Development of highly-performant Depleted Monolithic Active Pixel Sensors for future particle physics experiments

Year 1 Report

Jan Patrick Hammerich

Student No 201442864

Supervisor:

Dr Eva Vilella

University of Liverpool

Department of Physics

Contents

1 Introduction

In the search for new phenomena beyond the Standard Model of Particle Physics (SM), physicists are looking for increasingly rare processes. To make beyond SM (BSM) effects measurable, high rates to measure tiny deviations from theory due to quantum loops from BSM physics and high Centre-of-Mass energies to produce new unknown particles, are employed.
At the heart of most of these experiments is a tracking detector, which is measuring the decay products of the interactions. Tracking detectors consist of multiple layers of very granular sensors that provide the space points of the sensitive cells which were traversed by a charged particle. By combining these points, the trajectory of the particles can be reconstructed to distinguish different reactions close in space and time (pile-up) or to reconstruct displaced secondary vertices. In the presence of a magnetic field, the bending of the trajectories due to the Lorentz force can be exploited to measure the momentum of the particles.
Typically, the sensors are fabricated from silicon, which is made possible due to modern silicon technologies. Currently, only silicon sensors can achieve the strong requirements imposed by leading experiments. The constraints are formulated in terms of spatial resolution, which depends on the size of the sensitive unit, and radiation tolerance, the ability to perform well even after sustaining damage from the particles in the experiments which alters the properties of the sensor. Another constraint is the so called material budget, which describes the deflection of particles traversing the sensor and depends on the material and the thickness. Recently also the demand for a good time resolution of the sensors has arisen to help to suppress pile-up and improve reconstruction. On top of all these specifications, the detectors are expected to perform all these tasks with as little power as possible due to limited cooling and power distribution capacities.
Silicon pixel sensors are employed close to the interaction point where excellent spatial resolution is key. The most common type of these pixel detectors are so called hybrid detectors. They are manufactured by combining a passive sensor, a piece of silicon which contains segmented diodes, and a readout application-specific integrated circuit (ASIC) chip which reads out the sensor and digitises the information. By applying a high reverse bias voltage to the sensor, a depletion zone is formed by the diodes which rapidly collect the charge carriers freed by the ionising radiation. The connection between the sensor and the readout ASIC is realised by a process called bump bonding, in which many small solder balls are placed on one part of the assembly. Then the other piece has to be placed precisely on top and electrical connections are created. This production method is complex and time consuming which also makes it very expensive. In terms of material budget this limits the sensor due to the high Z material of the solder balls as well as the minimum thickness due to the required rigidity of both sensor and readout chip to perform the bump bonding. The size and positioning of the solder balls also constrain

the segmentation.
A concept that can overcome these limitations are evolutions of the monolithic active pixel sensors (MAPS). MAPS combine sensitive volume and readout electronics in a single chip, eliminating the need for bump bonds. They can only collect charge by diffusion, limiting their time resolution and rate capabilities as well as the radiation hardness. Since the charges are collected slower, they are much likelier to be trapped by defects in the silicon lattice caused by radiation damage, so they cannot contribute to the signal any more. However advancements in commercial CMOS processes allow for the creation of a depletion layer in the sensor to collect charge by drift. This was first demonstrated using low resistivity substrates and high depletion voltage (HV-MAPS)[1] and later with high resistivity and moderate voltages (DMAPS)[2]. The term depleted monolithic active pixel sensors (DMAPS) can be used to describe both implementations.

2 Physics of Semiconductor Sensors

2.1 Semiconductors

Semiconductors are characterised by their ability to transport electrical currents much worse than conductors like most metals, even though they are not completely isolating. The prime example is silicon. Silicon is a group IV element in the periodic table, meaning its outermost electron shell is occupied by 4 electrons in its ground state. When grown as a crystal, it forms a diamond-like lattice with each electron forming a covalent bond with one of its neighbours in a tetrahedron structure. In this state, all electrons are bound and the valence band, the highest occupied band, is completely full while the conduction band, the band above, is empty.
Since electrons are fermions, this is only true for T=0 K. Above absolute zero, the occupancy $\overline{n_i}$ follows the Fermi-Dirac distribution,

$$\overline{n_i} = \frac{1}{exp((\epsilon_i - \mu)/k_B T) + 1}, \tag{2.1}$$

with k_B being Boltzmann's constant, T the absolute temperature, μ the Fermi level, the middle between the valence and the conduction band, ϵ_i the energy of the state. This means that electrons can be excited to the conduction band and become free charge carriers. These electrons leave behind an empty state called hole. Holes can be filled by electrons from adjacent atoms, in turn leaving a new hole, allowing it to act as a positive free charge carrier. At room temperature the intrinsic charge density for silicon is $1.45 \times 10^{10}\,\mathrm{cm}^{-3}$ compared to the atomic density of $5.0 \times 10^{22}\,\mathrm{cm}^{-3}$ [3] which makes pure silicon effectively isolating.
The introduction of atoms of different groups change this picture significantly. If for example, an element of the group V like phosphorous is introduced to the lattice, it forms covalent bonds with its neighbours like silicon does. However the surplus 5th electron is only loosely bound and can easily transition into the conduction band, leaving behind a positive immobile ion. Materials which contribute electrons to the conduction band are called donors and the process of adding them to intrinsic silicon is called doping, in this case n-doping. Analogously, elements of group III, like boron, remove electrons and are called acceptors, producing p-doped material.
To emphasise the concentration of the doping + and - are added to describe the structure. An n+ region for example is heavily n-doped while a p- region is only lightly p-doped.

2.2 p-n Junction

The interface between a p-doped and an n-doped region is called p-n junction. Due to the different concentration of free charge carriers, electrons diffuse into the p material while holes diffuse into the n region. As they leave more and more ions behind, an electric field forms which induces a drift current in the opposite direction. At equilibrium, the diffusion current and the drift current due to the so called built-in voltage V_0 cancel out. In this state, the volume around the interface does not contain any free charge carriers as shown in Fig. 2.1 and is called the depletion zone. Charge carriers produced in the depletion zone drift to opposing directions due to the electric field. This ability to collect charges induced by ionising radiation make reverse bias diodes the basis of almost all silicon sensors in particle physics.

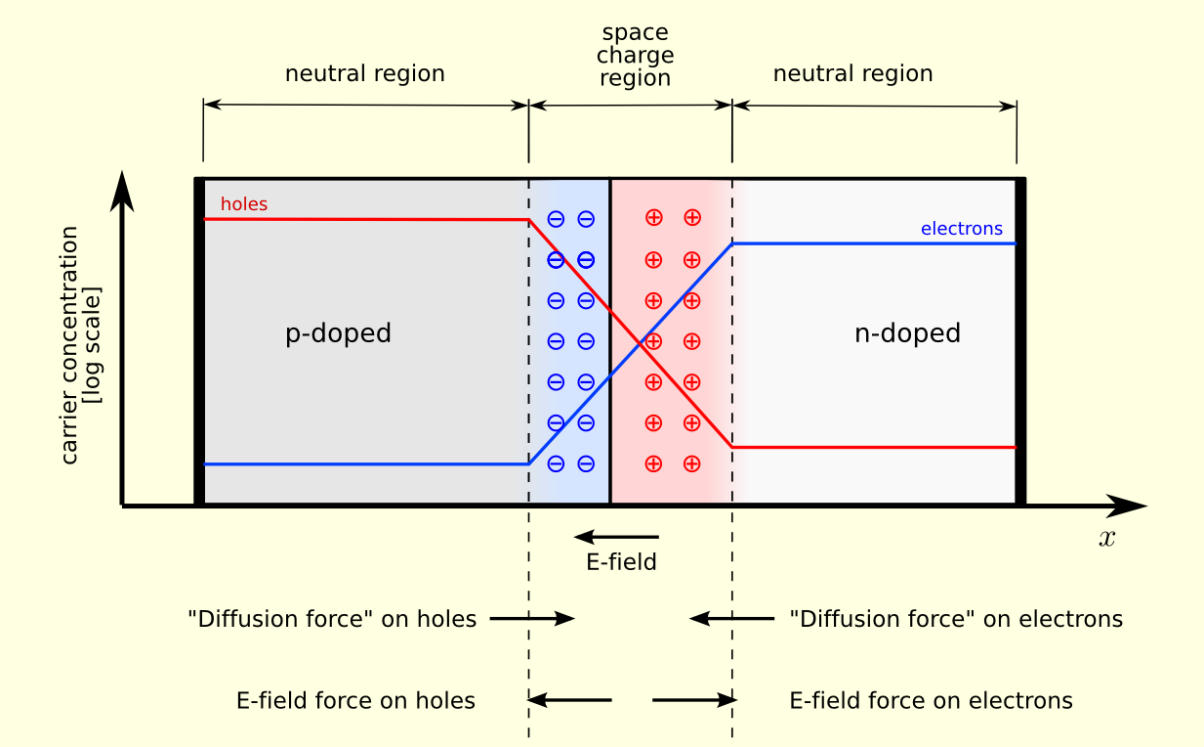

Figure 2.1: Equilibrium state of the p-n junction [4].

The width w of the depletion zone can be increased by applying a voltage V between the n and p bulk. If the doping concentrations N_A and N_D are known, w is given by [5]

$$w = \sqrt{\frac{2\epsilon_0\epsilon}{e}\frac{N_A + N_D}{N_A N_D}(V - V_0)}. \tag{2.2}$$

Here e is the electron charge, ϵ the vacuum permittivity, and ϵ_0 the relative permittivity of silicon. At high bias voltages ($V \gg V_0$) and asymmetric doping, i.e. $N_D \gg N_A$, the equation simplifies to

$$w = \sqrt{\frac{2\epsilon_0\epsilon}{eN_A}V}. \tag{2.3}$$

This means the higher the bias voltage and the smaller the doping concentration, the greater the depletion zone.
If a voltage is applied in the opposite direction, the depletion zone shrinks and eventually vanishes, making the junction fully conducting. This mode is called forward bias, while a voltage depleting the junction is called reverse bias. This component, conducting in one polarity and isolating in the other, is called diode in the field of electronics.

2.3 The MOSFET Transistor

The metal-oxide-silicon field effect (MOSFET) transistor is a 4-terminal structure and the basis of almost all modern semiconductor devices. There are two implementations of this structure the NMOS and the PMOS transistor which differ in the type of charge carriers they transport. A process which allows both NMOS and PMOS is called complimentary MOS or CMOS.
The terminals are called gate, bulk, source and drain. Charge carriers flow from the source to the drain while the mode of operation is controlled by the gate. In the following all voltages are defined to be positive and the bulk of the transistor is connected to the source.
An NMOS transistor consists of two n+ implants in a p-substrate. The area between the implants is covered by an insulating silicon oxide layer with the gate electrode on top. Depending on the applied voltages, the transistor has 3 operation modes: sub-threshold, linear, and saturation which are depicted in Fig. 2.2.
If the Gate-Source Voltage V_{GS} is lower than the threshold voltage V_{THR}, a technology parameter, no charges can flow from source to drain and the transistor is closed. This is called sub-threshold region.
In the case of $V_{GS} \geq V_{THR}$ and $(V_{GS} - V_{THR}) > V_{DS}$, the Drain-Source voltage, the transistor is in the so-called linear region. The positive charges in the gate push away holes in the bulk, allowing electrons to gather under the gate which is called inversion layer or channel. Through this channel, electrons can move from source to drain and the transistor is conducting. The higher V_{GS} the deeper the channel becomes which allows more electrons to travel through it. The current in the drain I_D is then given by:

$$I_D = \mu_n C_{OX} \frac{W}{L} \left((V_{GS} - V_{THR}) V_{DS} - \frac{{V_{DS}}^2}{2} \right) \tag{2.4}$$

μ_n is the mobility of electrons in silicon, C_{OX} the gate oxide capacitance per unit area, while W and L are the geometrical parameters of the gate, width and length respectively.

For small V_{DS}, the current is linear with V_{DS}, which means that the transistor operates like a resistor with a resistance controlled by V_{GS}, earning the region the other name of ohmic region.
In the case of $V_{GS} \geq V_{THR}$ and $V_{DS} > (V_{GS} - V_{THR})$ the transistor is in saturation mode. The channel becomes asymmetric, extending deeper into the bulk at the source while shrinking at the drain side until it no longer connects the electrodes. This point is called pinch off. While the electrodes are no longer connected in this mode, the electric fields are strong enough to ensure conduction. In this mode the current depends only little on V_{DS} and mostly on V_{GS}, which controls the length of the channel below the gate and with it the current. the current is described by:

$$I_D = \frac{\mu_n C_{OX}}{2} \frac{W}{L} (V_{GS} - V_{THR})^2 \big(1 + \lambda (V_{DS} - V_{DS,sat})\big) \tag{2.5}$$

The saturation Drain-Source Voltage $V_{DS,sat}$ and the channel-length modulation parameter λ are technology parameters. Since the current depends mostly on V_{GS}, the transistor acts as a voltage controlled current source in this mode.
For the PMOS, the description is analogue with electrons replaced by holes and p+ implants in an n bulk. However, the mobility of holes μ_h in silicon is lower than the mobility of electrons ($500\,\mathrm{cm^2\,V^{-1}\,s}$ compared to $1450\,\mathrm{cm^2\,V^{-1}\,s}$ [5]) which is why PMOS transistors have to be wider or shorter to achieve the same transconductance as an NMOS transistor.

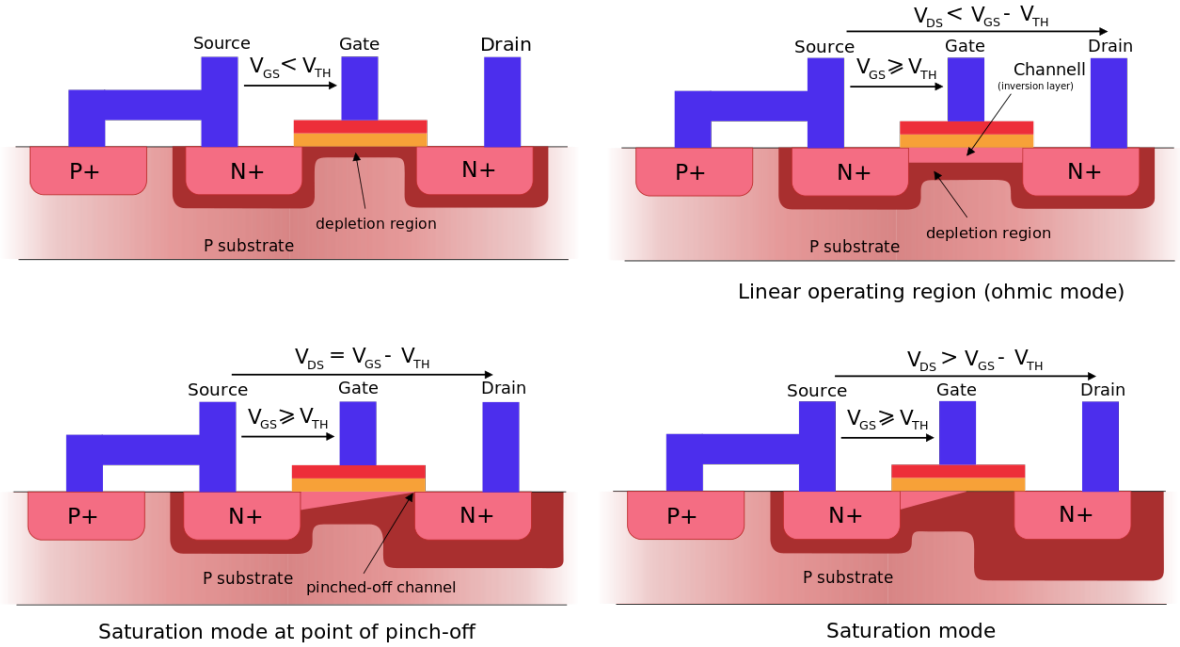

Figure 2.2: Operation modes of an NMOS transistor [6]. Top left: sub-threshold mode. Top right: linear mode. Bottom left: saturation mode at pinch-off. Bottom right: saturation mode.

3 Silicon Pixel Sensors

Pixel sensors are detectors which are segmented in two orthogonal dimensions. They have become common as silicon technology advanced and are now found in basically every digital camera. They can be roughly categorised in monolithic and hybrid devices.
Hybrids consist of a segmented sensor and a readout device with matching segmentation. This has the advantage that sensor and readout can be individually fabricated and optimised. For example in imaging applications the material can be chosen such that the sensitivity is at maximum for wavelengths used without the need to transfer the design of the readout to another process or technology. Typically, sensors are fabricated by doping silicon in a regular pattern, creating an array of diodes. Operated in reverse bias they collect charges induced by ionising radiation in the depletion zone. The readout usually takes the form of an ASIC specifically developed for the application. In each readout cell the collected charge from the partner sensor cell is amplified and digitised. To facilitate the connection a technique called bump bonding is used. First, small balls of solder material are deposited precisely on the contact of each channel on either the sensor or the readout chip. Then the other component has to be precisely positioned on top of the solder balls to ensure the 1-to-1 matching. With heat and pressure, an electrical connection is created. Due to their excellent position resolution and efficiency, hybrid pixel detectors are a staple for building vertex detectors in particle physics experiments. As mentioned in Ch. 1, their excellent performance comes with significant material and monetary budget. The solder technology further limits the achievable segmentation due to the minimum size and precision in positioning which can be achieved for the solder balls.
Advances in silicon processing allowed designers to implement an amplifier already on the sensor side of a hybrid which are called active pixel sensors (APS). A development from this concept are so called charge-coupled pixel detectors (CCPD) [7], which allow enough amplification such that the bump bonds could be replaced with for example dielectric glue, easing the production process.
Monolithic devices are created by implementing the sensitive volume and the readout electronics in the same silicon substrate. Most common are charge-coupled devices (CCD) which are widely used for cameras in commercial uses and also in astronomy applications. Here, charges are stored in each cell which is then digitised one by one in a dedicated circuit. They have excellent quantum efficiency and can be segmented even finer than hybrids but are limited in readout speed. High sensitivity to temperature changes and radiation damage make them less ideal for particle physics uses.
Another evolution of the APS concept places all readout electronics on the sensor side which eliminates the need for a readout chip. These sensors are called monolithic active pixel sensors (MAPS). These sensors however do not collect charge by drift but by diffusion. A deep epitaxial layer forms a potential well between the bulk and the top-

side implants. Confined by these barriers electrons diffuse until they reach the collection electrode. While charge collection is slow, the segmentation can be much smaller than hybrids. Examples of these sensors are the MIMOSA series of chips which are used in the STAR tracker [8] or the EUDET testbeam telescopes [9], or the ALPIDE sensor [10] which is used for the upgrade of the ALICE inner tracker. Since most of the bulk is passive, it can be thinned away to achieve sensor thicknesses of around 50 µm, reaching an unprecedented material budget.
Further, high-voltage CMOS processes allow to also include a deep well in the substrate, most commonly a n-well in a p-substrate, which acts as sensor diode, with the readout electronics isolated on top. This allows to apply a high voltage ($\mathcal{O}$(10 V to 100 V)) between the well and the substrate. This allows for a several µm deep depletion zone which is used as sensitive volume. The undepleted bulk can again be thinned away. These sensors are called high-voltage MAPS (HV-MAPS) [1], shown in Fig.3.1. Most prominent prototypes of HV-MAPS is the MuPix family which is developed for the Mu3e experiment [11]. Other experiments planning to use these sensors are the PANDA experiment at FAIR for the luminosity detector [12] and the P2 experiment [13].

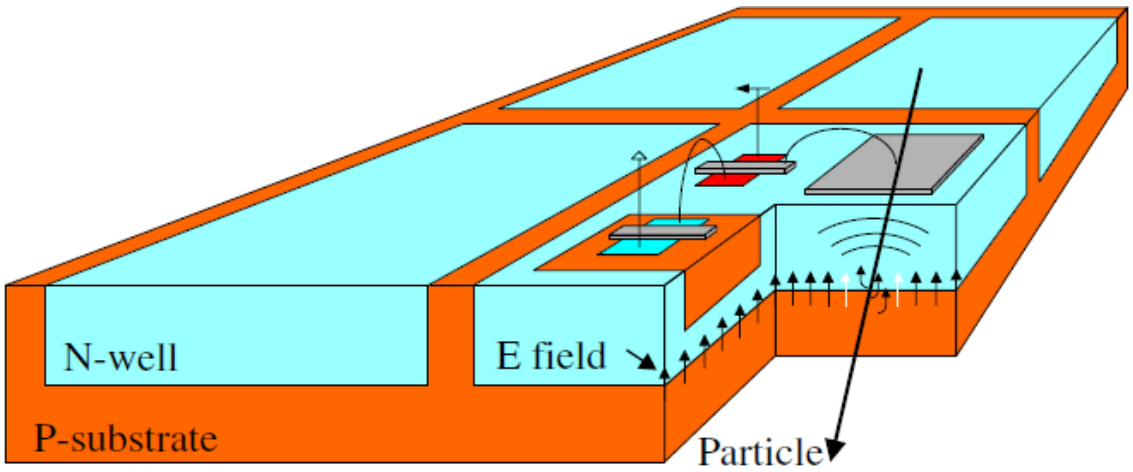

Figure 3.1: Sketch of the HV-MAPS concept[1].

4 DMAPS Development at UoL

The University of Liverpool has a long history in the research and development of silicon sensors for high energy physics. These efforts also include the design and characterisation of DMAPS prototypes. Recent contributions include the H35 Demo [14] which is a demonstrator in the ams H35 process which is a 3.3 V 0.35 µm technology [15]. This sensor was developed to demonstrate the capabilities of HV-CMOS in the frame of the ATLAS experiment. It is a 24.4 mm × 18.5 mm large chip containing monolithic and CCPD matrices with different pixel and digitisation architectures.
Currently, the HV-CMOS group at the University of Liverpool is leading the work package on HV-CMOS in the RD50 collaboration, which investigates radiation damage in silicon sensors and radiation hard technologies. In this framework, two multi-project wafer (MPW) ASICs have been produced which are called RD50-MPW1 and RD50-MPW2. These sensors have been fabricated the 150 nm 1.8 V high-voltage process by LFoundry [16]. RD50-MPW1 [17] is a fully monolithic chip consisting of two matrices. One is aimed at photon counting for medical applications while the other is designed for particle physics applications. This matrix features 78 × 40 pixel matrix with 50 µm × 50 µm pixel size. The readout is similar to the readout of the ATLAS FE-I3 chip [18], however hit discrimination and full hit digitisation are done in the pixel and not in the periphery like a lot of other HV-MAPS. This feature is possible due to the process which allows deep p-wells inside the deep n-well, see Fig. 4.1. These p-wells, here called PSUB, can be used to isolate shallow n-wells from the deep n-well. Without this implant all PMOS transistors would need to be placed in the deep n-well which would interfere heavily with the charge collection.

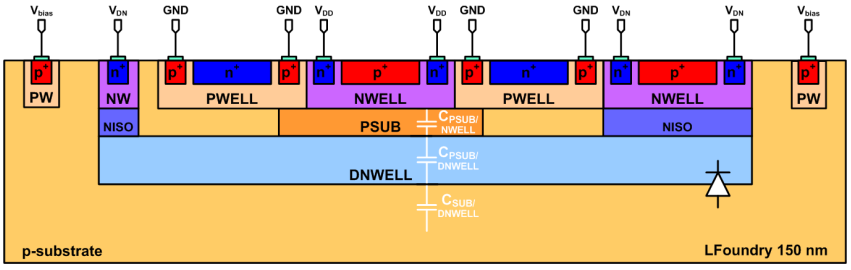

Figure 4.1: Well structure of the LF 150 nm process used for RD50-MPW1 [17] and RD50-MPW2.

RD50-MPW2 includes a 8×8 matrix of $60\,\mu m \times 60\,\mu m$ without a digital readout. This chip aims to reduce the high leakage current and increase the breakdown voltage observed in RD50-MPW1 as well as the characterisation of other circuits. The pixel matrix is divided in two flavours with different feedback schemes. The continuous feedback pixel use a constant reference current to linearly discharge the input and is widely used. The switched feedback flavour makes use of the high integration in the analogue pixel. The output of the comparator is connected to an additional current source which is triggered by it and allows for a much faster discharge. This in principle allows for a much higher rate capability of the pixel given a fast enough digital readout. RD50-MPW2 also hosts a number of other test circuits like a bandgap voltage reference and a radiation tolerant memory array.
Since 2019 the HV-CMOS group is supported by a UKRI Future Leaders Fellowship [19]. The research aimed at improving DMAPS for future applications. It is divided into three work packages:

- Reducing the pixel area below the current $50\,\mu m \times 50\,\mu m$ benchmark
- Improving the time resolution into the sub-nanosecond range
- Increasing the radiation tolerance beyond 10^{16} 1 MeV n_{eq}/cm^2

While most of the previously mentioned efforts are generic research, the HV-CMOS also contributes to developments for dedicated experiments. One of these applications is the study of how DMAPS could improve the measurement of the electric dipole moment of the proton. Another active topic is the development of an upgrade of the LHCb outer tracker, the so-called Mighty Tracker [20]. In this proposal, the area closest to the beam line is instrumented with DMAPS instead of the scintillating fibres, which are still used farther away. As a result both the spatial resolution as well as the radiation tolerance would be greatly improved.

5 Analogue Front-end Electronics

The analogue front-end electronics of the RD50 MPWs consists of a charge sensitive amplifier (CSA) with a feedback tree, a source follower (SF), and a high pass (HP) stage. To compensate for production variances and optimise the working point of the circuit several reverence voltages are generated on the chip by digital to analogue converters (DAC). It is shown schematically in Fig. 5.1.
The sensing diode is AC coupled to the input of the amplifier by the gate of a transistor. All other terminals are connected to the deep n-well, forming a capacitor. The deep n-well itself is biased by a dedicated circuit. It ensures that the potential is restored to its initial state after an event slow enough that the amplifier can generate a signal. To mimic charge depositions of real particles an injection circuit is implemented in the same way to form a capacitor between the n-well and the injection line. The CSA itself is realised as a single-ended folded cascode PMOS amplifier. The input PMOS transistor requires a dedicated supply voltage $VSSA$ for optimal performance. It is biased by the transistor controlled by the DAC VN such that a small change in the input causes a large change in the voltage between the two transistors. As result the cascode transistor controlled by $VNCASC$ stops conducting. The current sourced by the $VPBIAS$ transistor has nowhere else to go but the output which creates the signal.
To restore the amplifier to its working point a feedback circuit is required which discharges the output. This is realised by a current mirror. The transistor controlled by $VPFB$ sources a constant current which flows into a current mirror between the amplifier input and output. The current between the output and the feedback current source generates the gate voltage at the transistor between input and output. As consequence the same current flows between the two nets, discharging the output.
The SF is a buffer with gain 1 at lower frequencies. It ensures that enough current is sourced to load the filter capacity of the HP. At higher frequencies the gain becomes smaller resulting in a low pass characteristic. In other designs [14] the SF is also used to stabilise the feedback circuit. HP and SF together form a bandpass blocking unwanted frequencies from noise. The HP also has the advantage that the DC decoupling allows to choose a different baseline voltage BL as working point for the comparator such that the comparator and amplifier can be optimised independently.

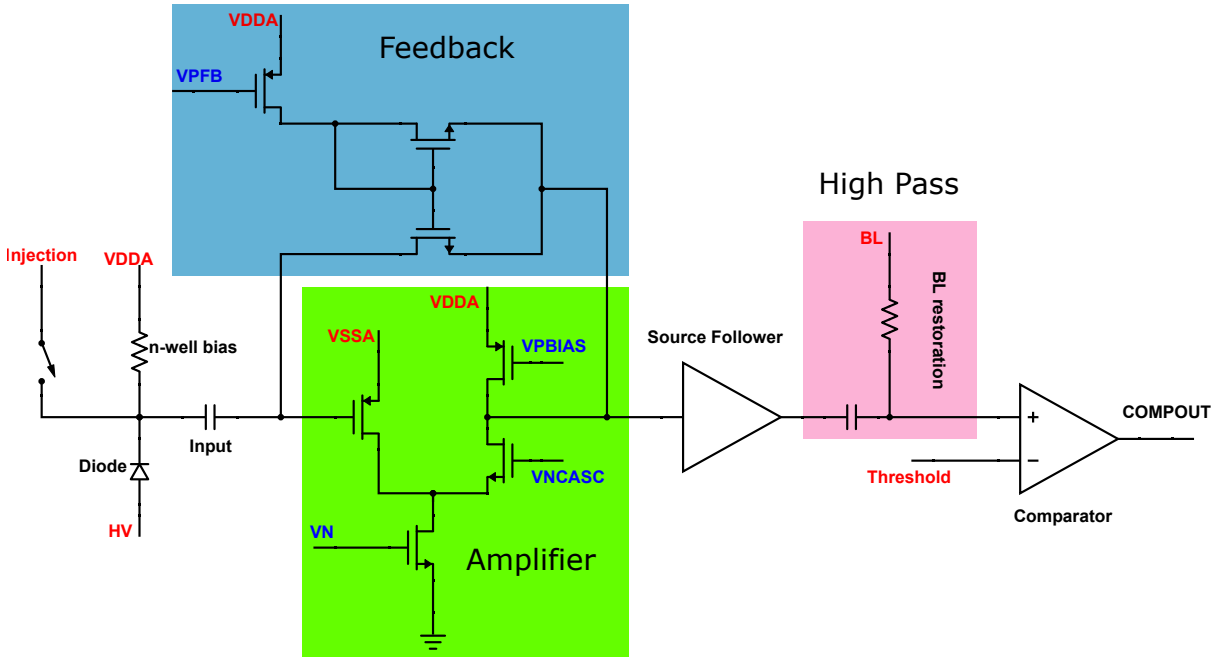

Figure 5.1: Sketch of the analogue front end electronics. Names in red are external voltages, names in blue are bias voltages generated by DACs. Black text indicates a function. The green part of the circuit is the CSA. The blue region highlights the feedback tree. The pink region forms the HP.

6 Developments for future Prototypes

This chapter describes studies already ongoing or proposed in the frame of this work. Different performance parameters are introduced to which improvements are either developed and simulated, or discussed. These topics are to be investigated in the future.

6.1 Time Resolution

One goal of the HV-CMOS group is to develop sensors with a time resolution of about 1 ns. To achieve this all contributions to the time resolution have to be examined.
The first contribution is the variance of the primary charge deposition. Clusters of electron-hole pairs are formed along the path of a traversing charged particle which fluctuate in size due to Landau fluctuations in the energy deposition. These fluctuations influence the signal of the collected charge. This process of converting energy deposition into electron-hole pairs is aided by the Fano effect. Due to the competition between different forms of energy transfer, charge carrier creation and lattice excitations, reduces the variance by the so called Fano factor (≈ 0.1 for silicon). The average energy required to create an electron-hole pair in silicon is 3.6 eV [5] compared to the band gap of 1.12 eV [5]. As result thinner depletion zones and with it less primary charge, which results in less variation, seem more desirable. However this imposes extreme limitations on the noise and performance of the amplifier.
Anisotropies in the electric field also play a role in the variance of the charge collection. Charges in regions with lower fields travel slower and arrive later at the electrodes compared to charges in high field regions. This fact favours larger pixels as the area with parallel field lines is larger relative to the area of the edges with high spatial dependence of the electric field. Also thinner depletion with higher total fields are advantageous since the total transit time is reduced.
As previously mentioned, all individual components of the circuit contribute noise to the signal. Noise adds additional variance to the signal which enforces a stricter trade-off between high efficiency or a low rate of fake hits induced by noise. These effects depend on the dimensions of the components, operating conditions, and the production process. Advanced simulation tools allow to model these effects to gauge their impact on the digitisation of the signal.
Another important source of diluted time measurements is called time-walk and is intrinsic to a binary readout. Smaller signals have a flatter slope than large signals which means they will cross the comparator threshold later even if both signals originate from the same point in time. Different methods to mitigate this effect exist and differ in complexity and rate of improvement. Adapting the electronics for steeper leading edge signals

is often difficult as the components are limited by their bandwidth. The additional power required for these amplifiers can also become too high to comply with the cooling systems of most experiments.
The most common and simplest way is measuring the time over threshold (ToT), the time between the signal exceeding the threshold and falling below it again. Since smaller signals will not only cross the threshold later but also for a shorter time, this relation can be exploited to correct the time walk. A similar method uses two comparators with different thresholds. One threshold is set as close to the baseline as possible while the other one is set just high enough to suppress noise signals. The lower threshold is used for measuring the time while the higher threshold is used for hit discrimination. A recent example is the MuPix8 [21] where a combination of this method with additional ToT correction showed very promising results [22]. Analogue time walk compensation schemes, which are for example implemented in the H35 Demo, are mostly disfavoured since they are very complex and sensitive to production variances. A more complex way to reduce time walk is the so called waveform sampling. Here, the signal is measured multiple times in short succession to capture the shape of the leading edge. These points can be used to extrapolate the time of the charge deposition to high precision. One prototype to make use of this concept is the LF-ATLASPix [23]. While this method is in principle superior to the others, it requires large amounts of logic to store and process the amount of information. This results in a requirement for high data bandwidth as well as a lot of inactive area where this logic is placed. The necessary calibration of the sampling circuits is also complex and time consuming.
The last contribution to the time resolution is the measurement itself, meaning the sampling. The resolution is intrinsically limited by the sampling clock frequency which determines the distinction between two time intervals. This contribution scales linearly with the size of the time binning t_{Bin} and is given by

$$\sigma_{Binning} = \frac{t_{Bin}}{\sqrt{12}}. \tag{6.1}$$

Higher clock frequencies are also not easy to achieve due to constraints in power and component bandwidth. An elegant solution to the problem of power dissipation is implemented in the Timepix3 [24], a readout chip for hybrids. While the main sampling clock is slow, the crossing of the threshold starts a local fast clock which only counts the cycles until the next cycle of the slow clock and is stopped afterwards. This allows for a very precise time of arrival (ToA) without the cost of requiring a fast clock at all times.

6.1.1 Trailing Edge Timing

Trailing edge timing describes the concept of optimising the timing of the trailing edge instead of the leading edge. The idea for this scheme is founded in the observation that, depending on the settings, the trailing edge in the switched feedback pixel of RD50-MPW2 is less affected by time walk than the leading edge. Since the switched feedback current is fixed and binary, the range of operation for this effect is limited. Motivated by

these findings a circuit was designed in this work which replaces the switched feedback with a dynamic current source to adapt for different operating conditions. This design is depicted schematically in Fig. 6.1.

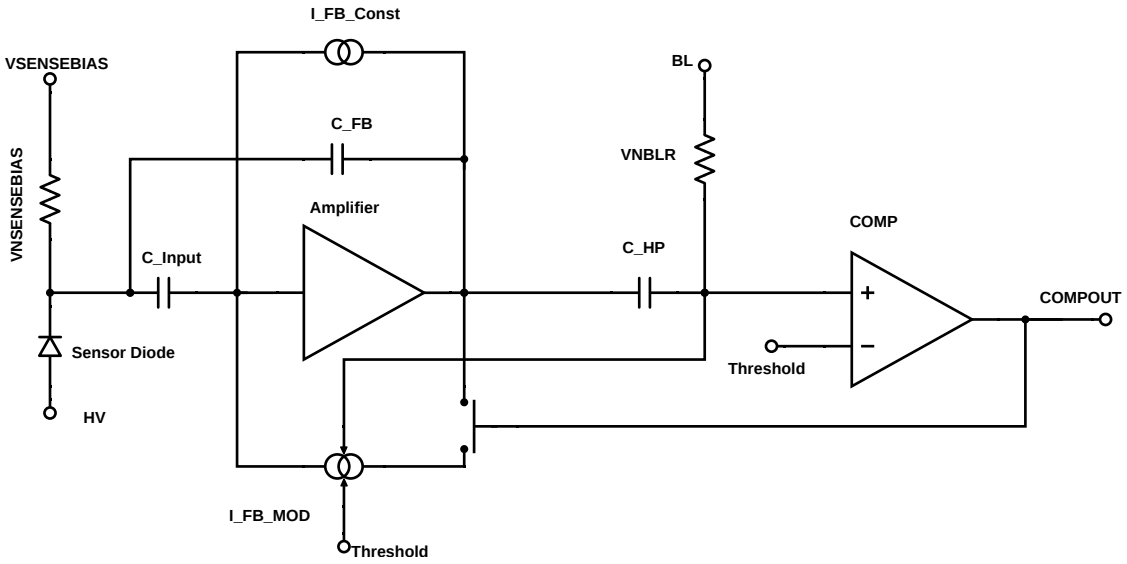

Figure 6.1: Schematic of the modulated feedback.

Conceptually, once the amplifier has reached the linear region, the information on the amount of collected charge is already present in the system. This means it can be used to shape the pulse such that it ends at the same point in time. For the proof of concept, a current source which is linear with the amplitude was chosen. This means that higher amplitudes cause an increased feedback current which in turn means the pulse is restored to the original operation point faster. In turn, smaller amplitudes have less feedback current delaying the point of crossing the threshold again. A second input in the form of the threshold voltage is used to stabilise the current source. The dynamic range of the feedback is externally configurable as less current is required for higher thresholds.
Fig. 6.2 shows simulation results of the modulated feedback current source. For input voltages close or below the threshold voltage the current output is constant, ensuring that the voltage is returned to the initial operation point. This is important since there is a delay between the crossing of the threshold and the comparator voltage, which could cause the circuit to oscillate if the signal rises again due to an insufficient discharge. This idea is however fundamentally limited by the constants of the circuit. For example signals far exceeding the linear region of the amplifier can no longer be dynamically compensated as the amplitude does not change any more. Other contributions are the time constants of the comparator. The first time constant is the delay between the threshold crossing and the generation of the output signal. The second constant is either the minimum time the comparator is active regardless of the input or the time constant to turn the comparator off after the signal has fallen below the threshold, whichever is shorter. If

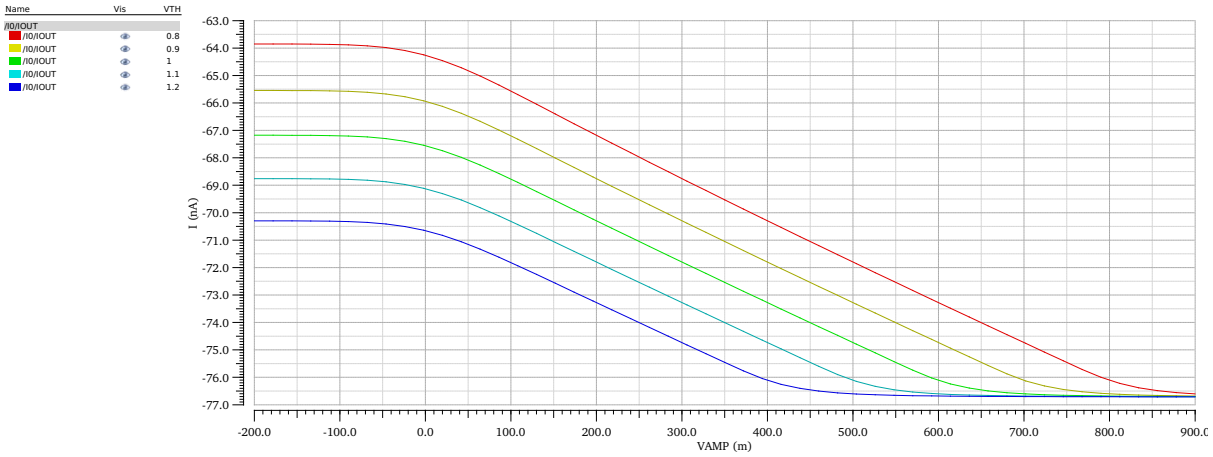

Figure 6.2: Simulation of the output current of the modulated feedback current source for different thresholds (VTH, in V). VAMP is the excess voltage over the threshold in V.

a signal is shorter than the sum of both these constants, the system is not linear and hence cannot compensate. This effect can be seen in Fig. 6.3 which shows the results of a transient simulation of the circuit. In this simulation, the amplifier is stimulated with a number of input electrons and the response is calculated. For small to medium number of electrons, the amplitude rises linearly. At large charge depositions the amplitude only increases minimally since the limit of the circuit is reached. The minimal charge deposition expected is about $80\,e/\mu\text{m}$ [5]. Fig. 6.3 also shows how the compensation has moved the trailing edges closer to each other compared to the leading edge. For the high number of input electrons ($\gtrapprox 12,000$), it can no longer compensate correctly.

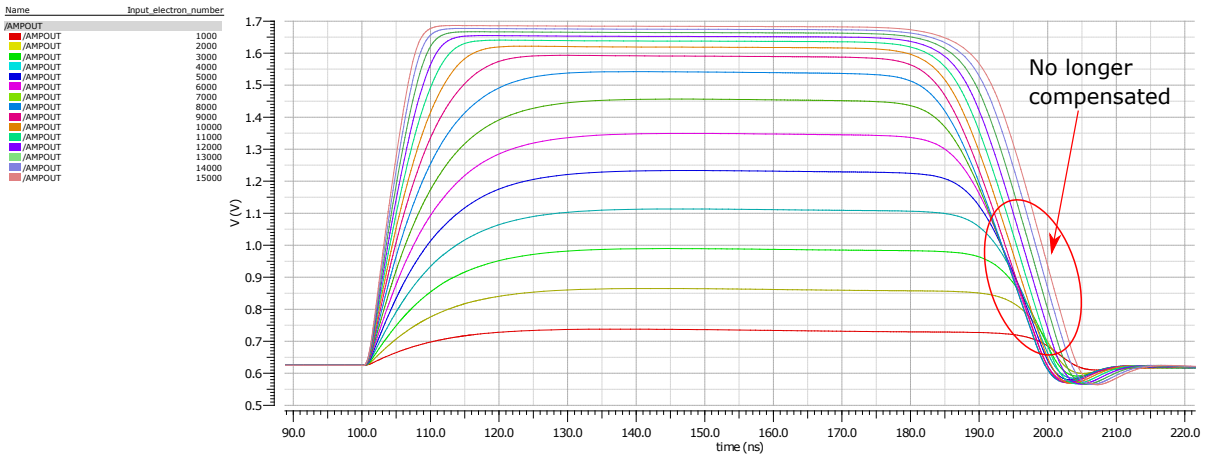

Figure 6.3: Transient simulation of the modulated feedback amplifier amplitude [V] for different number of input electrons. The region where the compensation no longer works is indicated.

The initial design has been studied in simulation on a schematic level. Important simulation results are summarised in Tab. 6.1. The power consumption is critical as most experiments have a strong limit on the cooling capacity. Compared to for example the upgrade of the ATLAS tracker the result would be in agreement with the assumed $500\,\mathrm{mW/cm^2}$ per sensor and below the cooling limit of $700\,\mathrm{mW/cm^2}$ [25]. For Mu3e it would be too high as the maximum for the experiment is $400\,\mathrm{mW/cm^2}$ with an assumed consumption of $250\,\mathrm{mW/cm^2}$ per chip [26]. The noise is another important figure as signal and noise have to be well separated enough to achieve high efficiency at a minimal rate of fake hits from noise. It can be extracted by a dedicated simulation in the form of a voltage. To convert this voltage to input electrons or equivalent noise charge (ENC), the slope of the amplitude in the linear region of the amplifier is used. Since only the schematic was used for the simulation, it is expected that the noise increases in the layout due to parasitic components.

Parameter	Unit	Value
Power Consumption per Pixel	µW	18
Power Consumption per Area	$\mathrm{mW/cm^2}$ ($60\,\mu\mathrm{m} \times 60\,\mu\mathrm{m}$ pixel size)	500
Noise	mV	10.4
Noise	ENC	120

Table 6.1: Key simulation results.

To estimate the performance on the time resolution the following method was used. The transient response of the analogue electronics was simulated for inputs between $1\,\mathrm{k}e$ and $10\,\mathrm{k}e$ in steps of $1\,\mathrm{k}e$. Then the difference in time Δt between the earliest and latest comparator signal (at 0.9 V, half the supply voltage) of these simulations was determined for both edges. The results for different thresholds (difference between the externally applied threshold voltage and the baseline voltage) are summarised in Tab. 6.2. While the timing in general worsens with increasing threshold, the performance of the trailing edge (TE) is much better than the leading edge (LE) and shows a slower rate of deterioration. At a high threshold of 80 mV the performance becomes drastically worse as the threshold is too high to work well with the lowest input charge, as shown in Fig. 6.4. The signal has already fallen below the threshold before the feedback modulation can take effect, showing the weak mode of the concept. This result shows that the shaping parameters have to be matched to the spectrum of the expected charge depositions to work optimally. Also it seems beneficial to still sample the time stamp of the LE to identify and correct cases where the feedback was not applied.
The effect of threshold trimming, the individual adaption of the comparator threshold for each individual pixel, also has to be studied. Trimming is necessary to compensate for variance between the pixels, which for example results in one pixel being much noisier than the others. If a pixel needs higher threshold to suppress noise while having the same amplitude, the comparator signal is created later which means the additional feedback is also added later, potentially adding an additional dispersion to the time resolution. On

the other hand, if the signal to noise ratio is good enough that a low threshold can be chosen without the need for trimming, the trimming in turn could be used to compensate dispersion in the timing performance of the pixel.

Threshold [mV]	Δt(LE) [ns]	Δt(TE) [ns]
35	7.5	2.0
50	10.9	2.3
65	15.9	2.8
80	27.5	31.2

Table 6.2: Simulation results for timing.

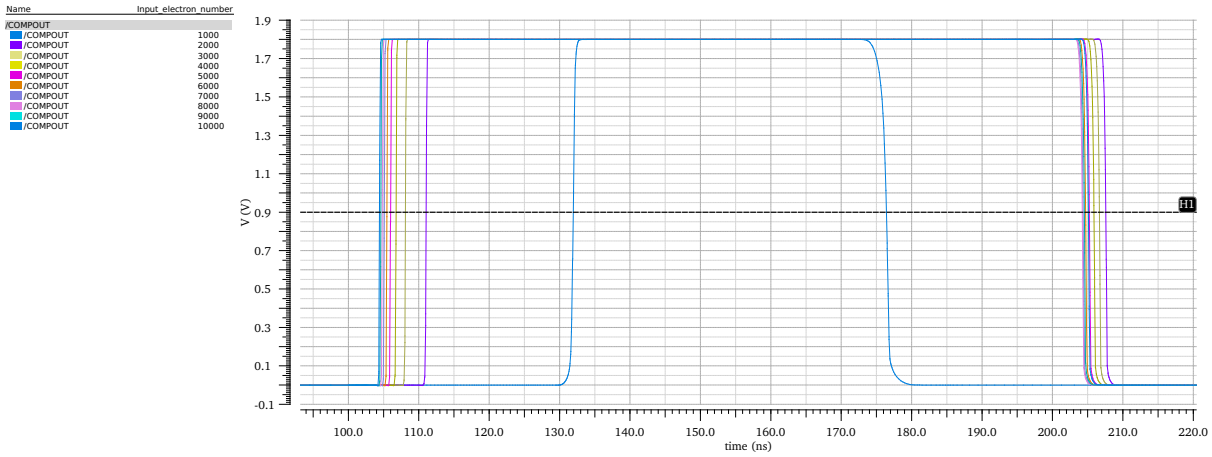

Figure 6.4: Comparator outputs at 80 mV threshold.

Since the rest of the analogue electronics are copied from RD50-MPW2, which were not designed with this goal in mind, it is expected that the performance can still be improved by changing parameters of the components or redesigning parts to fit this purpose. The amplifier and comparator of RD50-MPW2 were developed for a good time resolution on the rising edge, so it is unexpected that they are also optimal for the trailing edge. Also, components like the SF which are intrinsic asymmetric have to be investigated. A NMOS Source Follower, as used in RD50-MPW2, sources current faster than it can sink it. In the case of the comparator, the ToT dependency of the signal is commonly multiple clock cycles so that precision measurements of the trailing edge are not required. Once this concept has been proven in silicon and has been thoroughly characterised, further redesigns will be investigated.

6.2 Current Mode Signal Transmission

Most DMAPS suffer from the same problem for the internal signal transmission which is called cross-talk. Most architectures require long parallel lines with small spacing which creates a significant inter-line capacitance. If this capacitance becomes too large, voltage signals can be transferred to the neighbouring lines, creating false signals in the induced lines and a signal reduction in the original line. This effect has been shown even in small prototypes with analogue signal transmission [27] and becomes very signifiant in large scale sensors [28]. It was also found that RD50-MPW1, which uses a digital signal transmission, suffers from this effect [17].
A proposed solution is the usage of current mode signal transmission. In this concept, the signal is transmitted by a change of current in the line while the voltage difference is kept small. Since the capacitive coupling depends on the voltage change, the induced signal to other lines should be small. So far this scheme has only been implemented in the MuPix8 prototype [21] [1].
A challenge of this concept lies in the power management. As large DC connections are created in this scheme the separation of power domains has to be examined very carefully. Here, the LFoundry process with the deep p-well poses a distinct advantage since it allows for multiple power domains within the pixel, as shown in Fig. 4.1. Also the static power consumption when no signal is transmitted has to be low. While the constraints for digital signal transmission are more relaxed compared to analogue signal transmission, the current mode cannot be used for all data lines. Due to charge conservation, this scheme can only be used in either point-to-point connections or with multiple transmitters and a single receiver. For the distribution of for example the clock signal to all pixels a voltage signal is still required.

6.3 Spatial Resolution

High spatial resolution is one of the key demands by the next generation of particle physics experiments. The simplest way to improve the resolution is to reduce the pixel size. With a binary readout, the spatial resolution is then given by

$$\sigma_x = \frac{d_x}{\sqrt{12}}, \tag{6.2}$$

where d_x is the pixel size in x. Since the considerations are identical for y, only the x dimension is discussed. For some experiments rectangular pixels are more advantageous, requiring a dedicated treatment of both dimensions.
If the pixel is hit close to the edge of the pixel charge can be shared between the hit

[1] So far results on this circuit have only been presented in conferences and meetings for which the material is either not available or not public, and internal documents. No peer reviewed publication on this scheme has been published. Results are considered for a planned publication.

pixel and its neighbours. This is either due to charge being collected by the different pixels or by capacitive coupling between the pixels. In this case both pixel may pass the threshold and report a hit such that the crossing point of the particle can be set to the middle between both pixels reducing the error and improving the resolution. With charge information in the form of e.g. ToT the centre of gravity (CoG) for clusters of hits can be calculated:

$$x_{cog} = \frac{\sum_i x_i \times c_i}{\sum_i c_i}, \tag{6.3}$$

with x_i being the spatial coordinate of each pixel in the cluster and c_i the corresponding charge information. This method is especially important for hybrids and strip detectors with larger thickness. The thicker the depletion zone the more diffusion perpendicular to the drift direction takes place, making larger clusters more likely. In these regions magnetic fields have to be accounted for as they significantly influence the direction of diffusion and drift.
As one of the main goals of DMAPS development is to make the sensor as thin as possible the depletion zones are thin and diffusion is negligible. In turn, the maximum expected cluster size for traversing particles is four if the particle hits in the corner between four pixels if the diameter of the deposited charge is smaller than the pixel size. Depending on the geometry these regions of charge sharing might be small [28]. Most likely the shared charge is too small to pass the threshold if the distribution is very asymmetric or the deposited energy is small.
While smaller pixels are in general also more desirable, shrinking them poses a lot of challenges. All electronics have to fit into the deep n-well, which is smaller than the separation between the centres of two pixels to account for lateral depletion. Also all reference and supply voltages have to be connected to each pixel, which is especially critical for the supply voltages. As the line resistance can become substantial if the lines are too thin, a voltage drop and drop in performance can be the result. For sensors with digital logic in the pixel the placement of the electronics and routing of all connections are equally limiting but in theory allows for arbitrarily large matrices, neglecting voltage drops across the matrix due to finite resistance of the supply voltage lines. For sensors without logic in the pixel, point-to-point connections for each pixel to a digital partner cell are required, putting a strict limit to the size of the full matrix. In this scheme smaller pixels are possible due to the lower required space, however the ratio of active to inactive area, the so-called fill factor, becomes worse.
An idea proposed in this work to improve the spatial resolution while keeping the pixel size constant is based on Ref. [29]. The authors propose to take the charge deposition in neighbouring pixel for hit discrimination into account. For example if a pixel would register a hit, the neighbouring pixels would be informed and lower their threshold to detect the shared charge.

6.3.1 Charge Sharing Discrimination Logic

While this idea is conceptually simple and yields promising results in simulations in the reference, the implementation is not straightforward. The effect can however be emulated by combining it with the 2 threshold method for time resolution. The idea is the same as for the 2 threshold method, one high threshold for hit discrimination and a low one for timing, but the usage of the lower threshold comparator is extended. The setting of the hit flag is extended by the usage of logic gates such that either both thresholds have to be surpassed as in the 2 threshold scheme or the lower threshold in the discussed pixel as well as the higher threshold in one of the neighbouring pixels have to be exceeded. That means that a hit can also be registered if only the lower threshold is passed as long as one of the neighbours has a high charge deposition indicating the possibility for shared charge. Combined with a ToT information from the comparator with the lower threshold both an improved spatial resolution as well as an excellent time resolution could be achieved.
The concept is sketched in Fig. 6.5. A particle hits pixel 1 at the edge such that it also deposits charge (red circle) in the neighbouring pixels. Since the majority of the charge is collected in pixel 1 it is enough to pass both thresholds which would register as hit. In pixel 2 and 3 less charge is collected which is not enough to pass the high threshold. With the fanout of the hit information from pixel 1, both pixel would still be considered as hit. Pixel 4 does not receive enough charge to pass either threshold nor is it informed about a hit by its neighbours. The resulting cluster of size 3 has the CoG much closer to the actual particle crossing point than without this circuit. In that case the cluster size would be 1 and the CoG would be in the centre of pixel 1.
This scheme does not require much addition to the 2 threshold scheme besides a few logic gates and a fan out of the output of the high threshold comparator output. It is however a trade off in terms of power as the power consumption of the comparator is a significant contribution to the per pixel power consumption. However it would allow a simple way to trim the threshold of the lower comparator which is otherwise complicated. In this scheme, an arbitrary amount of charge larger than high threshold of the targeted pixel can be injected and the response of the neighbours can be measured to, e.g. reduce the amount of fake hits by a too small lower threshold.
The gain due to this logic has to be further studied and requires dedicated simulation frameworks like the Allpix^2 [30] framework in which this functionality has to be implemented.

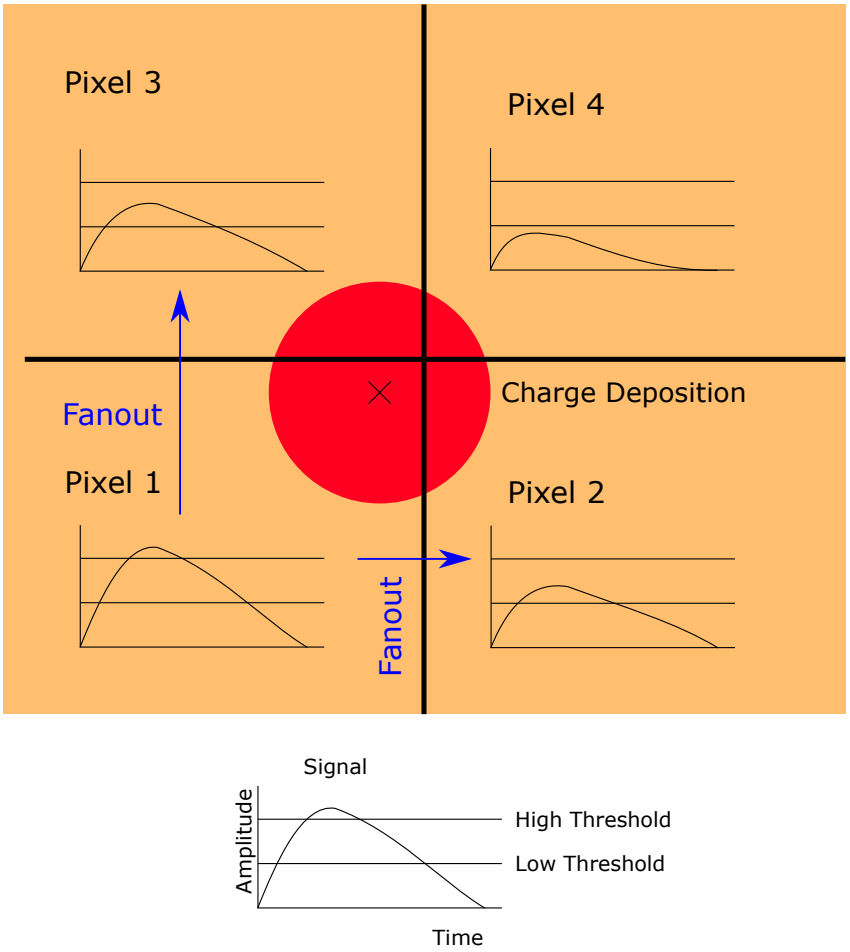

Figure 6.5: Sketch of charge sharing at the edge between pixel with induced signal.

7 Bibliography

[1] I. Perić. "A novel monolithic pixelated particle detector implemented in high-voltage CMOS technology". In: *Nucl.Instrum.Meth.* A582 (2007), p. 876. DOI: 10.1016/j.nima.2007.07.115 (cit. on pp. 4, 10).

[2] M. Havránek et al. "DMAPS: a fully depleted monolithic active pixel sensor—analog performance characterization". In: *Journal of Instrumentation* 10.02 (2015), P02013–P02013. DOI: 10.1088/1748-0221/10/02/p02013. URL: https://doi.org/10.1088%2F1748-0221%2F10%2F02%2Fp02013 (cit. on p. 4).

[3] R. Baker. *CMOS Circuit Design, Layout, and Simulation*. August 1997 (cit. on p. 5).

[4] Wikimedia Commons. *PN-junction equilibrium*. accessed 10 July 2015. URL: https://upload.wikimedia.org/wikipedia/commons/d/d6/Pn-junction-equilibrium.png (cit. on p. 6).

[5] Helmuth Spieler. *Semiconductor detector systems*. eng. Repr. Series on semiconductor science and technology ; 12 ; Oxford science publications ; Series on semiconductor science and technology. Oxford [u.a.]: Oxford Univ. Press, 2008 (cit. on pp. 6, 8, 15, 18).

[6] Wikimedia Commons. *Modes of operation of an N-type MOSFET*. accessed 19 May 2020. URL: https://en.wikipedia.org/wiki/File:MOSFET_functioning.svg (cit. on p. 8).

[7] I. Peric. "Hybrid Pixel Particle-Detector Without Bump Interconnection". In: *IEEE Transactions on Nuclear Science* 56.2 (2009), pp. 519–528 (cit. on p. 9).

[8] L. Greiner et al. "A MAPS based vertex detector for the STAR experiment at RHIC". In: *Nucl. Instr. Meth.* A 650 (2011). International Workshop on Semiconductor Pixel Detectors for Particles and Imaging 2010, pp. 68 –72. DOI: http://dx.doi.org/10.1016/j.nima.2010.12.006. URL: http://www.sciencedirect.com/science/article/pii/S0168900210027439 (cit. on p. 10).

[9] I.-M. Gregor. "EUDET-Pixel Telescope". In: *PoS* VERTEX2008 (2008), p. 034 (cit. on p. 10).

[10] Gianluca Aglieri Rinella. "The ALPIDE pixel sensor chip for the upgrade of the ALICE Inner Tracking System". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 845 (2017). Proceedings of the Vienna Conference on Instrumentation 2016, pp. 583 –587. DOI: https://doi.org/10.1016/j.nima.2016.05.016. URL: http://www.sciencedirect.com/science/article/pii/S0168900216303825 (cit. on p. 10).

[11] A. Blondel et al. *Letter of Intent for an Experiment to Search for the Decay $\mu \to eee$*. 2012 (cit. on p. 10).

[12] The PANDA collaboration. *TDR for the PANDA Luminosity Detector*. Technical Design Report (TDR). 2015. URL: https://panda.gsi.de/system/files/user_uploads/m.fritsch/RE-TDR-2015-001_1.pdf (cit. on p. 10).

[13] Dominik Becker et al. *The P2 Experiment - A future high-precision measurement of the electroweak mixing angle at low momentum transfer*. 2018. arXiv: 1802.04759 [nucl-ex] (cit. on p. 10).

[14] E. Vilella et al. "Prototyping of an HV-CMOS demonstrator for the High Luminosity-LHC upgrade". In: *Journal of Instrumentation* 11.01 (2016), pp. C01012–C01012. DOI: 10.1088/1748-0221/11/01/c01012. URL: https://doi.org/10.1088%2F1748-0221%2F11%2F01%2Fc01012 (cit. on pp. 11, 13).

[15] *ams AG*. URL: https://ams.com/ams-start (cit. on p. 11).

[16] *LFoundry S.r.l.* URL: http://www.lfoundry.com/ (cit. on p. 11).

[17] E. Vilella Figueras. "Recent depleted CMOS developments within the CERN-RD50 framework". In: (2019). URL: https://pos.sissa.it/373/019/ (cit. on pp. 11, 21).

[18] Ivan Perić et al. "The FEI3 readout chip for the ATLAS pixel detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 565.1 (2006). Proceedings of the International Workshop on Semiconductor Pixel Detectors for Particles and Imaging, pp. 178 –187. DOI: https://doi.org/10.1016/j.nima.2006.05.032. URL: http://www.sciencedirect.com/science/article/pii/S0168900206007649 (cit. on p. 11).

[19] UK Research and Innovation. *Listing of all round 1 Future Leader Fellowship cohorts*. accessed 23 May 2020. URL: https://www.ukri.org/funding/funding-opportunities/future-leaders-fellowships/meet-our-future-leaders-fellows/round-1-cohort/ (cit. on p. 12).

[20] Thomas Ackernley et al. "Mighty Tracker: Design studies for the downstream silicon tracker in Upgrade Ib and II". LHCb Internal. LHCb-INT-2019-007 (cit. on p. 12).

[21] Heiko Augustin et al. "MuPix8 - A large-area HV-MAPS chip". In: November 2018, p. 057. DOI: 10.22323/1.309.0057 (cit. on pp. 16, 21).

[22] Jan Hammerich. "Analog Characterization and Time Resolution of a large scale HV-MAPS Prototype". Master Thesis. Heidelberg University, October 2018. URL: "https://www.psi.ch/sites/default/files/import/mu3e/ThesesEN/MasterHammerich.pdf" (cit. on p. 16).

[23] Rudolf Schimassek, Felix Ehrler, and Ivan Perić. "HVCMOS pixel detectors — methods for enhancement of time resolution". In: October 2016, pp. 1–5. DOI: 10.1109/NSSMIC.2016.8069903 (cit. on p. 16).

[24] T Poikela et al. “Timepix3: a 65K channel hybrid pixel readout chip with simultaneous ToA/ToT and sparse readout”. In: *Journal of Instrumentation* 9.05 (2014), pp. C05013–C05013. DOI: 10.1088/1748-0221/9/05/c05013. URL: https://doi.org/10.1088%2F1748-0221%2F9%2F05%2Fc05013 (cit. on p. 16).

[25] ATLAS Collaboration. *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*. Tech. rep. CERN-LHCC-2017-021. ATLAS-TDR-030. Geneva: CERN, 2017. URL: http://cds.cern.ch/record/2285585 (cit. on p. 19).

[26] Frank Meier Aeschbacher, Marin Deflorin, and Lars Olivier Sebastian Noehte. *Mechanics, readout and cooling systems of the Mu3e experiment*. 2020. arXiv: 2003.11077 [physics.ins-det] (cit. on p. 19).

[27] H. Augustin et al. *MuPix7 - A fast monolithic HV-CMOS pixel chip for Mu3e*. 2016. arXiv: 1610.02210 [physics.ins-det] (cit. on p. 21).

[28] Lennart Huth. “A High Rate Testbeam Data Acquisition System and Characterization of High Voltage Monolithic Active Pixel Sensors”. PhD Thesis. Heidelberg University, December 2018. URL: "https://www.psi.ch/sites/default/files/import/mu3e/ThesesEN/DissertationHuth.pdf" (cit. on pp. 21, 22).

[29] Benjamin Nachman and Alex Spies. *Nonlocal Thresholds for Improving the Spatial Resolution of Pixel Detectors*. 2019. arXiv: 1903.01624 [physics.ins-det] (cit. on p. 22).

[30] Dominik Dannheim et al. *Combining TCAD and Monte Carlo Methods to Simulate CMOS Pixel Sensors with a Small Collection Electrode using the Allpix Squared Framework*. 2020. arXiv: 2002.12602 [physics.ins-det] (cit. on p. 23).