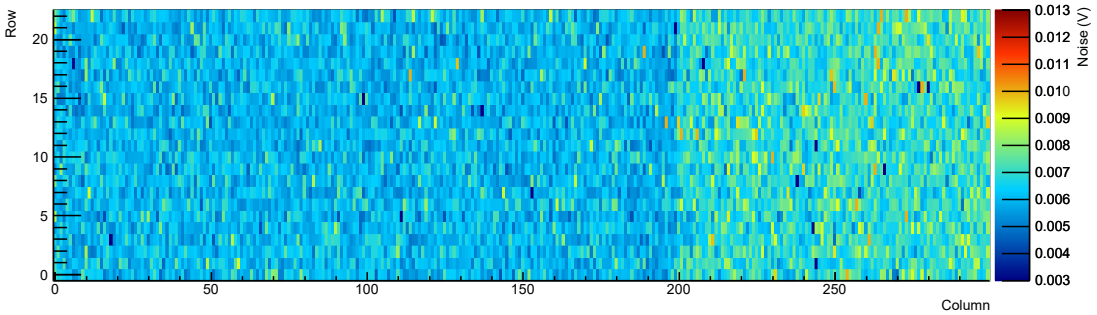
Figure E.22 shows that the noise level is independent of the distance to the bias block within each submatrix. The mean noises of the columns in the first two submatrices are not wide distributed, only few outliers can be spotted. The entries of each third submatrix however, are wider distributed (see figure E.31).

While the noise distribution in row direction of the Analog Matrix A is flat, noise is gaining amplitude with increasing row number in Matrix B. This increase can be spotted in every submatrix individually (see figure E.33).

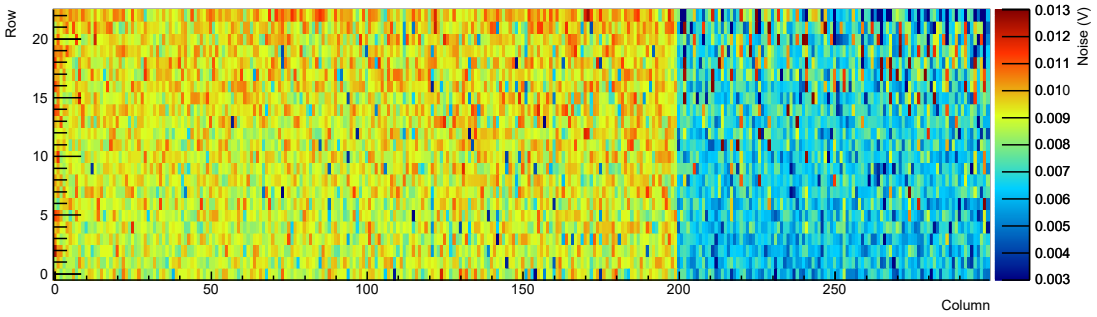
The combination of the signal data and the noise data results in figure E.24, an illustration of the signal-to-noise ratio of both Analog Matrices. The SNR of Analog Matrix A shows no obvious structure. The differences seen in the plots of signal and noise between pixels with linear and enclosed transistors in the feedback circuit have nearly vanished. The gain in signal comes with an increase of noise, at the same rate.

Figure E.25 shows the signal-to-noise ratio of each submatrix as histogram. Both the mean and the spread of all three submatrices of Analog Matrix A are very similar. The increased gain of linear transistors in the feedback loop affects both the signal and the noise in the same way. Linear transistors occupy less space, this is the only remaining benefit compared to enclosed transistors. Most experiments rate the better radiation tolerance of enclosed transistors higher, thus usage of linear transistors in the amplifier's feedback is not advisable.

Analog Matrix B has a homogeneous SNR in the submatrices one and two (Figure E.24b). The SNR of submatrix three is significantly lower, with additional drop in row direction.

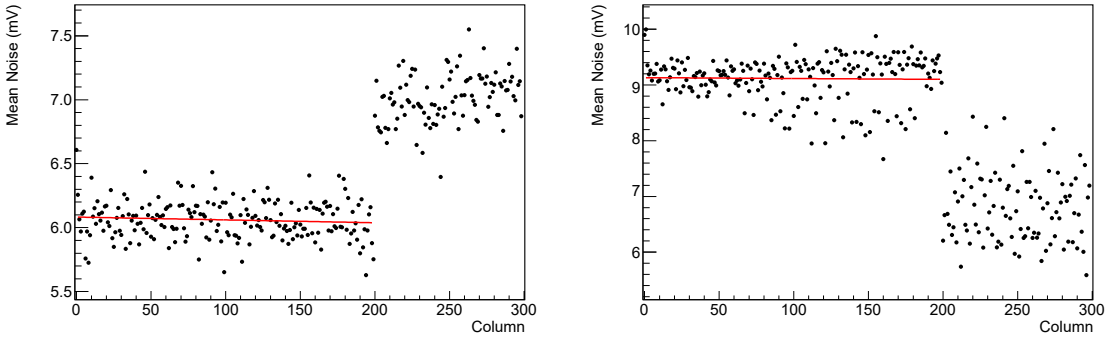


(a) Matrix A



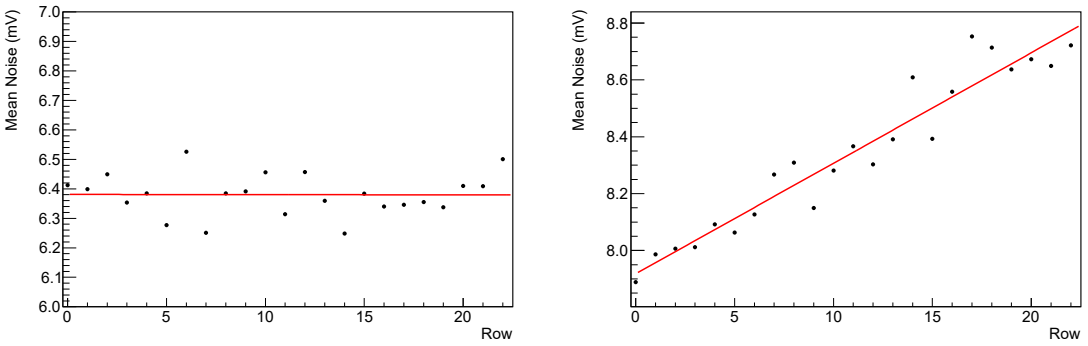
(b) Matrix B

Figure E.21: The analog response to illumination with ^{55}Fe-X-rays has been histogrammed. The standard deviation of each histogram is used as noise of the pixel. Just as in the signal plots, the third submatrix is easily distinguishable from the others.



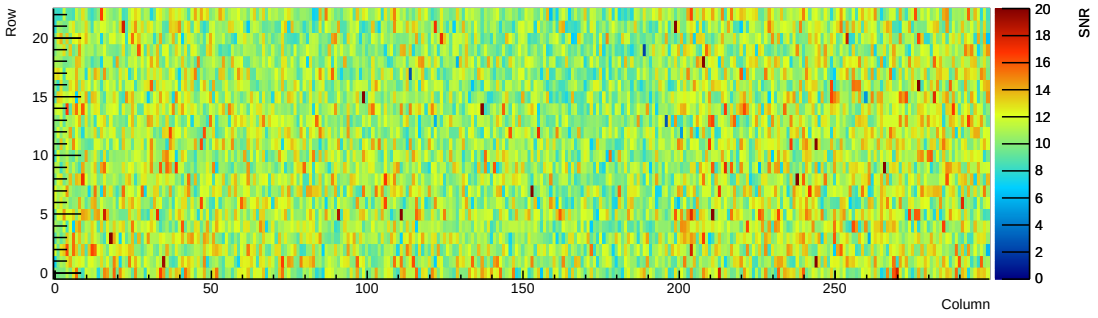
(a) Matrix A **(b)** Matrix B

Figure E.22: Within each submatrix the noise in column direction can be rated constant.

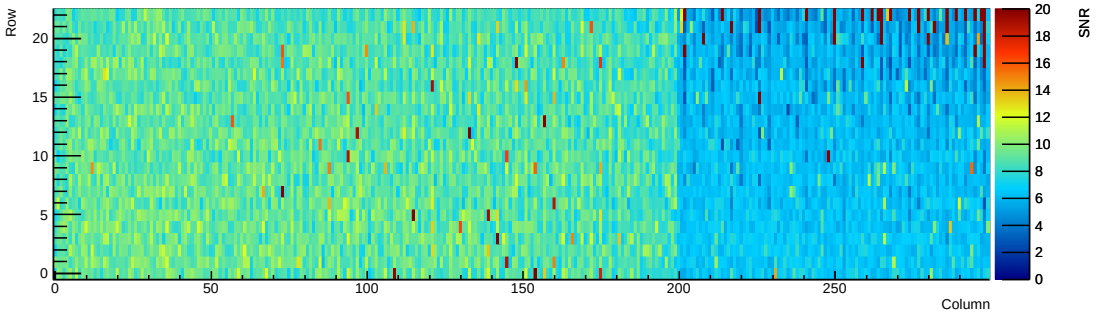


(a) Matrix A **(b)** Matrix B

Figure E.23: The noise of Matrix A shows no row dependency. The noise in Matrix B is gaining amplitude with increasing row.

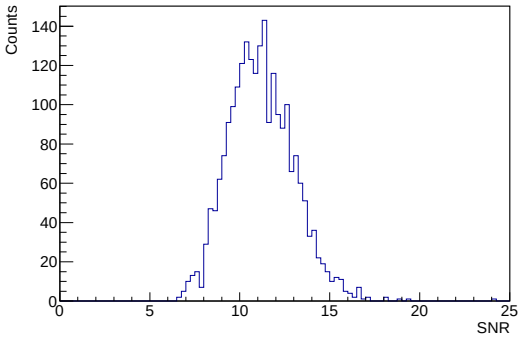


(a) Matrix A

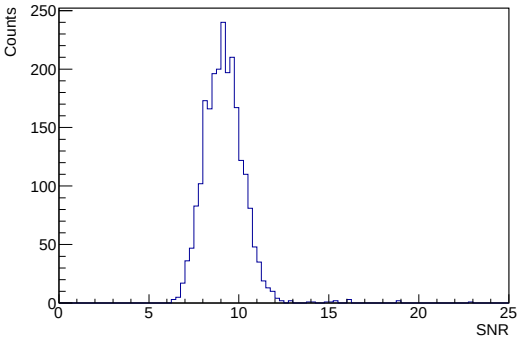


(b) Matrix B

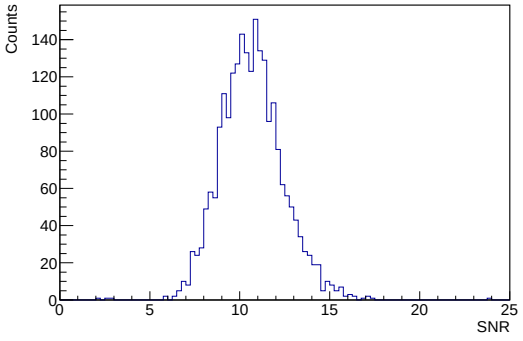
Figure E.24: One of the most important measures of the sensor performance is the signal-to-noise ratio (SNR). For both Analog Matrices SNR is of each pixel is drawn with the same color code. The stronger signal of the third submatrix of Matrix A was compensated by increased noise, resulting in an approximately constant SNR. The third submatrix of Matrix B has a smaller SNR, than the other submatrices, because the noise did not drop by the same fraction as the signal did.



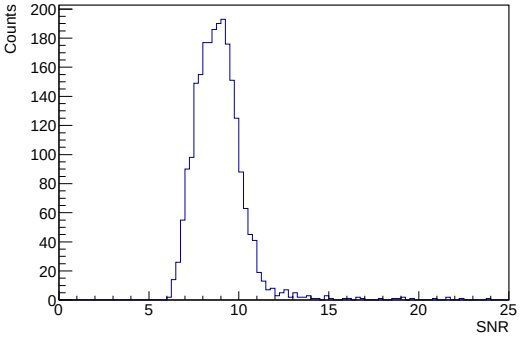
(a) Matrix A, submatrix one extra deep p-tub



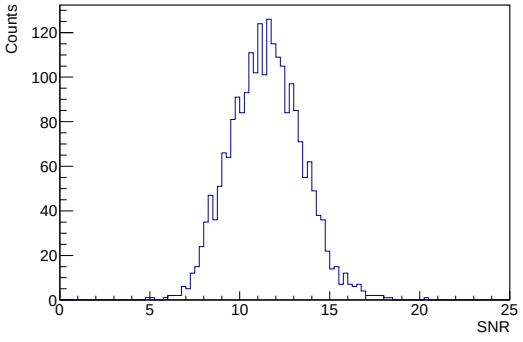
(b) Matrix B, submatrix one extra deep p-tub



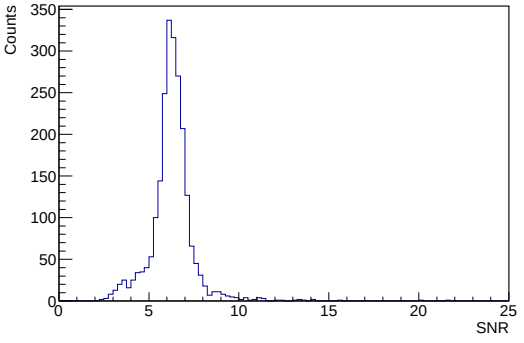
(c) Matrix A, submatrix two baseline design



(d) Matrix B, submatrix two baseline design



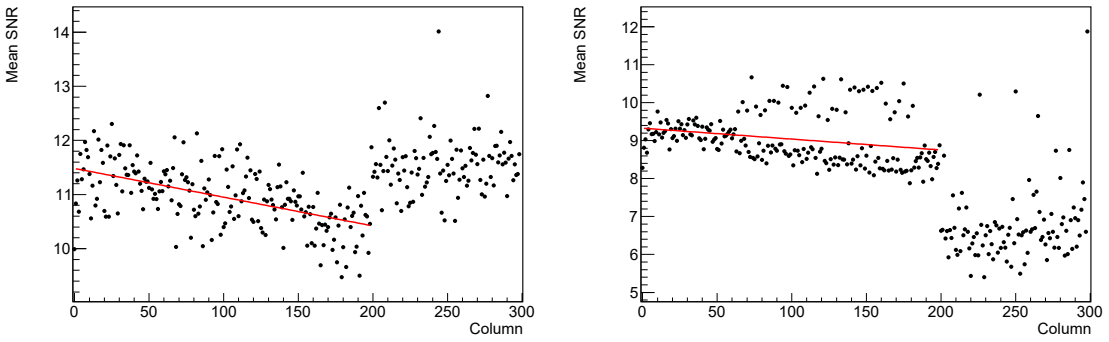
(e) Matrix A, submatrix three linear transistors



(f) Matrix B, submatrix three fast design

Figure E.25: The signal-to-noise ratio is the most important measure to compare the submatrices of each Analog Matrix. The first and second submatrix show a uniform behavior. The linear feedback transistors of the third submatrix of Matrix A have no effect on the SNR. The increased signal is compensated by the increased noise.
The third submatrix of Matrix B however, has a reduced SNR with respect to the submatrices one and two. The reduced gain of the faster feedback design has a small beneficial effect on the noise, but a strong negative effect on the signal.
The differences between Matrix A and B can be traced back to a difference in the main bias block.

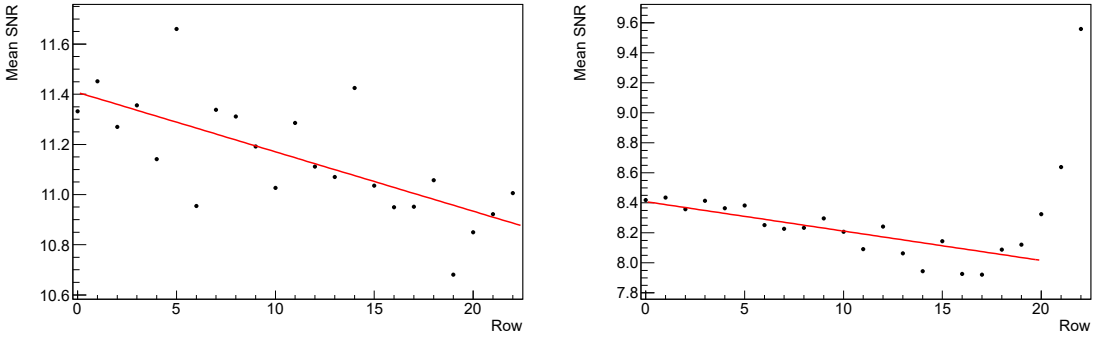
The lower signal in pixels with increased speed is not compensated by lower noise, therefore the SNR is reduced. Figure E.25 shows further, that SNR of the third submatrix is not really of Gaussian shape, but rather a narrow peak resting on a wide pedestal. The SNR of the faster submatrix is about 30% smaller than in the other submatrices. It is unlikely that the gain in speed is worth such a significant SNR drop. However, future designs might increase the overall SNR, rendering this trade-off a good deal.



(a) Matrix A **(b)** Matrix B

Figure E.26: The SNR shows a distinct drop in column direction. Both in Matrix A and B. This is due to the drop of signal along the column axis. The noise remains constant.

As expected from previous observations regarding behavior of signal and noise in column direction, the SNR drops by about 10% from pixels close to the bias block to pixels further away. This is displayed in figure E.26. The behavior is the same in Analog Matrices A and B. However, the mean SNR level in Matrix A is a bit higher, when comparing the first two submatrices, only.



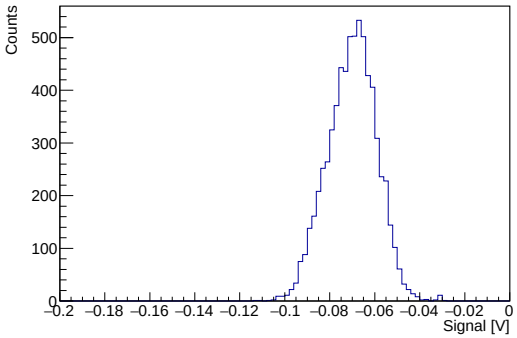
(a) Matrix A **(b)** Matrix B

Figure E.27: In row direction a significant drop in SNR can be identified, even though the row dependencies of signal and noise were inconclusive. Rows 21 to 23 of Matrix B indicate a jump in SNR, but this is just an artifact of failed fits.

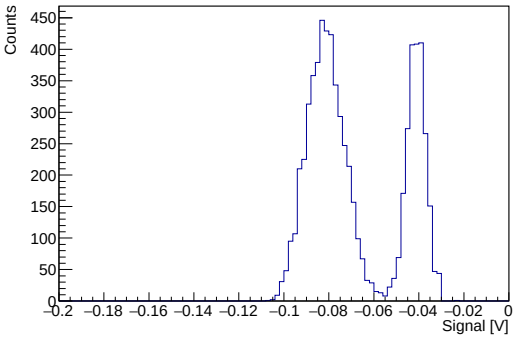
Figure E.27 shows the row dependency of SNR. While the SNR in Matrix A is continuously dropping, the drop in Matrix B is stopped in row 20. The rows 21 to 23 show greatly increased SNRs. Plotting each submatrix individually reveals that this effect has only occurred in the third matrix and is due to a signal rise (see figure E.34). It turns out that this increase is the result of some failed histogram fits in those rows in the third matrix.

Excluding this artefact, the constant drop in SNR remains. It mainly originates from increased noise, not from decreased signals. That might be a hint to the length difference of the signal lines from each individual pixel to the bond pad.

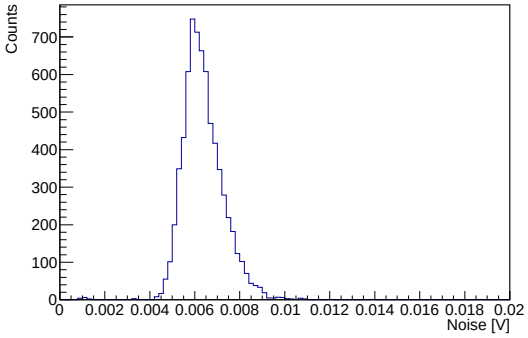
E.4 Signal-, noise- and SNR-histograms



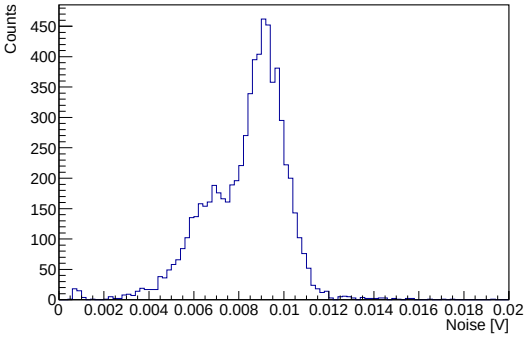
(a) Signal histogram of full Matrix Analog A



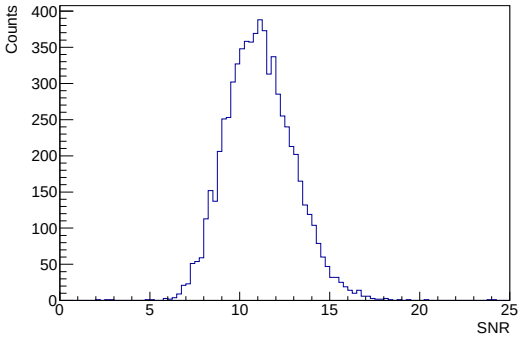
(b) Signal histogram of full Matrix Analog B



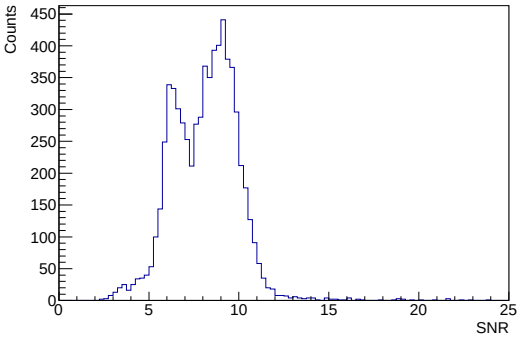
(c) Noise histogram of full Matrix Analog A



(d) Noise histogram of full Matrix Analog B

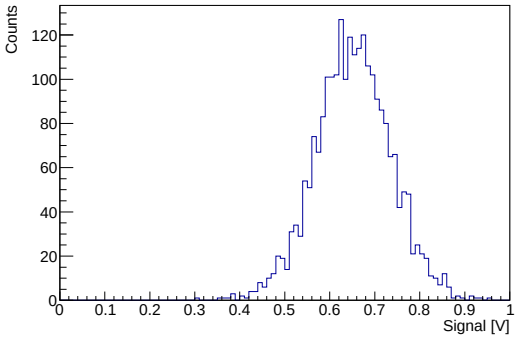


(e) SNR histogram of full Matrix Analog A

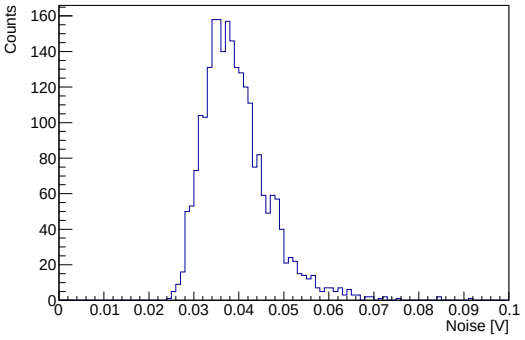


(f) SNR histogram of full Matrix Analog B

Figure E.28: Additional histograms of Analog Matrices A and B: ^{55}Fe signal, noise (signal variation of ^{55}Fe) and signal-to-noise ratio.

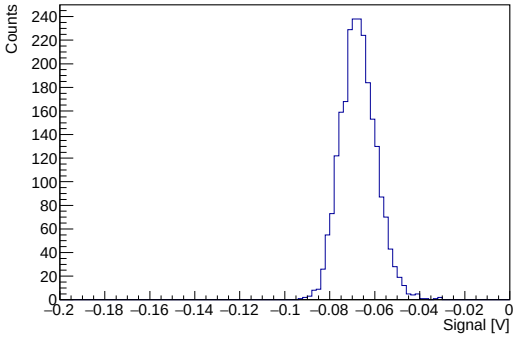


(a) Signal histogram of full NMOS Matrix

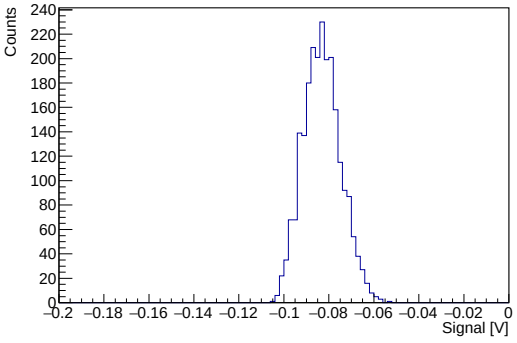


(b) Noise histogram of full NMOS Matrix

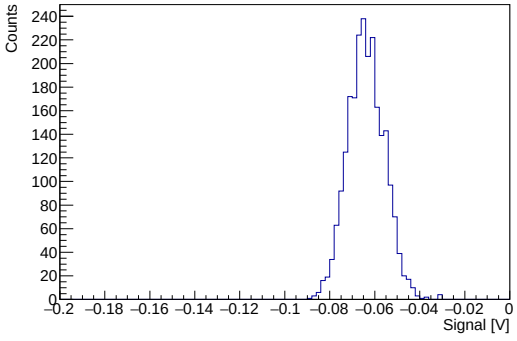
Figure E.29: Additional histograms of NMOS Matrix: ^{55}Fe signal and noise (signal variation of ^{55}Fe)



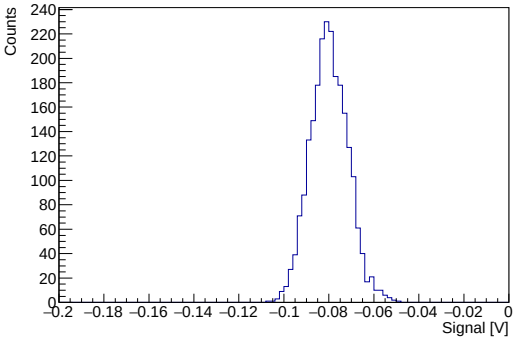
(a) Signal histogram of Analog Submatrix A1



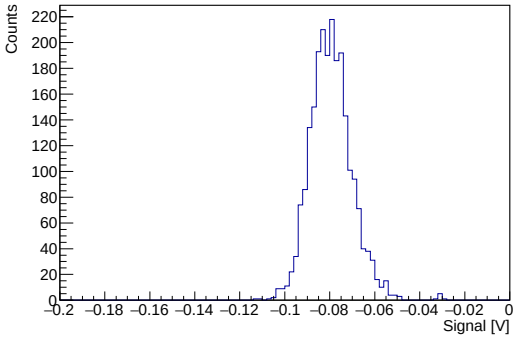
(b) Signal histogram of Analog Submatrix B1



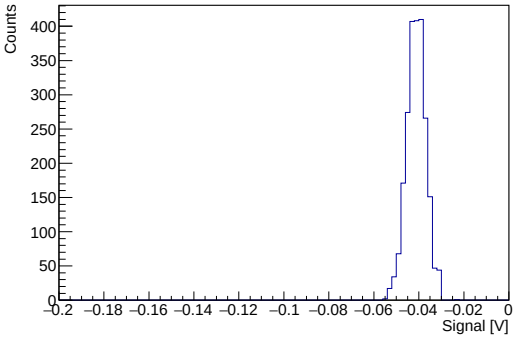
(c) Signal histogram of Analog Submatrix A2



(d) Signal histogram of Analog Submatrix B2

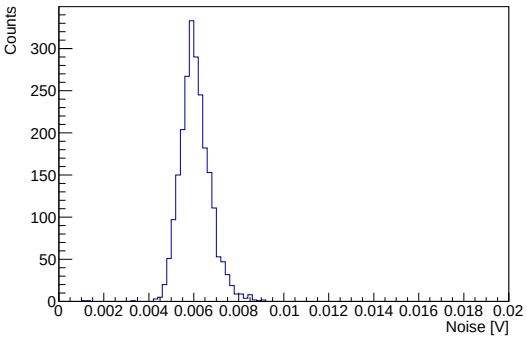


(e) Signal histogram of Analog Submatrix A3

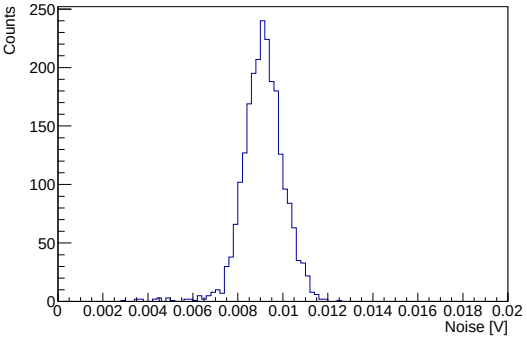


(f) Signal histogram of Analog Submatrix B3

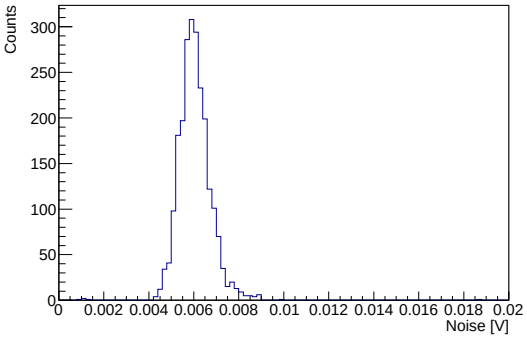
Figure E.30: Additional signal histograms of the three submatrices of Analog Matrices A and B.



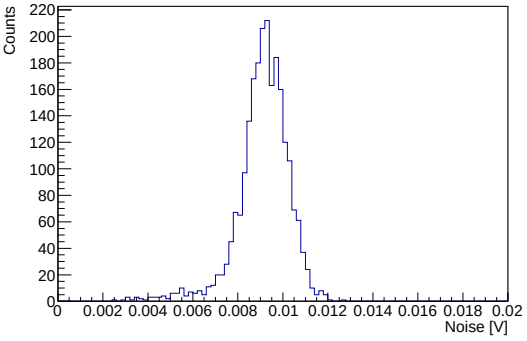
(a) Noise histogram of Analog Submatrix A1



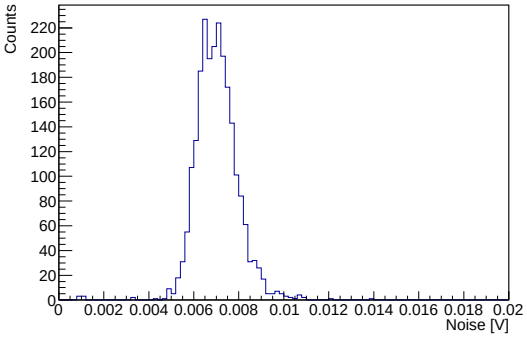
(b) Noise histogram of Analog Submatrix B1



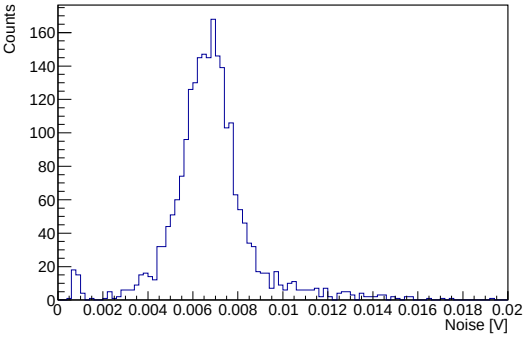
(c) Noise histogram of Analog Submatrix A2



(d) Noise histogram of Analog Submatrix B2



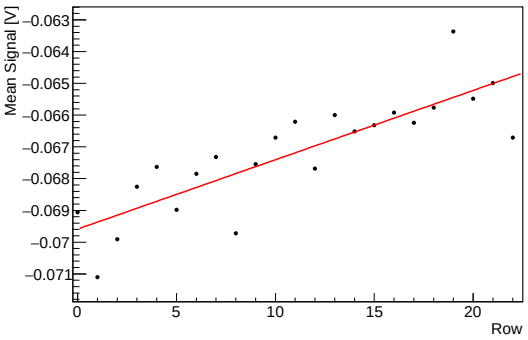
(e) Noise histogram of Analog Submatrix A3



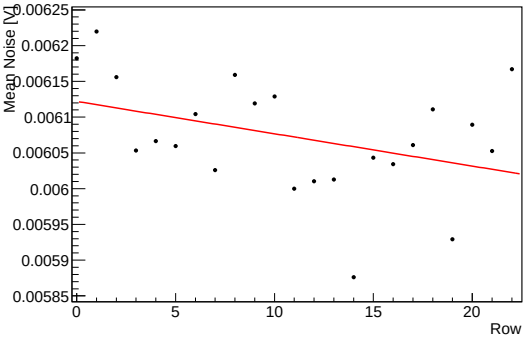
(f) Noise histogram of Analog Submatrix B3

Figure E.31: Additional noise (signal variation of ^{55}Fe) histograms of the three submatrices of Analog Matrices A and B.

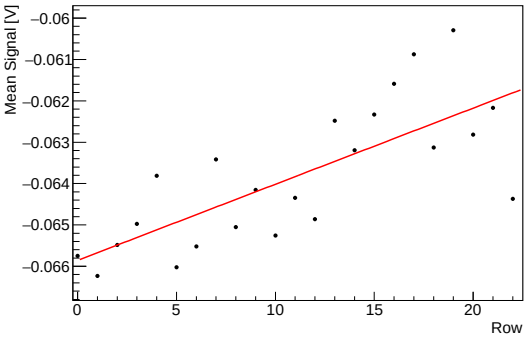
E.5 Variations in row direction



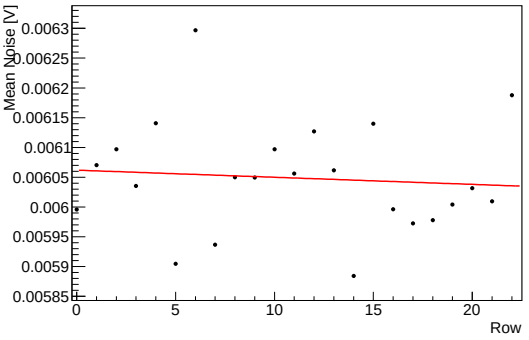
(a) Signal over row in submatrix A1



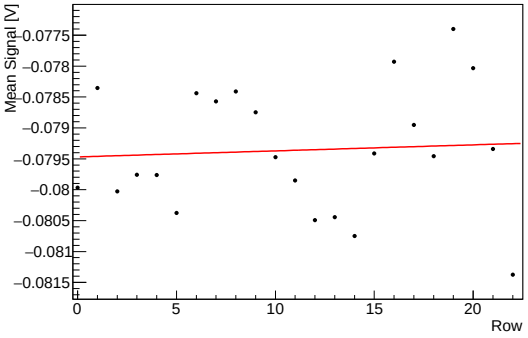
(b) Noise over row in submatrix A1



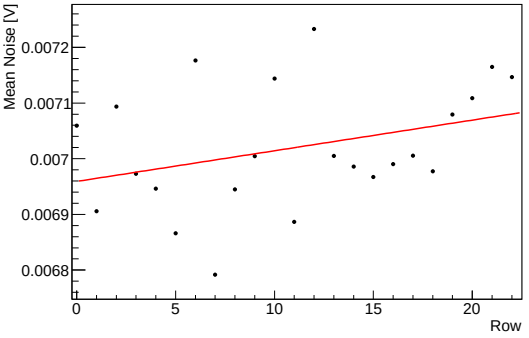
(c) Signal over row in submatrix A2



(d) Noise over row in submatrix A2

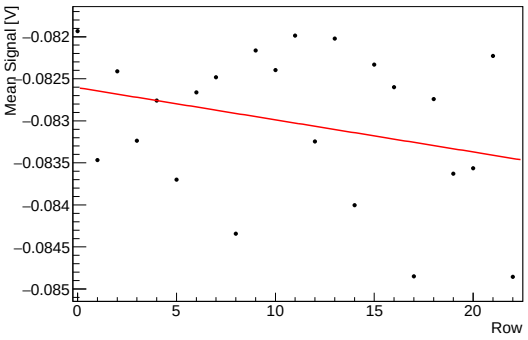


(e) Signal over row in submatrix A3

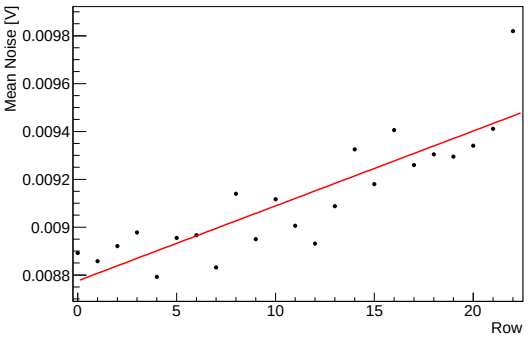


(f) Noise over row in submatrix A3

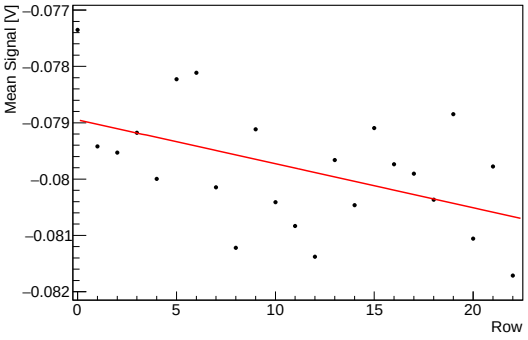
Figure E.32: Row-dependent signal and noise in submatrices of Analog Matrix A



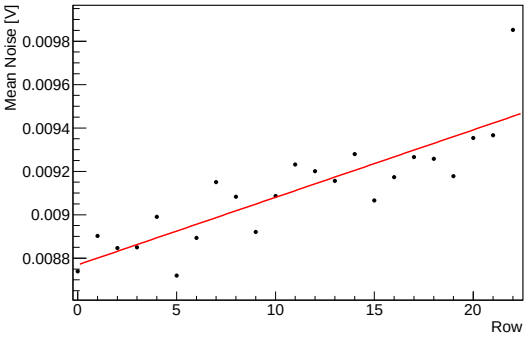
(a) Signal over row in submatrix B1



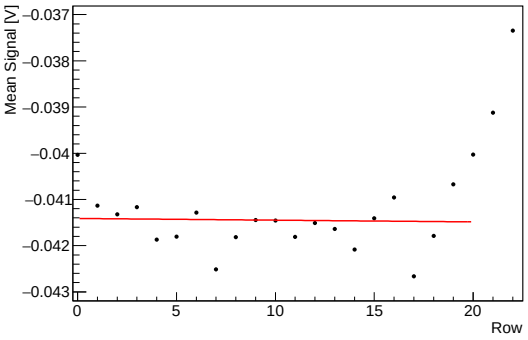
(b) Noise over row in submatrix B1



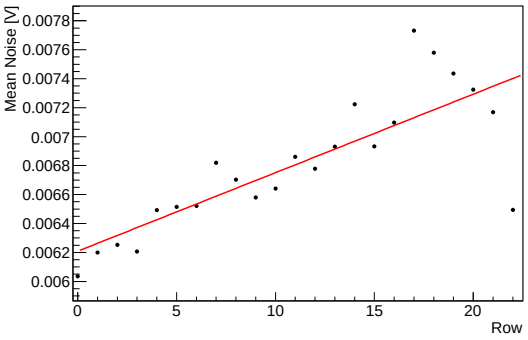
(c) Signal over row in submatrix B2



(d) Noise over row in submatrix B2

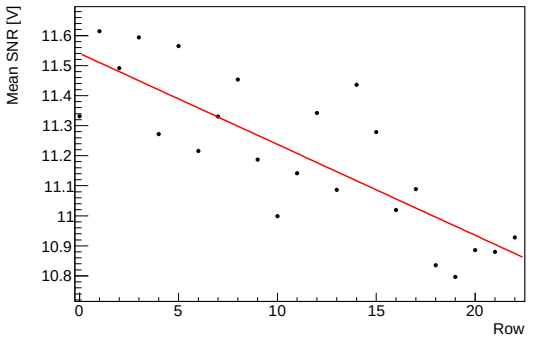


(e) Signal over row in submatrix B3

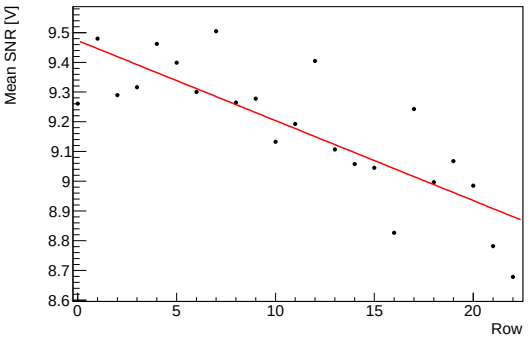


(f) Noise over row in submatrix B3

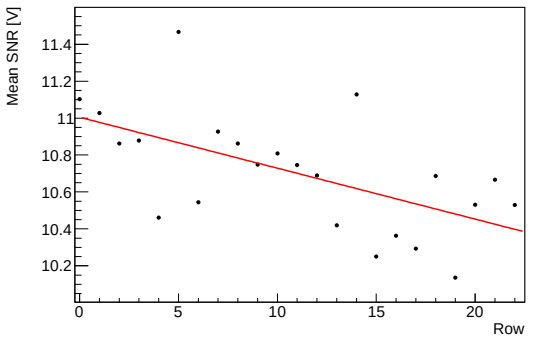
Figure E.33: Row-dependent signal and noise in submatrices of Analog Matrix B



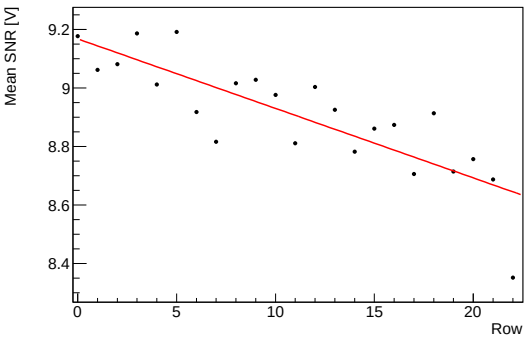
(a) SNR over row in submatrix A1



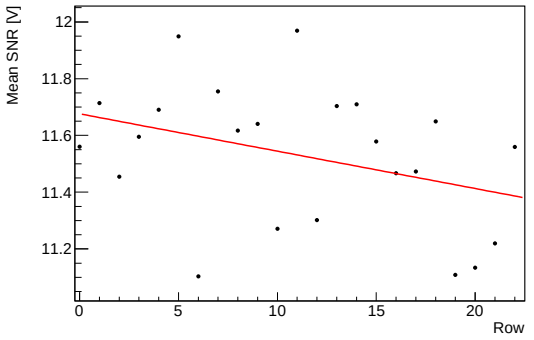
(b) SNR over row in submatrix B1



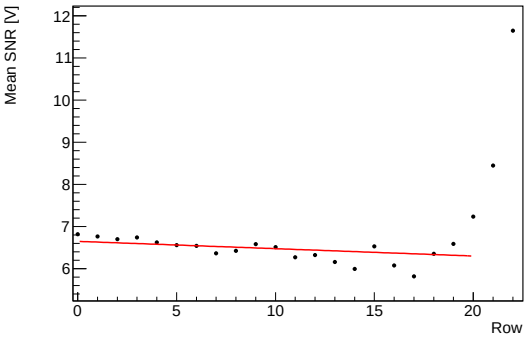
(c) SNR over row in submatrix A2



(d) SNR over row in submatrix B2



(e) SNR over row in submatrix A3



(f) SNR over row in submatrix B3

Figure E.34: Row-dependent signal-to-noise ratio in submatrices of Analog Matrix A and B

E.6 Strontium-90 hit map without collimation

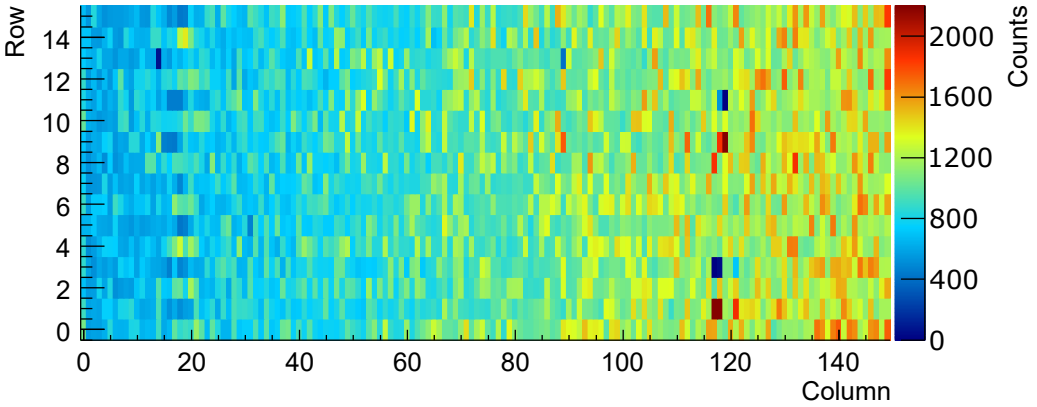


Figure E.35: Hit map of the NMOS Standalone Matrix without time-walk compensation. It has been illuminated with electrons from a ^{90}Sr source without collimation.

E.7 Optimal trimming target determination

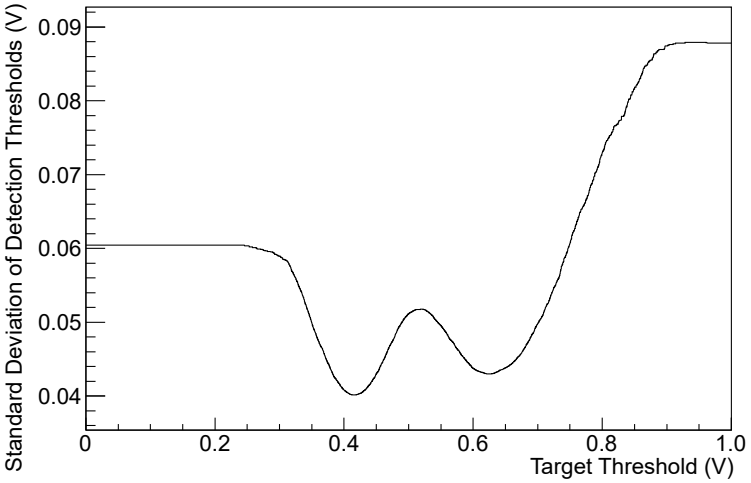


Figure E.36: The standard deviation of the detection threshold as a function of the target threshold is handy to identify the target threshold, which promises the best trimming result. The global minimum of this graph refers to the smallest distribution of detection thresholds. The double minimum indicates, that more trim bits could improve the trimming result. A smaller trim DAC had the same effect, but a rather large number of pixels would be rendered untrimmable.

In order to precisely determine the optimal trimming target threshold, every detection threshold value between 0 and 1 V is tried as trimming target with a step size of 0.1 mV. The result of each try is a detection threshold histogram, such as figure 7.26b or figure 7.27. For each of these histograms the standard deviation is calculated, as this is a good measure of trimming quality. Figure E.36 displays the standard deviation as a function of the trimming target threshold following the standard procedure described above. A low value means a small spread of S-curves, while a large value means a big spread.

It was anticipated that for very low and very high target thresholds the distribution is wide spread (bad trimming result), because only a small fraction of S-curves can be pushed by the trimming values to these extreme detection threshold. Consequently a minimum

somewhere in the middle is to be expected. However, the occurrence of a double minimum is curious.

In order to explain this behavior, two things have to be recalled: Only four trim values per pixel are available. And the spread of the S-curves of some untrimmed pixels is large, thus the *PTrim* settings has to be considerably high.

In the minimum at 0.4 V, nearly 3/4 of all pixels receive the RAM value 1. The second minimum has more than half of the pixel RAMs set to 2. That means, that despite the total spread of all untrimmed S-curves, the majority of all pixels have very similar detection thresholds. Targets which lay between these two minima can not be properly matched by the four given trim values. Depending on which value is chosen, they either exceed or fall below the target by large. A smaller setting of the trim DAC or more trim values could push the local maximum down to a new minimum.

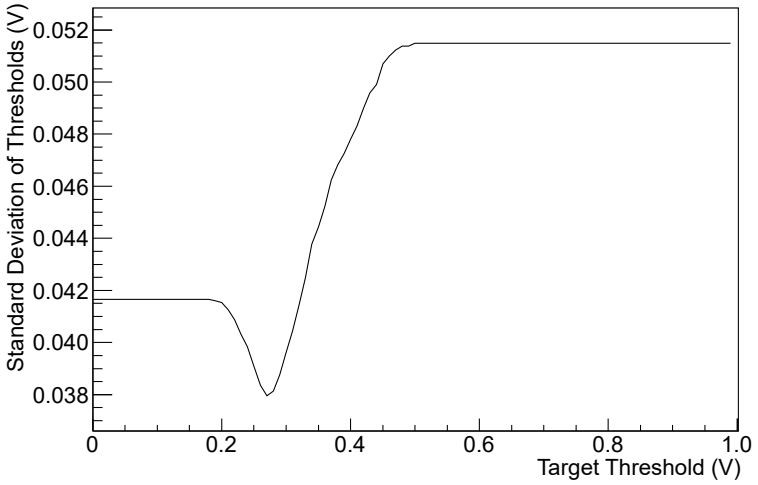
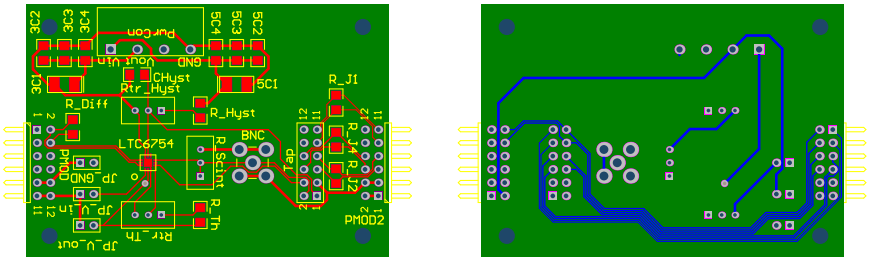


Figure E.37: The setting of the trim DAC was set to 15 for this measurement. Compared to the previous plot, the two local minima have merged into a global minimum, eliminating the local maximum. This is due to a smaller step size of the trim values. The prize to pay for this improvement is a large amount of untrimmable pixels.

Figure E.37 shows the same graph for a measurement with smaller *PTrim* (corresponds to figure 7.27). The local maximum has vanished in favor of a single global minimum. This indicates that the shift applied to an S-curve by changing the trim RAM value by ±1 is sufficiently small.

E.8 High speed comparator for particle triggering



(a) Top **(b)** Bottom

Figure E.38: The layout of the comparator PCB. It translates the scintillator signal to an LVDS signal which is understood by the FPGA.

In order to convert the analog signal of a scintillator and its photomultiplier tube to a signal that can be understood by an FPGA, a commercial high-speed, low latency comparator is used [130]. The layout of the custom PCB supporting the comparator is shown in figure E.38. The threshold, the input resistance and the hysteresis of the output pulse can be adjusted by variable resistors (trimmers). The comparator's LVDS output is fed into the same FPGA that controls the device under test, where the reference time stamp is generated and stored.

F Characterization setup for ATLASpix1 and MuPix8

The characterization setup for ATLASpix1 and MuPix8 is based on the Multi-purpose Adapter Board system, which has been described in the previous section. The bias voltage generation and injection pulse generation is provided by the system. Only an ASIC-specific carrier board is missing to complete the hardware.

The firmware needs a FastReadout module tailored to the readout state machine of ATLASpix1 and MuPix8, which receives the serial data from the sensor. The data are being parallelized and forwarded to buffers for transmission to the computer. The software consists of generic communication and configuration classes, configured for ATLASpix1 or MuPix8, and sensor-specific functions. A graphic user interface is added for convenient operation.

F.1 Carrier PCB

The three matrices of ATLASpix1 (see chapter 11) have been produced on the same reticle as MuPix8 (see chapter 13). The initial dicing plan did not involve separation of ATLASpix1 and MuPix8. Therefore, a combined PCB was required on which both sensors can be tested independently from each other. An individual PCB would result in many lost sensors as for each assembled ATLASpix1 sensor a MuPix8 would be sacrificed and vice versa.

The electrical separation of the matrices should be complete, including supply voltages and configuration, in order to avoid single point failure, even though both sensors need the same or very similar off-chip periphery.

Both ATLASpix1 and MuPix8 have the possibility of internal bias generation. Compared to the earlier H35Demo, this is an entirely new feature, and thus untested. In order to not fully rely on these circuits, their output voltages can be overwritten by external voltages.

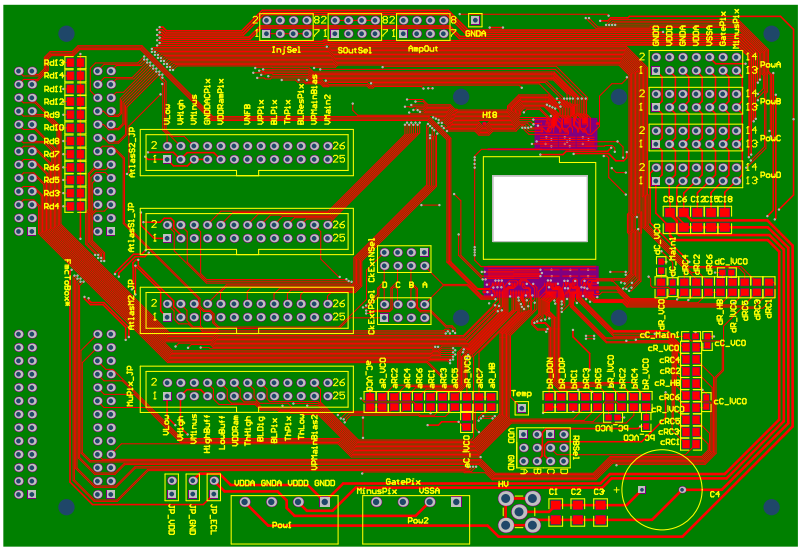


Figure F.39: Top side layout of the combined carrier PCB for ATLASpix1 and MuPix8. Eight jumper arrays allow separation of certain matrices from power and voltage supply.

Figures F.39 and F.40 show the layout of the PCB designed for the two sensors ATLASpix1 and MuPix8. The four sensor matrices are to be glued on top of the hole in the PCB. The hole is necessary to minimize material in the path of particles passing through the sensor. The approximate outline of the sensor is indicated by a yellow line around the hole.

The connection to the rest of the hardware is established via four box headers bottom side. Power connectors can be found on the lower edge. The power lines are not routed to the

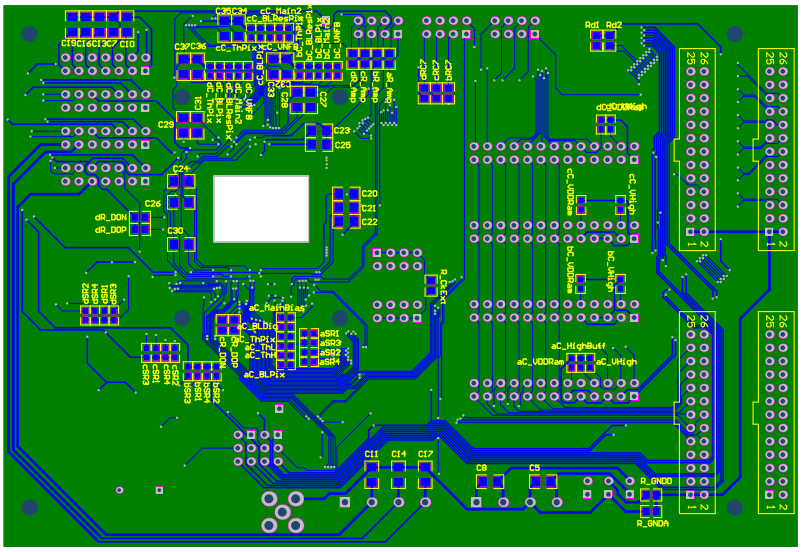


Figure F.40: Bottom side layout of the combined carrier PCB for ATLASpix1 and MuPix8. The connection to the Multi-purpose Adapter Board is established via four box headers on the right.

sensor directly, but over jumpers on the top right side. Removing a jumper, cuts a certain matrix off a certain line.

Four arrays on the left top side are used to override or probe bias voltages generated on any of the matrices.

The top edge pin headers are used to probe debug signals: analog output of the in-pixel amplifier and output of the configuration shift-register as well as jumpers to distribute the test signal injection to a specific matrix.

F.2 Software and firmware

The underlying scheme of both firm- and software have been developed from the H35Demo project. Basically all modules and functions underwent improvements, but the functionality stayed the same.

Figure F.41 shows a simplified block diagram of the ATLASpix1 and MuPix8 software. The biggest novelty is the encapsulation of all sensor and setup configuration properties in classes, which connect the properties of a DAC or configuration bit with a name and generate the configuration bit-vector on request, instead of manipulating the bit-vector directly. This improves readability and speeds up the development of new control software and functions.

Figure F.42 shows a screen shot of main functions and configuration of the ATLASpix1 and MuPix8 control software. Manipulation widgets for DAC and configuration bits, VoltageBoards and InjectionBoard are represented here, as well as the monolithic readout.

The second tab in figure F.43 houses more complex functions, e.g. an automated threshold tuning of the entire matrix and others.

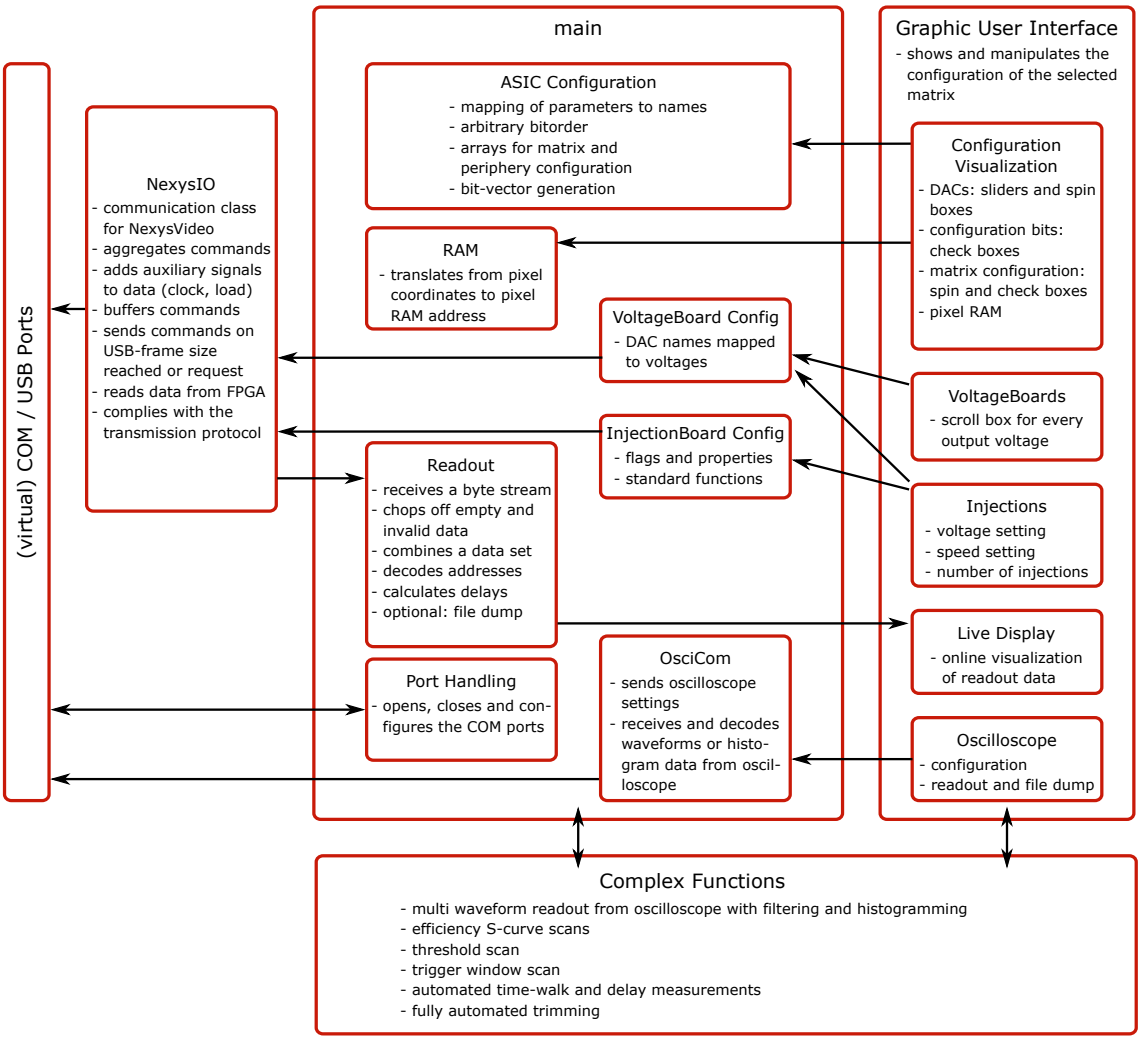


Figure F.41: Block diagram of the computer software. It is an updated version of the diagram shown and described in chapter 6.3 and figure 6.5.

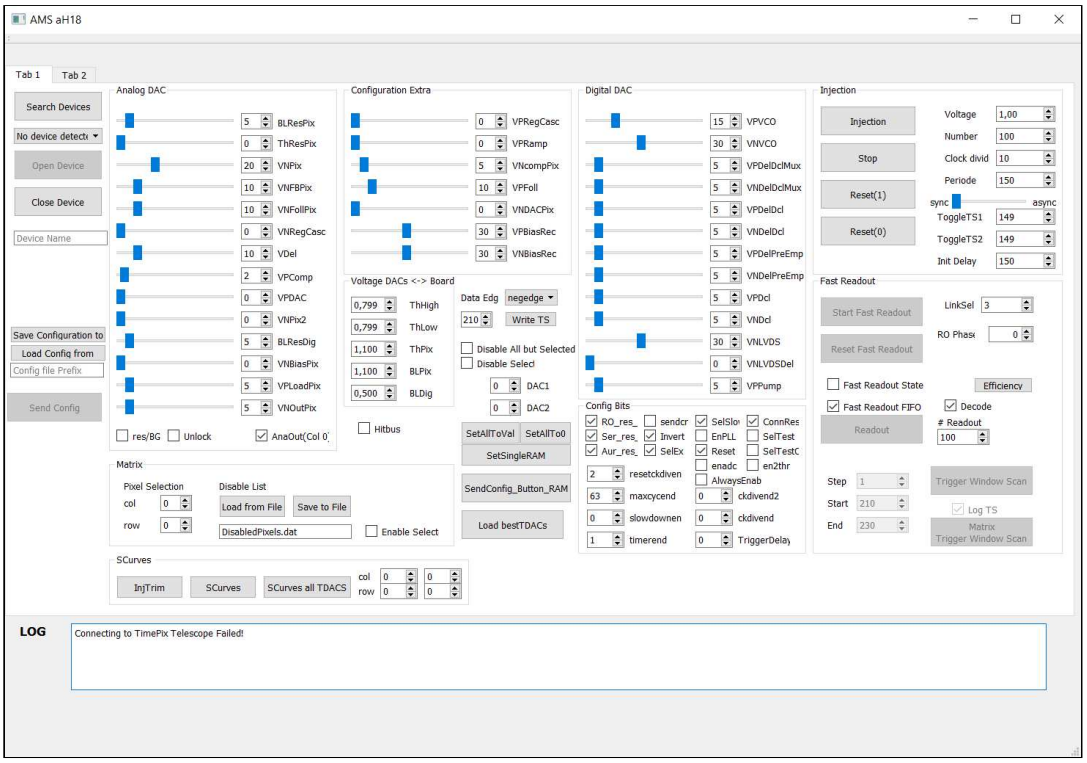


Figure F.42: Screen capture of the control software for the combined project ATLASpix1 and MuPix8. Tab1 contains the basic configuration and MAB controls.

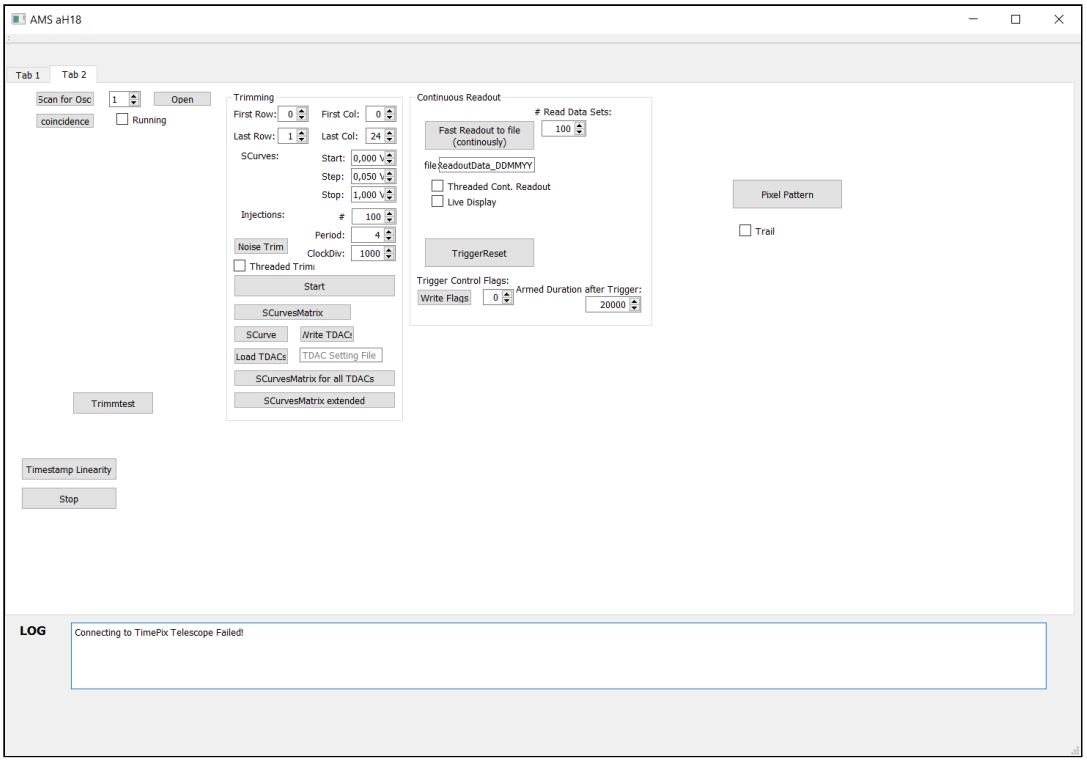


Figure F.43: Screen shot of Tab2 of the control software. Complex functions are located here.

G TRISTAN setup standard configuration

The standard configuration was theoretically determined by simulation of the readout ASIC or the silicon drift diode by FBK respectively:

setting	DAC name	default value
regulated cascode	PRegC	20
amplifier feedback	NFB	1
amplifier load	PLdAmp	20
source follower	NSF	10
amplifier 1	NAmp	40
amplifier 2	NAmp	40

Drift ring voltage is set to −60 V and backside voltage to −80 V. The KIT TRISTAN Integrated Circuit is powered with 3.3 V main voltage (VDDA) and 2.2 V amplifier voltage (VSSA).

H Sensor production costs

The presented cost estimations are based on numbers from actual invoices without taxes. Only the costs for sensor and readout are included. Costs for module assembly are expected to be comparable, additional costs for cooling are excluded.

H.1 Hybrid sensor

The following estimation of costs for a hybrid sensor assembly is based on a planar sensor production for CMS detector ($\approx 3.9\,\mathrm{m}^2$). The respected costs are sensors, readout chips (ROCs) and assembly (bump bonding). Costs are calculated from total per module (1 sensor and 4 ROCs, 16 cm^2) and square meter:

	total	per module	per m^2
planar sensors	2 368 662.11 €	970.76 €	606 726.98 €
ROCs	2 798 672.50 €	830.48 €	519 041.64 €
assembly	3 187 311.33 €	1 584.37 €	990 230.61 €
sum	8 354 645.95 €	3 385.60 €	2 115 999.22 €

H.2 HV-CMOS sensor

The costs of HV-CMOS are estimated by the example of a sensor production for Mu3e ($\approx 2.5\,\mathrm{m}^2$). The three main contributions to the total costs are: raw wafer, implementation and post-processing (thinning and dicing). The costs are calculated from the bill total per wafer, reticle and square meter:

	total	per wafer	per reticle	per m^2
wafer	11 700 €	90 €	1.88 €	4 687.50 €
implementation	169 000 €	1 300 €	27.08 €	67 708.33 €
post-processing	29 250 €	225 €	4.69 €	11 718.75 €
sum	209 950 €	1 615 €	33.65 €	84 114.58 €

Publications

Published

Publications by or with contribution from the author of this thesis. The author list of publications with more than three authors have been reduced to the leading author and the thesis author.

[1] F. Ehrler *et al.*, "Characterization results of a HVCMOS sensor for ATLAS," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 936, pp. 654 – 656, 2019. Frontier Detectors for Frontier Physics: 14th Pisa Meeting on Advanced Detectors.

[2] F. Ehrler, I. Perić, and R. Schimassek, "HVCMOS pixel detectors first measurements on the reticle size prototype for the ATLAS pixel layers," in *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*, pp. 1–8, Oct. 2016.

[3] F. Ehrler *et al.*, "High-voltage CMOS detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 824, pp. 400 – 401, 2016. Frontier Detectors for Frontier Physics: Proceedings of the 13th Pisa Meeting on Advanced Detectors.

[4] I. Perić, F. Ehrler, *et al.*, "High Voltage CMOS Active Pixel Sensor," *IEEE Journal of Solid-State Circuits*, 2021.

[5] R. Schimassek, F. Ehrler, *et al.*, "Test results of ATLASPIX3 — A reticle size HVCMOS pixel sensor designed for construction of multi chip modules," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 986, p. 164812, 2021.

[6] H. Zhang, F. Ehrler, R. Schimassek, and I. Perić, "Measurement results on capacitively coupled particle detector with PHOTON readout chip," *Journal of Instrumentation*, vol. 15, pp. P09041–P09041, Sep. 2020.

[7] K. Arndt, F. Ehrler, *et al.*, "Technical design of the phase I Mu3e experiment," 2020.

[8] Y. Han, F. Ehrler, *et al.*, "Study of CMOS strip sensor for future silicon tracker," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 981, p. 164520, 2020.

[9] M. Kiehn, F. Ehrler, *et al.*, "Performance of the ATLASPix1 pixel sensor prototype in ams aH18 CMOS technology for the ATLAS ITk upgrade," *Journal of Instrumentation*, vol. 14, pp. C08013–C08013, Aug. 2019.

[10] M. Prathapan, F. Ehrler, *et al.*, "Towards the large area HVCMOS demonstrator for ATLAS ITk," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 936, pp. 389 – 391, 2019. Frontier Detectors for Frontier Physics: 14th Pisa Meeting on Advanced Detectors.

[11] H. Augustin, F. Ehrler, *et al.*, "MuPix8 — Large area monolithic HVCMOS pixel detector for the Mu3e experiment," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 936, pp. 681 – 683, 2019. Frontier Detectors for Frontier Physics: 14th Pisa Meeting on Advanced Detectors.

[12] H. Augustin, F. Ehrler, *et al.*, "Performance of the large scale HV-CMOS pixel sensor MuPix8," *Journal of Instrumentation*, vol. 14, pp. C10011–C10011, Oct. 2019.

[13] B. Hiti, F. Ehrler, *et al.*, "Charge collection in irradiated HV-CMOS detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 924, pp. 214 – 218, 2019. 11th International Hiroshima Symposium on Development and Application of Semiconductor Tracking Detectors.

[14] I. Perić, F. Ehrler, *et al.*, "A high-voltage pixel sensor for the ATLAS upgrade," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 924, pp. 99 – 103, 2019. 11th International Hiroshima Symposium on Development and Application of Semiconductor Tracking Detectors.

[15] D. Sultan, F. Ehrler, *et al.*, "Electrical characterization of AMS aH18 HV-CMOS after neutrons and protons irradiation," *Journal of Instrumentation*, vol. 14, pp. C05003–C05003, May 2019.

[16] M. Kiehn, F. Ehrler, *et al.*, "Performance of CMOS pixel sensor prototypes in ams H35 and aH18 technology for the ATLAS ITk upgrade," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 924, pp. 104 – 107, 2019. 11th International Hiroshima Symposium on Development and Application of Semiconductor Tracking Detectors.

[17] R. Schimassek, F. Ehrler, *et al.*, "Monolithic sensors in LFoundry technology: Concepts and measurements," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 936, pp. 679 – 680, 2019. Frontier Detectors for Frontier Physics: 14th Pisa Meeting on Advanced Detectors.

[18] R. Blanco, F. Ehrler, *et al.*, "HVCMOS Monolithic Sensors for the High Luminosity Upgrade of ATLAS Experiment," *Journal of Instrumentation*, vol. 12, pp. C04001–C04001, Apr. 2017.

[19] K. Kanisauskas, F. Ehrler, *et al.*, "Radiation hardness studies of AMS HV-CMOS 350 nm prototype chip HVStripV1," *Journal of Instrumentation*, vol. 12, pp. P02010–P02010, Feb. 2017.

[20] I. Perić, F. Ehrler, *et al.*, "Status of HVCMOS developments for ATLAS," *Journal of Instrumentation*, vol. 12, pp. C02030–C02030, Feb. 2017.

[21] R. Schimassek, F. Ehrler, and I. Perić, "HVCMOS pixel detectors — methods for enhancement of time resolution," in *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*, pp. 1–5, 2016.

[22] C. Tamma, F. Ehrler, *et al.*, "The CHESS-2 prototype in AMS 0.35 µm process: A high voltage CMOS monolithic sensor for ATLAS upgrade," in *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*, pp. 1–5, 2016.

[23] Z. Liang, F. Ehrler, *et al.*, "Study of built-in amplifier performance on HV-CMOS sensor for the ATLAS phase-II strip tracker upgrade," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 831, pp. 156 – 160, 2016. Proceedings of the 10th International "Hiroshima" Symposium on the Development and Application of Semiconductor Tracking Detectors.

[24] A. Affolder, F. Ehrler, *et al.*, "Charge collection studies in irradiated HV-CMOS particle detectors," *Journal of Instrumentation*, vol. 11, pp. P04007–P04007, Apr. 2016.

[25] B. Huffman, F. Ehrler, *et al.*, "Radiation hardness of two CMOS prototypes for the ATLAS HL-LHC upgrade project.," *Journal of Instrumentation*, vol. 11, pp. C02005–C02005, Feb. 2016.

[26] V. Fadeyev, F. Ehrler, *et al.*, "Investigation of HV/HR-CMOS technology for the ATLAS Phase-II Strip Tracker Upgrade," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 831, pp. 189 – 196, 2016. Proceedings of the 10th International "Hiroshima" Symposium on the Development and Application of Semiconductor Tracking Detectors.

[27] I. Perić, F. Ehrler, *et al.*, "Overview of HVCMOS pixel sensors," *Journal of Instrumentation*, vol. 10, pp. C05021–C05021, May 2015.

[28] I. Perić, F. Ehrler, R. Leys, and R. Blanco, "HVCMOS pixel sensors," in *2015 IEEE SENSORS*, pp. 1–4, 2015.

[29] M. Prathapan, F. Ehrler, *et al.*, "ATLASpix3 : A high voltage CMOS sensor chip designed for ATLAS Inner Tracker," *PoS*, vol. TWEPP2019, p. 010, 2020.

[30] A. Schöning, F. Ehrler, *et al.*, "MuPix and ATLASPix – Architectures and Results," *PoS*, vol. Vertex2019, p. 024, 2020.

[31] M. Prathapan, F. Ehrler, *et al.*, "Design of a HVCMOS pixel sensor ASIC with on-chip readout electronics for ATLAS ITk Upgrade," *PoS*, vol. TWEPP2018, p. 074, 2019.

[32] H. Augustin, F. Ehrler, *et al.*, "MuPix8 - A large-area HV-MAPS chip," *PoS*, vol. Vertex2017, p. 057, 2018.

[33] D. Wiedner, F. Ehrler, *et al.*, "Readout Electronics for the First Large HV-MAPS Chip for Mu3e," *PoS*, vol. TWEPP-17, p. 099, 2018.

Bibliography

[1] I. Perić, "A novel monolithic pixel detector implemented in high-voltage CMOS technology," in *2007 IEEE Nuclear Science Symposium Conference Record*, vol. 2, pp. 1033–1039, 2007.

[2] M. Kiehn, F. Ehrler, *et al.*, "Performance of the ATLASPix1 pixel sensor prototype in ams aH18 CMOS technology for the ATLAS ITk upgrade," *Journal of Instrumentation*, vol. 14, pp. C08013–C08013, Aug. 2019.

[3] H. Kolanoski and N. Wermes, *Teilchendetektoren*. Springer-Verlag, 2016.

[4] F. Hartmann, *Evolution of Silicon Sensor Technology in Particle Physics*. Springer-Verlag, 2017.

[5] P. Ferrando, "SIMBOL-X: A new generation hard x-ray telescope," *Proc. SPIE Conf.*, vol. 5168, pp. 65–76, Jan. 2003.

[6] G. F. Knoll, *Radiation detection and measurement*. New York, NY: Wiley, 3. ed., 2000.

[7] E. B. Paul, *Nuclear and particle physics*. Amsterdam: North-Holland, 1969.

[8] V. Lepeltier, "Review on TPC's," *Journal of Physics: Conference Series*, vol. 65, p. 012001, Apr. 2007.

[9] R. Santonico and R. Cardarelli, "Development of resistive plate counters," *Nuclear Instruments and Methods in Physics Research*, vol. 187, no. 2, pp. 377–380, 1981.

[10] J. Seguinot and T. Ypsilantis, "Photo-ionization and Cherenkov ring imaging," *Nuclear Instruments and Methods*, vol. 142, no. 3, pp. 377–391, 1977.

[11] J. Kemmer and G. Lutz, "New detector concepts," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 253, no. 3, pp. 365–377, 1987.

[12] Wikipedia, "Boule." `https://commons.wikimedia.org/wiki/File:Monokristalines_Silizium_für_die_Waferherstellung.jpg`. Accessed: 2020-09-30.

[13] R. Sauer, *Halbleiterphysik: Lehrbuch für Physiker und Ingenieure*. München: Oldenbourg, 2010.

[14] G. Lutz, *Semiconductor Radiation Detectors: Device Physics*. SpringerLink, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

[15] F. Bloch, "Über die Quantenmechanik der Elektronen in Kristallgittern," *Zeitschrift für Physik*, vol. 52, pp. 555–600, Jul. 1929.

[16] R. Eber, *Investigations of new Sensor Designs and Development of an effective Radiation Damage Model for the Simulation of highly irradiated Silicon Particle Detectors*. PhD thesis, Karlsruhe Institute of Technology, 2013.

[17] J. R. Chelikowsky and M. L. Cohen, "Electronic structure of silicon," *Phys. Rev. B*, vol. 10, pp. 5095–5107, Dec. 1974.

[18] W. Setyawan and S. Curtarolo, "High-throughput electronic band structure calculations: Challenges and tools," *Computational Materials Science*, vol. 49, no. 2, pp. 299 – 312, 2010.

[19] F. Ehrler, "Development of active CMOS sensors for particle physics experiments," Master's thesis, Karlsruhe Institute of Technology, 2015.

[20] S. Hunklinger, *Festkörperphysik*. Studium, München: De Gruyter, 4. aufl. ed., 2014.

[21] A. Affolder, F. Ehrler, *et al.*, "Charge collection studies in irradiated HV-CMOS particle detectors," *Journal of Instrumentation*, vol. 11, pp. P04007–P04007, Apr. 2016.

[22] F. Thuselt, *Physik der Halbleiterbauelemente*. Springer-Verlag, 2018.

[23] M. Namordi and H. Thompson, "Aluminum-silicon Schottky barriers as semiconductor targets for EBS devices," *Solid-State Electronics*, vol. 18, no. 6, pp. 499 – 508, 1975.

[24] Wikipedia, "MOSFET drain current vs. drain-to-source voltage." `https://commons.wikimedia.org/wiki/File:IvsV_mosfet.svg`. Accessed: 2020-09-30.

[25] H. Bethe, "Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie," *Annalen der Physik*, vol. 397, no. 3, pp. 325–400, 1930.

[26] F. Bloch, "Zur Bremsung rasch bewegter Teilchen beim Durchgang durch Materie," *Annalen der Physik*, vol. 408, no. 3, pp. 285–320, 1933.

[27] K. Nakamura and Particle Data Group, "Review of Particle Physics," *Journal of Physics G: Nuclear and Particle Physics*, vol. 37, p. 075021, Jul. 2010.

[28] L. Landau, "On the energy loss of fast particles by ionization," *J. Phys. (USSR)*, vol. 8, pp. 201–205, 1944.

[29] Wikipedia, "Landau distribution." `https://commons.wikimedia.org/wiki/File:Landau_distribution.svg`. Accessed: 2020-09-30.

[30] G. Molière, "Theorie der Streuung schneller geladener Teilchen II. Mehrfach- und Vielfachstreuung," *Zeitschrift Naturforschung Teil A*, vol. 3, pp. 78–97, Feb. 1948.

[31] V. L. Highland, "Some practical remarks on multiple scattering," *Nuclear Instruments and Methods*, vol. 129, no. 2, pp. 497 – 499, 1975.

[32] G. R. Lynch and O. I. Dahl, "Approximations to multiple Coulomb scattering," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 58, no. 1, pp. 6 – 10, 1991.

[33] A. Dierlamm, "The CMS Outer Tracker Upgrade for the HL-LHC," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 924, pp. 256 – 261, 2019. 11th International Hiroshima Symposium on Development and Application of Semiconductor Tracking Detectors.

[34] I. Perić, F. Ehrler, *et al.*, "Status of HVCMOS developments for ATLAS," *Journal of Instrumentation*, vol. 12, pp. C02030–C02030, Feb. 2017.

[35] W. Snoeys, "Monolithic CMOS sensors for high energy physics," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 924, pp. 51 – 58, 2019. 11th International Hiroshima Symposium on Development and Application of Semiconductor Tracking Detectors.

[36] M. Garcia-Sciveres, "The RD53A Integrated Circuit," Tech. Rep. CERN-RD53-PUB-17-001, CERN, Geneva, Oct. 2017. https://cds.cern.ch/record/2287593.

[37] austriamicrosystems AG, "0.35 µm CMOS process allows voltages from 3.3 V to 120 V." https://archive.eetasia.com/www.eetasia.com/ART_8800417940_480200_NP_8512baee.HTM, 2006. Accessed: 2020-10-07.

[38] S. Hunklinger, *Festkörperphysik*. Oldenbourg, 2. ed., 2009.

[39] N. S. Saks, M. G. Ancona, and J. A. Modolo, "Radiation Effects in MOS Capacitors with Very Thin Oxides at 80°K," *IEEE Transactions on Nuclear Science*, vol. 31, no. 6, pp. 1249–1255, 1984.

[40] LHC collaboration, "The Large Hadron Collider." https://home.cern/science/accelerators/large-hadron-collider. Accessed: 2020-10-07.

[41] P. Mouche, "Overall view of the LHC. Vue d'ensemble du LHC." http://cds.cern.ch/record/1708847, Jun. 2014.

[42] ATLAS collaboration, "The ATLAS experiment." http://atlas.cern/. Accessed: 2020-10-07.

[43] CMS collaboration, "The CMS experiment." https://cms.cern/. Accessed: 2020-10-07.

[44] ALICE collaboration, "The ALICE experiment." http://aliceinfo.cern.ch/. Accessed: 2020-10-07.

[45] LHCb collaboration, "The LHCb experiment." http://lhcb.web.cern.ch/. Accessed: 2020-10-07.

[46] C. Lefèvre, "The CERN accelerator complex. Complexe des accélérateurs du CERN," Dec. 2008. http://cds.cern.ch/record/1260465.

[47] G. Apollinari *et al.*, *High-Luminosity Large Hadron Collider (HL-LHC): Preliminary Design Report*. CERN Yellow Reports: Monographs, Geneva: CERN, 2015.

[48] CMS collaboration, "How CMS weeds out particles that pile up." https://cms.cern/news/how-cms-weeds-out-particles-pile. Accessed: 2021-01-20.

[49] P. Azzi *et al.*, "Standard Model Physics at the HL-LHC and HE-LHC," 2019.

[50] X. C. Vidal *et al.*, "Beyond the Standard Model Physics at the HL-LHC and HE-LHC," 2019.

[51] M. Cepeda *et al.*, "Higgs Physics at the HL-LHC and HE-LHC," 2019.

[52] J. Pequenao, "Computer generated image of the whole ATLAS detector," Mar. 2008. http://cds.cern.ch/record/1095924.

[53] J. Pequenao, "Computer generated image of the ATLAS inner detector," Mar. 2008. https://cds.cern.ch/record/1095926.

[54] ATLAS Collaboration, *ATLAS inner detector: Technical Design Report, 1*. Technical Design Report ATLAS, Geneva: CERN, 1997. https://cds.cern.ch/record/331063.

[55] ATLAS Collaboration, *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*. No. CERN-LHCC-2017-021. ATLAS-TDR-030, Sep. 2017. http://cds.cern.ch/record/2285585.

[56] J. Pequenao, "Computer Generated image of the ATLAS calorimeter," Mar. 2008. https://cds.cern.ch/record/1095927.

[57] ATLAS Collaboration, *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical Design Report ATLAS, Geneva: CERN, 1996. https://cds.cern.ch/record/331061.

[58] ATLAS Collaboration, *ATLAS tile calorimeter: Technical Design Report*. Technical Design Report ATLAS, Geneva: CERN, 1996. https://cds.cern.ch/record/331062.

[59] J. Pequenao, "Computer generated image of the ATLAS Muons subsystem," Mar. 2008. http://cds.cern.ch/record/1095929.

[60] ATLAS Collaboration, *ATLAS muon spectrometer: Technical Design Report*. Technical Design Report ATLAS, Geneva: CERN, 1997. http://cds.cern.ch/record/331068.

[61] ATLAS Collaboration, *ATLAS magnet system: Technical Design Report, 1*. Technical Design Report ATLAS, Geneva: CERN, 1997. https://cds.cern.ch/record/338080.

[62] P. Jenni, M. Nordberg, M. Nessi, and K. Jon-And, *ATLAS Forward Detectors for Measurement of Elastic Scattering and Luminosity*. Technical Design Report ATLAS, Geneva: CERN, 2008. https://cds.cern.ch/record/1095847.

[63] D. Scannicchio, "ATLAS Trigger and Data Acquisition: Capabilities and commissioning," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 617, no. 1, pp. 306 – 309, 2010. 11th Pisa Meeting on Advanced Detectors.

[64] P. Jenni, M. Nessi, M. Nordberg, and K. Smith, *ATLAS high-level trigger, data-acquisition and controls: Technical Design Report*. Technical Design Report ATLAS, Geneva: CERN, 2003. https://cds.cern.ch/record/616089.

[65] ATLAS Collaboration, "Expected Tracking Performance of the ATLAS Inner Tracker at the HL-LHC," Tech. Rep. ATL-PHYS-PUB-2019-014, CERN, Geneva, Mar. 2019. https://cds.cern.ch/record/2669540.

[66] H. Pernegger, "Specification for ATLAS CMOS monolithic sensor." internal document, 2016.

[67] FE-I4 collaboration, *The FE-I4B Integrated Circuit Guide, version 2.3*, 2012. https://indico.cern.ch/event/261840/contributions/1594374/attachments/462649/641213/FE-I4B_V2.3.pdf.

[68] R. Schimassek, "Evaluation of CMOS Sensors as Particle Detectors at HL-LHC," Master's thesis, Karlsruhe Institute of Technology, 2016.

[69] I. Perić *et al.*, "The FEI3 readout chip for the ATLAS pixel detector," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 565, no. 1, pp. 178 – 187, 2006. Proceedings of the International Workshop on Semiconductor Pixel Detectors for Particles and Imaging.

[70] E. Vilella, R. Casanova, and I. Perić, *H35DEMO documentation*, 2016.

[71] T. Noulis, G. Fikos, S. Siskos, and G. Sarrabayrouse, "Noise Analysis of Radiation Detector Charge Sensitive Amplifier Architectures," CERN, Jan. 2008.

[72] National Nuclear Data Center, Brookhaven National Laboratory, "NuDat 2.7," 2019. https://www.nndc.bnl.gov/nudat2/.

[73] S. Banerjee *et al.*, "Silicon Bethe Bloch Muon (Geant) Proton (Geant) Electron (Geant) $\beta\gamma$," in *CMS NOTE 1999/056*, 2007.

[74] A. Chilingarov and RD50, *Recommendations towards a standardisation of the macroscopic parameter measurements - Part II: MIP CCE measurements*, 2004. `http://rd50.web.cern.ch/rd50/doc/Internal/rd50_2004_001.doc`.

[75] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE Journal of Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1439, 1989.

[76] K. R. Lakshmikumar, R. A. Hadaway, and M. A. Copeland, "Characterisation and modeling of mismatch in MOS transistors for precision analog design," *IEEE Journal of Solid-State Circuits*, vol. 21, no. 6, pp. 1057–1066, 1986.

[77] K. Akai, K. Furukawa, and H. Koiso, "SuperKEKB collider," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 907, p. 188–199, Nov. 2018.

[78] KEK, "SuperKEKB collider achieves the world's highest luminosity." `https://www.interactions.org/press-release/superkekb-collider-achieves-worlds-highest-luminosity`. Accessed: 2020-10-16.

[79] Belle II Collaboration, 2020. `https://www.belle2.org/archives/`.

[80] T. Abe *et al.*, "Belle II Technical Design Report," 2010. `https://arxiv.org/abs/1011.0352`.

[81] H.-G. Moser, "The Belle II DEPFET pixel detector," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 831, pp. 85–87, 2016. Proceedings of the 10th International "Hiroshima" Symposium on the Development and Application of Semiconductor Tracking Detectors.

[82] CERN, "Belle II VXD Open Workshop." `https://indico.cern.ch/event/810687`. Accessed: 2020-10-16.

[83] F. Ehrler and I. Perić, "Belle II VXD Open Workshop: HV-CMOS." `https://indico.cern.ch/event/810687/contributions/3451978/attachments/1876758/3090759/CMOSBelleII_CERN_v3.pdf`, Jul. 2019.

[84] J. N. Bader, "Construction and Characterization of a Detector Module made of Monolithic HV-CMOS Senors," Master's thesis, Karlsruhe Institute of Technology, 2020.

[85] I. Perić *et al.*, *Description of the ATLASPIX_Simple and ATLASPIX_M2 Preliminary v1*, 2017.

[86] A. X. Widmer and P. A. Franaszek, "A DC-Balanced, Partitioned-Block, 8B/10B Transmission Code," *IBM Journal of Research and Development*, vol. 27, pp. 440–451, Sep. 1983.

[87] A. Blondel *et al.*, "Letter of Intent for an Experiment to Search for the Decay $\mu \to eee$," Jan. 2012.

[88] A. Blondel *et al.*, "Research Proposal for an Experiment to Search for the Decay $\mu \to eee$," 2013.

[89] Y. Fukuda *et al.*, "Evidence for oscillation of atmospheric neutrinos," *Phys. Rev. Lett.*, vol. 81, pp. 1562–1567, 1998.

[90] Q. R. Ahmad *et al.*, "Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino Observatory," *Phys. Rev. Lett.*, vol. 87, p. 071301, Jul. 2001.

[91] The Royal Swedish Academy of Sciences, "The Nobel Prize in Physics 2015." `https://www.nobelprize.org/prizes/physics/2015/press-release/`.

[92] A. de Gouvea and P. Vogel, "Lepton Flavor and Number Conservation, and Physics Beyond the Standard Model," *Prog. Part. Nucl. Phys.*, vol. 71, pp. 75–92, 2013.

[93] B. Murakami, "Impact of lepton-flavor violating $Z^{'}$ bosons on muon $g-2$ and other muon observables," *Phys. Rev. D*, vol. 65, p. 055003, Jan. 2002.

[94] U. Bellgardt *et al.*, "Search for the decay $\mu^+ \rightarrow e^+e^+e^-$," *Nuclear Physics B*, vol. 299, no. 1, pp. 1 – 6, 1988.

[95] J. Adam *et al.*, "New Limit on the Lepton-Flavor-Violating Decay $\mu^+ \rightarrow e^+\gamma$," *Phys. Rev. Lett.*, vol. 107, p. 171801, Oct. 2011.

[96] A. K. Perrevoort for Mu3e Collaboration, "Status of the Mu3e Experiment at PSI," *EPJ Web of Conferences*, vol. 118, p. 01028, 2016.

[97] A. Weber and I. Perić, *Documentation MuPix8 (preliminary version 1)*, 2017.

[98] Linear Technology Corporation, *LTC6754 – 14-Bit Rail-to-RailMicropower DAC in MSOP*, 1998. `https://www.analog.com/media/en/technical-documentation/data-sheets/1658f.pdf`.

[99] Texas Instruments Incorporated, *LMV721-N/LMV722 10MHz, Low Noise, Low Voltage, and Low Power Operational Amplifier*, 2013. `http://www.ti.com/lit/ds/symlink/lmv721-n.pdf`.

[100] F. Ehrler *et al.*, "High-voltage CMOS detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 824, pp. 400 – 401, 2016. Frontier Detectors for Frontier Physics: Proceedings of the 13th Pisa Meeting on Advanced Detectors.

[101] B. Huffman, F. Ehrler, *et al.*, "Radiation hardness of two CMOS prototypes for the ATLAS HL-LHC upgrade project.," *Journal of Instrumentation*, vol. 11, pp. C02005–C02005, Feb. 2016.

[102] Z. Liang, F. Ehrler, *et al.*, "Study of built-in amplifier performance on HV-CMOS sensor for the ATLAS phase-II strip tracker upgrade," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 831, pp. 156 – 160, 2016. Proceedings of the 10th International "Hiroshima" Symposium on the Development and Application of Semiconductor Tracking Detectors.

[103] K. Kanisauskas, F. Ehrler, *et al.*, "Radiation hardness studies of AMS HV-CMOS 350 nm prototype chip HVStripV1," *Journal of Instrumentation*, vol. 12, pp. P02010–P02010, Feb. 2017.

[104] B. Hiti, F. Ehrler, *et al.*, "Charge collection in irradiated HV-CMOS detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 924, pp. 214 – 218, 2019. 11th International Hiroshima Symposium on Development and Application of Semiconductor Tracking Detectors.

[105] M. A. Green and M. J. Keevers, "Optical properties of intrinsic silicon at 300 K," *Progress in Photovoltaics: Research and Applications*, vol. 3, no. 3, pp. 189–192, 1995.

[106] R. K. Institut, "Neue Zahlen zu Krebs in Deutschland." https://www.rki.de/DE/Content/Service/Presse/Pressemitteilungen/2019/16_2019.html, Dec. 2019.

[107] J. Meikle, "How particles can be therapeutic." Physics Web, 2003. Accessed: 2020, August.

[108] C. Schoemers *et al.*, "The intensity feedback system at Heidelberg Ion-Beam Therapy Centre," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 795, pp. 92 – 99, 2015.

[109] I. Perić. personal communication, 2019.

[110] V. Lobashev *et al.*, "Direct search for mass of neutrino and anomaly in the tritium beta-spectrum," *Physics Letters B*, vol. 460, no. 1, pp. 227 – 235, 1999.

[111] C. Weinheimer *et al.*, "High precision measurement of the tritium β spectrum near its endpoint and upper limit on the neutrino mass," *Physics Letters B*, vol. 460, no. 1, pp. 219 – 226, 1999.

[112] KATRIN collaboration, "KATRIN design report," 2004. https://www.katrin.kit.edu/publikationen/DesignReport2004-12Jan2005.pdf.

[113] M. Kleesiek *et al.*, "β-Decay spectrum, response function and statistical model for neutrino mass measurements with the KATRIN experiment," *The European Physical Journal C*, vol. 79, Mar. 2019.

[114] M. Arenz *et al.*, "First transmission of electrons and ions through the KATRIN beamline," *Journal of Instrumentation*, vol. 13, pp. P04020–P04020, Apr. 2018.

[115] T. Houdy *et al.*, "Hunting keV sterile neutrinos with KATRIN: building the first TRISTAN module," *Journal of Physics: Conference Series*, vol. 1468, p. 012177, Feb. 2020.

[116] R. Adhikari *et al.*, "A White Paper on keV sterile neutrino Dark Matter," *Journal of Cosmology and Astroparticle Physics*, vol. 2017, pp. 025–025, Jan. 2017.

[117] G. Mention *et al.*, "Reactor antineutrino anomaly," *Phys. Rev. D*, vol. 83, p. 073006, Apr. 2011.

[118] P. Lechner *et al.*, "Silicon drift detectors for high count rate X-ray spectroscopy at room temperature," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 458, no. 1, pp. 281 – 287, 2001. Proc. 11th Inbt. Workshop on Room Temperature Semiconductor X- and Gamma-Ray Detectors and Associated Electronics.

[119] Fondazione Bruno Kessler IRIS group, "Dodecagonal SDD." TRISTAN meeting, Jul. 2017.

[120] I. Perić, "KIT TRISTAN Integrated Circuit." TRISTAN meeting, May 2017.

[121] M. Guthoff *et al.*, "Geant4 simulation of a filtered X-ray source for radiation damage studies," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 675, pp. 118 – 122, 2012.

[122] ATLAS collaboration, "Letter of Intent for the Phase-II Upgrade of the ATLAS Experiment," 2012.

[123] I. Bediaga, "LHCb VELO Upgrade Technical Design Report," Nov. 2013.

[124] L. Linssen *et al.*, "Physics and Detectors at CLIC: CLIC Conceptual Design Report," 2012.

[125] B. Adams *et al.*, "Letter of Intent: A New QCD facility at the M2 beam line of the CERN SPS (COMPASS++/AMBER)," 2019.

[126] H. Zhu, "Vertex and Tracking Detectors for the Circular Electron Positron Collider (CEPC)," *PoS*, vol. Vertex2019, p. 046, 2020.

[127] CMS collaboration, "Computer generated image of CMS." `https://cms.cern/detector`. Accessed: 2020-09-30.

[128] A. Tauro, "ALICE Schematics," tech. rep., May 2017.

[129] The LHCb Collaboration, "The LHCb Detector at the LHC," *Journal of Instrumentation*, vol. 3, pp. S08005–S08005, Aug. 2008.

[130] Linear Technology Corporation, *LTC6754 – High Speed Rail-to-Rail Input Comparator with LVDS Compatible Outputs*, 2015. `https://www.analog.com/media/en/technical-documentation/data-sheets/6754f.pdf`.

[131] F. Ehrler *et al.*, "Characterization results of a HVCMOS sensor for ATLAS," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 936, pp. 654 – 656, 2019. Frontier Detectors for Frontier Physics: 14th Pisa Meeting on Advanced Detectors.

[132] F. Ehrler, I. Perić, and R. Schimassek, "HVCMOS pixel detectors first measurements on the reticle size prototype for the ATLAS pixel layers," in *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*, pp. 1–8, Oct. 2016.

[133] I. Perić, F. Ehrler, *et al.*, "High Voltage CMOS Active Pixel Sensor," *IEEE Journal of Solid-State Circuits*, 2021.

[134] R. Schimassek, F. Ehrler, *et al.*, "Test results of ATLASPIX3 — A reticle size HVCMOS pixel sensor designed for construction of multi chip modules," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 986, p. 164812, 2021.

[135] H. Zhang, F. Ehrler, R. Schimassek, and I. Perić, "Measurement results on capacitively coupled particle detector with PHOTON readout chip," *Journal of Instrumentation*, vol. 15, pp. P09041–P09041, Sep. 2020.

[136] K. Arndt, F. Ehrler, *et al.*, "Technical design of the phase I Mu3e experiment," 2020.

[137] Y. Han, F. Ehrler, *et al.*, "Study of CMOS strip sensor for future silicon tracker," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 981, p. 164520, 2020.

[138] M. Prathapan, F. Ehrler, *et al.*, "Towards the large area HVCMOS demonstrator for ATLAS ITk," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 936, pp. 389 – 391, 2019. Frontier Detectors for Frontier Physics: 14th Pisa Meeting on Advanced Detectors.

[139] H. Augustin, F. Ehrler, *et al.*, "MuPix8 — Large area monolithic HVCMOS pixel detector for the Mu3e experiment," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 936, pp. 681 – 683, 2019. Frontier Detectors for Frontier Physics: 14th Pisa Meeting on Advanced Detectors.

[140] H. Augustin, F. Ehrler, *et al.*, "Performance of the large scale HV-CMOS pixel sensor MuPix8," *Journal of Instrumentation*, vol. 14, pp. C10011–C10011, Oct. 2019.

[141] I. Perić, F. Ehrler, *et al.*, "A high-voltage pixel sensor for the ATLAS upgrade," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 924, pp. 99 – 103, 2019. 11th International Hiroshima Symposium on Development and Application of Semiconductor Tracking Detectors.

[142] D. Sultan, F. Ehrler, *et al.*, "Electrical characterization of AMS aH18 HV-CMOS after neutrons and protons irradiation," *Journal of Instrumentation*, vol. 14, pp. C05003–C05003, May 2019.

[143] M. Kiehn, F. Ehrler, *et al.*, "Performance of CMOS pixel sensor prototypes in ams H35 and aH18 technology for the ATLAS ITk upgrade," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 924, pp. 104 – 107, 2019. 11th International Hiroshima Symposium on Development and Application of Semiconductor Tracking Detectors.

[144] R. Schimassek, F. Ehrler, *et al.*, "Monolithic sensors in LFoundry technology: Concepts and measurements," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 936, pp. 679 – 680, 2019. Frontier Detectors for Frontier Physics: 14th Pisa Meeting on Advanced Detectors.

[145] R. Blanco, F. Ehrler, *et al.*, "HVCMOS Monolithic Sensors for the High Luminosity Upgrade of ATLAS Experiment," *Journal of Instrumentation*, vol. 12, pp. C04001–C04001, Apr. 2017.

[146] R. Schimassek, F. Ehrler, and I. Perić, "HVCMOS pixel detectors — methods for enhancement of time resolution," in *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*, pp. 1–5, 2016.

[147] C. Tamma, F. Ehrler, *et al.*, "The CHESS-2 prototype in AMS 0.35 µm process: A high voltage CMOS monolithic sensor for ATLAS upgrade," in *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*, pp. 1–5, 2016.

[148] V. Fadeyev, F. Ehrler, *et al.*, "Investigation of HV/HR-CMOS technology for the ATLAS Phase-II Strip Tracker Upgrade," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 831, pp. 189 – 196, 2016. Proceedings of the 10th International "Hiroshima" Symposium on the Development and Application of Semiconductor Tracking Detectors.

[149] I. Perić, F. Ehrler, *et al.*, "Overview of HVCMOS pixel sensors," *Journal of Instrumentation*, vol. 10, pp. C05021–C05021, May 2015.

[150] I. Perić, F. Ehrler, R. Leys, and R. Blanco, "HVCMOS pixel sensors," in *2015 IEEE SENSORS*, pp. 1–4, 2015.

[151] M. Prathapan, F. Ehrler, *et al.*, "ATLASpix3 : A high voltage CMOS sensor chip designed for ATLAS Inner Tracker," *PoS*, vol. TWEPP2019, p. 010, 2020.

[152] A. Schöning, F. Ehrler, *et al.*, "MuPix and ATLASPix – Architectures and Results," *PoS*, vol. Vertex2019, p. 024, 2020.

[153] M. Prathapan, F. Ehrler, *et al.*, "Design of a HVCMOS pixel sensor ASIC with on-chip readout electronics for ATLAS ITk Upgrade," *PoS*, vol. TWEPP2018, p. 074, 2019.

[154] H. Augustin, F. Ehrler, *et al.*, "MuPix8 - A large-area HV-MAPS chip," *PoS*, vol. Vertex2017, p. 057, 2018.

[155] D. Wiedner, F. Ehrler, *et al.*, "Readout Electronics for the First Large HV-MAPS Chip for Mu3e," *PoS*, vol. TWEPP-17, p. 099, 2018.

Acknowledgments – Danksagung

This dissertation would not have been possible without the support of many people, to whom I would like to express my gratitude!

Diese Dissertation wäre ohne die Unterstützung zahlreicher Menschen nicht möglich gewesen, denen ich meinen tiefen Dank aussprechen möchte!

Als erstem möchte ich Herrn Prof. Dr. Marc Weber für die Gelegenheit danken, am Institut für Prozessdatenverarbeitung und Elektronik meine Doktorarbeit durchzuführen und mir als Referent zur Seite zu stehen.

Prof. Dr. Ivan Perić gilt meine Dankbarkeit dafür, dass er mir immer sowohl mit Rat und Tat zur Seite stand, mir aber auch großen kreativen Freiraum gab, um meine Arbeit zu entwickeln. Ohne ihn und seine Führung des KIT-ADL würde es diese Arbeit nicht geben. Außerdem danke ich ihm für die Übernahme des Korreferats.

Allen derzeitigen und ehemaligen Mitstreitern vom KIT-ADL Rudolf Schimassek, Alena Weber, Roberto Blanco, Richard Leys, Horacio Mateos, Mridula Prathapan und Hui Zhang möchte ich für die gute Zusammenarbeit, viele große und kleine Hilfen und natürlich für das angenehme Arbeitsklima danken.

Allen Kollegen am IPE möchte ich danken für die große Hilfsbereitschaft und die zur Verfügung gestellte Infrastruktur. Dies gilt insbesondere für die Teams in den Werkstätten und im Reinraum: Uwe Bauer, Alexander Bacher, Peter Hoffmann Benjamin Leyrer, Bernhard Oßwald, Tibor Piller und Peter Schöck. Marc Schneider möchte ich dafür danken, dass er zur Lösung von Problem immer Tipps und Materialien beigesteuert hat.

Gleiches trifft auf die Mitarbeiter der Hardware-Gruppe im Institut für Experimentelle Teilchenphysik zu. Besonders bei Bestrahlungen und aufwändigen Messungen war die Hilfe und Expertise von Alexander Dierlamm, Tobias Barvich, Stefan Heindl, Andreas Nürnberg Stefan Maier, Hans-Jürgen Simonis, Pia Steck und vielen weiteren unersetzlich.

Den Mitarbeiterinnen der Administration, Christiane Buchwald-Kayser, Antje Martin, Saskia Pulch und Tamara Wezel möchte ich danken, dass sie mir immer bei den vielfältigen Herausforderungen von Papieren geholfen und den Rücken freigehalten haben.

Bedanken möchte ich mich bei KSETA für die Unterstützung in der Zeit meiner Arbeit und besonders für die Organisation von Workshops und Kursen, die so wertvoll für den wissenschaftlichen Austausch sind.

Abschließend möchte ich mich herzlich bei meiner Familie Klaus, Verena, Julian und Sighild für die immerwährende Unterstützung bedanken.